2009 INTERNATIONAL SYMPOSIUM ON

Aviation Psychology

DAYTON,OH / APRIL 27-30 2009

ISAP

WRIGHT STATE UNIVERSITY

Human Effectiveness DIRECTORATE

For Your Convenience, The Table Of Contents Is Linked To The Body Of The Document And Is Keyword Searchable.

Table Of Contents

Airport and Runway Safety

Work Analysis for Future Aviation

Aiding the General Aviation Pilot

Accident Analysis and Operational Errors

Cognitive Engineering: Visualization

Air Traffic Control

Posters

Training and Simulation

Safety Reporting

Task Load Effects

Representation in Different Perceptual Modalities

University Aviation Degree Programs

Aerospace Automation

Flight-Path Displays

Training

Posters

Air Traffic Challenges

Factors Affecting Cabin Crew Performance

Air and Space Displays

Aeronautical Decision Making

ISAP Keynote

THE PSYCHOLOGY OF AVIATION SURPRISE: An 8 YEAR UPDATE REGARDING THE NOTICING OF BLACK SWANS

Christopher D. Wickens
Alion Science & Technology, Boulder, Colo.
& University of Illinois Human Factors Division

We describe the limitation that people have in noticing very unexpected, surprising "off-nominal", or black swan events, as reflected in the psychology of change blindness; and how this limitation can compromise aviation safety. We then describe a three phase program of research examining pilot response to these black swan events, using (1) a meta-analysis to reveal the miss rate in noticing black swans, (2) a model of visual attention to predict this miss rate, and (3) the same model to make predictions regarding the safety impact of NextGen technology and procedures.

In 2006, an Embraer Legacy business jet and a commercial 737 passenger aircraft collided in mid air over Brazil (Command of Aeronautics, 2006). The 737 was seriously damaged, crashed, and all lives were lost. While, as in any fatal aircraft accident, there were many factors responsible, one of the most critical is that the transponder on the Embraer was not sending its position, such that a TCAS alert in the 737 would have registered the impending collision and an evasive maneuver could have taken place. At an earlier time in the flight history, there is good evidence from air traffic control communications that the Legacy **was** transponding (in response to interrogation) so, at some time prior to the collision, the communications system within the Legacy must have become disabled, and a display within the cockpit changed its state to signal this event, one of great significance and importance, but one that the pilots on board apparently failed to notice. We will emphasize below that this "failure" is one that is quite understandable given the frailties of human attention. We point out here that this provides a prototypical example of the criticality to aviation safety, of noticing unexpected, and often not very salient "off-nominal" events or, to use the term coined by Taleb (2007), "black swan" events.

This issue then is a key element of the "psychology of surprise", which was the focus of my talk at this symposium in 2001 (Wickens, 2001). Since that time, two key elements have led me to revisit this theme, 8 years later. First, at that time, I expressed regret at the lack of much valid data, in realistic environments, that could help aviation psychologists understand pilot (and controller) response to the black swans. A good deal more of such data exist now, and will be summarized below. Second, we are entering a period when revolutionary changes in the airspace are forecast, as reflected by the proposed procedures and equipment that are embodied in the next generation of the airspace or NextGen (JPDO, 2008). Such changes are designed to increase the productivity in the airspace, while preserving levels of safety. Predictive models are being developed to demonstrate the assumed productivity (i.e., capacity) benefits of procedures like merging and spacing, self separation, RNAV, and equivalent visual operations. It is important however that valid models also be developed to predict the safety implications of these productivity enhancements; we argue here that safety concerns **in an already very safe system** must, by definition, be associated with 'black swan" unpredictable events. If such events were predictable, then their consequences would have been mitigated. Any such model will be unlikely to predict **when** such an event will occur, but it should be able to predict the conditions that might make a black swan more possible, as well as key features of the human response to the event, which is the focus of this address.

If we examine the human (e.g., aviation worker) response to very surprising events, we can identify three categories of processes where the response might break down:

1. In noticing (or failing to notice) the triggering event. The Embraer mid-air collision provided such a prototypical example, with the apparent failure to notice the display change signaling the cessation of position broadcast. Another example would be the runway overshoot crash on take-off at Lexington Kentucky (NTSB, 2007), where pilots failed to notice important cues that they had lined up for approach on the wrong runway.
2. In diagnosing. Although airspace workers may notice that things are not right in a timely fashion. (This is, after all, the process that alarm systems support), they may not fully understand the nature of the unexpected problem: a correct situation assessment. Pilots in the CFIT accident near Cali Columbia were aware of a navigational problem, but did not understand, until too late, the course of their trajectory relative to the mountains.
3. In selecting and executing appropriate procedures. A number of aircraft accidents in the previous decade, associated with the flight management system (Dornheim, 1995), were attributed to pilots, "fighting" the autopilot. For example in the Air China crash at Nagoya Japan, the pilot and autopilot were imposing opposite forces on the plane's elevators, until an abrupt pitch up attitude caused a stall.

Of these three stages of pilot information processing, I focus on the first – noticing – for two important reasons. First, it is a well-defined safety bottleneck. Jones and Endsley (1996) have surveyed the literature and found that stage 1 situation awareness breakdowns (which can roughly translate to failure of noticing and/or perception) account for 76% of SA related errors in aviation. Second, such failures directly reflect the psychology of change blindness or inattentional blindness (Simons & Levin, 1997; Rensink, 2002), a striking phenomenon well researched in basic psychological laboratory, that "scales up" remarkably well to applied worlds of driving, flying, and process supervision. (Carpenter, 2001, Martens, 2007; Wickens Thomas & Young, 2000, Sarter, Mumaw & Wickens, 2007, Stelzer & Wickens, 2006).

The phenomenon of change blindness, whereby people are quite insensitive to noticing changes or events in the world around them has three characteristics, and change blindness will be more prevalent to the extent that all three are present:

- The event occurs away from foveal vision, and bear in mind that at any given time, only about 0.02% of the visual world occupies a pilot's foveal vision.
- The event is relatively subtle or non-salient (e.g., not a flashing light, but just the appearance or disappearance of a visual object or displayed element).
- The event is unexpected: Here what is meant by "unexpected" can range between events such as a conflict alert in an ATC facility, which occurs rarely, but the controller assumes that one could happen at any time, and events such as the offset of the transponder signal, for which there may be no expectation whatsoever. We have referred to these as "unexpected" versus "truly surprising" events, and Taleb (2007) has called them gray swans and black swans respectively.

In addition to these three features contributing to change blindness, data suggest that the failure to notice will be amplified as resources are withdrawn from monitoring by dual task conditions, such as those prevalent in the descent phase of flight, particularly in single pilot operations.

How prevalent is change blindness in these circumstances? In a study of cockpit displays of traffic information to support self-separation, Stelzer and Wickens (2006) observed that pilots failed to notice over 80% of changes to flight trajectories of nearby aircraft on the (fairly cluttered) display in front of them, given that they were also engaged in a primary flight task (miles in trail distance keeping), and that the events themselves were not signaled by a warning such as a flash; but rather just a change in a digital data tag (for altitude), or movement direction across the display (for heading). In another study, pilots were inferred to 'miss" around 50% of flight mode annunciator changes in a full mission simulator, at least as this miss rate was inferred from the absence of a visual fixation on the FMA following the change (Sarter et al., 2007). Importantly, in both of these cases, we can refer to the events as gray swans, not black swans, since participants were very much aware that such changes could occur in the context of the experiment.

For extrapolation to aviation safety our research team was more concerned with the actual "black swan" events, for which we suspected that noticing rate might be lower (even though the base rate of such events would also be, by definition, drastically lower). However our challenge was to find **statistically reliable** estimates of such a miss rate, and of the causal effects that could moderate it. The challenge here of course is that by definition, in any experiment, from the perspective of the pilot once such an event occurs once it is no longer totally unexpected (and therefore no longer a black swan). Hence the response to the event can occur only once per pilot per experiment, and this "low N" often thwarts the efforts of researchers to extract statistically reliable data regarding a black swan response.

In order to overcome this challenge to statistical power, in the first of three elements of our research program (Gore et al, 2009), we turned to the technique of meta-analysis (Rosenthal, 1991), in an approach described in detail in Hooey et al. (2009). Here, we identified in the literature every aviation study we could find that used a relatively realistic flight simulation along with licensed pilots, and at some point in the experiment presented a truly surprising, safety-critical black swan event. For example an investigation of synthetic vision systems for landing may present, on the final trial of the experiment, a runway incursion, after several sessions of incursion-free landings (Wickens et al., 2009).

The output of this meta-analysis produced a series of "effects" on off-nominal miss rates that supported our understanding of the safety concern that they engender. We found that overall about 1/3 of the pilots missed these events, and one study (Thomas & Wickens, 2004) was able to attribute such misses in part to pilot scan strategies: those who tended to look less frequently where the event occurred, were less likely to notice it. Importantly, we also found at least four factors that affected this miss rate in a statistically reliable fashion when pooled over studies. Our analyses revealed that pilots had a higher miss rate (MR) for the off nominal event when:

- A black swan outside world event was to be detected while they were flying with a HUD (MR = 0.36) versus without a HUD (MR = 0.27).
- The event was a truly surprising black swan (MR = 0.48) rather than an unexpected gray swan (MR = 0.29).
- The unexpected event occurred down on the instrument panel (MR = 0.39) than out the window ((MR = 0.29)
- An outside-the-cockpit black-swan event occurred while pilots were flying with a head down highway in the sky display (HITS : MR = 0.45) rather than without a HITS (MR = 0.22).
- The off-nominal event was an erroneous clearance delivery and it was delivered by data link alone (MR = 0.69) rather than redundantly with data link and voice (MR = 0.38).

While such relatively low levels of performance might well be considered disconcerting for aviation safety, we also recognize that such misses will occur quite infrequently, since the base rate of these off-nominal black swan events is, by definition, exceedingly low (but not impossible). However one of the ironies of automation (whose failures may often be considered black swan events) is the ironic fact that the rarer the event is, the less expected it becomes, and hence the greater is the likelihood of missing it (Bainbridge, 1983). Furthermore, the results from these high-fidelity flight simulations certainly replicate what is now well known regarding change blindness and inattentional blindness in the real world (Rensink, 2002: Simons & Levin, 1997; Sarter et al., 2007; Stelzer & Wickens, 2007; Wickens & Alexander, 2009; Wickens et al., 2000). That is, people simply do a poor job of noticing changes (events) when these are unexpected, are not salient and occur outside of foveal vision; all conditions that typified the events analyzed in our meta-analysis.

Of course such empirical data as those reported above, while of value in explaining potential concerns with current day (e.g., HUD) and near-future (e.g., HITS) technology, does not inform us of the miss rate of future NEXGEN systems and concepts. For this we must turn to computational modeling (Foyle & Hooey, 2008), which constituted the second element of our 3-element research effort. In this element we developed a model of noticing rare visual events, called N-SEEV. SEEV describes the four components that drive visual scanning around the workplace, typical of the cockpit: Salience, Effort (conservation) Expectancy and Value. Then, in the context of this normal steady-state allocation of visual attention, a to-be-noticed event occurs, in this context the “off-nominal event” (e.g, the onset of a warning signal in the cockpit). Now the noticing (“N”) component of N-SEEV predicts how long the eye will take to land on the location of the event and/or, the probability that it may not be noticed at all (miss rate), or noticed before some deadline.

N-SEEV actually has several parameters (See Wickens et al., 2009); but can effectively drive a simulated eyeball around a simulated cockpit, in a way that does an adequate job of mimicking actual pilot scanning (Sarter et al., 2007) and capturing the variance in noticing time across different display concepts (Nikolic et al., 2004). What we did in the second component of our study was to program N-SEEV to mimic the conditions in which pilots confronted the black or gray swan events, across three of the dichotomous comparisons revealed by out meta analysis: event location, event expectancy (black vs. gray swans) and the presence or absence of a HITS in the cockpit. We correlated the model-predicted miss rate with the observed miss rate from the pilot-in-the-loop simulation (the output of the meta-analysis), and found that our model could predict the actual miss rate of all six conditions in the three contrasts within 14%, and four of the six within 7%.

For those of us interested in the value of computational models in human factors (Foyle & Hooey, 2008) these findings were of great importance because they revealed a good degree of empirical validation of the models in predicting new data. Thus equipped with a model that we believe is valid, the third element of our research was to demonstrate how the model could apply to making predictions of vulnerabilities for certain NextGen procedures and equipment.

For example, these model prediction runs revealed the degree of advantage gained (in noticing in-cockpit warnings) by positioning those warnings close to the primary flight display; but also the high cost, to noticing out-of-the-window black swan events, associated with placing heavy visual demands on the pilot to monitor a CDTI in a self-separation procedure, and particularly with the enhanced cognitive demands when there is a simultaneous engine failure. Here our model predicts that around 80% of these events will be missed.

It is important to reiterate that the likelihood of such occurrences are by definition extremely rare. But, as others have noted (e.g., Taleb, 2007), just because they may be rare, they are not impossible, and so human factors practitioners should be concerned with ways to mitigate these examples of black swan change blindness. In this regard, another advantage of this (and other) computational models emerges: It is a matter of only a few minutes to change parameters of the model (e.g., those associated with an added visual alert, or with re-positioning a display to a HUD location), and a new (and presumably lower) miss rate can be calculated, to signal the safety-advantage of the mitigation.

Of course models are not the panacea for aviation safety. It is important to realize that:

- they are only as good as their validity, and validity itself can only be achieved from empirical pilot-in-the-loop data.

- They can rarely account for all factors (without becoming unwieldy and over-complex). For example N-SEEV only predicts the behavior of a single pilot, and it is unclear how a pair of eyes in the commercial cockpit, with collaborative scan strategies might mitigate some of these effects.
- A model like N-SEEV can rarely predict **when** a miss of a black swan event will occur; or precisely what form the black swan itself will take, but only the circumstances that drive this miss rate up and down and, by extension, how to reduce this rate.

Finally, in conclusion, while we will clearly never be able to eliminate either the occurrence of off-nominal events or the challenge to human attention of noticing them, by understanding what those challenges are, and predicting the circumstances in which they may be amplified, we can go a long way in helping pilots and controllers cope with the "psychology of surprise".

Acknowledgements

The ideas expressed and data presented in this paper could not have been done without the contributions of my senior collaborators on two NASA research projects: Brian Gore, Becky Hooey, Jason McCarley and Angie Sebok. In addition, Kelly Steelman-Allen, Shaun Hutchins, Julie Bzostek, Corey Koenicke, and Ellen Salud all made vital contributions to model development and data collection. Captain Rick Shay of United Airlines provided vital professional input to the projects. The two NASA projects were NASA's Airspace Systems – NGATS ATM: Airspace Project (NRA # NX08AE87A) to San Jose State University; subcontracted to Alion Science: MA&D Operations (Dave Foyle, Scientific technical manager) and NASA ROA 2007: Integrated Intelligent Flight Deck Technologies, NNX07AV97A, to the University of Illinois, and also subcontracted to Alion Science: MA&D Operations (Jeff Mulligan, Scientific-technical manager).

References

Carpenter, S. (2001). Sights Unseen. *APA Monitor, 34,* 54-57.

Bainbridge, L (1983). Ironies of Automation. *Automatica, 19,* 775-779.

Command of Aeronautics (2006) final report A-0-22 CENIPA /2008. Brazil. Command of Aeronautics, General Staff of the Aeronautics, Aeronautical Accident Investigation and Prevention Center (CENIPA). (2008). Aeronautical Accident, PR-GTD and N600XL, B-737 8EH and EMB-135 BJ Legacy, 29 September 2006.) Brazilia, Brazil: Cenipa

Dornheim, M. A. (1995, January 30). Dramatic incidents highlight mode problems in cockpit. *Aviation Week and Space Technology*, 57-59.

Foyle, D. C. & B. L. Hooey (Eds.). (2008). *Human performance modeling in aviation*. Boca Raton, FL: Taylor and Francis/CRC Press.

Gore, B.F., Hooey, B.L., , Wickens, C.D., Sebok, Koenecke, C., Salud, E., Hutchins, S & Bzostek, J. (2009a). *Identification of pilot performance parameters for human performance models of off-nominal events in the NextGen environment*. (NASA Final Report and Deliverable. ROA 2006: Airspace Systems – NGATS ATM: Airspace Project, NRA # NNX08AE87A), San Jose, CA: San Jose State University

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 of Aviation Psychology*. Dayton, OH: Wright State University.

JPDO (2009). Joint Program & Development Office. Next Generation Air Transportation. Integrated work plan. Version 1.0.: Washington D.C.: FAA..

Martens, M. (2007). *The failure to act on important information: where do things go wron*g. Dissertation: Free University of Netherlands: Amsterdam: The Netherlands.

Nikolic, M. I., Orr, J. M., & Sarter, N. B. (2004). Why pilots miss the green box: How display context undermines attention capture. *The International Journal of Aviation Psychology,* 14, 39-52.

NTSB (2007). Attempted takeoff from wrong runway. Aircraft accident report. Washington D.C.: National Transportation Safety Board.

Rosenthal, R. (1991). *Meta-anallytic procedures for social research*. Newbury Pk, CA: Sage Press.

Rensink, R.A. (2002). Change detection. *Annual Review of Psychology, 53*, 245-277.

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

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Science, 1*(7), 261-267.

Steltzer E.M. & Wickens, C.D. (2006). Pilots strategically compensate for display enlargements in surveillance and flight control tasks*. Human Factors, 48*, 166-181.

Taleb, N (2007). *The Black Swan. The impact of the highly improbable*. NY: Random House.

Thomas, L. C., & Wickens, C. D. (2004). Eye-tracking and individual differences in off-normal event detection when flying with a synthetic vision system display. *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: Human Factors and Ergonomics Society.

Wickens, C.D. (2001). Keynote address: Attention to safety and the psychology of surprise. *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus, OH: Dept. of Aerospace Engineering, Applied Mechanics, and Aviation, Ohio State University.

Wickens, C.D. & Alexander, A. L. (2009). Attentional tunneling and task management in synthetic vision displays. *International Journal of Aviation Psychology, 19*.

Wickens, C.D., Thomas, L.C., & Young, R. (2000). Frames of reference for display of battlefield terrain and enemy information: Task-display dependencies and viewpoint interaction use. *Human Factors, 42*, 660-675.

Wickens C.D., Sebok, A., Bzostek, J., Steelman-Allen, K., McCarley, J. & Sarter, N. (2009). NT-SEEV: A model of attention capture and noticing on the Flight Deck. *Proceedings of the 15th International Symposium of Aviation Psychology*. Dayton, OH: Wright State University.

Simulator Motion…It Rocks! (Or Maybe Not)

Bob Jacobs

University of Illinois Engineering Psychology Ph.D. (1976)

I want to briefly share some of our early work at the University of Illinois Aviation Research Laboratory relating to the nature and role of motion cueing and its relationship to pilot performance in flight training, skill evaluation, and flight instrument utilization. First, though, I'd like to offer some personal testimony about Stan Roscoe, the director of the Lab, so that you can appreciate the extraordinary environment we were provided in which to pursue our research and to learn.

Stan was one of a kind. As a scientist, as an educator, and as the leader of a research organization, Stan always set the bar at its highest limit. He insisted that we who worked for him, studied under him, and helped him to contribute to the understanding of how best to combine human beings and technology in systems reach well beyond "good enough". For Stan, whether the task at hand was human factors engineering support to a major aerospace program, the conduct of experimental studies in aviation psychology, or the sharing of our scientific activities with sponsors and peers, a clear focus, professional quality, and absolute integrity were required without compromise.

Like my colleague Larry Scanlan, I had the privilege of working for Stan at Hughes Aircraft for several years before following him back to the University for graduate studies. During that time, I was constantly amazed at Stan's incredible energy, his patience, and his persistence as he directed our work in the Display Systems and Human Factors Department. In most aerospace companies, human factors engineering was just one of the "illities" – a backwater discipline staffed by individuals of modest aspirations performing work that was required by the customer to check off the contractual "boxes", but in truth not of very high interest to company leadership. This because like training, logistics, or safety, human factors was regarded as an annoying constraint to main product line technical innovation and certainly not fertile ground for the growth of business and profits. Across the industry, the human factors organizations were on the lowest link of the food chain when it came to support for corporate sponsored independent research and development, capital investment, or other expenditures.

Stan made things very different for us at Hughes. First, he constantly reminded us and others up and down the chain of command how important our work was to producing systems that would deliver maximum man-machine system performance to our customers. He taught us to find out what was going on in every corner of the company and to aggressively market our technical knowledge and research capabilities to programs inside the company that could benefit. The result was that at Hughes, the human factors organization was, I believe, held in much higher regard for its contributions to program success than was common elsewhere.

Second, Stan early on recognized the importance of simulation as a tool for supporting system engineering, and developed a center of expertise focused primarily on man-in-the-loop simulation for research and concept demonstration. This was a unique resource within the company, and it became an important tool for the attraction of work both within the company and from customers outside. Although quite unusual for what would typically be regarded as a support service by other companies in the aerospace business, at Hughes, the human factors activity was a very successful direct support contractor to many DoD customers - system developers as well as research oriented agencies.

Third, Stan believed strongly in hiring strong people and developing the talents of those who worked for him. Stan must have been the record holder for securing a disproportionate share of Hughes' generous educational support for graduate studies for his people, and the department enjoyed one of the highest ratios of graduate degreed professionals in the company. Larry Scanlan and I were both to become beneficiaries of his efforts to secure graduate fellowships for us – in both cases unprecedented at Hughes because they provided full time study for an extended period at a University far away from southern California.

Lastly, Stan appreciated the importance of communication skills for his staff. Even those in very junior positions were put before customers to present their work, and given the opportunity to participate in the preparation of proposals and reports so that we could learn how to share our ideas verbally and in writing to his high professional standard. The fact, as we knew all too well, that he had been an undergraduate English major in college, was constantly on our minds as we prepared our materials.

I offer this description of how Stan ran his organization at Hughes because when he returned to the University of Illinois to reactivate the Aviation Research Laboratory, he brought these same values and beliefs to the directing of the Lab. Stan was able to create the Engineering Psychology program and establish parity for the discipline within the Psychology Department with the other more traditional pursuits there. He brought relationships with the government agencies that sponsor research in aviation psychology with him when he arrived, and quickly built up a portfolio of funded research contracts that made ARL a going concern. Stan's appreciation for the importance of simulation as a focal point and tool for the research program continued, and he developed a sponsorship relationship with a major manufacturer of general aviation simulation systems that resulted in the laboratory obtaining a state-of-the-art simulator for its work. He also succeeded in finding the resources to provide real-time computing capability and access to aircraft to support our experiments.

Stan's standards for high quality staff remained unchanged as well. The Psychology Department of the University of Illinois is held in very high regard and is very selective in its acceptance of graduate students. At the time of my admission, the acceptance ratio was around 5% of the applicants. Over and above this, Stan insisted that for a graduate student to become part of the research staff at the ARL, a minimum of a private pilot's license was a prerequisite, and more advanced certification was desired. Many of us were commercial pilots and flight instructors when we became part of the ARL research family.

Stan continued to insist upon high standards of communication skills for his people at the University as well. Unlike most research organizations on campus, Stan scheduled annual program reviews for our research sponsors and professional colleagues. Graduate students were expected to prepare professional conference/journal quality papers for presentation of their work to an audience that traveled to the University from all across the country. These were quite formal affairs, but Stan considered them to be learning experiences as significant as any of the academic work or research. The papers were published in a proceeding of the meeting, and often submitted for further publication in professional journals.

Student researchers were also required to write proposals and technical reports describing their proposed work and results to meet contractual requirements of our sponsors. These documents were also required to meet a high standard of quality. Stan considered them to be practice for our eventual professional counterpart activities.

In many respects, the laboratories operated as if it was a research and development enterprise, but with the overlay of University academics and periodic turnover of the staff researchers as new candidates were accepted and others completed their degrees and moved on.

Now in that time, high fidelity flight simulators were a very expensive commodity. High performance visual systems ran about $1M per channel, and synergistic 6 post motion systems were also significant cost drivers. Stan recognized and wrote that in the case of the motion systems, the cost impact arose from more than just the cost of the motion platform and its driving software. These systems were very maintenance intensive, but in addition required a large volume of space within which to operate –read bigger building, consumed a lot of power, and were thought to pose a safety risk and so required costly risk mitigation. Our research simulator had a simpler pitch/roll motion system, but even that device was significantly more expensive to purchase than a non-moving counterpart.

The prevailing wisdom among simulator manufacturers than was that the contribution of various aspects of simulator fidelity to the overall effectiveness of the systems for flight training or flight proficiency assessment was not known, so under the presumption that higher fidelity correlated with higher transfer effectiveness and/or higher predictive validity for flight checks, simulator users were advised to purchase as much fidelity as possible. One simulator manufacturer at the time marketed its products with the slogan "uncompromising realism", as if that ensured the purchaser would realize maximum return on investment.

The problem with that theory, at least in the motion dimension of fidelity, is that even the best synergistic platforms are only capable of limited physical excursion on each axis; so sustained acceleration cannot be simulated. Instead, these systems can be used to cue supra-threshold linear and rotational acceleration over a very limited range, and then must be restored to a neutral state ideally subliminally so that the simulator occupant does not notice the transition. This is not the same as the set of motion cues experienced by the occupant of a maneuvering aircraft which can sustain accelerations through vast displacements – so even the best motion simulation does not produce "uncompromising realism".

We began talking about an alternative design philosophy for simulation – one that we called "selective-fidelity". The concept was to invest in cue realism in visual, audio, whole body motion, tactical feedback, etc. when the cues could be demonstrated to contribute to the transfer effectiveness or predictive validity of the simulator experience, but not to spend money on aspects of the simulator design where no relationship could be shown to its value as an environment for training or testing. Of course to put this strategy into practice, it became important to gain an understanding of the relationship between the nature and fidelity of these cues and the effectiveness of the simulator in its design mission. Understanding the role of motion cueing in this respect became a major research thrust for ARL during the early 1970's.

A further question on the table had to do with the role that motion cues had to play in aircraft flight control related response. Was the perception of motion an alerting cue – one that merely triggered a process of interpretation of instrument indications leading to formulating a response? Or was the cue an essential input to the response formulation itself – were the magnitude and direction of the motion cue characteristics when processed in concert with some sort of operative dynamic model of the aircraft control loop, determinants of the characteristics of the response?

Recall that at that time, ARL was the beneficiary of substantial support from a major manufacturer of simulation systems, including the general aviation simulator that the lab intended to employ to address these questions. When it became known that it was our intention to try to quantify the role of motion cues in determining the effectiveness of simulators as training or testing environments, our benefactor, who realized a meaningful proportion of its revenue through the sale of motion systems, was none too happy. I recall that there were extensive discussions between Stan and our point of contact at the company in which it was suggested that:

1. Perhaps we ought not to be doing this research, or, alternatively,

2. Perhaps we ought to perform the study at their company facility or at an Air Force simulation facility that they operated, where they could provide “help”.

To his great credit, and at some risk to the continuing support of the lab, Stan resisted both suggestions. I regard this episode as one of many instances of Stan putting scientific integrity before political considerations.

In the few years that followed, we conducted three sets of experiments that focused on the issue.

One of our graduate researchers, Fuat Ince, conducted a research application oriented study in which various formats of attitude indicators (moving horizon, moving airplane, frequency separated, kinalog) were tested in a disturbed roll-tracking task under various conditions of motion. Error in tracking and especially control reversals were measured. Ince found that there was a reliable interaction between the nature of the motion cueing and tracking performance, and that there was also a significant difference in the frequency of control reversals in recovery from unknown attitudes across motion conditions. Interestingly, the results showed that tracking performance most closely matched performance in an aircraft when the simulator was operating with washout motion, but that control reversals were minimized when the simulator was set to present sustained bank and pitch cues. This suggests that the role of motion cueing extends beyond the alerting role postulated earlier, and that to some degree at least the directionality of the motion in the roll axis helps to produce an initial roll control response in the right direction.

Lt. Col. Jeff Koonce, a graduate student at the lab with the Air Force Institute of Technology Ph.D. program, investigated the role of motion cueing with respect to a second domain of simulator application – predictive validity of ground based flight proficiency testing. In his experiment, experienced instrument rated pilots were given two flight proficiency checks in the simulator on successive days, followed by an check flight in an aircraft. Simulator check rides were conducted under three motion conditions (no motion, sustained motion, and washout motion). Jeff found that, as would be expected, performance improved with each succeeding check ride. Test subjects made more errors in the simulator without motion, made fewer with sustained motion, and performed best with washout motion. The order of results was consistent from day one to day two in the simulator. But when the check ride was conducted in the aircraft, the no-motion group performed reliably better than the other two indicating a differential impact of motion condition on any learning that may have been taking place over the three sets of trials. Flying the simulator without motion is harder – pilots have to concentrate more intensely on the instruments without motion cues to aid them. The suggestion here is that perhaps that greater effort resulted in measurably different skill gain in transfer to the aircraft.

To test whether this might have been the case, in my own dissertation study, I examined the role of motion cues as a factor in the transfer effectiveness of the simulator in an abbreviated primary flight-training curriculum. I will not detail the procedures for establishing control of such variables as instructional technique or subject aptitude; only reassure that these were accounted for. Four groups of subjects, none of whom had any previous flight experience as either pilot or passenger and thus had no expectations for the motion cues in an aircraft or simulator, were trained to private pilot proficiency standards on a series of maneuvers under instrument conditions. To make the task more challenging, a complex airplane with retractable landing gear and a controllable pitch propeller was used. A control group received all of its training in the aircraft, repeating each of a series of successively difficult maneuvers that involved at first simple maneuvers such as maintaining heading and altitude, and progressing to more complicated "Charlie" patterns in which the subject had to calculate headings, climb and descend then maintain target altitudes, make standard rate turns alternating left and right through 90 or 270 degrees, adjust power, retrim the aircraft, etc. All of this was performed under an instrument hood which would pose a significant challenge to even a certified private pilot. Subjects repeated each maneuver task until two successive trials were performed to exit criterion (private pilot) standards.

The experimental groups went through the same training sequence, but received instruction in the simulator first under one of three conditions of motion then were tested in the same way in the aircraft. One group trained without motion, while the other two groups experienced washout motion or a special hybrid condition in which the simulator provided washout motion but with a random directionality. The latter condition was introduced because it provided an alerting cue, but not a dependable polarity and so could not be relied upon as a parameter from which to formulate a directional response.

I found that when comparing the performance of the various experimental groups to the control group performance, that motion produced higher transfer effectiveness in the simulator and that washout motion was best in terms of skill gain rate. However, an examination of the uptake rate resulting under the various conditions of motion and no motion, when adjusted for the respective costs of motion and non-moving simulators produced a surprising conclusion. For this particular set of flight skills, the most cost effective strategy for training is to utilize the non-moving simulator for a longer period of time to reach exit criteria rather than to achieve it faster in the washout motion condition.

The comparison of the washout motion to the random washout condition was not definitive for every dimension of performance measured, but generally it indicated that motion cues provide a reliable alert to a need to take a control action, but that the students cannot utilize the magnitude or direction of the motion perception to decide what should be done. This is consistent with what every flight instructor attempts to teach – trust the instruments, not you own senses.

These motion studies had an impact in the simulation industry – some of us became rather widely known – or perhaps infamous would be a better choice of words – because of them. Today, many general aviation training devices do not move but deliver cost effective flight skills as compared with training exclusively in the aircraft. Further, they enable the practice of certain types of tasks that would be too risky or expensive for initial training in flight.

Motion has a place in flight simulation, but it is application specific and the design of a simulator must take into account for what it will be used to develop the best cueing environment for the device. Since leaving the University, I have gone on to develop many hundreds of simulators, some with motion systems, some without. For those in which motion is the right choice, I say "Simulator Motion Rocks!" In other cases, we don't need it.

Thank you for listening, and thank you Stan.

STAN AND THE MOON ILLUSION:
Drilling Down at One End of the Human Systems Integration Elephant

Hector M. Acosta, Ph.D, CHFP
711th Human Performance Wing, Human Performance Integration
Air Force Research Laboratory, Brooks City-Base, San Antonio, TX

Dr. Stanley Nebuchadnezzar Roscoe invested decades researching a mystery born of observations separated by over two millennia. Characteristically, he addressed this peculiar problem set based on its Total System Performance implications, inspiring students and colleagues along the way. It was during his dissertation research that Stan found that pilots attempting landings using periscopic displays needed about 20% magnification to avoid landing long and hot. He did not know "the why" of his observation or its pragmatic solution. To Stan's great frustration, nothing in the existing literature explained the phenomena. Years later, Stan and his students hit upon a mechanism that might have mediated similar perceptual errors and interventions--including no less than the classic conundrum, the Moon Illusion. Much research followed. This presentation is about a few of his many long term influences and my work with him as his last doctoral student and current keeper of the flame.

I currently work as a consultant promoting the processes of Human Systems Integration (HSI) in requirements definition, development, acquisition, and sustainment in the U.S. Air Force. The goal of HSI is to optimize Total System Performance while minimizing Life-cycle System costs. I am very comfortable in this role, owing much of that comfort to the abiding influences of Dr. Stanley N. Roscoe.

When I first met Stan, he was already well-established as one of the great names in Aviation Psychology. My first graduate seminar with him introduced me to a man with an easy, disciplined comfort at applying the scientific method to solving practical problems affecting human performance. He did this while almost casually (often grinning like a pirate) grappling with the complexities of Total Systems. Due in no small part to his influence, I have never, ever, been able to approach a topic of research without trying to take a total systems view, from the operator out and from the operationally relevant environment in.

Stan's view of aviation psychology's role was both simple and hugely inclusive. In his seminal work, *Aviation Psychology* (1980), he defines the role of applied psychologists as behavioral engineers and offers a variation of a whole systems view of the aviation research domain in his Exhibit 1.1 of that publication. An adaptation of that view is presented at Figure 1.

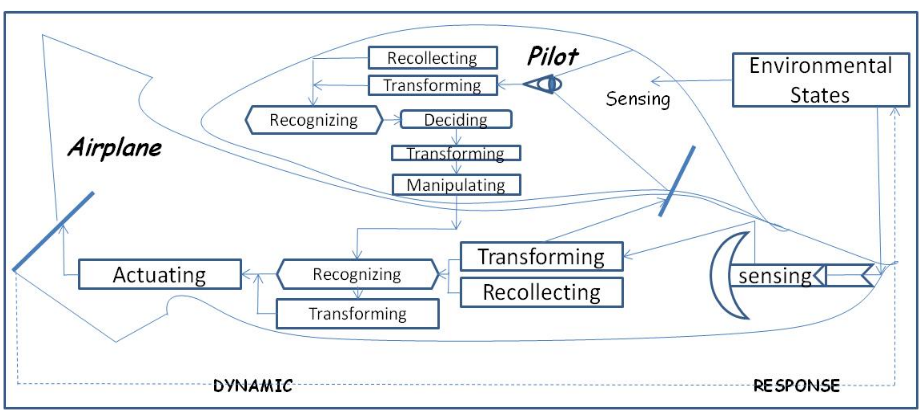


Figure 1. Adaptation of Stan's functional model of a pilot-airplane system (Roscoe, 1980, p. 4)

In a paper presented to Air University, you find his influence in the model representing my simple conceptual framework for developing the U.S. Air Force's first computer-based aircrew selection and classification Basic Attributes Testing system, Figure 2 (Acosta, 1985). The model served well and the system was fielded. Later, when

Distribution Statement A: Approved for public release; distribution is unlimited. 311 HSW/PA No. 09-164, 21 Apr 2009 working with Stan and his long-time research and engineering associate, Mr. Louis Corl, in preparation for my adventures in the worlds of visual accommodation, size perception and the Moon Illusion, I again reflected his influence in yet another conceptual model (Acosta, 1997), Figure 3.

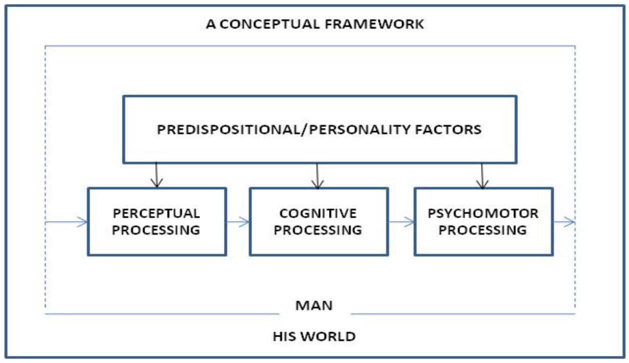


Figure 2. Stan's Total System emphasis reflected in Acosta's1985 conceptual model as a basis for the original Basic Attributes Testing system.

The emphasis in this later model was on sensation and perception and their critical translation into information driving goal-oriented behavior. The Total System emphasis here was the expectation that perception, veridical or not, is very much affected by the integration of inputs from multiple components in the human visual system. Stan and I were to have many discussions about powerful *perceptual illusions* being the product of sensory perceptual systems operating correctly, but under specifiable conditions effecting data-driven errors, an instantiation of the ubiquitous: *garbage in, garbage out*. The challenge to the Human Factors Engineer then is to understand human perceptual processes well enough to intervene.

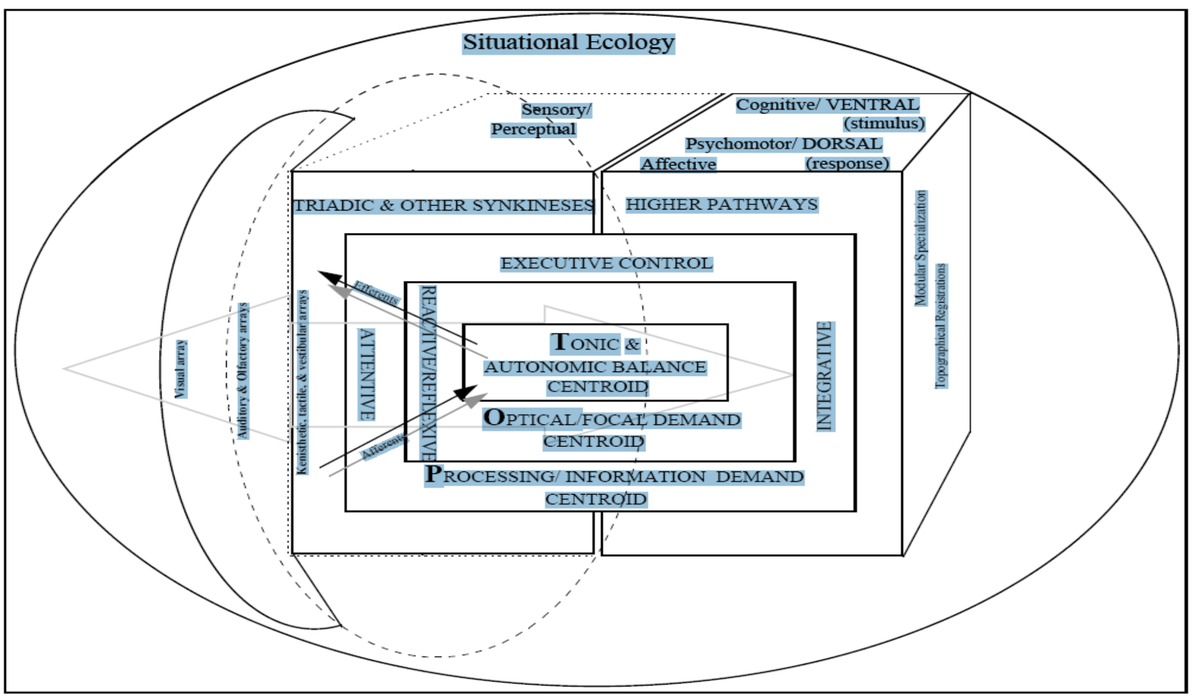


Figure 3. Stan's Total System influence in a maturing Oculomotor Perceptual model (Acosta, 1997, p. 81).

The Moon Illusion

A minimal definition of the moon illusion might state that it consists of the misperception of the size of the moon (also the sun and constellations) that varies depending on its elevation relative to the earth's natural horizon. The change is illusory, easily demonstrated by projection and measurement or using photography. It is perceived to be largest near the horizon and to diminish in apparent size as it rises toward the zenith. There is reasonable debate as to whether the primary locus of the illusion is a too large moon on the horizon or a too small moon overhead.

For millennia, the mystery of the moon illusion has defied explanation. As summarized by Plug and Ross (1989):

- **Aristotle** (384-322 BC) proposed that distance and air density cause a mirror or lens effect (refraction theory), so that the rising and setting moon, sun, and constellations appear larger than overhead.
- **Ptolemy** (ca. 142 AD) also suggested an atmospheric refraction effect. He further noted that observation on the horizon is "the usual, normal and, therefore, a more correct, condition of vision" while overhead viewing is unusual and difficult resulting in erroneous viewing (angle of regard theory).
- **Ibn al-Haytham** (11th century) proposed that the size of an object is judged by combining its visual angle with its known distance. Distance can only be judged accurately when **an uninterrupted sequence of intervening bodies (a texture gradient)** is present (Intervening Objects Theory).
- **Greaves** (ca. 1638), **Castelli** (ca. 1630) both refuted variations of the refraction theory. Greaves, impressed by apparent differences between size in Egypt versus England, measured the real angular size of the sun at various elevations and found no change. Castelli did the same for constellations.
- **Gassendi** (1636-1642) and **Anonymous** (possibly Bourdelot, ca. 1672) proposed some kind of oculomotor mediation of the phenomenon. Gassendi proposed a physiological optics-based enlargement of the retinal image for the horizon versus the elevated moon. His hypothesis was that there was less brightness near the horizon causing an enlarged pupil and with it an enlarged percept of size. Bourdelot is thought to have attempted to explain the dilation effect by saying that it caused a flattening of the lens and a simultaneous lengthening of the projection distance (lens to retina). While this position was widely discounted by the mid-18th century, it awaited Young to disprove the basic mechanics proposed.
- **Berkeley** (1709) held that both size and distance were judged from various learned cues. Among these were **aerial perspective** (increasing faintness and loss of color contrast with distance). Berkeley proposed that aerial perspective, variable under different atmospheric conditions, was *the* dominant determinant for the enlarged appearance on the horizon. Expanding on his *learned cue* theme, Berkeley proposed reduced size constancy with elevated **angle of regard.** Across a large variety of experiments, culminating over two and half centuries later with the work of Kaufman and Rock (1962), the effects of angle of regard seem to: 1. confirm an optimization of the human perceptual system to an *upright straight ahead* angle of regard; 2. result in a very small, relatively insignificant, degradation in size constancy accuracy with departures from this orientation (i.e., not nearly enough to account for the illusion); and 3. support the conclusion that the presence and angular distance from a textured gradient beneath the moon was the dominant stimulus condition associated with the illusion.

None of the above, however, provided a satisfactory candidate mechanism ("process") for generating the illusion.

Tonic Focus and Visual Accommodation

Visual accommodation is the oculomotor mechanism involving the eye's flexible lens that permits normal human eyes to adjust focus over a range of up to about 15 diopters in the very young, approaching zero diopters of dynamic range for individuals above the age of 50. The majority of this optical power is necessary to focus on objects near to us, converging light to form useful images on our retinas. Because of the inverse relationship between units of optical power and distances in the real world, only one diopter separates the degree of convergence of light needed to focus the details of an object at 1 meter and the power to converge parallel rays from objects at infinity. So, only about 1/7 of a diopter change is needed for precise refocus, for example, from an object at 6 meters to an object at 60 meters. All pretty straightforward.

The human visual system is, however, much more complex than a simple optical system. It is in this complexity that Stan and his students found measureable results that suggested at least one mechanism for mediating the moon illusion and Stan's periscopic dilemma. As detailed by Roscoe (2008) and Acosta (2004), a series of studies in the literature and investigations by Stan, his students and colleagues found that:

1. We rarely focus accurately or steadily
2. There are tonic biases of visual focus and convergence and that humans tend to regress to these tonic states

whenever focal demand (including temporal variation) is low
3. Imbalanced parasympathetic and sympathetic innervations are involved
4. Neurophysiological evidence points to cascading perception via a Dorsal, fast perceptual processing pathway (magnocellular flow) and a Ventral, slower pathway (both magno and parvocellular)
5. There are a variety of conditions, typically found in contact, contact-analog and imaging displays that are likely to effect changes in both judgments of angular size and measured distance of accommodation.

These findings were consistent with the notion that visual accommodation was a candidate mechanism in mediating perceptual errors like Stan's periscopic micropsia and, possibly, the moon illusion. His conclusions included the hypothesis that tonic or resting accommodation "competes" with the stimulus quality of the object of interest in the visual scene (i.e., its focal demand) so that our distance of focus is always the competitive product of these two. Leibowitz, Hennessey & Owens (1975) had demonstrated that: 1) under conditions of poor illumination or empty field (low texture gradient) conditions the eye tended to lapse toward each individual's tonic focus; and 2) that this tonic focus distance is reasonably stable and readily measureable. Sample distributions of measured tonic focus were roughly normal with a mean at about one meter.

Stan's fundamental hypothesis was that the moon appears larger when it is nearer the horizon because focal demand is higher and our distance of focus is more likely to be correctly distant. Alternatively, micropsia, requiring magnification to correct, results from poor focal demand, as experienced when the moon is separated from the texture gradient of the distant horizon and the resulting lapse toward tonic focus.

A series of univariate experiments had demonstrated that either focal demand and/or perceived size could be manipulated using selected independent variables to include: Cue availability (Holway and Boring, 1941; Iavecchia, Iavecchia, and Roscoe, 1983); Stimulus size (Gilinsky, 1954); Cue locations (Kaufman and Rock, 1962); Stimulus contrast (Hamilton, 1964); Stimulus quality (Roscoe, Olzak, and Randle, 1976; Simonelli, 1979); Mandelbaum effects (Benel, 1979); Optical distance (Enright, 1989; and Roscoe, Corl, and Couchman, 1994).

A Multivariate Psychophysical Experiment

The stage had been set for Stan to convince me to choose "something simple" to address for my long-delayed dissertation research. THE NEXT TEN YEARS resolved themselves into a multivariate psychophysical experiment that manipulated variables relevant to operational display design while measuring both distance of focus (accommodation) and perceived size. Based on my requirements, Lou Corl, Stan's generous and brilliant partner in Illiana Aviation Sciences, Ltd., designed and built a testing system we dubbed TOBE (transportable optical bench, experimental). The TOBE system supported a good number of functions to include: the binocular projection of an artificial moon onto a bounded view of a real New Mexico vista, and measurement of line of sight comparative moon size, pupil size, ambient light levels, and distance of focus using a computer-controlled polarized vernier optometer. The system supported all the frames and mounting devices needed to conduct the overall experiment.

The independent variables selected included four "target" variables and four "scene" variables. A fractional factorial experimental design permitted the manipulation of 3 levels of all 8 variables, as summarized at Table 1.

Table 1. *Independent variables and levels manipulated in dissertation main experiment (Acosta, 2004).*

Target Variables		Scene Variables	
TSZ: Moon Size (degrees):	0.25, 0.5, 2.0	**VSR**: Filter (optical density):	1.7, 2.8, 4.0
TDS: Optical Distance (diopters):	0.167, 0, -.0167	**FOV**: Field of View: IPD ±	5, 10, 22.5 degrees
TLM: Luminance (candellas/m2):	3.8k, 0.8k, 0.2k	**MND**: Screen Level:	Clear, Grid, Screen
TFD· Focal Demand Contrast:	0%, 30%, 90 %	**EYP**· Tilt (degrees):	-10, 0, 10

Findings

Stan's continuous insistence on precision and attention to total system detail, not to mention a legacy of solid research, paid off. The experiment demonstrated reliable effects of all independent variables on both perceived size and distance of focus, most strikingly represented in the robust mutual linear effect of optical density filters on both primary dependent variables (Figure 4). Solid linear effects on accommodation where present despite the fact that projected moon distances where all within a range of 1/3 diopter and all at or beyond 6 meters, the clinical standard for approximating optical infinity.

There were also reliable interactions among all of the independent variables and an inescapable observation that visual accommodation could only account for part of the variable size perceptual phenomenon as reflected in the spread among bivariate means relating distance of focus to measured relative moon size (Figure 5). The perceptual drivers for the moon illusion and the micropsia of Stan's periscopic display must include oculomotor components, but, clearly include higher visual processing components. That tangled story is left to a future detailed presentation.

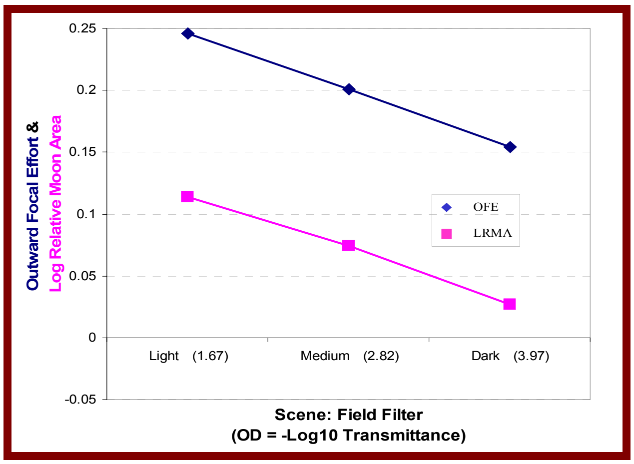


Figure 4. Effects of full field optical density filters on both focal distance and relative moon size judgments.

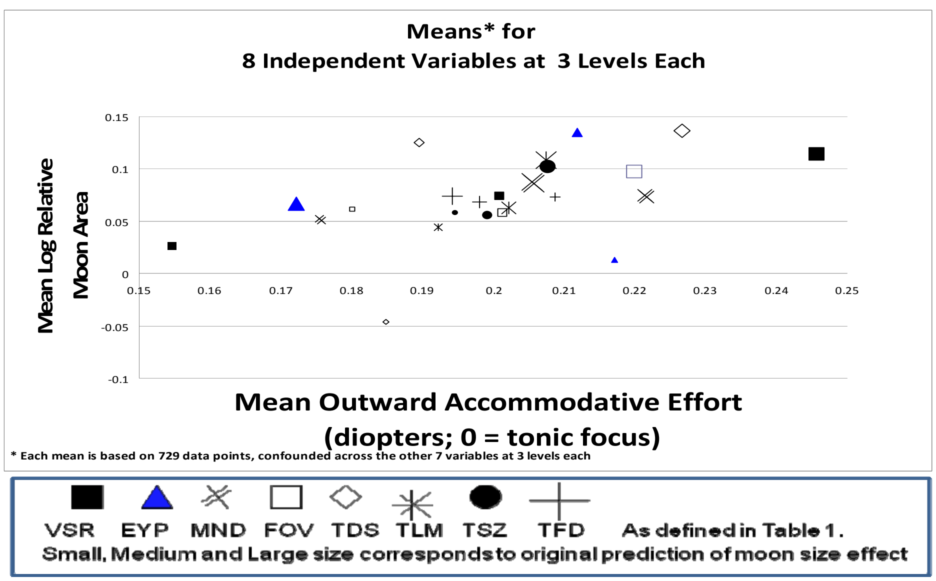


Figure 5. Bivariate plot of measured relative moon size as a function of outward focus (Acosta, 2004).

Conclusion

Stan saw his 87th year satisfied that his work relating visual accommodation to the perception of the size of distant objects had been well done and that our latest effort had pulled much of his prior work into a cohesive and comprehensible framework. To the very end he was excited about the next great idea and insistent that we needed to get on with the next important bit of research and that—with that twinkle in his eye—the answers were out there—ripe for the picking. That was Stan.

References

Acosta, H. M. (2004). *Eight factors affecting focal distance and the moon illusion.* Unpublished doctoral dissertation, New Mexico State University, Department of Psychology.

Acosta, H. M. (1997). *Multifactor determinants of visual accommodation as a critical intervening variable in the perception of size and distance: Phase I report* (Tech. Report AL/HR-TR-1996-0137). Brooks AFB, TX: Armstrong Laboratory.

Acosta, H. M. (1985). *Development of the Basic Attributes Testing System (BATS): Computer-based aircrew selection and classification research* (Report 85-0015). Maxwell AFB, AL: Air University, Air Command Staff College

Benel, R. A., and Benel, D. C. R. (1979). Accommodation in untextured stimulus fields (Eng Psy-79-1/AFOSR-79-3). Champaign: University of Illinois at Urbana-Champaign, Department of Psychology.

Enright, J. T. (1989). The eye, the brain, and the size of the moon: Toward a unified oculomotor hypothesis for the moon illusion. In M. Hershenson (Ed.), *The moon illusion* (pp. 59-121). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gilinsky, A. S. (1954). Perception of size of objects at various distances (Tech. Report 54-92). Lackland AFB, TX: Air Force Personnel and Training Research Center.

Hamilton, J. E. (1966). Luminance and the moon illusion. *American Journal of Optometry*, 43, 593-604.

Holway, A. H., and Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *American Journal of Psychology*, 54, 21-37.

Iavecchia, J. H., Iavecchia, H. P., and Roscoe, S. N. (1983). The moon illusion revisited. *Aviation, Space, and Environmental Medicine*, 54, 39-46.

Kaufman, L., and Rock, I. (1962). The moon illusion. *Scientific American*, 207, 120-136.

Leibowitz, H. W., Hennessy, R. T., and Owens, D. A. (1975). The intermediate resting position of accommodation and some implications for space perception. *Psychologia*, *18*, 162-170.

Plug, C., and Ross, H. E. (1989). Historical review. In M. Hershenson (Ed.), *The moon illusion* (pp. 5-27). Hillsdale, NJ: Lawrence Erlbaum.

Roscoe, S.N. (2008). A flight by periscope and where it landed, *Human Factors***,** Vol. 50, No. 3, June 2008, pp. 361-367.

Roscoe, S.N. (1980). *Aviation Psychology*. Ames, Iowa: The Iowa State University Press.

Roscoe, S. N., Corl, L., and Couchman, D. H. (1994). Eye accommodation: Use of vernier optometers in behavioral research. *International Journal of Aviation Psychology*, 4, 359-371.

Roscoe, S. N., Olzak, L. A., and Randle, R. J. (1976). Ground referenced visual orientation with imaging displays: Monocular versus binocular accommodation and judgments of relative size. In Proceedings of the AGARD Aerospace Medical Panel Specialists Meeting, Athens, Greece (A6.1-A6.9). Neuilly-sur-Seine, France: North Atlantic Treaty Organization.

Simonelli, N. M. (1979). The dark focus of visual accommodation: Its existence, its measurement, its effects (Tech. Report BEL 79-3/AFOSR-79-7). Las Cruces, NM: New Mexico State University,Behavioral Engineering Laboratory.

HUMAN FACTORS IN THE GROUND-SUPPORT OF SMALL UNMANNED AIRCRAFT SYSTEMS

Alan Hobbs, Ph.D.*
San Jose State University Foundation, NASA Ames Research Center, Moffett Field, CA.

Stanley R. Herwitz, Ph.D.
UAV Collaborative, NASA Research Park, Moffett Field, CA.

Glen R. Gallaway, M.S.
Federal Aviation Administration, Washington, DC.

A growing body of research has been directed at the human factors of Unmanned Aircraft System (UAS) flight operations, yet up to now, virtually no attention has been given to the human factors of UAS maintenance. The aim of the current research program was to identify the challenges facing the maintainers of small unmanned aircraft systems. Unlike their counterparts in conventional aviation, UAS maintenance technicians are responsible for the functioning of an entire system, comprising airborne and ground-based components. Challenges include absent or poor maintenance documentation, the need to make frequent decisions about salvaging components, difficulties in troubleshooting software problems, the maintenance of radio control model aircraft components, and the potential unfamiliarity of UAS maintenance personnel with the culture and practices of the aviation industry. A "dirty dozen" list of UAS human factors is proposed.

Unmanned aircraft range from small inexpensive, hand-launched micro air vehicles such as micro-electric helicopters to large, high-altitude-long-endurance vehicles such as the Global Hawk. In between these extremes are a vast array of vehicles and systems. As well as military applications, unmanned aircraft systems (UAS) have many potential non-military uses, including law enforcement, firefighting, traffic monitoring, aerial photography, agriculture, search and rescue, border surveillance, wildlife monitoring, power-line inspection, minerals exploration and homeland security activities. At present, concerns about collision avoidance are holding back the operation of unmanned aircraft in civilian airspace (Flight Safety Foundation, 2005). Assuming that this issue can be resolved, small, inexpensive unmanned aircraft may become a common sight.

The most rapid growth in the emerging civil UAS sector may occur with small systems, defined here as those in which the aircraft weighs less than 100 lbs. Technological developments, such as miniaturization of sensor equipment and autopilot systems, and developments in battery technology, are allowing small unmanned aircraft to perform tasks that would have previously required large, expensive aircraft. Large unmanned systems are generally maintained by specialist maintenance technicians. However, small commercial UAS are frequently operated by generalist teams of multi-skilled individuals who perform all ground tasks including assembly, flight preparation, in-flight operation, and maintenance. Throughout this paper, the terms "maintenance personnel" or "maintainer" are used to refer to anyone who maintains a UAS, even though the individual also may perform other roles as a member of the UAS operating team.

The nascent UAS industry has an accident rate significantly greater than that of conventional aviation (Williams, 2004) and human factors are emerging as major challenges to be resolved (McCarley & Wickens, 2005; Cooke, Pringle, Pedersen & Connor, 2006). If unmanned aircraft are to be permitted to share civilian airspace with conventional aircraft, it will be necessary to understand the human factors associated with these vehicles. Rather than eliminating the potential for human error, the removal of the on-board pilot may transfer some of the risk of human error to personnel on the ground, including maintenance technicians. Furthermore, tele-operated transport systems such as unmanned aircraft may be especially vulnerable to maintenance error due to the absence of an on-site operator able to respond rapidly to an anomalous situation.

A large amount of information has been published on human factors of airline maintenance, much of it based on FAA-sponsored research (Johnson, 2006). Issues such as stress, distraction, and poor access for maintainability are now widely identified as hazards in conventional aircraft maintenance. While recognizing that these issues also apply to UAS maintenance, this research was focused on issues that uniquely affect UAS maintenance.

The research approach

The objective of the research program described in this paper was to identify the human challenges in maintaining small UAS. For the current purposes maintenance was defined as any activity performed on the ground before or after flight to ensure the successful and safe operation of the system. This definition covers a wide range of ground-support activities including assembly, fuelling, updates to software, and pre-flight testing. As this was an area that had not been examined previously, the research involved the gathering of qualitative information that would enable broad issues to be identified. Three approaches were used to gather data. First, a series of site visits were made to UAS maintenance or manufacturing facilities, and UAS flight operations were observed. Second, structured interviews were conducted with UAS maintenance personnel. These interviews focused on the qualifications and skills of maintenance personnel and the challenges they face in the course of their work. Details of this stage of data collection can be found at Hobbs and Herwitz (2006) and Herwitz and Hobbs (2006). In a second round of interviews, questions focused on the specific tasks performed on UAS components, including ground systems such as computers. A summary of the results can be found in Hobbs and Herwitz, (2009). The sections below outline some of the key differences between the job of UAS maintainer and that of a conventional aircraft mechanic.

Emerging human factors in UAS maintenance

The task of maintaining an unmanned aircraft system. A significant difference between UAS maintenance and conventional aircraft maintenance is that the UAS maintainer is responsible for a complete system, comprising the aircraft, a diverse set of ground-based equipment, and the links between these elements, (see Figure 1). While the aircraft may be the most obvious element of the system, the ground based elements also require attention and maintenance.

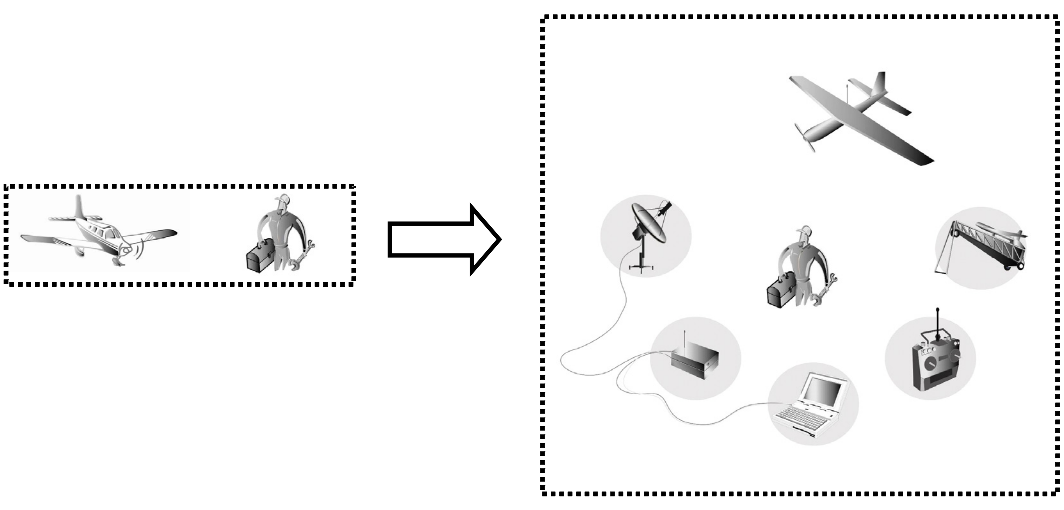


Figure 1. In conventional aviation, the aircraft maintenance technician is responsible for the airworthiness of an aircraft, whereas the UAS maintenance technician is responsible for a system comprised of diverse elements, including the aircraft, radio transmission equipment, modems, computers, and in some cases, handheld controllers and launch/recovery equipment.

Key differences between a UAS and a conventional aircraft are:

- Commercial "off the shelf" desktop or laptop computers are likely to be part of flight system.
- System elements frequently assembled and disassembled between flights.
- Modular construction facilitates repair by replacement and shipping of components to specialist repair facilities.
- Unmanned flight is not possible without functioning avionics and/or communication equipment.
- Some UAS components were originally intended for radio control model aircraft, and have limited reliability data.
- Payload is more likely to interfere with operation of aircraft, e.g., through electromagnetic interference.

Shift of risk. The introduction of unmanned aviation shifts the balance of risk in ways that must be understood by maintenance personnel. In conventional aviation, the safety risks associated with flight are in large part borne by the people who receive the benefit of flight, i.e., flight crew and passengers. Sometimes referred to as "shared fate," a threat to the safety of a conventional aircraft is also a threat to the occupants of the aircraft.

In unmanned aviation, the beneficiaries of the flight remain on the ground, and the safety risks are borne largely by non-involved individuals -- occupants of conventional aircraft, people under the flight path of the aircraft, and property owners. With no on-board lives at risk, the maintenance person is not necessarily conducting maintenance for the safety of specific identifiable individuals, but for the safety of the community as a whole. The public tends to demand especially high safety standards for technologies that are new, are not well understood, and where exposure to risk is involuntary (Slovic, 2000). For these reasons, there may be a low public tolerance of incidents involving unmanned aircraft, even when the consequences are limited to property damage.

Diverse skill and knowledge requirements. The UAS maintainer, whether a specialist or generalist, requires a skill set beyond the traditional skill and knowledge requirements of aviation airframe and powerplant mechanics. In addition to the maintenance of an engine and airframe, a UAS technician can be expected to interact with computer systems, micro autopilots, radio communication equipment, modems, and, in some cases, satellite phones. Ensuring the data link between the ground control station and the aircraft takes on a level of criticality not present in conventional aviation because the loss of communication is more likely to result in the loss of the aircraft.

Lack of direct feedback on aircraft performance. In conventional aviation, the on-board pilot has a direct experience of aircraft performance via the handling qualities of the aircraft, as well as sounds, vibrations, and even smells. With no on-board pilot, UAS maintenance personnel lack a key source of information about aircraft performance. To some extent, automated in-flight monitoring provides an alternative source of detailed information. However automated monitoring systems can at times provide an overwhelming volume of precise data with relatively little consolidated information.

Maintenance and fault diagnosis of IT systems. For most small UAS, the "cockpit on the ground" is a standard laptop or desktop computer exposed to the hazards of outdoor operations such as moisture, dust and temperature extremes. Computer system administration tasks now take on flight safety importance because system failures, such as screen lockups or software slowdowns that would be minor irritations in an office environment, can present significant hazards if they occur during a flight (Hobbs & Herwitz, 2008).

Fault diagnosis in software-based systems can be significantly more difficult than with electromechanical systems. Mysterious, ill-defined faults such as computer slowdowns, screen freezes, or radio frequency interference are sometimes resolved without the UAS technician understanding why the fault occurred, and whether their actions corrected the underlying problem or merely removed the symptoms. System re-boots are common responses to computer problems as illustrated in the following incident report (Hobbs & Herwitz, 2008).

"The desktop computer, which was serving as the ground control system, locked up while the unmanned aircraft was in flight. The PC-based computer was housed in the ground control station trailer. The only alternative was to re-boot the computer, and this took about two to three minutes before command-and-control was reestablished. The unmanned aircraft's flight path, however, was already uploaded so there was no effect on the flight sequence."

Model aircraft culture. The personnel who maintain small UAS tend to have a background in radio-controlled model aircraft or engineering, and relatively few have experience in commercial aviation maintenance. These personnel may possess attitudes to risk that are significantly different to those held by qualified aircraft maintenance technicians. For example, they may be accustomed to operating without formal procedures or checklists, and may be unfamiliar with the ethics and standard practices of aircraft maintenance.

Task performance in the absence of documentation. Document design has been identified as a critical performance shaping factor in conventional aviation maintenance (Drury, Sarac, & Driscoll, 1997). Small UA generally have rudimentary flight manuals, however many are delivered without maintenance documentation. Users generally develop their own maintenance checklists and procedures to guide routine tasks such as system assembly, and scheduled pre-flight checks. However, for troubleshooting and corrective maintenance, maintainers may have no choice but to rely on "knowledge in the head" or "trial and error".

Salvage decisions. Compared to conventional aircraft, small unmanned aircraft are more likely to experience damage caused by events such as hard landings, contact with water, or landing in trees. Unmanned aircraft also tend to be less waterproof than conventional aircraft leading to a greater chance of water damage to internal components. To a greater extent than in conventional aviation, UAS maintenance personnel will be required to make judgments about the salvage, testing and re-use of components from damaged UA. In the case of modular aircraft designs, an apparently undamaged modular unit may have an unseen defect.

Repetitive assembly and handling. In contrast to conventional aircraft, most small unmanned systems are designed to be reassembled and disassembled before and after each flight, necessitating the frequent connection and disconnection of electrical, fuel and data systems. The probability of an error during a single connection task may be relatively low, in the order of 0.001 (Kirwan, 1994). However UAS maintenance personnel are exposed to this risk on a regular basis, and consequently the chance of an assembly error or maintenance-induced damage may become significant over the course of months or years. The following example illustrates an assembly error involving a small hand-launched unmanned aircraft:

"After departure the unmanned aircraft performed unusually slow rates of turn to the right and tight turns to the left and struggled to track as designated by the operator. Approximately seven minutes into the flight, the outboard section of the right wing separated from the centre wing section. The aircraft immediately entered a rapid clockwise spiral before impacting the ground. The most likely explanation for the crash was that the outboard section of the right wing was incorrectly attached during pre-flight assembly and from launch it flew with difficulty until the wing section eventually separated." (Hobbs & Herwitz, 2008).

Risk associated with maintenance or disturbance of ground equipment while missions are underway. The cockpit of a conventional aircraft is beyond the reach of maintenance personnel once the aircraft is in flight. In contrast, the ground station of a UAS is always accessible to maintenance personnel on the ground. They may be required to perform corrective maintenance while a flight is underway, or may carry out other actions that could potentially impact system performance. For example, an in-flight problem may require troubleshooting of ground equipment, the checking of cables, or a re-start of the ground control computer. A maintenance technician interacting with a live system requires a clear understanding of the operational implications of the planned intervention. The technician must also consider the potential effects of errors, whether mistakes such as misdiagnosing a fault, or simple slips such as tripping over a cable. Even a brief interruption to a computer's power supply can have an extended impact if it leads to a slow re-boot sequence.

Conclusion

Technological developments have increased the capabilities of unmanned aircraft systems to the point where they can now potentially serve a large range of non-military purposes. Despite the absence of an on-board pilot, human factors are emerging as key issues in this sector. As automation decreases the role of humans as direct physical controllers of unmanned aircraft, it is possible that maintenance and other ground-support activities will become increasingly important.

The maintenance of unmanned aircraft systems introduces a new set of human factors in addition to those that apply in conventional aviation maintenance. The "Dirty Dozen" list has been widely used to educate airline maintenance technicians about human factors (Dupont, 1997). Table 1 contains a proposed "UAS Dirty Dozen" intended to raise awareness of the emerging maintenance human factors in small UAS operations. Each of the 12 issues is illustrated with an example of a dangerous attitude or situation. This list will be updated as more is learned about this topic.

Table 1: ***A "Dirty Dozen" for small UAS maintenance.***

Issue	*Example*
1. Mysterious software faults	*I don't know why the software did that. I'll just re-boot it. I'll just swap the card.*
2. Lack of checklists for routine tasks	*I don't need a checklist, I do this procedure all the time.*
3. Assembly and handling	*I've assembled this system hundreds of times.*
4. Laptop maintenance	*Need to check your email? Use the ground control laptop.*
5. Awareness of risk to public	*No-one's life is at stake here.*
6. Salvage decisions	*We can re-use that component, it doesn't look damaged.*
7. Payload interference with aircraft	*This is just a small change to the payload*
8. End-to-end connectivity	*All the individual components are working, I guess it will work when we connect it all up.*
9. Disclosure and sharing of information	*I don't want my competitors to know about this problem. I don't want the FAA to find out.*
10. Trial and error repair and troubleshooting	*Not sure how this goes back, but that looks right.*
11. Frequency management	*No one else seems to be using this frequency.*
12. Disturbance of ground equipment during flight	*Let's move the ground control computer into the shade.*

Confidential reporting systems such as NASA's Aviation Safety Reporting System (ASRS) have been valuable sources of human factors information in conventional aviation. The emerging UAS industry, where safety issues are least understood, lacks a confidential incident reporting system. Any future UAS reporting system must include maintenance personnel. In the course of discussions with UAS operators, it became apparent that concerns about commercial confidentiality and FAA enforcement action are currently suppressing the open disclosure of incidents, which in turn may make it difficult for the UAS industry to learn from experience.

Acknowledgements

This research was funded by the FAA under inter-agency agreement DTFA01-01-X-02045 between the Federal Aviation Administration (FAA) Human Factors Research and Engineering Division and the National Aeronautics and Space Administration (NASA) Aeronautical Safety and Human Factors. We would like to thank the many UAS operators and maintainers who agreed to be interviewed. We also acknowledge the assistance provided by Maj. Keirin Joyce of the Australian Defence Force.

References

Cooke, N. J., Pringle, H.L. Pedersen, H.K & Connor, O. (Eds.). (2006). Human factors of remotely operated vehicles. San Diego: Elsevier.

Drury, C., Sarac, A, & Driscoll, D. (1997). Documentation design aid development. Phase 4 progress report. Washington, DC: Federal Aviation Administration. (http://hfskyway.faa.gov).

Dupont, G. (1997). The Dirty Dozen Errors in Maintenance. Proceedings of Eleventh Federal Aviation Administration Meeting on Human Factors Issues in Aircraft Maintenance and Inspection. Washington, DC: Federal Aviation Administration/Office of Aviation Medicine. (http://hfskyway.faa.gov)

Flight Safety Foundation. (2005, May). See what's sharing your airspace. Flight Safety Digest, 24, (5), 1–26.

Herwitz, S., & Hobbs, A. (2006). A Review of the Resources and Facilities Required to Conduct Maintenance of Unmanned Aircraft Systems. Report to Federal Aviation Administration under inter-agency agreement DTFA01-01-X-02045. Moffett Field, CA: NASA Ames Research Center.

Hobbs, A., & Herwitz, S. (2006). Human challenges in the maintenance of Unmanned Aircraft Systems. Report to Federal Aviation Administration under inter-agency agreement DTFA01-01-X-02045. Moffett Field, CA: NASA Ames Research Center.

Hobbs, A., & Herwitz, S. (2008). Maintenance challenges of small unmanned aircraft systems. - A human factors perspective. Final report to Federal Aviation Administration under inter-agency agreement DTFA01-01-X-02045. Moffett Field, CA: NASA Ames Research Center.

Johnson, W. B. (2006, September). Keeping it real with maintenance human factors at FAA. Overhaul and Maintenance, 26.

Kirwan, B. (1994). A practical guide to human reliability assessment. London: Taylor & Francis.

McCarley, J. S., & Wickens, C. D. (2005). Human factors implications of UAVs in the national airspace. Technical Report AHFD-05-05/FAA-05-01. Atlantic City, NJ: Federal Aviation Administration.

Slovic, P. (2000). The perception of risk. London: Earthscan.

Williams, K.W. (2004). A summary of unmanned aircraft accident/incident data: human factors implications. Technical Report No. DOT/FAA/AM–04/24. Washington, DC: U.S. Department of Transportation, Federal Aviation Administration, Office of Aerospace Medicine.

* alan.hobbs@nasa.gov

The Technological, Financial, and Social Realities That Are Defining the Aircraft Mechanic of Tomorrow

Glen R. Gallaway
Federal Aviation Administration
Washington, DC, USA

(Disclaimer: THE FINDINGS AND CONCLUSIONS IN THIS DOCUMENT HAVE NOT BEEN FORMALLY DISSEMINATED BY THE FEDERAL AVIATION ADMINISTRATION AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY FAA OR DEPARTMENT OF TRANSPORTATION VIEW, DETERMINATION, OR POLICY)

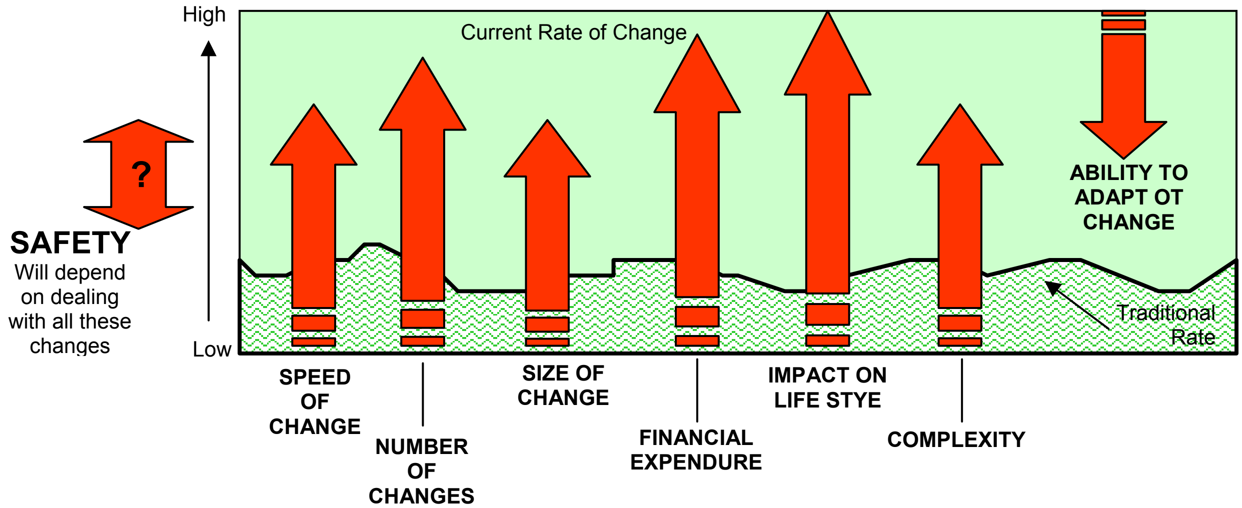
Today it is becoming increasingly difficult to describe maintenance roles because an enormous number of once stable factors affecting the maintenance person and process are changing. Technology changes like software based aircraft, air/ground/satellite/airport system integration, highly complexity systems, and other technology issues are not currently part of a maintenance person's normal skill set but are already part of aircraft maintenance needs.

Many of the change issues that are considered to be "TECHNOLOGY" initiated but usually are caused by changing financial and/or social requirements that has occurred. This is a circular result in that when technology changes occur it further drives financial and social changes. I feel that because this cycle has been going on for so long that people have accepted the spiral and failed to keep track of the state of the industry. All the changing factors must be identified and dealt with as a basis for redefining the role of "Aircraft Maintenance Person".

In the field of aircraft maintenance, major change has been occurring for years and much larger changes are on the horizon. Traditionally the changes occur in the technology aspect of aircraft maintenance and technology is definitely experiencing the most far reaching and complex changes in its history. This includes the extensive computerization of aircraft systems, major increases in the use of automation, support systems complexity, and the tightly coupling of air/ground/satellite/airport systems that is under develop as part of the NextGen (next generation) airspace initiative. [Note: This article will concentrate on computerization technologies in the maintenance task]. Accompanying these changes are some dark clouds of business issues that are making the maintenance function more of a challenge. These issues include (but are not limited to): In the labor arena there is purported to be a shortage of maintenance personnel; lack of interest in the profession because they can make more money in car dealerships; lack of interest in the work because of the inherently difficult working conditions; FAA training requirements for certification don't match the new technologies so maintenance people haven't developed the needed skills after completing authorized training; Younger people are not interested in this field partially due to the strict rule enforcement; The new technologies are driving skills and training requirements ever higher which is a cost and time issue for the potential maintenance person. Then there are substantial changes also taking place in the businesses involved with building aircraft and support systems; the airlines; and the maintenance / repair stations. These business issues include: The ever increasing cost of building, implementing, and operating new technology based aircraft; business viability due to competition, fuel cost, labor, and reduced demand. Because numerous foreign countries are able to perform aircraft maintenance at substantially lower cost than in the US, 70 % of our transport aircraft now have heavy maintenance performed off shore (reducing demand of US ma and introduces potential for problems that would be prevented in US repair centers. Add to this list of issues is the difficulty of the regulators to keep up with all these technical and sociotechnical issues and synthesize guidance for the future.

Because so many of the technical, labor, business, and regulatory sub-fields of aircraft maintenance have traditionally operated and been managed independently of each other, it has been difficult for any one organization to grasp how the whole maintenance process works. This also means that there has been minimal knowledge developed about how changes in one sub-field have impacted others and how to adjust for change. This may be one reason that the industry as a whole has not been aware that all the changes were occurring in the individual subfield areas and therefore has not assessed the collective major negative impacts the changes were having on the industry. A number of people saw the convergence of problems coming, knew that a broad infrastructure solution was needed, but were not in a position to address more than their piece of the problem. Worse yet, the changes that were occurring in these so called independent sub-fields often were having a ripple effect economically, personnel wise, and technology wise that were only identified after the ripple has affected the next sub-field.

And to put the impact of the current changes are having in perspective, it is important to note that in the past we might only have to deal with small/slow change over long periods of time, and now we are now dealing with accelerated changes in everything. Table 1 gives us a sense of the rate of and size of changes that are occurring with a result that maybe we are not able to adapt to these changes in a way that will ensure that me meet aircraft maintenance needs now and in the future. How well these changes are managed will also impact the safety of the aircraft and related systems that must be maintained.



Seven Measurement Scales of Change

1. SPEED of CHANGE – When change occurs faster that people can understand and adapt.
2. SIZE OF CHANGE – The larger the change the larger the increased potential for problems adapting and accepting change.
3. NUMBER OF CHANGES – As number that concurrently change, the more difficult they will be to address.
4. ABILITY TO ADAPT TO CHANGE – As other measures increase, the ability and resources to deal with the changes will decrease.
5. IMPACT ON LIFE STYLE– Changes outside one's believes, interest, acceptance.
6. COMPLEXITY – Complexity and technology that is outside one's ability to understand, knowledge, or learn.
7. FINANCIAL – Aviation costs continue to climb and can quickly skyrocket with market changes like fuel costs.

Figure 1 Safety Will Rise or Fall Depending On How Well the Changes Are Addressed

The Big Picture and Collective Impact

Because there are so many sub-fields that make up the actual process of aircraft maintenance, the industry has had difficulty determining the collective impact of the individual sub-field changes on the process. And because of the lack of a consolidated view, it has been hard to know what the problems are, determine their magnitude, know the extent of the problem, and what action should be taken. So it is obvious that a big picture view is necessary to deal with the changing aircraft maintenance field. This means that we must look across all the changes occurring in the maintenance field and find approaches that will direct it's future. Figure 2 shows some of the sub-field areas that need to be looked at and tied together to establish the big picture of maintenance sub-fields that need to be tied together to provide the state of the industry view.

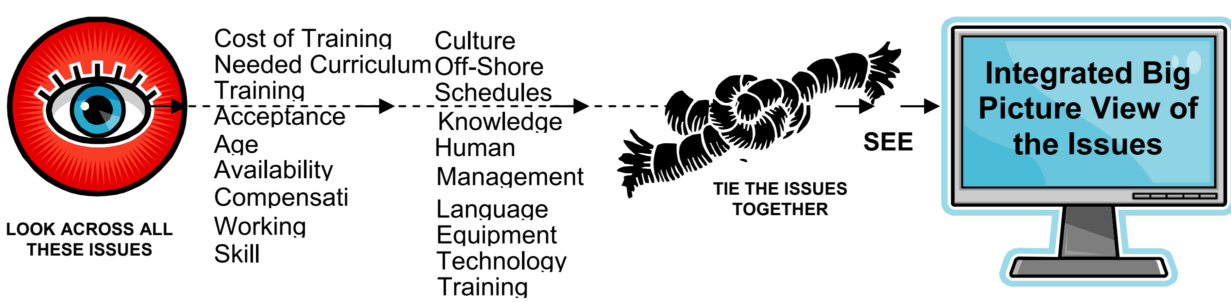


Figure 2. Integration of Issues to Create the State of the Industry View.

Major Changes Have Occurred in Maintenance Personnel's Job

By looking at a high level history of aircraft maintenance we can see how changes in aircraft roles, technology, financial issues, and personnel both caused or altered what was thought of as the process of aircraft maintenance. As we go through this history, notice that the changes that occur are usually increasing the complexity of the process, require increasingly more skilled maintenance personnel, were trying to reduce the amount of maintenance needed, and more recently trying to use automation to eliminate large portions of the high tech maintenance process.

In the good old days an aircraft mechanic's job was well defined. Everybody knew the job was those tasks that were performed by the people with the big wrench in there hand. When something needed fixin' they would have specific procedures to perform the repairs. Over the years the wrench would not fit on all the electric parts that began to show up on aircraft. Then came electronics and the wrench was only good for tapping on those new fangled radios and guidance boxes. Although tapping sometimes worked to fix lose connections, the mechanic had to develop new skills to work on these devices. So mechanics, electronic device vendors, and device repair shops, teamed up to service and repair these devices (an increasing level of complexity to repair). These electronic devices began the process of sending electric messages through the aircraft between devices and those messages could break and had to be fixed. To this point the changes did not drastically change the maintenance job. Then some more new contraptions called computers were being used to do tasks on aircraft. Now the wrench had to be kept away from these devices because tapping a computing device might damage it. What to do. The mechanical skills and the electronic skills that the mechanic had developed were not enough to service and repair computers. Until this point the work the mechanic had to do was usually physical things that they could see, touch, measure, and replace. Although the computer has the "see and touch" component of work they now have a few new wrinkles. They had programs (software) that performed functions, but they could not directly be observed. They had to be dealt with by indirect means through things called displays, keyboards and other test equipment that may also be software based. Service and repair of these computer devices where further confounded by stuff that the programs worked on called data. The data could also break and had to be serviced and repaired. These computer contraptions where so neat (and held great potential to that people began using them for everything. The designers found that they could add many functions in one device could have many computer devices work together and pass data around to be processed and perform many aircraft functions. There were many reasons why these ideas were so attractive including provided new capabilities and at perceived cheaper cost. On the down side, the users often had more difficulty using the devices, and servicing and repairing them was becoming an increasing challenge for the mechanic and the computer specialists that were now a part of the maintenance organizations. The mechanic's dilemma was that the mechanic did not have the skills and training to deal with computer devices. So shifting the work to the computer specialist currently seems to be an appropriate solution for this changing technology.

Computers now are being used as the basic tool to introduce automation to the aircraft from the cockpit to the maintenance process. From a users view, work is supposed to be easier with automation. This is also true for maintenance people when the automation is working but introduces increased problems when the automation fails and it has to be repaired. This emerging philosophy of "automate everything" brings use to today's dilemma of understanding what should, how should, when should, why should automation be employed and how do we best use the human maintainer in the maintenance process of the future.

Software and Automation in Maintenance Task - Make No Assumptions

The software industry has shown that automation technology is becoming increasingly more competent, but is it ready to take over large portions of the maintenance process for the next generation of aircraft and the next generation of the integrated ground/air/satellite/airport systems? Here are a few points that suggest that this may be at best a practical solution that should be cautiously applied.

There is a movement in the development of the next generation of aircraft and air/ground/satellite/airport integration systems maintenance to think of software and automation as the primary solution for improving effectiveness, reducing cost, and reducing direct human involvement (assumed to be a good thing for a number of reasons). A little evaluation of the viability of this concept will help us realize that we should not be in a hurry to eliminate the role of the human maintenance person.

Software Is Never Completely Tested – All Problems Are Never Identified

Because of the complexity and variation of the functions performed by software, data used, and users' interactions there is much of a software application's operation that never is evaluated/tested. This means that many of the problems and errors encountered in use have not been identified by the programmers and therefore have no error correction provided. The human is therefore the only resource available to resolve those issues and they are often the more obscure, difficult to understand types of errors.

Automation Will Only Be a Particle Solution

Some system developers think that the maintenance of aircraft computer systems (and linked devices) should be completely automated including the monitoring of operations and self correct any software or data problem and compensate for any hardware problem. These systems would also provide explicit instructions for the humans to intervene in the rare case that the computer cannot deal with the problem.

The automation solutions that are applied to aircraft systems must provide quick, accurate, and safe control and/or problem resolution. The following paraphrased statement suggests that the state-of the-art solutions of our advanced computer programs are helpful but not necessarily going to replace the aircraft maintenance person for a while. If current automation is going to depend on probability and conjecture based solutions as described below, it probably will not meet the safety levels required in the aircraft industry.

> Current traditional automation and "artificial intelligent programming techniques are in a transition from narrow, carefully defined domains to real-work situations in which systems learn to deal with complex data and adapt to uncertainty. Today, systems can perform useful work in a very large and complex world. Because these small [software] **agents don't have a complete representation of the world, they are uncertain about their actions. So they learn to understand the probabilities of various thinks happening,** they learn the preferences [of users] and costs of outcomes and, perhaps more important, they become self-aware" (Anthes, 2009).

The rest of the story about how bright these computers can be today is also suggested by Anthes in his concluding statement which follows: "We still hope that some time in the future computers will be as intelligent as we are but it's not a problem we'll solve in 10 years. It may take over 100 years." (Anthes, 2009)

With these inputs I am going to assume that the human is going to play a large role maintaining the next generation of aircraft and air/ground/satellite integration operations.

The Place of Automation and Human in Future Aircraft Maintenance Systems

Way back in 1983 L. Bainbridge stated the following purpose of automation. It is to replace human manual control, planning and problem solving by automatic devices and computers. But some of her colleagues pointed out: "even highly automated systems, such as electric power networks, need human beings for supervision, adjustment, maintenance, expansion and improvement." Therefore one can draw the paradoxical conclusion that automated systems still are man-machine systems, for which both technical and human factors are important.' It was suggested that the increased interest in human factors among engineers reflects the irony that the more advanced a control system is, so the more crucial may be the contribution of the human operator. (Bainbridge, 1983)

Non Automation of Maintenance for the Newly Complex Computerized Everything

There are also aircraft maintenance issues when automation is not incorporated into the highly integrated, complex computer systems.

Limited Attempted Implementation

There are probably only a few systems that are being designed with these goals in mind. Organizations that are making components (that may or may not work in a system environment) will not automatically implement any or the same maintenance processes as the next organization (currently no industry standard). Consequently the devices probably won't have the same service and repair process.

No Required Implementation Means Initial Cost Lower / Operational Cost and Life Cycle Costs Much Higher / Safety will be an Issue

With no pressure to implement "design-for-maintainability" and/or maintenance automation, many organizations that only build/sell devices/system will not be inclined to spend the money to implement. If they are directed to deal with life cycle costs, then implementation will be of great benefit. With no designed for maintenance automation, the task of maintenance requires a human solution and the labor, cost and safety problems will grow during the life cycle of the device/system.

Maintenance Systems Need Maintenance and Automated Systems May Be Unable To Do the Job

A key point is that the when the self maintenance automation doesn't work or the guided service and repair systems have problems (which they will) they will have to be serviced by another smarter computer and/or a human. Now we have added another level of complexity. The complexity of the problems that will be turned over to the human could make it very difficult and time consuming to make corrections.

The Challenge – Preparing for the Future

Most of this paper has discussed both the state of the aircraft maintenance industry and about the thinking about replacing people maintainers with automation of one type or another. I have suggested that it is highly unlikely that our very sophisticated computer systems located in aircraft, aircraft support systems, airport systems, air traffic control systems, and all the other related systems will be able to eliminate the human maintainer in the near future. It is projected that the traditional description of the maintenance function probably will have to drastically change. Many tasks not traditionally attributed to the current maintenance person's job (including dealing with complex computer functions, configuration management, software/hardware version updating and control, data management, data repair, security or maintenance systems and data, interfacing with electro mechanic systems and traditional maintenance, etc., etc.) may become part of this job. Someone is already doing these functions on modern aircraft but not under the traditional maintenance job classification.

The Rest of the Story – Financial and Social Realities That Are Defining the Aircraft Mechanic of Today and Tomorrow

As you were reading about the changes occurring in the maintenance process and industry, how often did you translate the technology drivers into the underlying financial, personnel, and social underpinnings of those technologies? If like many people, you seldom or never thought about the causation of change which is often a result of financial, personnel, social, or other practical reasons that result in technology and process change. This is a circular argument in that when financial, personnel, and social changes drive technology changes, these technology changes further drive additional financial and social changes. Because this cycle has been going on for so long people have accepted the spiral and pay little attention to the causal issues and the cumulative state of the industry caused by the changes.

Financial and Social Drivers to Develop a State of the Industry

There are numerous issues that will bring substantial financial pressure to our industry to identify the state of the industry and to develop solutions to resolve those issues. Among those issues are: Technological complexity may be advancing faster than the maintenance process can address it with the financial impact it will be difficult to economically maintain aircraft from a life cycle view. There may not be enough people being trained with the skills that will be needed for new technology and there will be a great shortfall of trained maintenance personnel with new and traditional skills. This shortage could impact maintaining the market or limit its expansion. Automation will be employed to reduce some costs but from a systems view in a life cycle environment it could actually be more costly to maintain highly automated systems.

As the aging maintenance workforce retires the work ethic, objectives, and style is changing. This change is having far ranging impact on accomplishing safe, effective repair. Because of salary pressures on airlines, maintenance personnel can find much better salaries at auto dealerships so the airlines are not competing. Work conditions and schedules are issues that were accepted by older workers but are not tolerated by younger works. A huge number additional social issues have developed over the years and many will have to be resolved before the social aspect of the maintenance job will be seen as a desirable.

What Needs to be Done?

Traditional thinking would suggest that we identify the problems and fix them so they work as well as they used to. But the "used to" has gone away. This realization should lead the industry to think of the maintenance process needing a new beginning. By evaluating the needs of the customer (system, devices, airlines, repair stations, maintenance people, etc.) (Gallaway, 2006, 2007) a picture of the maintenance industry role for the future can be built. This would then be followed buy specific action plans to deliver what was needed. By openly addressing the financial, personnel, and social requirements through the development of the new process, the new process can meet the industry *need.*

Working on the Future

The maintenance industry no longer has the option to continue on it's lassie fair path any longer if the maintenance industry wants to provide aircraft maintenance services in the future. The conclusion is that this industry must benchmark its state, identify the current and future needs for maintenance, determine what the maintenance person will need to know and be able to do, develop standards for work sharing between automation and humans, look at alternatives to meeting maintenance service, and begin the shaping and resolving of issues to meet safety and business needs, and identify how the process should be managed

The FAA Flight Standards Division has initiated a research program to start the process. This work will be supported by representatives from industry, government, academia, labor. It will also be well grounded in financial, social and regulator requirements of the industry.

References

Anthes, G. (2009) AI Comes of Age, *Computerworld* Vol. 43, No 4. 16-19.

Bainbridge, L. (1983) Ironies of automation - Increasing levels of automation can increase, rather than decrease, the problems of supporting the human operator. Automatica, **19**, 775-779. Reprinted in : (1987) Rasmussen, J., Duncan, K. and Leplat, J. (eds.) New Technology and Human Error, Wiley, Chichester, 276-283.

Gallaway, G. (2005) "A Sociotechnical Enterprise That Supports the Software-Based Aircraft – Snapshot of a New Reality" HF in ODAM VIII Conference.USA..

Gallaway, G. (2006) Linking organizational, managerial, administrative, and financial business issues to human factors science. 16th Congress IEA,.

Gallaway, G. (2007) An Operational and Marketing Infrastructure for the Human Factors And Ergonomics Professions -- A Call To Action. *Meeting Diversity in Ergonomics.* ,Elsevier Ergonomics Book Series, Pikaar, R. et al Eds,.

Gallaway, G.(2007) "Accounting for Human Requirements and Cost in Software Development" A class segment presentation for FAA60004200 Training Class - *Basics of Software Cost Estimating*, USA.

DESIGNING WORK-CENTERED SUPPORT FOR DYNAMIC MULTI-MISSION SYNCHRONIZATION

Emilie Roth
Roth Cognitive Engineering
Brookline, MA

Ronald Scott
BBN Technologies
Eden Prairie, MN

Randall Whitaker
Northrop Grumman
Information Technology
Fairborn, OH

Thomas Kazmierczak
C5T Corporation
O'Fallon, IL

Robert Truxler
BBN Technologies
Cambridge, MA

John Ostwald
BBN Technologies
Cambridge, MA

Jeffrey Wampler
Air Force Research Laboratory, Human Effectiveness Directorate (AFRL/HECS)
WPAFB, OH

We describe recent developments in an ongoing program to design work-centered C2 support for a military airlift organization. Work-centered design tailors support to the cognitive and collaborative demands of the work. A coordinated suite of visualizations was developed to support synchronized replanning in response to dynamically changing conditions by revealing the inter-relationships and constraints across multiple missions distributed in time and space. Support included the ability to perform 'what if' simulations across multiple missions so as to assess the impact of a change in one mission on other missions. An empirical evaluation was conducted comparing target user replanning performance on multi-mission synchronization problems when using the new tool vs. a comparison tool representative of their current computer systems. The results revealed statistically significant improvements in solution times (three times faster), quality of solution (a third as many errors), situation awareness and workload, reinforcing the value of work-centered design.

This paper describes recent developments in an ongoing program to design work-centered command and control (C2) support for a military airlift organization (Scott et.al, 2005; Wampler et al., 2005; Roth et al., 2006; Roth et al., 2007). Work-centered design (WCD) focuses on developing visualizations and decision aids that are adapted to the cognitive and collaborative demands of the work (Eggleston and Whitaker, 2002; Eggleston, 2003). The WCD approach emphasizes acquisition and analysis of work domain knowledge to (1) identify key tasks requiring supportive intervention, (2) discover critical aspects of each such task, and (3) create visualization and control features tailored to facilitating the task from the decision maker's point of view. Work aiding is provided through a combination of visualizations that enable practitioners to directly perceive work goals, affordances and constraints (*representational aiding*) and *direct aiding* whereby machine intelligence is used to synthesize and present needed information in the context of work visualizations (e.g., alerting to problems or suggesting solutions). These innovations are collectively referred to as work-centered support services (WCSS). This paper reports on the design and evaluation of such WCSS concepts to enable more effective multi-mission synchronization during dynamic mission replanning.

Overview of the Context of Work

The military airlift organization is an air operations center (AOC) responsible for planning, scheduling and tracking of airlift and air refueling missions worldwide. Missions are initially planned by mission planners. Twenty-four hours prior to a planned mission launch, responsibility for the mission is transferred to a C2 center (referred to as the 'execution floor'), which is then responsible for handling any last minute changes and problems that might arise during mission execution. Up to 90% of missions are not executed as planned. Handling last minute changes is a complicated activity that must take into account issues such as balancing competing airlift demands; ensuring

88 ABW PA Cleared 3/16/2009; 88 ABW-2009-0900

diplomatic clearances (DIPs) for landings in and over-flights of foreign nations; considering airfield, cargo and aircrew constraints; and providing for aircraft refueling requirements (e.g., in-air refueling). One of the most constraining factors is parking MOG (Maximum On Ground) at an airfield. This is the maximum number of aircraft that can park at an airfield. Execution floor personnel do not currently have an effective way of visualizing all these factors and how they are related in order to understand the full impact of mission changes that might arise during execution and what further repercussions any possible solutions might have. Inefficient re-planning can be very costly and include the costs for fuel, wasted crew time, and delayed troop and cargo movements, for example.

The WCD team has been studying the organization and activities of mission planning and execution floor personnel, and developing WCSS for them, since 2004 under an advanced technology development program entitled Work-centered Interface Distributed Environment (WIDE). In Spiral 1 a timeline prototype designed to support execution floor personnel in identifying repercussions of temporal changes in missions was developed and evaluated (Roth, Stilson, Scott, Whitaker, Kazmierczak, Thomas-Meyers, and Wampler, 2006). In Spiral 2 the prototype was expanded to include additional visualizations to improve the ability to perform dynamic replanning to support 'pop-up' requirements (e.g., when an aeromedical evacuation mission is needed; or aircraft malfunction requires a re-allocation of resources to support a high priority mission). These capabilities were shown to positively enhance situation awareness (SA) and dynamic replanning (Roth, Scott, Whitaker, Kazmierczak, Forsythe, Thomas, Stilson and Wampler, 2007).

In this latest phase of the program (Spiral 3) the WIDE prototype was expanded to support the ability of execution floor personnel to maintain SA of the *relations among multiple missions*. The objective was to enable execution personnel to consider constraints across multiple missions simultaneously as they create and replan missions in response to changing conditions (e.g., a delay due to aircraft malfunction or weather).

Overview of Knowledge Acquisition Activities and Results

The focus of knowledge acquisition and design in Spiral 3 was on a subgroup of C2 execution floor personnel that are responsible for planning and managing missions for a set of 'tails' (aircraft) that are resident in theatre (Theatre Direct Delivery of TDD). TDD was selected because it represents a 'microcosm' that encompasses the range of mission planning activities at the AOC including tail resource management, mission planning, mission monitoring during execution and dynamic mission replanning. While TDD provided the central focus, the goal was to develop capabilities that would generalize to all AOC personnel engaged in dynamic mission replanning. The Knowledge acquisition (KA) activities in Spiral 3 involved a combination of structured interviews, field observations and feedback on early design prototypes. The results revealed that major challenges included:

- Understanding and dealing with repercussions of mission changes (either delays or missions leaving early), both for the original mission and other missions (e.g., follow-on missions that were scheduled to use the same tail or other missions that were planning to pass through the same airfields).
- Understanding and dealing with repercussions of sudden reductions in anticipated resources available to satisfy mission movement requirements. This includes dealing with cases where the number of available tails is unexpectedly reduced (e.g., because of a broken tail, or a need to take-away a tail for a higher priority mission) and cases where airfields were temporarily closed or the number of parking spots at an airfield were temporarily reduced (e.g., because of construction work or a broken tail).

Handling these situations required understanding:

- The purposes of the missions and their relative priority (e.g., what cargo and passengers were they carrying? What is the priority of the cargo being carried? What was the formal priority level of the mission?).
- Individual mission characteristics and constraints (e.g., the mission itinerary, constraints on the type of aircraft suitable for mission needs, crew duty day constraints, airfield operating hour constraints, DIPS constraints, cargo constraints such as the required delivery date, and whether the mission is already in delay).
- Constraints across multiple missions (e.g., does this mission have a follow-on mission that is planned to use the same tail or that is otherwise linked to the mission so that a delay in one will force a delay in another? Does this mission go through an airfield that has a potential MOG problem so that the mission will not be able to land because there are no landing slots available or if it lands will cause a MOG situation making it impossible for another aircraft to land at the airfield?).

The results of the KA made clear that the current legacy tools provide limited support for these cognitive aspects of mission planning and replanning. Currently personnel utilize multiple, unintegrated, systems that include a database tool that presents updated mission information in tabular format; a Station Coordinator Tool that provides information on tails scheduled to land at a given airfield (one airfield at a time); and a 'homegrown' static visualization developed in Visio by the TDD personnel themselves that allows them to keep track of missions, what tail each has been assigned to, and their status, by manually updating the information as it comes in (See Figure 1).

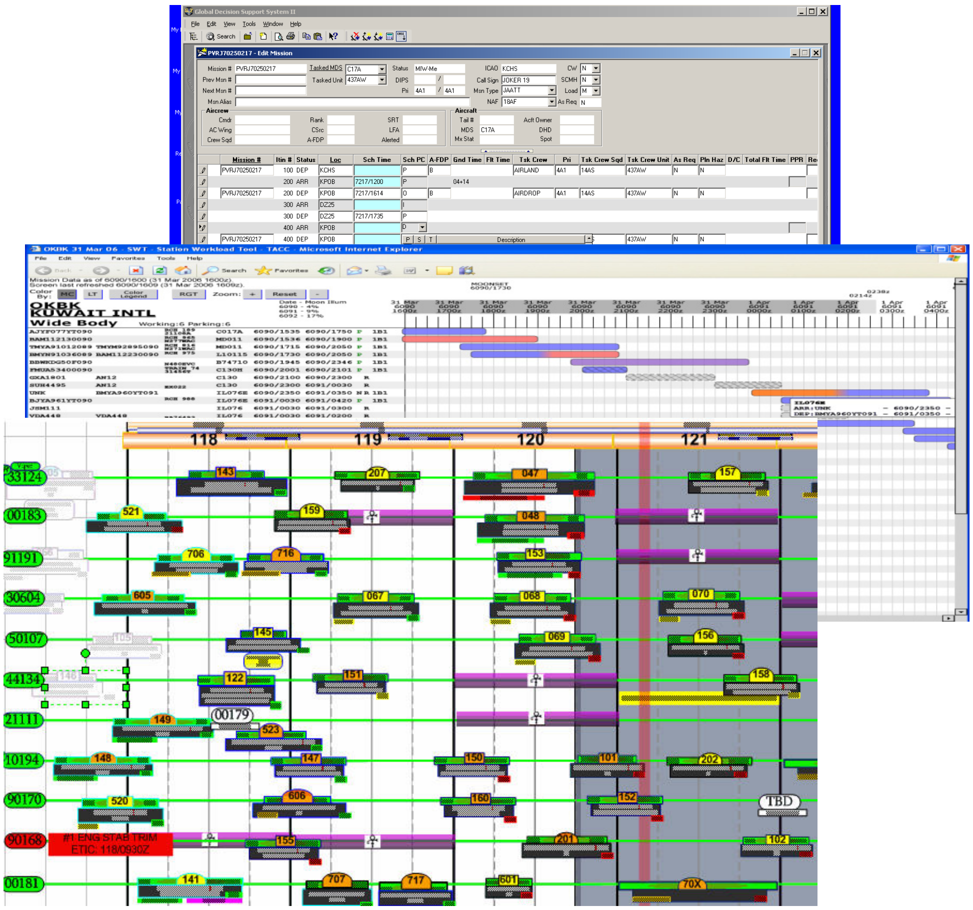


Figure1. Representative screens from software tools currently used by TDD personnel to plan and monitor missions. These tools are unintegrated and provide limited visibility into the repercussions of mission changes.

Analysis of the KA results suggested opportunities for more effective cognitive support for mission monitoring and dynamic replanning. This included:

- A dynamic 'Big Picture' mission-level overview visualization that is analogous to the static Visio display in terms of cognitive support, but that is automatically updated, and flags repercussions of changes within and across missions.

- Visualizations that enable simultaneous consideration of multiple constraints across missions during mission planning and replanning. These would facilitate rapid assessment of single mission constraints (e.g., crew duty day; Required Delivery Date; OPS hours); and multi-mission constraints (e.g., MOG); and would provide the ability to perform 'what-if' analyses to assess the repercussions of changes within and across missions.

Overview of Spiral 3 Design

The support requirements identified based on the KA and analysis activities were used to guide the development of WIDE Spiral 3 display concepts. A suite of three coordinated views were developed to support synchronized replanning by revealing the interrelationships and constraints across multiple missions distributed in time and space. Users are able to make changes to one or more missions in 'what if' simulation mode and see the impact of those changes on other missions. The three views include:

- *a multi-aircraft timeline display.* This view allows users to see mission details (individual sorties, crew, airfield, cargo) organized on a timeline. Users can see relationships among missions (both for a single tail and across multiple tails) and assess effects of changes to one or more missions on other missions.
- *a multi-airfield timeline display.* This view allows users to see all the tails going into and out of a specific airfield. The time duration during which a given aircraft is on the ground at that airfield is highlighted, but the rest of its mission (as well as subsequent missions) is also shown. This allows the user to de-conflict MOG situations more easily by enabling the user to move missions and see the impact on both that mission (e.g., exceeding crew duty day) as well as on other missions (e.g., creating a MOG situation for another mission.)
- *a multi-aircraft mission-level overview display.* This mission-level timeline view allows the user to maintain high level SA of how tail resources are being allocated to missions, as well as the objectives and status of the missions they are currently monitoring (e.g., itinerary, cargo, priority, delivery date requirements). It provides a dynamic 'Big Picture' equivalent of the static Visio display currently used by TDD personnel, allowing users to make changes in 'what if' simulation mode and immediately see the repercussions across missions.

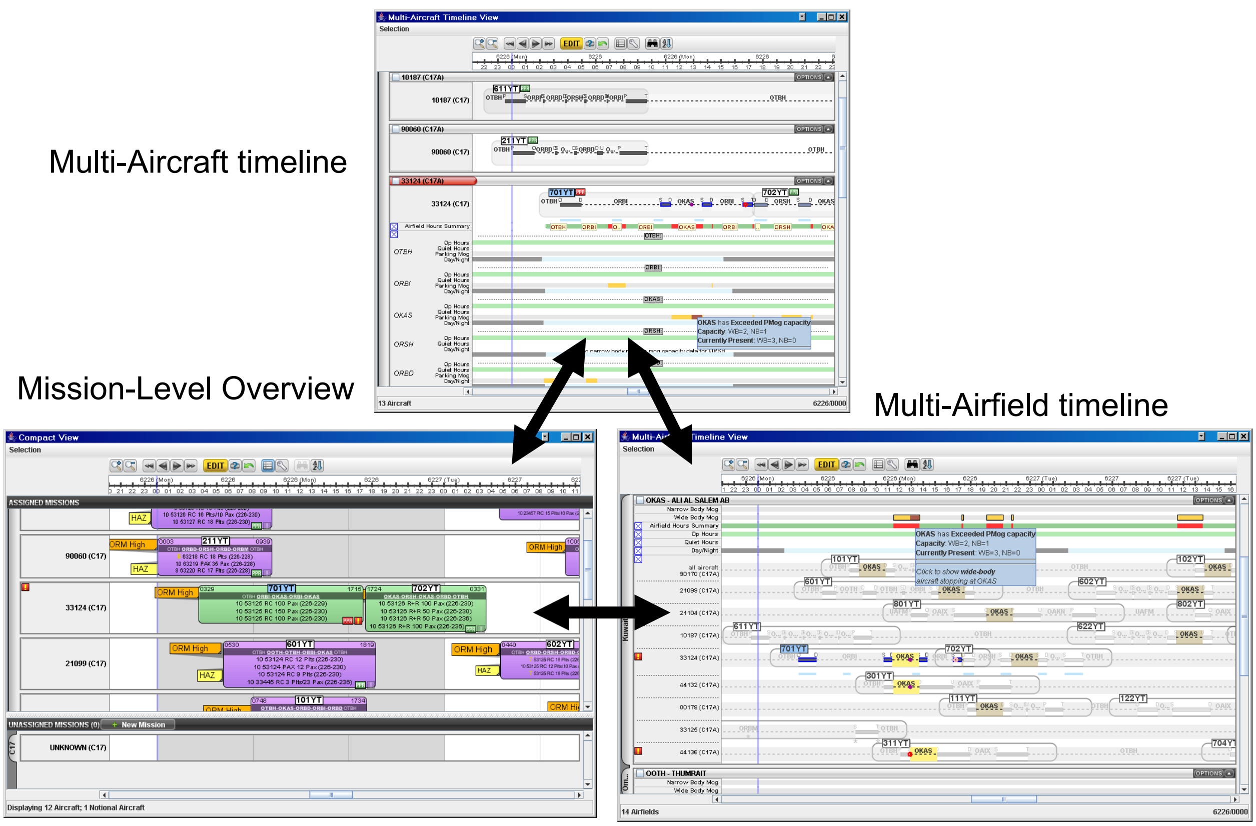


Figure 2. Snapshots of the three main WIDE Spiral 3 prototype views. These views are dynamically linked and allow users to make changes in 'what if' simulation mode and see the impacts across missions and views.

These three separate displays are dynamically linked so that any change that is made in one view (using a 'what if' simulation mode) is immediately reflected in the other views. (See Figure 2).

Spiral 3 Evaluation

An empirical evaluation of the WIDE Spiral 3 prototype was performed. Ideally the WIDE Spiral 3 designs would be compared to the current legacy tools being used on the execution floor. However a variety of technical and organizational obstacles precluded this option. Consequently we developed 'information equivalent surrogates' to use as a baseline comparison. The surrogates were equivalent to their legacy counterparts in the sense that they provided the same information and required similar integration of information across displays to come up with problem solutions. Execution floor participants in both the pilot study and the main study readily accepted the surrogate displays as reasonable equivalents to the corresponding displays on the floor.

The evaluation study compared the performance of 12 execution floor personnel using the WIDE Spiral 3 prototype to their performance using the information equivalent surrogate displays on comparable multi-mission synchronization problem scenarios. Scenarios described mission changes, and required the participants to assess the impact of those changes on a given mission, it's follow-on mission(s) and other missions passing through the same airfields, and to generate a revised plan that resolved the mission problems across all the missions involved. A variety of objective and subjective measures of performance were collected including:

- Solution times (i.e., time to assess impact of mission changes and time to generate a revised plan that eliminated all problems across missions);
- Errors in evaluating repercussions and/or generating a solution that eliminated all problems across missions;
- Self-rating of SA;
- NASA-TLX ratings of workload;
- Evaluations of usability, usefulness, and impact as measured via 8-point Likert-rating scale questions included in a post-test questionnaire.

A within-subjects experiment design was used. Each test participant experienced both the WIDE Spiral 3 condition and the Surrogate condition. They were presented with three mission replan scenarios to solve in each of the display conditions. The scenarios used in each condition were comparable but different, and were counterbalanced so that each scenario was presented in both conditions across participants. For example, if participant 1 saw scenario 1 with the WIDE Spiral 3 displays then participant 2 saw scenario 1 with the information equivalent surrogate displays. The order of the two test conditions was also counterbalanced. Half the participants went through the WIDE Spiral 3 condition first and half went through the information equivalent surrogate condition first.

The WIDE Spiral 3 prototype, when compared to the Surrogate Displays, resulted in:

- Significantly faster solution times (mean of 79 vs. 232 seconds; $F=19.89$, $p < .001$)
- Significantly fewer errors (mean of 10% vs. 32% errors; $F=9.71$; $p < .01$)
- Significantly better situation awareness as measured by self-reported ratings of SA (Paired T-test with $p < .05$ for all elements of SA);
- Significantly lower workload as measured by NASA TLX ratings (Paired T-test with $p < .05$ for effort, performance, temporal workload and mental workload. No statistical difference for frustration or physical workload.)

Participant responses on the final questionnaire reinforced the objective performance results. Participants consistently gave the WIDE Spiral 3 prototype high ratings (mean above 6.4 on 8-point Likert rating scale questions with 1 = extremely negative and 8 = extremely positive end-points). This included questions that asked about usability (6.7), usefulness (7.2), learnability (6.8), impact on own work (7.4) and impact on the organization (7.3). Participants also provided high ratings on questions relating to the ability of the WIDE Spiral 3 prototype to improve the quality of decisions relative to performance with 'existing TACC tools and practices'. These questions used an 8 point scale where 1 = not at all effective and 8 = extremely effective. Mean ratings were 6.8 or above on these questions including questions that asked about: reducing time to come up with a solution (7.3), reducing errors (7.4), minimizing negative repercussions on other missions (6.9), and generally improving solution quality 'because it is faster and easier to investigate multiple alternative options'(7.4).

Conclusions

The evaluation established the performance benefits that can be expected from implementation of the WIDE Spiral 3 prototype. Participants were able to solve representative multi-mission synchronization problems three times faster and with less than one third the number of errors. SA and workload self-report measures reinforced those results as did participant responses on the final questionnaire Likert-rating scale questions. Participants indicated that the WIDE prototype allows them to have significantly better SA of the impact of mission changes on own and other missions with significantly less workload than is possible with their current tools and practices on the floor. These improvements in performance are likely to translate into increased efficiency in terms of time required to come up with a mission plan, and increased mission replan quality, in terms of reduced mission delays, fewer mission cancelations and improved asset utilization to meet the AOC objectives.

The results reinforce the value of work-centered design approaches. As in the case of earlier WIDE Spirals, this project provides an illustrative example of the methods and performance benefits of WCD approaches. WCD progresses from knowledge acquisition through analysis and design to development and evaluation with particular attention to the cognitive requirements of and demands on the focal decision maker. By developing systems finely tuned to the cognitive and collaborative requirements of C2 work it is possible to achieve performance improvements that can have substantive positive impact on organizational objectives.

Acknowledgments

The research, prototype development, and evaluation was funded by the Air Force Research Laboratory (AFRL) Human Effectiveness Directorate at Wright Patterson AFB, OH and sponsored by Air Mobility Command (AMC), at Scott AFB, IL. We are indebted to AMC personnel at Scott AFB for their willingness to participate in our work analysis activities, and their assistance in carrying out and participating in the prototype evaluation

References

Eggleston, R. G. (2003). Work-Centered Design: A Cognitive Engineering Approach to System Design. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting (pp. 263-267).* Santa Monica, CA: Human Factors and Ergonomics Society.

Eggleston, R. G. and Whitaker, R. D. (2002). Work-Centered Support Systems Design: Using organizing frames to reduce work complexity. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeti*ng (pp 265-269). Santa Monica, CA: Human Factors and Ergonomics Society.

Roth, E M., Scott, R., Whitaker, R., Kazmierczak, T., Forsythe, M., Thomas, G., Stilson, M. and Wampler, J. (2007). Designing decision support for mission resource retasking. *Proceedings of the 2007 International Symposium on Aviation Psychology*, April 23 – 26, 2007, Dayton, Ohio.

Roth, E. M., Stilson, M., Scott, R., Whitaker, R., Kazmierczak, T., Thomas-Meyers, G. and Wampler, J. (2006). Work-centered Design and Evaluation of a C2 Visualization Aid. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 255- 259). Santa Monica, CA: Human Factors and Ergonomics Society.

Scott, R., Roth, E. M., Deutsch, S. E., Malchiodi, E., Kazmierczak, T., Eggleston, R. Kuper, S. R. and Whitaker, R. (2005). Work-Centered Support Systems: A Human-Centered Approach to Intelligent System Design*. IEEE Intelligent Systems*, vol. 20, issue 2, pp. 73-81.

Wampler, J., Whitaker, R., Roth, E., Scott, R., Stilson, M. and Thomas-Meyers, G. (2005). Cognitive Work Aids for C2 Planning: Actionable Information to Support Operational Decision Making. In *Proceedings of the 10th International Command and Control Research and Technology Symposium (June, 2005*).

TOWARDS A MEANINGFUL PRESENTATION OF FMS TRAJECTORY INFORMATION FOR TACTICAL SELF-SEPARATION

S.B.J. Van Dam, M. Mulder, M. M. (René) van Paassen
Delft University of Technology, Faculty of Aerospace Engineering
Kluijverweg 1, 2629HS Delft, The Netherlands

In the context of future airspace management concepts, the flight crew will need tactical navigation support for airborne self-separation. Applying ecological interface design principles, a state-based navigation tool was designed that uses functional information overlays that show how traffic and aircraft performance constrain the horizontal maneuvering. The state-based system has been enhanced with a visualization of intent information from the flight plan trajectory (Van Dam, Paassen, & Mulder, 2007). This paper discusses in detail the exchange of intent information using ADS-B. It presents some promising ideas to show intent in a more meaningful and pilot-intuitive way, particularly focusing on the impact of mode transition from trajectory control mode (Flight Management System) to target state control mode.

Background

In order to give pilots effective support for airborne aircraft self-separation, an Airborne Separation Assistance System (ASAS) support tool was designed to give the flight crew insight into which maneuvers best deal with conflict situations (Van Dam, Mulder, & Paassen, 2008). The design is aimed at showing the reasoning of the automation that deals with the separation problem, and promoting pilot traffic awareness. Applying the Ecological Interface Design principles (Vicente & Rasmussen, 1992), functional information is presented via overlays that show pilots how horizontal maneuvering possibilities are constrained. Maneuver constraints originate from limits to the own aircraft performance (internal constraints), and limits imposed by the environment, i.e., the surrounding traffic (external constraints). The display is usable without the use of explicit maneuver commands. This approach promotes the preservation of travel freedom in a flexible airspace environment, and also facilitates full integration with other navigation support tools.

In the state-based display, the 'eXtended Airborne Trajectory Planner (XATP)(Appleton, Mulder, & Paassen, 2006), the 'State Vector Envelope' (SVE) overlay shows a speed-heading maneuver space that is mapped on the existing Navigation Display (ND), Figure 1. The orange color of the Forbidden Beam Zone (FBZ)-layer informs pilots that the aircraft will enter the Protected Zone (PZ) of an intruder aircraft within the next five minutes. If the FBZ is red, separation will be lost within 3 minutes. The SVE overlay shows the range of feasible aircraft maneuvers in terms of target heading-speed states, thus, separation is maintained by steering the own 'state vector' out of the FBZ. Cooperative conflict resolution is realized by steering out the FBZ in the direction of the closest FBZ boundary, while efficient resolution are realized by staying away from the FBZ origin. A detailed description of the domain analysis, display design and pilot maneuver strategy can be found in (Van Dam et al., 2008). A design for the vertical plane can be found (Heylen, Mulder, Van Dam, & Paassen, 2008). For a general overview and discussion of ecological interfaces applied to vehicle motion applications, consult (Paassen, Amelink, Borst, Van Dam, & Mulder, 2007).

The state-based XATP design predicts aircraft motion by extrapolating the current state (position, speed and heading) of the own aicraft and the surrounding traffic. In realistic traffic situations however, the aicraft trajectory is controlled according the flight plan managed by the Flight Management System (FMS). The state-based display concept was adapted to take into account planned trajectory changes within the prediction horizon of ASAS systems, leading to a preliminary intent-based display (Van Dam et al., 2007). The design assumes availability of the FMS flight plan and Mode Control Panel (MCP) - Flight Control Unit (FCU) target states through the use of Automatic Dependent Surveillance-Broadcast (ADS-B)(RTCA, 2002; Barhydt & Warren, 2002). This paper presents more profound research on how exactly ADS-B technology supports the exchange of intent information. It also analyzes the differences between the state-based display and the proposed preliminary intent display (Van Dam et al., 2007), in particular how these designs shape pilot traffic awareness and affect pilot maneuver strategies. Based on this analysis, FBZ maneuver constraint areas are categorized and a new FBZ color-symbology is detailed with the aim to improve pilot understanding of these areas, particularly when FMS trajectory control mode is disengaged when maneuvering with the MCP/FCU target state control. Finally a proposal for a new intent display is described.

Figure 1: State-based display (XATP)

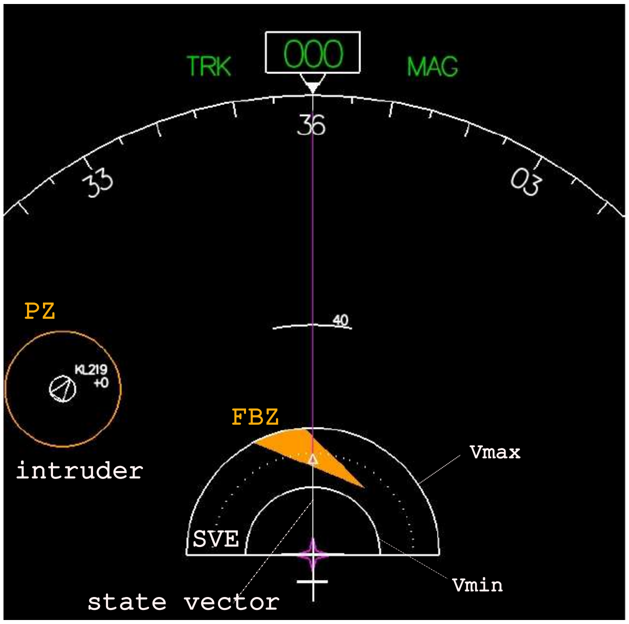
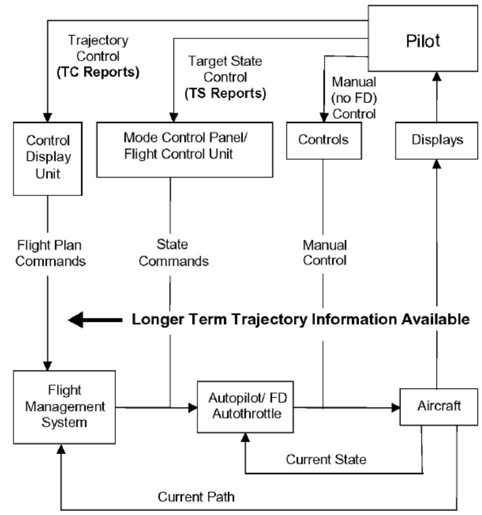


Figure 2: Aircraft control states (Barhydt and Warren, 2002)



ADS-B technology: Trajectory Change (TC) and Target State (TS) Reports

The ADS-B transponders are used to enable airborne data communication between aircraft in each other's vicinity. In addition to current state information the messages can also contain intent information. The transmitting aircraft must ofcourse support FCU-MCP modes to acquire target state commands and FMS-RNAV mode to get the flight plan information of the waypoints where trajectory changes are made. The requirements regarding the message contents are laid down in a RTCA report (Barhydt & Warren, 2002) and is used as a guideline. Without going into further detail it is assumed that the capacity and update rates of the system are sufficient to properly support an intent-based separation assistance tool. There are multiple types of data messages that are sent through ADS-B. Aircraft state reports include actual position and speed information that is used by the no-intent XATP system. For intent messages, two message types exist. First, the Trajectory Change report gives information on the aircraft's FMS flight plan. The Target State report provides information about the aircraft's target state commands, e.g., target heading entered by the pilot in order to make an autopilot controlled turn. Figure 2 presents an overview of aircraft control states (Barhydt & Warren, 2002).

The FMS system is a navigation aid database that contains intent information in the form of waypoints. The information of a waypoint is detailed in a so-called 'Trajectory Change Point'(TCP). Up to four TCPs are defined in one 'TC report'. TC report cycle numbers make it possible to distinguish between TCPs and they define the sequence order of the waypoints for reconstructing the flight trajectory. Figure 1 lists the elements provided in a TC report. Included are waypoint elements such as Time-To-Go (TTG), position, turn radius, track to TCP, track from TCP, and the command/planned flag for different TC types, e.g. a Fly-By turn or a Direct-to-Fix transition. TC reports can only be sent when the FMS is enabled and the aircraft is flying in accordance with the flight path depicted by the FMS. In case the pilot uses the the FCU-MCP to command a autopilot maneuver, the FMS is disabled. From then on all TC reports will have the flag type set on 'Planned' instead of the 'Command' indicating that the listed waypoints are not 'active'. With the FMS disabled, additional TS reports are sent out, containing the MCP target heading. The elements of a TS Report are also given in Table 1. This table is adapted from (Barhydt & Warren, 2002).

With respect to conflict situations in this research it is assumed that pilots fly in FMS mode while the are confronted with a separation problem. After analysis of the situation, the pilot manipulates the MCP to initiate the resolution maneuver. the FMS is automatically disconnected and TS reports representing the heading change will be available from that moment on, while TC reports are also available containing a "planned" typeset flag, informing the ASAS system about the 'FMS-disengaged' status . When the conflict is resolved and both aircraft have passed

Table 1: *Selection of Trajectory Change (TC) and Trajectory State (TS) Report elements*

	Element	TC Content	TS Content	Bits
ID	1	Participant Address	idem	24
	2	Address Qualifier	idem	
TOA	3	Time of Applicability	idem	6
TCR number	4	TCR sequence number	2	
TCR version	5	TCR cycle number	6	
TTG	6	Time To Go	idem	6
Horizontal information	7a	Horizontal data available and TC Type	Target Source Indicator	2
	7b	TC Latitude	Target Heading or Track Angle	16
	7c	TC Longitude	Target Heading or Track Indicator	16
	7d	Turn radius	Horizontal Mode Indicator 8	
	7e	Track to TCP	-	8
	7f	Track from TCP	-	8
	7g	Horizontal command/planned flag	-	1
Vertical information	8a	Vertical data available and TC Type	Target Source Indicator	2
	8b	TC Altitude	Target Altitude	12
	8c	TC Altitude Type	Target Altitude Type	2
	8d	Vertical Command/Planned Flag	Vertical Mode Indicator	1

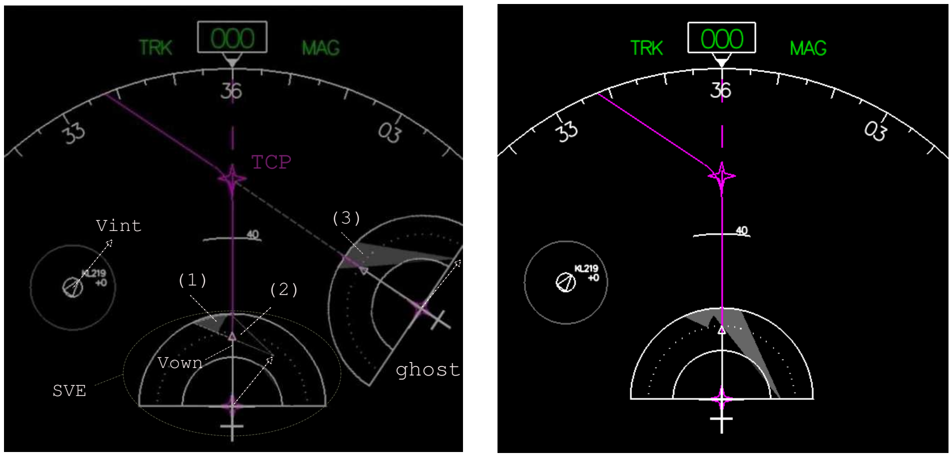
each other, the pilot will initiate the path recovery maneuver by flying a Direct-to to the closest TCP waypoint on the FMS flight path. The FMS is updated and activated again while the TS reports are suspended. At present, the design of the intent display will focus on the scenario where the pilot identifies a conflict situation along the flight plan and manipulates the MCP settings to resolve a conflict situation, hereby disengaging the FMS.

Comparing the state-based and intent-based display

Figure 1 and Figure 3 show an example conflict situation as presented on the ‘state-based display’ and the ‘intent-based display’ respectively. Based on the state-based display pilots assume separation is lost within three to five minutes if no maneuver is performed. With the intent display, the current state vector lies outside the FBZ’s. Therefore the FBZ is in grey color. No loss of separation will happen if both aircraft continue according the FMS flight plan. In this case, the intent display clearly enhances conflict awareness. Figure 3(a) shows how the intent display is the result of mapping two SVE’s on each other. The main SVE shows the constraint area indicated by (1) and visualizes maneuvers that would lose separation before the TCP is passed by. This time is labeled t_{TCP}. The other part of the FBZ indicated by (2) represents all states that would result in a loss of separation when the TCP is already passed by. Since the aircraft turns away at the TCP, this conflict is resolved and area (2) is therefore not shown on the intent display. The borderline between area (1) and (2) is part of a circle with the FBZ origin as it’s center and is referred to as ‘break-circle’ (Van Dam et al., 2007). The FBZ on the ghost SVE indicated by (3) is created using a ghost image for the own aircraft that shows the current fictive position and speed vector of the own aircraft as if it were flying already with constant speed and heading of the TCP.

When using the state-based display no information cues are available to determine wether the FMS turn will resolve the conflict situation or not. The intent display does show the pilot that there will not be a conflict situation when flying according the flight plan. On the other hand, pilots flying with the intent display are more likely to be unaware of how the conflict situation would look like when the FMS is disengaged. This lack of insight could lead to dangerous mode change situations where the FMS flight plan is suddenly abandoned by one of the aircraft. In the example situation, the SVE image will jump from the SVE in Figure 3(b) to the SVE in Figure 1.

On the intent display, Figure 3(b), the pilot can see that he is not able to turn to the right. This area does not exist on the state-based display as it is related to intended path after the TCP turn. Thus, this area is in fact conflict-free if one would maneuver before passing by the TCP. Pilots using the intent display are likely to (mis)interpret this area as a instantaneous no-go zone, i.e., they would not consider a maneuver to the right as a valid maneuver option to instantly resolve a problem.



(a) Constructing the intent display (b) First version of intent display

Figure 3: Intent-based XATP display for the example conflict situation with own intent.

Table 2: *Typification quadrant of FBZ maneuver constraint areas*

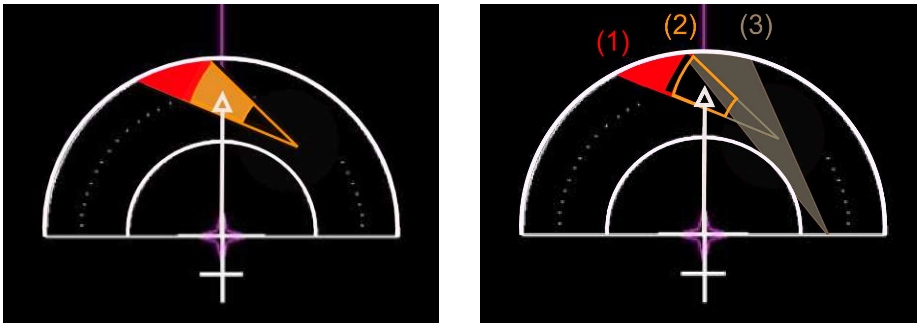
	Before first TCP	**After** first TCP
State-Based *current*	(1) **'pre-SB'**: always conflict Yes	(2) **'post-SB'**: FMS-disabled conflict No
Intent-Based *current*	(-) not applicable (ghost) -	(3) **'post-IB'**: FMS-enabled conflict Yes

Different types of FBZ maneuver constraint areas

It is clear that the introduction of FMS intent information result in different types of FBZ maneuver constraint areas. The current display formats create confusion, especially when disengaging the FMS. What does a particular FBZ-area shown inside the SVE actually means to the pilot? Is it possible to steer into this area right now? Will it disappear when the FMS is disabled? Will it's color change? Each display triggers different pilot behavior and conflict awareness.

In order to take away confusion about the interpretation of FBZ's areas, the different kinds of FBZ-areas are typified. Two parameters can be defined to make a distinction. A clear difference exist between areas generated from 'state-based' position and velocity information as opposed to areas generated from intent-based TCP position and velocity information (provided in the TC reports). The former is the most physical constraint where as the latter takes into account planned aircraft behavior from the FMS. A second parameter splits FBZ-areas in areas that result in a maneuver that looses separation 'before' (pre-TCP) and 'after' (post-TCP) the TCP is passed by. These two parameters make up a typification quadrant for FBZ maneuver constraint areas, Table 2.

First, 'State-based pre-TCP' areas (1) are areas created using the actual position, speed and heading to calculate the FBZ, AND showing that part of the FBZ that applies to maneuvers that lose separation before t_{TCP}. This FBZ constraint type is considered the most important constraint type. It is always visible on both state-based and intent-based displays. Second, the 'state-based post-TCP area type (2) is complementary with type (1) in the sense that it captures the state-based conflicts that occur after t_{TCP}. This area is currently not shown on the intent-display. In situations where the FMS is disengaged the display suddenly shows this area. In general, pilots unaware of the location of this type of constraints can not predict if a conflict would be triggered if one of the aircraft would ignore the next TCP and fly straight on. Third, 'Intent-based pre-TCP areas are fictive and not relevant. Fourth, 'Intent-based post-TCP' constraints (3) represent the FBZ-areas created by using intent-based information AND applied to maneuvers that lose separation after t_{TCP}. It is relevant to situations where both aircraft are FMS-enabled. It gives pilots a preview on how the FBZ areas look like from the point of view of the new state vector after the TCP turn (which is



(a) New FBZ color-coding for state-based display (b) New design of the intent display

Figure 4: The new color-coding for FBZ and the new version of the intent display

thus mapped on the current state vector in the SVE). If the current state vector lies inside a 'post-IB' FBZ-area (3), it means separation will be lost after the TCP maneuver is made. Without these constraints shown, pilots would be unable to predict if a conflict would appear after t_{TCP}.

The visualization of urgency

The idea to split up the FBZ shape into a 'pre-TCP' and 'post-TCP' is also usable to enhance, in fact, change the traditional color-coding of the FBZ's. Traditionally, the entire FBZ would be drawn red (or orange), reflecting the urgency of the conflict created by the current state vector. If the intruder's PZ would be intruded within three minutes the FBZ would be entirely red, if within five minutes, orange. If it would take more than five minutes, it would be shown in grey. If the state vector would not lie inside the FBZ, the FBZ would also be filled up in grey if a maneuver moving the state vector inside the region would trigger a conflict (with a predicted loss of separation within 5 minutes). Using the break-circle principle however, one single FBZ can be split up in time-intervals according the urgency color coding. For the example situation, this would results in a state-based display like Figure 4(a). The area representing maneuvers that lead to loss of separation further than 5 minutes away is not filled with any color, but lies by definition inside the FBZ shape. By applying this drawing convention, the grey zones are no longer used. 'Conflict zones' will always be orange or red, also when the current state vector lies outside the FBZ area.

A pilot-suited meaningful visualization of maneuver constraints

With the new insights regarding FBZ constraint area types and 'urgency' color coding, a new display design proposal can be made. The display is aimed at supporting the way pilots deal with a conflict problem when flying in FMS mode, interpreting the situation, and then go to MCP-mode to resolve the problem. The comments below are directly applicable to areas (1), (2) and (3) indicated on the display figures in Figure 3(a) and Figure 4(b).

First of all, *pilots need to be able to identify beyond any doubt if separation will be lost or not at all times.* This can be achieved by only filling areas with color or grey when a conflict exists. In both FMS-enabled and disabled mode, FBZ type (1) should be visible and given most importance. Therefore these constraint areas are brightly coloured in red and/or orange, Figure 4(a). If the FMS is engaged, FBZ type (2) areas do not apply and should not be filled. FBZ type (3) areas do apply to the current trajectory prediction and should be visible on the display. If the pilot switches to AP mode, the FMS is disengaged and FBZ type (2) should be brightly visible in the same way as FBZ type (1) constraints while FBZ type (3) should not be filled. Secondly pilots, when identifying a conflict situation in FMS mode, *should be aware which areas are instantly constraining the aircraft maneuver options* when disengaging the FMS, i.e., pilots should be aware that FBZ type (2) will appear and FBZ type (3) will disappear. Creating awareness about the type (3) constraint can be achieved by always showing intent-based constraints in grey. Grey areas are only shown when the FMS is enabled and inform pilot about intent-based (post-TCP) conflicts. Pilots will learn to take into account grey when predicting conflicts along the FMS trajectory, and will learn to ignore grey when they need to instantly come up with a conflict resolution maneuver. Informing pilots about the location of type (2) constraints, while in FMS-mode, creates awareness about the maneuver constraints when the FMS is disengaged. This awareness can be achieved by clearly depicting the FBZ contour. Type (2) areas are always the unfilled FBZ areas next to the colored type (1) area, see final design in Figure 4(b).

Concluding remarks

Based on FMS Trajectory Change Point information, the typical constraint representation of a conflict, the FBZ, is split up. This leads to a higher number of FBZ shapes on the display, Figure 3(b). Given the different nature of some of this areas, the original intent display proposal in (Van Dam et al., 2007) creates confusion, especially when disabling the FMS. A typification quadrant was set up to define the different types of FBZ-areas, Figure 2. Based on on the differences between each type, a more straightforward FBZ symbology is proposed so that pilots can clearly understand the meaning of each FBZ-area's shown on the display. It allows pilots to quickly perceive how their maneuver space is constraint when flying FMS enabled as well as FMS disabled, Figure 4(b).

The situation example in this paper has been chosen fairly simple in order to address a complex problem domain in an understandable way. The ideas expressed in this paper are however expandable to more complex situations, including multi-conflict scenario's, situations with intruder TCP point, situations with more than one TCP point. In the future, the interception of FMS trajectory after recovering from the conflict resolution maneuver will be treated. Even more display design can be enhanced by using target heading information of the Target State reports when switched to MCP target state commands (Van Dam et al., 2007). A pilot experiment will be set up were the display will be evaluated online using more complex multi-conflict situations.

References

Appleton, R., Mulder, M., & Paassen, M. M. van. (2006, August). Comparison of two interfaces for supporting pilots in airborne self-separation tasks. In *Aiaa guidance, navigation and control conference & exhibit.* Keystone (CO), USA: AIAA.

Barhydt, R., & Warren, A. (2002). *Development of intent information changes to revised minimium aviation system performance standards for automatic dependent suveillance roadcast (rtca/do-242a)* (Tech. Rep. No. NASA/TM-2002-211669). NASA.

Heylen, F. M., Mulder, M., Van Dam, S. B. J., & Paassen, M. M. van. (2008, August). Design and evaluation of a vertical separation assistance display. In *Aiaa guidance, navigation & control conference and exhibit, honolulu, hawaii.*

Paassen, M. M. van, Amelink, M. H. J., Borst, C., Van Dam, S. B. J., & Mulder, M. (2007, April). The chicken, the egg, the workspace analysis, and the ecological interface. In *2007 isap international symposium on aviation psychology.* Dayton, USA.

RTCA. (2002). *Minimal aviation system performance standards for automatic dependent surveillance broadcast (ads-b)* (Tech. Rep. No. DO-242A). RTCA Special Committee-186, Federal Aviation Authorities.

Van Dam, S. B. J., Mulder, M., & Paassen, M. M. van. (2005). Eid of a pilot support system for airborne separation assurance. In *2005 isap international symposium on aviation psychology.* Oklahoma City, USA.

Van Dam, S. B. J., 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, 38,* 1221-1233.

Van Dam, S. B. J., Paassen, M. M. van, & Mulder, M. (2007, April). Visualization of maneuver constraints for airborne self-separation: Use of intent information. In *2007 isap international symposium on aviation psychology.* Dayton, USA.

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

AN INTERFACE FOR INBOUND TRAFFIC ROUTE PLANNING

Ewout S. van Dijk, M. M. (René) van Paassen, Max Mulder
Aerospace Engineering – Delft University of Technology
Delft, The Netherlands

Mariska I. Roerdink
Air Traffic Control the Netherlands
Schiphol Oost, The Netherlands

It is expected that, with increasing automation, the emphasis in the air traffic controllers' work will shift from tactical control towards supervision and planning of aircraft trajectories. To support this work, a planning interface for area controllers has been developed. The interface uses a normal Plan View Display (PVD) supplemented with a Time-Space Diagram (TSD), that visualizes the travel of the incoming aircraft across their planned track. With the constraints on speed and timing as given in the TSD, the interface permits direct manipulation of the arrival time within these constraints. Using a simulation of air traffic, the interface was tested in an experiment. The results indicate that the interface can be used to manage traffic efficiently, but that maintaining a coherent mental picture using both the TSD and the PVD is still difficult.

Air traffic control (ATC) is a complex task, today still performed by human beings with relatively little support from automation. Current required competences for ATC personnel are for example summarized by the ATC Performance Model, developed at ATC the Netherlands (LVNL) (Oprins, Burggraaff, & Van Weerdenburg, 2006; Oprins & Schuver, 2003). With increasing traffic, and without changes to the current system, demands on ATC personnel can only become higher.

An important trend in research programs for future ATC systems (Anon., 2007; Dlugi et al., 2007) is the shift from the current tactical, sector-based air traffic control to strategic, trajectory-based Air Traffic Management (ATM). As an example of this shift, a possible scenario for future ATM in Area Control Center (ACC) sectors is considered. It is expected that, with the application of more fixed routing in the Terminal Maneuvering Area (TMA), the transfer of approaching aircraft to the TMA will have to adhere to stricter timing requirements. Also, in order to increase flight efficiency, holding patterns should be avoided, and aircraft timing will have to be adjusted with speed instructions. A concept is developed in which the ACC controller creates a 4D arrival trajectory for approaching aircraft, and implements this trajectory, optionally using speed requests to an adjacent sector. An interface to support Air Traffic Controllers (ATCo's) in this task has been designed.

Display Design

Inbound Traffic Management

For this study, a hypothesized future situation regarding inbound traffic management by Area Control will be described, taking planned procedures around Amsterdam Airport Schiphol (AAS) as a starting point. This scenario uses 3D fixed routes in the TMA with merging of traffic in the ACC. Aircraft must arrive at one of the Initial Approach Fixes (IAF) on the border of the TMA at co-ordinated times, and only limited modifications to the arrival time are applied in the TMA (Figure 1).

Since the ACC airspace, especially for the case of AAS, is limited, much could be gained from cooperation with adjacent sectors to change the timing of arriving aircraft. To make such adjustments feasible, the new display will display arriving aircraft as soon as they are available in the system. Using the presentation on the display, an ATCo can determine whether a request to an adjacent center is useful and feasible.

Time-Space Diagram

In order to control the arrival planning of inbound aircraft, ATCo's need a tool to consider the traffic in four dimensions; the spatial path and the temporal dimension. The main display currently used, the Plan View Display (PVD), offers only support for prediction over a limited time span, sufficient for an experienced ATCo to merge aircraft into a separated stream over the entry point to the TMA, but not sufficient for creating an arrival plan and

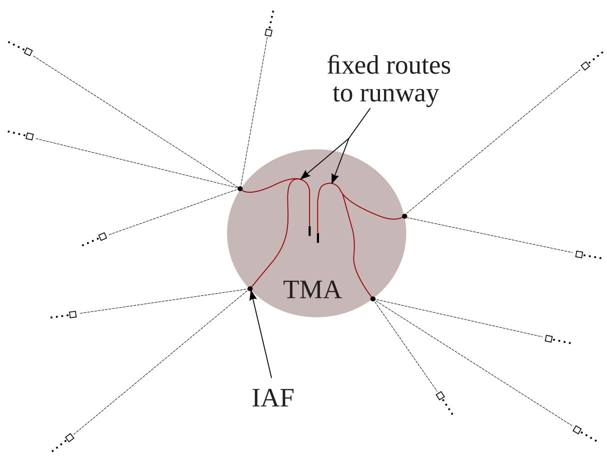


Figure 1: 3D fixed routes inside the TMA, starting at Initial Approach Fixes. Outside the TMA, aircraft arrive from all directions and are merged by ACC.

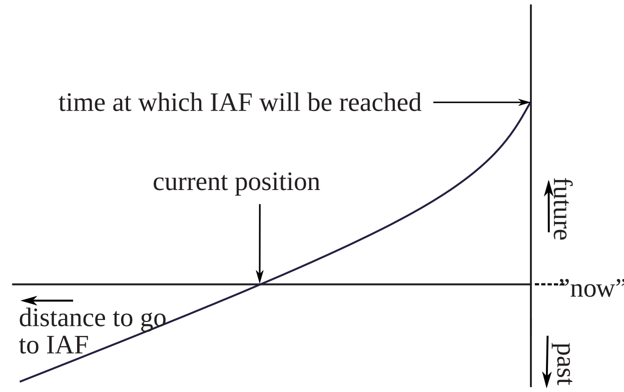


Figure 2: A time-space diagram of an aircraft flying to an IAF

issuing speed or heading vectors at or outside the ACC boundary to create a planning for entry into the TMA or solve upcoming conflicts when merging close to the TMA boundary.

As a starting point of the new interface, therefore, a Time-Space Diagram will be used. This kind of diagram has been tested in Eurocontrol's PHARE-project (Jorna, Pavet, Van Blanken, & Pichancourt, 1999) and by Delft University of Technology for assisting the ATCO in planning and monitoring Continuous Descent Approaches in the TMA (Tielrooij, in 't Veld, Mulder, & van Paassen, 2008). The principle of the time-space diagram is shown in Figure 2, for an aircraft which is on its way to the IAF. The horizontal axis shows the distance to go before the IAF is reached, the IAF can be imagined to be at the right side of this axis. The vertical axis is a time line. This makes the horizontal axis 'now', everything above it 'the future' and everything below it 'the past'. The time-space line moves downward in time, making the intersection with the horizontal axis move to the right: the aircraft flies in the direction of the IAF.

Aircraft Constraints. At AAS, aircraft generally enter the ACC at a high altitude, descending from upper airspace controlled by Eurocontrol or horizontally from adjacent centers. When a straight path to the IAF is planned, control of the aircraft speed is the only option to modify the arrival time. Speed control is of course constrained by the aircraft properties, resulting in upper limits on Mach number and lower and upper boundaries on Calibrated Airspeed (CAS). Since the maximum altitude in the AAS ACC sectors is limited to FL 245, the Mach limit does not need to be taken into control. When the ATCo selects a particular aircraft in the interface, the CAS limits for this aircraft are added to the TSD. For aircraft still in the adjacent sector, a double prediction is presented; one assuming that instructions are given by the ATCo in the adjacent sector, and one assuming that instructions are issued after the aircraft enters the own sector. In this way the ATCo can determine whether a request to an adjacent sector is feasible and useful.

For the implementation of the deceleration and descent behavior of the aircraft, the performance envelopes of the three aircraft implemented in the simulation were compared. It was determined that a descent flight path angle of 2 degrees was an acceptable value for all considered aircraft.

Separation Constraints. In principle, aircraft paths will be planned straight to the IAF. For aircraft that cross such a path, or converge on these paths, it is possible to calculate "forbidden zones" in the TSD. These zones are specific to the path of a considered aircraft, and indicate time and path combinations that will result in a conflict with another aircraft. Figure 3 shows how an aircraft crossing the path of another aircraft results in a forbidden zone.

Creating or solving a conflict by changing a speed is visualized by bending a time-space line in or out of the conflict zone. For example, when moving the right part of the time-space line of aircraft 1 in Figure 4 up, it will at

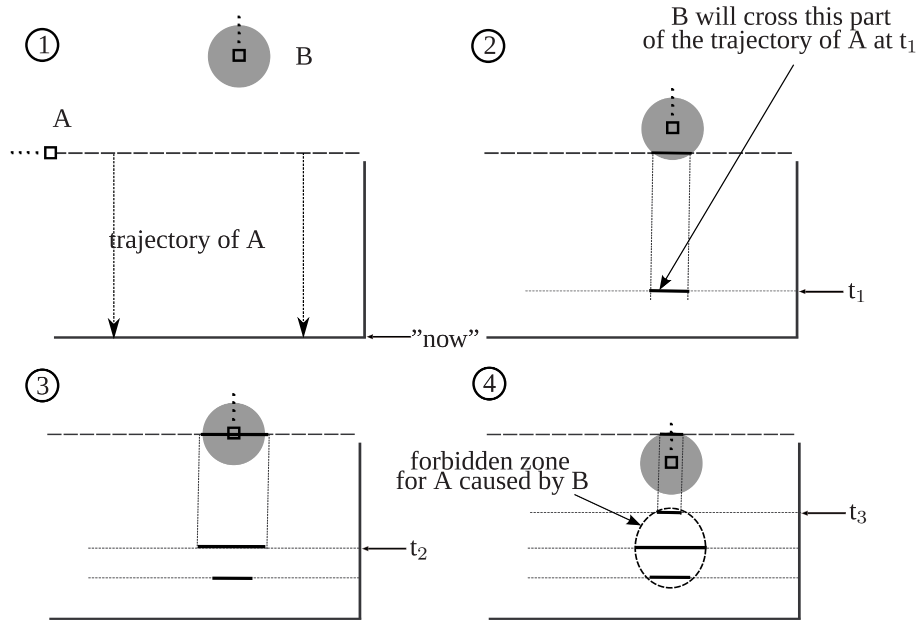


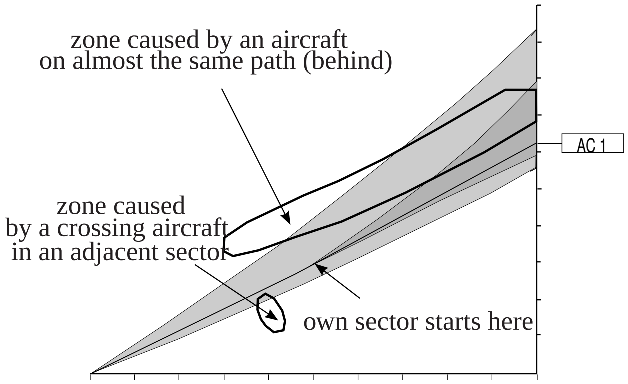
Figure 3: Detection of a forbidden zone in the TSD of aircraft A, caused by aircraft B when both trajectories are in the horizontal plane. Graphs 1 to 4 show a progression in time, a plan view is given at the top of each graph and its representation in the TSD at the bottom.

some point enter the forbidden zone above it. This implies, that by delaying the aircraft (slowing it down), the aircraft behind it starts overtaking it and a conflict will occur. This indicates a major advantage of the direct manipulation principle: it becomes immediately clear if a forbidden zone is crossed, when dragging the label. In this way, the constraints of the work domain are mapped on the interface. Since meaningful behavior (adjusting the time dimension) is also visible on the interface, direct manipulation is possible.

Inbound Planning Interface. While the time-distance lines in the TSD show the possible speed profile, and on the time axis, the possible arrival times of the aircraft, the planning interface needs to also show the constraints of the total planning process. Aircraft could be guided to the IAF's with appropriate separation, but their different speeds could result in them running into each other in the TMA. Furthermore, Figure 1 shows that the merging of two streams of traffic takes place in the TMA as well. Both aspects need to be taken into account by ACC when planning at which time the different aircraft should cross an IAF. The designed interface supports this process.

If aircraft performance, route and weather are known, an estimate can be made of the time between reaching the IAF and lining up with the runway. Assume that this is equal to ten minutes for a certain aircraft under certain conditions. On the right side of the TSD, the time at which this point is reached, could then be marked 'ten minutes above the arrival time at the IAF'. There, the time required before the next aircraft may arrive, can be expressed by a vertical bar. This time, i.e. the height of the bar, depends on the speed of both aircraft, as well as their wake vortex categories. This is shown in Figure 5, in which all bars have been shifted down the time line by the minimum travel time in the TMA for clarity. The bars that are not aligned with the arrival time at the fix indicated in the TSD represent aircraft that arrive, in this case, at the southernmost fix, and have a longer travel time in the TMA.

Path manipulation. In addition to manipulation with the arrival time, and thereby changing the speed of the aircraft, the ATCo is also given the opportunity to change the aircraft path in the plan view interface. Changes are applied to the basic (straight-in) path by adding a waypoint to a path. The changes applied to the lateral path are presented in the TSD as well. See Figure 6 for an example.



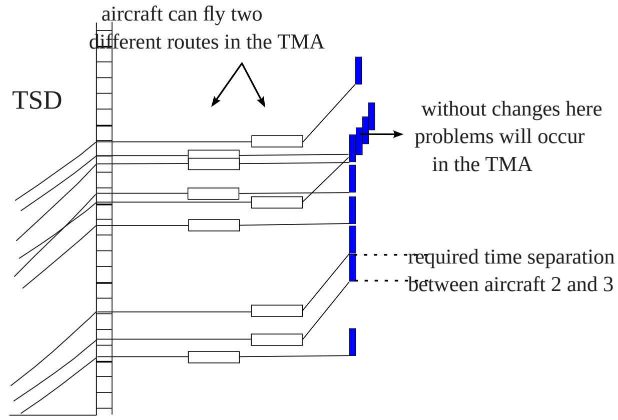


Figure 4: The TSD including 'forbidden zones' and margins in time

Figure 5: The time line on the right side of the TSD, indicating arrival times at the final merge point and required separation at this point

Experiment

An evaluation of the display and operational concept was carried out. The main purpose of the experiment was to investigate whether the interface would allow a safe and efficient planning of the inbound traffic, and to identify problem areas and possible

Experiment Set-Up

Equipment and subjects. The experiment was programmed on a laptop computer. The TSD was shown on the laptop screen, and the PVD was shown on an additional display connected to the laptop. Figure 7 shows a screen shot of the two displays. Ten subjects participated in the experiment, five of whom were active air traffic controllers, with experience ranging from 4 to 26 years. The other five subjects were research staff and students.

Scenarios. Four scenarios were created, with aircraft coming from the North, East and South and entering through one of the two IAF's (see Figure 7). Aircraft were kept at initial altitude before descent to the IAF with a 2° flight path. A mix of three aircraft types (Boeing 737-800, Boeing 777-200 and Airbus 320-212) was used, the simulation was based on BADA data (Nuic, 2004). Aircraft had to be delivered to one of the IAF's with time intervals of 1.7 min. Scenarios 1 and 2 were for familiarization, with low traffic rates, scenarios 3 and 4 had a high traffic rate (15 aircraft in 21 minutes), with scenario 4 being the most difficult.

Procedure An experiment session began with a 15-minute briefing. Using scenario 1, the working of the interface was explained, and after explanation subjects could practice with scenario 1. After subjects indicated they felt comfortable with the task, the other scenarios were presented. If at some points subjects had problems with the task, hints were offered by the experimenter. When all aircraft in the scenario had been provided with a plan, the simulation was run in fast-forward to show the results. Total time per subject was approximately one hour. After the runs, subjects completed a questionnaire, scoring statements on a four-point scale (agree, partly agree, partly disagree, disagree) and answering a number of open questions.

Results

This test of the interface should be considered as a first evaluation of a work in progress. The scenarios were fairly short, and in particular scenario 4 started in a state that was not representative of the traffic situation at, for example, a hand-over.

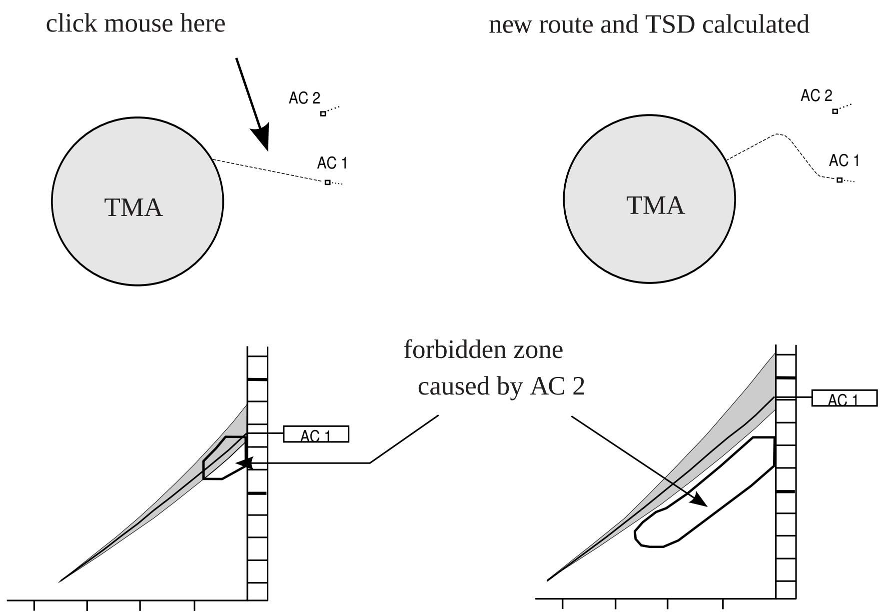


Figure 6: Re-routing aircraft 1 with a mouse click leads to immediate recalculation of the TSD

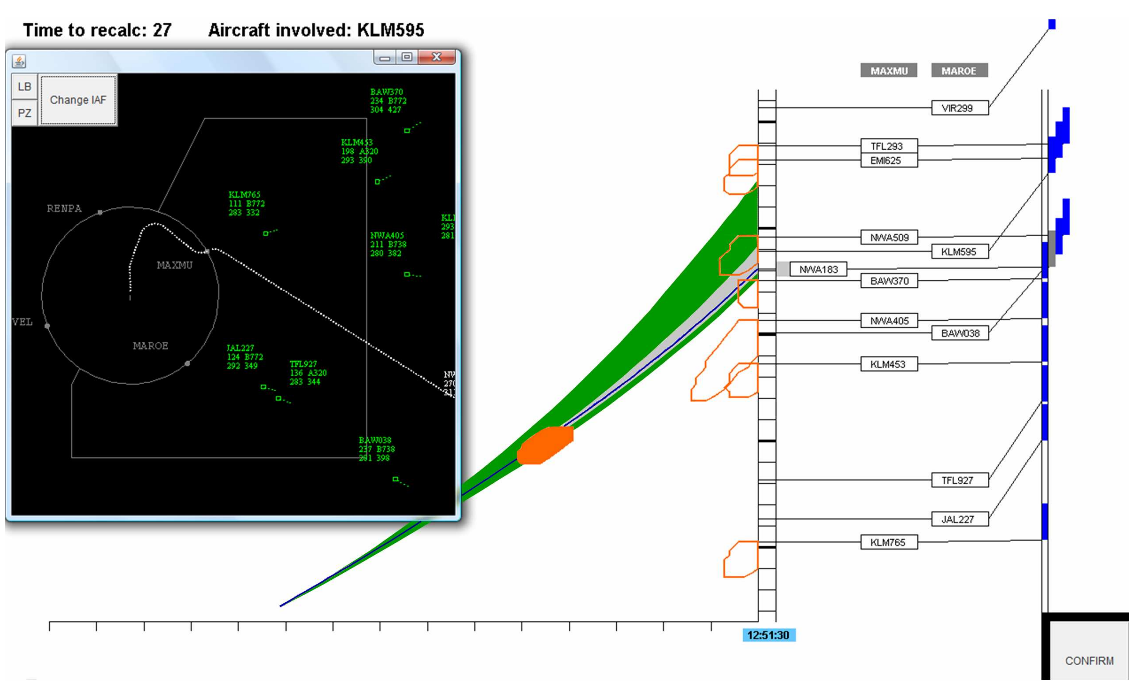


Figure 7: The TSD and PVD with the NWA183 selected

Safety and efficiency. The subjective impression of safety was tested by means of the statement "I can handle traffic safely". All subjects agreed or partly agreed, with the exception of two ATCo's who disagreed. These felt they were lacking the "mental picture" of the traffic situation. Complaints were mainly about the problem of integrating information from the two displays. All subjects agreed or partly agreed that they could handle traffic efficiently.

Interface use. Most subjects indicated that the TSD became their primary tool for the planning. Creating a plan was started on the TSD, and completed on the PVD when the need arose. Two of the ATCo indicated that it was difficult to interpret the conflict zones on the TSD, and that a better link to the PVD would be needed. The majority of the ATCo's indicated the need to also use the vertical path of the aircraft for separation. The possibility to request a speed change in the adjacent sector was very much valued.

Conclusions

The objective of the present work was to investigate the creation of a path planning tool air traffic control. The display combination of the extended TSD and PVD enable a human controller to create an efficient arrival planning. The main problem is still the integration of the information from the PVD and TSD to create a single mental picture of the traffic situation. The presentation on the TSD of the constraints of the work domain facilitate direct manipulation of the flight parameters in the search for a solution. A focal point for the future work is the increased (visual) integration of the information on the two displays, and the visualization of the constraints on the PVD, making path manipulations in the PVD as easy as the speed manipulations in the TSD.

Acknowledgments

The authors would like to thank the subjects who participated in the evaluation, and acknowledge the support of Hans Huisman of LVNL the Netherlands for arranging the internship position of the first author.

References

Anon. (2007). *Concept of operations for the next generation air transportation system. version 2.0.* (Report). Joint Planning and Development Office. http://www.jpdo.gov/library/NextGen_v2.0.pdf.

Dlugi, O., Astheimer, T., Baldoni, C., Bara, J. M., Brown, R., Cato, R., et al. (2007). *The atm target concept* (Deliverable 3 No. DLM-0612-001-02-00). Toulouse, France: SESAR Consortium. http://www.eurocontrol.int/sesar/gallery/content/public/docs/DLM-0612-001-02-00.pdf.

Jorna, P. G. A. M., Pavet, D., Van Blanken, M., & Pichancourt, I. (1999, August). *Phare ground human machine interface (ghmi) project* (Summary Report Nos. PHARE/NLR/GHMI-7/FR/1.0, DOC 99-70-02). Brussels, Belgium: Eurocontrol. http://www.eurocontrol.int/phare/gallery/content/public/documents/99-70-02ghmi.pdf.

Nuic, A. (2004, July). *User manual for the base of aircraft data (revision 3.6)* (Technical/Scientific Report No. EEC Note No. 10/04). Brétigny-sur-Orge, France: Eurocontrol Experimental Centre. http://www.eurocontrol.int/eec/public/standard_page/ACE_bada_documents_36.html.

Oprins, E. A. P. B., Burggraaff, E., & Van Weerdenburg, H. (2006, July). Design of a competence-based assessment system for air traffic control training. *International Journal of Aviation Psychology*, *16*(3), 297-320.

Oprins, E. A. P. B., & Schuver, M. (2003). Competentiegericht opleiden en beoordelen bij lvnl [competence based training and assessment at lvnl]. *HUFAG nieuwsbrief* (6), 2-4.

Tielrooij, M., in 't Veld, A. C., Mulder, M., & van Paassen, M. M. (2008). Assisting ATC in Metering, Sequencing and Merging Arrival Streams in Three-Degree Decelerating Approaches. *Proceedings of the 27th European Annual Conference on Human Decision Making and Manual Control, Delft, The Netherlands, June 12-13*.

TESTING A MULTIDIMENSIONAL NONVERIDICAL AIRCRAFT COLLISION AVOIDANCE SYSTEM, EXPERIMENTS 3, 4

William R. Knecht
Maneuver Space Technologies,
Pocasset, OK, USA

Veridical displays represent realistic scenes. State spaces are *non*veridical displays representing *n*-dimensional information. This research tests an aircraft separation maintenance display based on a nonveridical state space. In two experiments, licensed general aviation pilots flew flight scenarios, trying to deviate as little as possible from a pre-assigned course while still maintaining standard enroute separation from traffic. Flight performance using only a veridical cockpit display of traffic information with conflict alert capability was compared to performance augmented by a 4D non-veridical state space collision avoidance system. Results suggest that nonveridical display enhances operator performance on an aircraft separation maintenance task.

The present research examines an aircraft separation assurance display based on a nonveridical state-space. The term *veridical* means "coinciding with, or representing, physical reality." *State spaces* are *non*veridical representations common to engineering. A state space can be constructed from any quantifiable features, and can describe the state of a dynamic, multidimensional system at some current or future time *t*.

Motivation for This Work

Background. Currently, U.S. commercial aircraft do not fly point-to-point, but follow segmented jet routes in enroute airspace (the "long-haul" airspace starting about 40 miles [64 km] from airports). These jet routes add unnecessary travel distance and time. By enabling direct flight from departure to destination, airlines could lower fuel use by up to 6% (Operations Research and Analysis, 1998). Full implementation of direct flight will require advanced technology to minimize enroute air traffic conflicts (Krozel, 2000).

Enroute "conflicts" are defined as any two aircraft approaching within 5 nautical miles (nm) *and* 1,000 ft (9.3 km/304.8 m) of each other. Direct routing is expected to increase the base conflict rate because it transforms air traffic control (ATC) from a 2D spacing problem into a 3D spacing problem, increasing airspace complexity and conflict probability (Azuma, Neely, Daily, & Correa, 1999; Xing & Manning, 2005).

To minimize conflicts, veridical displays of traffic information have been developed, including map-view ATC displays and cockpit displays of traffic information (CDTI), 2D conflict resolution displays (Johnson, Battiste, & Holland, 1999), coplanar displays (Pekela & Hilburn, 1998; Thomas & Wickens, 2005), and 3D veridical displays (Canton, Refai, Johnson, & Battiste, 2005; Granada, Quang Dao, Wong, Johnson, & Battiste, 2005; Naikar, 1998).

Some systems make use of separation-maintenance technology to predict and even help resolve conflicts between aircraft. Cockpit variants of veridical *collision avoidance systems* (CAS) have been developed (Johnson & Battiste, 1999; van Gent, Hoekstra, & Ruigrok, 1998). The most widely known is TCAS ([Traffic Alert and Collision Avoidance System], Kuchar & Yang, 2000). However, TCAS has a short time lookahead. Strategists must now focus on systems with lookaheads sufficiently long to allow gentle aircraft maneuvers.

Veridical displays have difficulty displaying certain kinds of maneuver information. In response, researchers have turned to nonveridical display. For instance, NASA's En Route/Descent Advisor (Green & Vivona, 2001) allows aircraft spacing by positioning individual traffic icons on a slider representing desired arrival time-at-destination. Van Dam, Appleton, Mulder, and van Paassen (2006) tested a nonveridical CDTI allowing speed+heading combination maneuvers. Both devices have demonstrated their effectiveness on difficult air traffic scenarios.

Maneuver space. Knecht and Smith, (2001) proposed the concept of *maneuver space* (MS). Maneuver space has been defined by the military as "the physical space within which one can maneuver." Now, MS is redefined as *a 4D state space unique to each aircraft, dimensionalized by that aircraft's a) heading, b) speed, c) altitude, and d) available maneuver time. MS represents all conflictual and non-conflictual maneuvers achievable by that aircraft within a fixed period of time, given the obstacles predicted along each potential maneuver's hypothesized path*.

Maneuver space is a maneuver **hypothesis-tester**. It has seven key attributes:

1) Each translucent cube inside MS represents **one maneuver** (one autopilot setting of heading, speed, altitude).
2) Therefore, moving within MS represents **resetting the autopilot**.
3) **Colored MS represents "unsafe" maneuvers** (predicted to yield separation failure).
4) **Color** represents **available maneuver time** (minutes until separation failure).
5) **3D MS-center** represents **current autopilot setting**.
6) Therefore, **no avoidance is needed unless MS-center is colored.**
7) Maneuvers involving multiple conflicts are colored for the single conflict *closest in time*.

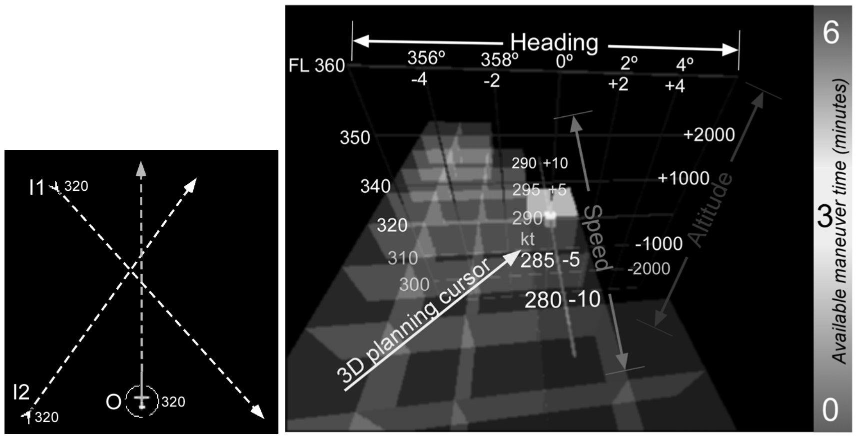


Figure 1. (Left) Three aircraft, all traveling 290 kt indicated airspeed (.78 mach) at flight level 32,000 (FL 320). The pilot's own ship (ownship, O) must maneuver to avoid two intruders (I1, 12). (Right) A view of the resulting 4D MS.

In Figure 1, the entire translucent, colored structure is called a *conflict region* (CR)—a set of numerically contiguous maneuvers predicted unsafe by a *conflict probe* (Kuchar & Yang, 2000), given a specified lookahead time.

Preliminary Development

A 4-Dimensional Collision Avoidance System (4CAS) was coded by the author. In "Experiment 1" (Knecht, 2007), eight general aviation (GA) pilots flew nominal straight-line courses threatened by traffic. Comparing CDTI-only trials (with no conflict alert or resolution capability) to 4CAS+CDTI trials, with 4CAS present, average path length, maneuver onset time, and duration of pilot deviations were significantly shorter, maneuver complexity was lower, and enjoyability-of-use was reported as significantly greater. In "Experiment 2" (Knecht, 2008), using four matched-pair, mirror-image scenarios with higher traffic density, with 4CAS present, 12 GA pilots averaged shorter path lengths, smaller deviations from path, greater minimum separations, shorter maneuver onset time, fewer (and briefer) pilot deviations, fewer types and numbers of maneuvers made, and reported greater ease of avoiding traffic.

System improvements were made. Below, results of the latest-generation system are reported.

Experiment 3

Method

Participants. Twelve GA pilots volunteered with informed consent, nine male, three female. Median age was 46.0 (range 20-69, mean 45.2, SD 15.5), median flight hours 995 (range 100-13300, mean 2025, SD 3645). All pilots held a private license, eight held instrument ratings, five were certified as both Certified Flight Instructor (CFI) and Certified Flight Instructor-Instrument (CFII), five held Commercial ratings. One held the Air Transport Pilot (ATP) rating. All received $50US for participating.

Apparatus. A part-task flight simulator similar to Knecht (2008) was used, based on Microsoft Flight Simulator (FS2004), with its Boeing 737-400 model and Artificial Intelligence (AI) Traffic. This simulated enroute air traffic and generated recordable latitudes, longitudes, headings, ground speeds, and vertical speeds for ownship and traffic.

The CDTI. A custom CDTI displayed a top-down, moving map of physical space, with ownship occupying display-center. Traffic was depicted as chevrons aimed in the direction of travel. Text data tags showed traffic flight level (FL). Zoom buttons allowed selectable map widths/heights of 5-200 statute miles ([sm], 8-322 km). In CDTI-only mode, pilots clicked directly on the B737 autopilot to maneuver.

The CDTI updated and wrote data to file every 2.5 s., except during a *pilot deviation* (PD, [FAA, 2006]), that is during failure to maintain 5 NM/1,000 ft aircraft separation. Then, sampling rate increased to 25 Hz.

Experiment 3's CDTI differed from that of Experiments 1-2, in that its traffic icons were also linked to the 4CAS conflict probe. When ownship separation was threatened, CDTI traffic icons were also colored by time to contact, using the same color scheme as 4CAS. The intent was to present a more challenging, fairer comparison of the two displays, in that the CDTI now alerted for separation failure (although not for possible solutions).

4CAS. 4CAS showed the MS and CRs corresponding to real-time traffic. Each CR's translucent, colored, cubes depicted autopilot settings predicted to lose separation with traffic within 6.0 minutes. Cube color represented available maneuver time (minutes-to-predicted separation failure). Colors were based on three anchor RGB values, with intermediate values linearly interpolated. A color/time reference bar was displayed under the MS.

The MS was rotatable around its vertical and horizontal axes. A 3D planning cursor moved within MS, allowing selection of avoidance maneuver. To resolve a conflict, users simply positioned the 3D cursor in a black "safe" region of MS and then hit the "Execute" button. This reset the B737 autopilot, initiating the maneuver. The 3D cursor was translucent, and stayed put after maneuver planning. A smaller cube represented real-time values of heading/speed/altitude. After maneuver completion, the display recentered itself to again represent current autopilot settings as occupying MS-center. A message box displayed "NO MANEUVER NECESSARY," changing to the alert "MANEUVER!!" as necessary.

Task. The overall task was to stay generally on-course (path+altitude), deviating for traffic as necessary, returning to course when clear of traffic. A red dot at the end of the nominal flight path signified the "destination." For greater accuracy, program shutdown was automatic, triggered by point-of-closest approach to destination.

Experimental design. Repeated-measures were used, with scenario presentation order counterbalanced by Latin squares. Half the 12 pilots started in the CDTI-only condition, flying the first four scenarios, followed by a short break, followed by the CDTI+4CAS condition using mirror-image scenarios in the same presentation order. The remaining pilots ran similarly, but with the CDTI+4CAS first. Pilots were not told they would repeat scenarios.

Flight scenarios. Like Experiment 2, Experiment 3 employed straight-and-level "primary conflict" traffic generated via the custom Traffic Creation Utility. FS2004's AI Traffic mode was used only to create distractor and blocking traffic for a single "standard background." During experimental trials, unique primary traffic was added to the standard background to create each individual traffic scenario.

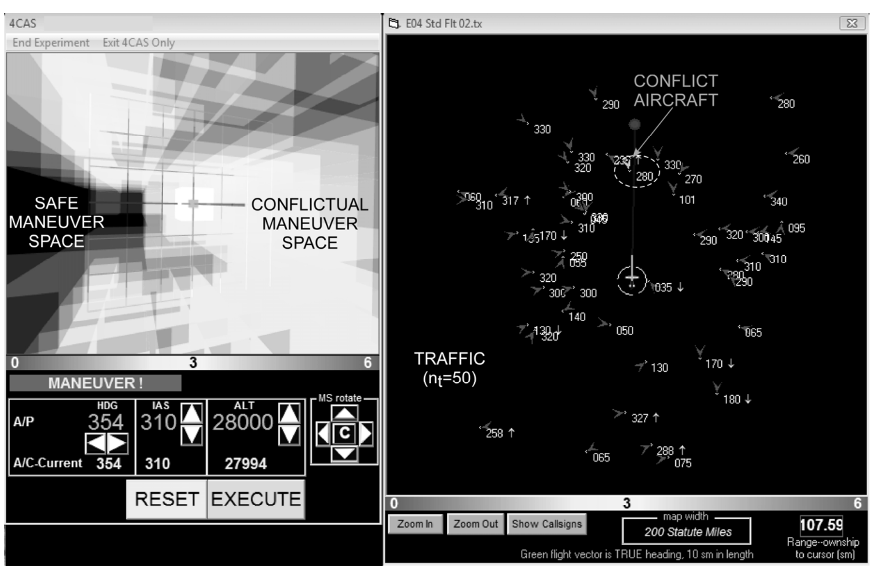


Figure 2. (Left) Annotated view of 4CAS display; (Right) CDTI, showing traffic from Experiment 4.

Figure 2 depicts scenario 2 (annotated, from Experiment 4). All Experiment 3 scenarios began in mid-flight, at 32,000 ft (FL 320), indicated airspeed (IAS) of 280 kt (.76 mach). Scenarios emulated enroute free flight (RTCA,

1995) in that aircraft were not restricted to normal odd-or-even flight levels by thousands (no "East-West Rule").

The CDTI portrayed the ownship flying nearly north (354º). Each of the five 10-min base scenarios had a mirror-image (generated by affine transform) for use as the repeated measure. All primary traffic (generated by the Traffic Creation Utility) converged toward the ownship straight and level from various angles, shaping the conflict.

All scenarios were "close calls" in both heading and altitude. To test false alarms, one mirror-pair contained a near-conflict, but technically required no avoidance.

Within the CDTI's maximum viewable area, each scenario maintained traffic density of 10-12 primary aircraft (median 11) plus an additional 11-16 secondary, distractor/blocking aircraft (median 12.5)—approximately double Experiment 1's primary traffic density, and triple its overall density. One participant, a professional FAA ATC instructor, judged the overall traffic densities as "moderate" (his word) compared to real-life, everyday enroute traffic.

Dependent measures. These are shown in Table 1 and detailed in Knecht (2007, 2008).

Training. Training was brief, about 25-30 min. Pilots received a one-page instruction sheet describing the task. They next received a one-page description of the CDTI and one for 4CAS, as appropriate. They then practiced on two training scenarios as desired before starting data collection. After completion of three test scenarios, pilots received a short break, and then retrained similarly for the second half.

Results

Table 1 summarizes relative performance of CDTI-only trials versus 4CAS+CDTI trials for 12 participants x 5 trial-pairs each = 120 total trials. Distributional non-normalities dictated nonparametric statistics (Hollander & Wolfe, 1999)—Wilcoxon's paired-ranks test, with McNemar's test for false alarms. DVs 2-9 reflect matched-scenario pair difference scores (4CAS+CDTI trial – CDTI-only trial). DVs 2-4, 6-9 are significant in favor of 4CAS.

Table 1. Experiment 3, CDTI-only vs. CDTI+4CAS trials.

	Dependent variable (DV)(1)	Median, (mean), or *n* (CDTI-only)	Median, (mean), or *n* (CDTI+4CAS)	P (2-tail)	P (1-tail)
		Efficiency Measures			
1	False alarms(2)	*n*=4	*n*=1	.250	
2	Unnormalized path length (sm)(3)	64.533	64.450	.0001	
3	Normalized 3D path length (std units--SU) (3)	12.146	11.971	.008	
4	3D maximum deviation from path (SU) (3)	1.048	1.040	.085	.043
5	Rmin (scenarios w no PDs, *n*=74, SU)	1.097	1.233	.078	
		Safety Measures			
6	Rmin (scenarios w ≥1 PD / pair, *n*=22, SU)	.987	1.240	.006	
7	Maneuver onset time (sec)(3)	43.4	35.0	.031	
8	Pilot deviations (experiment-wide counts)	*n*=13	*n*=1	.008	
9	Pilot deviations, average duration (sec)	(19.1)	(7.6)	.010	

(1) Measures 2-7 compare matched scenario pairs.
(2) Computed only for the 2 scenarios per pilot where maneuver was unnecessary (*n*=24)
(3) Computed only for the 8 scenarios where maneuver was necessary (*n*=96)

Rmin is the scenario-wide 3D normalized minimum range between ownship and traffic (Knecht and Hancock, 1999) where *x*- and *y*-differences reflect lateral separation (NM), and *z* reflects altitude differences (ft). *Rmin* can be used bimodally, as a measure of efficiency when separation is legal, and as a measure of safety when separation fails. Used as an efficiency measure, only *error-free* scenarios were averaged (no PDs). Less separation therefore implies greater efficiency, with no violation of mandated separation. Used as a safety measure, only *error* scenarios were averaged (those with PDs). Therefore, more separation implies greater safety.

Individual differences. Given that the CDTI now gave conflict alert, nine of 12 pilots in the CDTI-only condition independently discovered an interesting *maneuver titration* strategy. For example, a pilot might start a turn to solve a conflict. If, after completing that turn, the CDTI still showed conflict, the turn was increased by a degree or two, "titrating" the maneuver until the traffic icon changed color to indicate conflict resolution.

In many cases, maneuver titration proved efficient—sometimes more efficient than using 4CAS, if DV5, (Table 1) is all we consider. However: 1) DV2-3 were significant in favor of 4CAS, whereas DV5 was only a trend in favor of the CDTI; 2) Titration appeared significantly less safe (DV6-9); 3) If the pilot picked an inefficient maneuver to start with (e.g., a left turn instead of a more-efficient right turn)—then, titration exacerbated that inefficiency.

Experiment 4

Method

Participants. Eight licensed GA pilots volunteered with informed consent, seven male, one female. Median age was 50.0 (range 38-61, mean 48.6, SD 7.1), median flight hours 650 (range 138-1503, mean 729, SD 581). Four held instrument ratings, one was a Certified Flight Instructor (CFI), two held Commercial ratings. All received $50US.

Apparatus. The apparatus of Experiment 3 was used, with one exception: 4CAS was enhanced to subtract ownship maneuver execution time from time to contact. Maneuver execution time data were collected for a wide range of off-nominal maneuvers (±45º heading, ±35 kt IAS, and ±4000' altitude). Separate *h,s,a* modeling functions were parameterized by minimizing least-squares fit to FS2004 performance data. Log functions were selected to represent heading and speed changes. A linear function was selected for altitude changes. Modeling functions were coded into the 4CAS/CDTI time-to-contact algorithm, and allowed estimation of maneuver execution time to <10 sec accuracy.

Task, experimental design, dependent measures and training. These were similar to Experiment 3.

Flight scenarios. These were similar to Experiment 3, with a few exceptions. First, ownship starting altitude was lowered to 28,000' to allow more headroom, with initial speed set at 310 kt IAS (.76 Mach). Second, with false alarm rate having been explored in Experiments 1-3, the no-conflict scenarios were deemed unnecessary. Four mirror-image conflict scenario pairs were therefore tested per pilot.

Finally, traffic density was more than doubled from Experiment 3. Primary traffic was 24-28 aircraft (median 25.5), plus an additional 20-27 secondary, distractor/blocking aircraft (median 23.5). One participant, a professional FAA ATC tower control instructor, judged the overall densities as "heavy" (his word) compared to real-life, East-coast traffic (itself some of the U.S.' heaviest traffic).

Results

Table 2. Experiment 4, CDTI-only vs. CDTI+4CAS trials.

	Dependent variable(1)	Median, (mean), or *n* (CDTI-only)	Median, (mean), or *n* (CDTI+4CAS)	P (2-tail)
	Efficiency measures			
1	Unnormalized path length (sm)	65.998	65.628	.002
2	Normalized 3D path length (std units, SU)	12.717	11.714	.0002
3	3D maximum deviation from path (SU)	1.287	.696	.00004
4	Rmin (scenarios w no PDs, *n*=38, SU)	1.075	1.105	.872
	Safety Measures			
5	Rmin (scenarios w ≥1 PD / pair, *n*=26, SU)	.944	1.047	.028
6	Maneuver onset time (sec)	36.6	29.1	.001
7	Pilot deviations (experiment-wide counts)	*n*=17	*n*=4	.028
8	Pilot deviations, average duration (sec)	(20.7)	(4.0)	.002

(1) Measures 1-6 compare matched scenario pairs.

Table 2 summarizes the relative performance of CDTI-only trials versus 4CAS+CDTI trials for 8 participants x 4 trial-pairs each = 64 total trials. Measures 1-3, 5-8 are significant in favor of 4CAS.

Individual differences. Here, six of eight individuals titrated their maneuvers in the CDTI-only condition. With a correct initial guess, the results were generally good. However, incorrect guesses led to far more effort with far poorer results. Given the high traffic density and complexity, incorrect guesses were common.

Discussion

Veridical means "coinciding with, or representing, physical reality." *Maneuver space* is defined here as a 4D *non*veridical state space unique to each aircraft, dimensionalized by that aircraft's a) heading, b) speed, c) altitude, and d) available maneuver time. Maneuver space represents conflictual and non-conflictual maneuvers achievable by that aircraft within a fixed period of time, given obstacles predicted along each potential maneuver's path.

This work constitutes Experiment 3 and 4 in a series of tests of a nonveridical, MS-based 4D collision avoidance system called 4CAS. 4CAS is not meant to replace veridical traffic displays—merely to augment them.

In Experiment 3, 12 licensed GA pilots flew five matched-pair, mirror-image scenarios with traffic and geometry similar to Experiment 2. To provide a more competitive comparison, the CDTI was enhanced to add conflict alert (but not resolution) capability. The CDTI+4CAS condition showed performance superiority over the baseline CDTI for three out of five dependent measures of maneuver efficiency, and four of four measures of maneuver safety.

In Experiment 4, eight licensed GA pilots flew four matched-pair, mirror-image scenarios with very heavy traffic (median=49)—double that of Experiment 3. Maneuver execution time was subtracted from the available maneuver time on both displays. The CDTI+4CAS condition showed performance superiority over the baseline CDTI for three out of four dependent measures of maneuver efficiency, and four of four measures of maneuver safety.

Taken together, the entire series of four experiments suggests that human operators can safely, effectively use such a 4D nonveridical aircraft maneuver safety display.

Acknowledgments

This project was funded by a Lindbergh Foundation grant. Sincere thanks go out to the pilot participants.

References

Azuma, R., Neely III, H., Daily, M., & Correa, M. (1999). Visualization of conflicts and resolution in a "free flight" scenario. *Proceedings of IEEE Visualization '99* (pp. 433-436, 557).

Canton, R., Refai, M, Johnson, W.W., & Battiste, V. (2005). Development and integration of human-centered conflict detection and resolution tools for airborne autonomous operations. *Proceedings of the 13th International Symposium on Aviation Psychology* (pp. 115-120). Columbus, OH: Ohio University Press.

Federal Aviation Administration. (2006). *Order 7110.65R Air traffic control.* Washington, DC: Author.

Granada, S. Quang Dao, A., Wong, D., Johnson, W.W., & Battiste, V. (2005). Development and integration of a human-centered volumetric cockpit situation display for distributed air-ground operations. *Proceedings of the 13th International Symposium on Aviation Psychology.* (pp. 279-284). Columbus, OH: The Ohio University Press.

Green, S.M., & Vivona, R.A. (2001, August). *En route Descent Advisor Concept for Arrival Metering.* Guidance, Navigation, and Control Conference, Montreal, Canada.

Hollander, M., & Wolfe, D.A. (1999). *Nonparametric statistical methods* (2^{nd} ed.). New York: Wiley.

Johnson, W.W., Battiste, V., & Holland, S. (1999). A cockpit display designed to enable limited flight deck separation responsibility. *SAE Transactions, Journal of Aerospace, 108*(1), 1041-1062.

Knecht, W.R. (2007). Testing a nonveridical aircraft collision avoidance system: Experiment 1. *International Journal of Applied Aviation Studies*, 7(1), 60-82.

Knecht, W.R., & Hancock, P.A. (1999). Separation maintenance in high-stress free flight using a time-to-contact-based cockpit display of traffic information. *Proceedings of the 4th Annual Meeting of the Human Factors and Ergonomics Society*. (pp. 16-20). Santa Monica, CA: Human Factors and Ergonomics Society.

Knecht, W.R., & Smith, K. (2001) The manoeuvre space. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics.* Aldershot: Ashgate.

Krozel, J. (2000). *Free flight research issues and literature search.* Los Gatos, CA: Seagull Technology.

Kuchar, J.K., & Yang, L.C. (2000). A review of conflict detection and resolution modeling methods. *IEEE Transactions on Intelligent Transportation Systems. 1*(4), 179-89.

Naikar, N. (1998). *Perspective displays: A review of human factors issues.* Retrieved December 13, 2006 from Australian Government, Dept. of Defense, Defence Science and Technology Organisation Web site: http://www.dsto.defence.gov.au/ publications/2069/DSTO-TR-0630.pdf.

Pekela, W.D., & Hilburn, B. (1998, Sept). *Air traffic controller strategies in resolving free flight traffic conflicts: The effect of enhanced controller displays for situation awareness.* AIAA and SAE World Aviation Conference, Anaheim, CA.

Radio Technical Commission for Aeronautics (January, 1995). *Report of the RTCA board of directors' select committee on free flight.* RTCA, Incorporated, Washington, DC.

Thomas, L.C., & 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 49^{th} Annual Meeting*. (pp. 40-44). Santa Monica, CA: Human Factors and Ergonomics Society.

Van Dam, S.B.J, Appleton, R., Mulder, M., & van Paassen, M.M. (2006, September). *Comparison of two interfaces for supporting pilots in airborne self-separation tasks.* International Conference on Human-Computer Interaction in Aeronautics, Seattle, WA.

van Gent, R.N.H.W., Hoekstra, J.M., & Ruigrok, R.C.J. (1998, May). *Free flight with airborne separation assurance: A man-in-*

the loop simulation study. International Conference on Human-Computer Interaction in Aeronautics, Montreal, Canada.

HOW EFFECTIVE IS ITEM BANK TESTING OF PILOT TRAINING APPLICANTS IN REDUCING TEST PREPARATION EFFECTS?

Oliver Zierke
German Aerospace Center (DLR)
Aviation and Space Psychology
Hamburg, Germany

In the selection of aviation personnel, special test preparation has become an emerging problem. Specific test preparation aims at raising the probability to master a certain test rather than developing the underlying ability. Knowledge tests are particularly susceptible to the problem of test preparation. One strategy to counter this problem is the use of comprehensive item banks for testing in knowledge domains. In 2005 over 770 student pilot applicants participated in an evaluation study of two item bank tests, an English language test and a test of physical knowledge. A conventional test form as well as an item bank test form were given to each subject. Consequently, both test forms were compared in a repeated measures design. The test preparation effects, correlations with school grades, and prognostic validity of both tests were analyzed. It is shown that item bank testing reduces test preparation effects and enhances construct validity.

Parallel to the rapid changes in the aviation business, a new challenge in the selection of student pilots must be realised: The problem of test preparation offered through new media, such as the internet, or by commercial training institutes. This kind of specific test preparation aims at raising the probability to master a certain test rather than developing the underlying ability. For a pilot training applicant, the successful accomplishment of a selection procedure can result in sponsored flight training, financed by a few larger commercial airlines. Compared to a private pilot training, such sponsorships can provide a suitable applicant with several ten thousand Euros worth of training. Testing in aviation is therefore referred to as “high stake testing”. Thus, it is quite understandable that applicants are willing to try everything to prepare optimally, and a test preparation market has evolved for satisfying this need. In Germany at least four commercial institutes, one commercial CD with training material, and two internet chat rooms exist exclusively for the preparation for the DLR (German Aerospace Center) test. For the applicant, as well as for the preparation institute, it matters little whether the student actually improves his aptitude or general knowledge or whether he simply improved his ability to solve one specific test. The latter would be the case if an applicant has access to the questions of a test, e.g. a technical comprehension test, prior to taking it. He could possibly memorize these very items and their correct solutions without any in-depth understanding of the subject.

Test preparation effects are defined as achieving higher scores without real knowledge of the underlying domain. These effects lead to an overestimation of the ability of a dishonestly prepared candidate. This example demonstrates the threat of test preparation for selection in aviation business, because test fairness and test validity can be compromised. This can lead to incorrect selection decisions, which in the long run may have an impact on aviation safety. For this reason countermeasures are necessary.

Countermeasures

The problem of specific test preparation concerning aptitude tests is answered by constructing new tests regularly and by offering detailed pre-information and own training material to the applicants (Huelmann & Oubaid, 2004). Countermeasures concerning knowledge tests differ from those for aptitude tests. Knowledge tests are frequently used for licensing purposes (Impara, 1995) or for measuring basic requirements for an apprenticeship. Therefore, knowledge tests play a prominent role in aviation psychology. The problem of test preparation is of particular importance for knowledge tests, because it is not difficult for applicants to publish via internet memorized items from the test after completing the examination and to provide future applicants with preparation material. A method to counter the preparation problem for knowledge tests is the use of comprehensive item banks instead of fixed tests. Using item banks lowers the predictability of items for test takers and thus may encourage them to prepare for the whole subject of the test rather than merely for the known individual items.

The approach of DLR

At the German Aerospace Center, item banks were installed for the knowledge domains of physics, mechanical comprehension, mathematics, and English language. For every individual test form, items are randomly drawn from the item bank, while maintaining a balance of item difficulty, test standard deviation and reliability for all forms (Figure 1). This procedure is based on a method developed by Gibson and Weiner (1998) and leads to different test forms for each applicant.

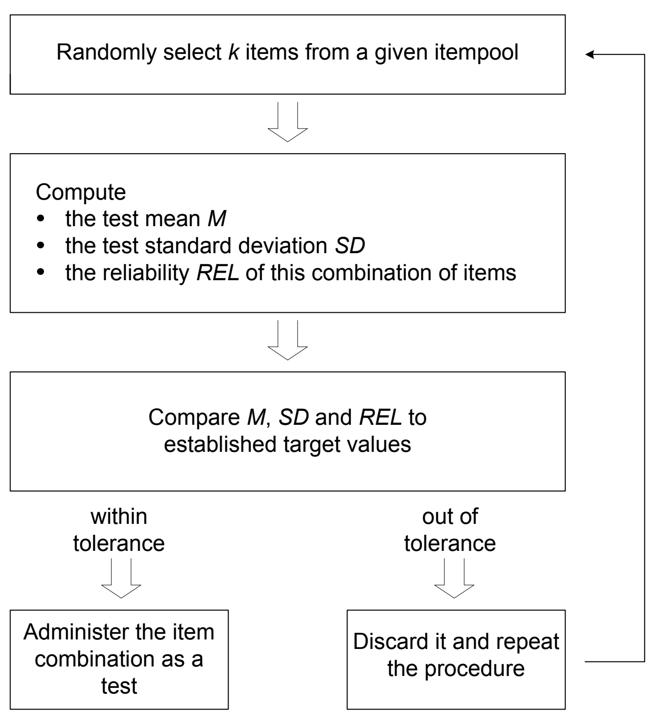


Figure 1. The procedure of test assembly

Method

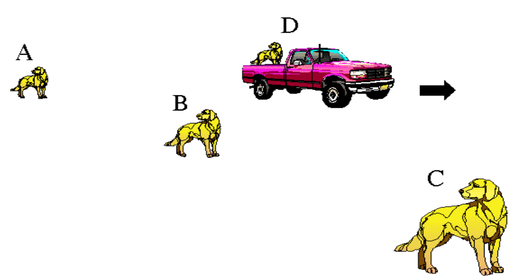
An evaluation study of item bank testing as a means of reducing test preparation effects was conducted. In 2005 over 770 student pilot applicants participated in this study. The English item bank was composed of four parallel tests which were active during the past in the DLR pilot selection. This assembly resulted in an item bank comprising 204 items. A single test drawn out of this item bank consisted of 60 items and had an internal consistency of Cronbach`s $\alpha = .90$. The item bank of physical knowledge consisted of 104 completely new items. A resulting single test comprised 40 items and had an internal consistency of Cronbach`s $\alpha = .78$. Item examples are shown in Figure 2. A conventional test form as well as an item bank test form were given to each subject. Consequently, both test forms were compared in a repeated measures design.

English Test

To participate ..... the Olympics must be a real thrill.

1) on
2) in
3) by
4) at

Physical Knowledge Test



“Which dog hears the sound of the pickup truck in the highest frequency?”

Figure 2. Item examples for English and Physical Knowledge Test

Subjects were requested in a questionnaire to disclose any commercial preparation. In this study a test preparation effect is calculated as the mean difference in test scores between the two groups of candidates, one which was commercially prepared and another which was not. The following hypotheses have been addressed:

Hypotheses

1. Test preparation effects are smaller for item bank tests when compared with those of conventional tests.
2. Item bank tests show larger correlations with school grades than the conventional tests do.
3. Item bank tests show higher prognostic validity than conventional tests.

Results

Test preparation

For the English test 34 of 451 applicants disclosed they have attended a commercial preparation course. This makes a preparation rate of 7.5%. For the Physical Knowledge Test only 16 of 314 applicants disclosed a commercial preparation course. This means a preparation rate of 5.1%. The English and the Physical Knowledge Test have been administered at different times, thus the difference in the preparation rates is explainable.

Hypothesis 1

The first two hypotheses were confirmed completely. Item bank testing reduces test preparation effects for both the English and the Physical Knowledge Test (see Figure 3 and 4). For both tests the ANOVA interaction effect became highly significant with $F(1, 449) = 40.0$ for English, and $F(1, 312) = 46.4$ for Physics. This resulted in a medium effect for the English test ($f = 0.30$) and a large effect for the Physical Knowledge test ($f = 0.39$).

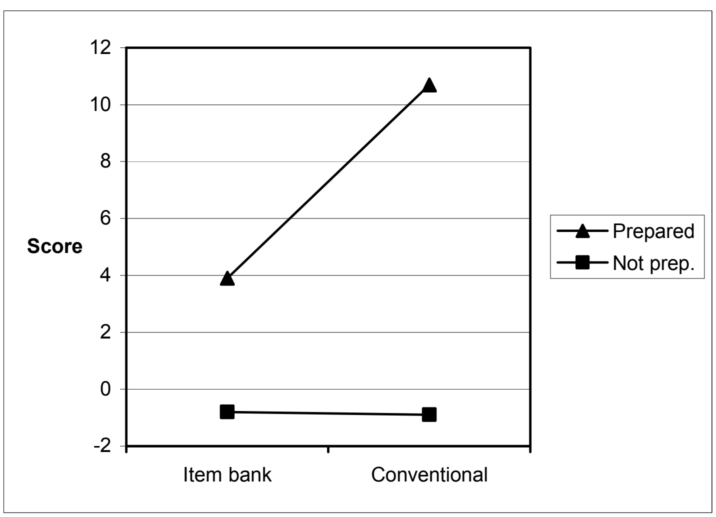


Figure 3. Test preparation effects for English Test

For not specially prepared applicants there was no difference, whether they got an item bank test or a conventional test form. In both tests they reached nearly the same result. In contrast, specially prepared applicants achieved much higher scores in the conventional tests, presumably because they already knew some items. Therefore, item bank tests yield more realistic measurements of aptitudes for prepared applicants in particular.

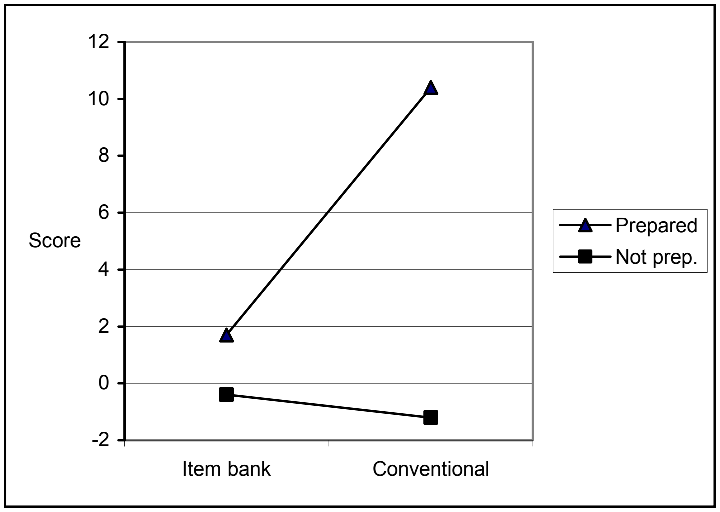


Figure 4. Test preparation effects for Physical Knowledge Test

Hypothesis 2

Item bank testing enhances construct validity in form of correlations with school grades. The respective correlation for the item bank test is significantly higher than for the conventional test (r = .51 vs. r = .43, N = 379, p < .01 for the English test and r = .39 vs. r = .25, N = 266, p = .01 for the Physical Knowledge test). That means that item bank tests measure more true variance than conventional tests.

Hypothesis 3

For a definite confirmation concerning the third hypothesis, the data could not be interpreted clearly because too few applicants were recommended for pilot training to calculate stable correlations.

Discussion

Item bank testing turned out to reduce test preparation effects in the selection of pilot training applicants. This is an important result because commercial test preparation is a challenge to test fairness as well as to test validity. Consequently, the second question was whether reduced test preparation effects will improve test validity. Indeed, item bank testing raised the correlations with school grades as an aspect of construct validity, which means that item bank tests measure more true variance than conventional fixed tests. This result is not surprising. It seems obvious that large item banks reduce the predictability of items for prepared applicants and thus improve the quality of measurement in terms of test fairness and validity. It has never been shown before how effective item banks are in contrast with conventional tests. The item banks reduced test preparation effects, although not to zero. Prepared applicants are still better than not specially prepared ones. Why? The question is whether these differences are true differences, e.g. if prepared candidates really learned and understood more than the unprepared group. If so, they must achieve better results. It seems plausible that candidates who invest more time in their preparation are on average more motivated and consequently achieve better results. Therefore, we should not aim for tests yielding equal results for prepared and unprepared applicants. Rather, we should ensure that possible differences between both groups represent true differences. With regard to the third hypothesis further research is needed to learn more about the effects of item bank testing on the prognostic validity of knowledge tests.

References

Gibson, W. & Weiner, J.A. (1998). Generating random parallel test forms using CTT in a computer-based environment. *Journal of Educational Measurement*, *35*, 297-310.

Huelmann, G. & Oubaid, V. (2004). Computer assisted testing (CAT) in aviation psychology. In: K.-M. Goeters (Ed.), *Aviation psychology: practice and research* (pp. 123-134). Aldershot: Ashgate.

Impara, J.C. (Ed.) (1995). *Licensure testing: purposes, procedures, and practices.* Lincoln: Buros institute of mental measurement.

VIGILANT WARRIOR™: A SELECTION TOOL FOR VIGILANCE PERFORMANCE

David E. Weldon, Ph. D.
JXT Applications, Incorporated
Beavercreek, Ohio
Clark Shingledecker, Ph. D.
Wright State University
Dayton, Ohio

In this paper, we describe an individual differences model of vigilance performance—the ability to maintain one's focus of attention and remain alert for prolonged periods of time—and summarize our model evaluation research. Our goal was an automated test battery (*Vigilant Warrior*™) that could be employed to select personnel with superior abilities for assignment to critical vigilance duties. Thus, we conducted extensive laboratory research to identify an optimal set of vigilance predictors and validate them against a simulated, real-world, electronic-display, battlefield-monitoring task with high vigilance requirements. The results confirmed that an objective, Short Vigilance Task (SVT), coupled with analytic skill and stress-coping measures, could account for 33% or more of the criterion variance. Moreover, the SVT was the most powerful predictor in the battery. Analytic skill and situational variables contributed to vigilance performance, but to a lesser degree. *Vigilant Warrior*™ is currently receiving extensive field testing in military settings.

Vigilance is the ability to maintain one's focus of attention and remain alert for prolonged periods of time. As such, vigilance is a key cognitive attribute for exceptional performance over a wide range of work domains where the ability to detect and respond to relatively rare and sometimes obscure events must be sustained despite lengthy duty requirements. Tasks requiring a high degree of vigilance are an integral to warfare. In addition to conventional visual monitoring activities, the modern warfighter is likely to engage in computer-mediated monitoring tasks associated with control of aircraft, missiles, unmanned aerial vehicles, or combat robots, and perform detection tasks in efforts to counter enemy threats. Past research has shown that individuals vary widely in their capacity to be vigilant in these situations. Therefore, a need exists to identify and selectively assign individuals with exceptional vigilance performance capabilities to critical jobs with high, sustained attention demands. This paper summarizes the theoretical basis for the development of *Vigilant Warrior*™: a new personnel selection battery designed to identify individuals who display exceptional vigilance performance. It also describes the results of research conducted to refine and validate the predictive abilities of the *Vigilant Warrior*™ battery.

A Model for Development of a Vigilance Selection Test Battery

Previous attempts to identify measures or factors reflecting differences among individuals that reliably predict vigilance performance have been largely unsuccessful. One likely reason for this failure is that approaches that were taken to the problem were typically based solely on single personality characteristics. We developed the *Vigilant Warrior*™ test battery to remedy this shortfall by adopting a multidimensional view of the prediction problem, guided by current theoretical treatments of vigilance and a by a broad examination of past vigilance research findings. This perspective raises the possibility that improved vigilance prediction may be possible by combining information derived from classical personality variables with measures of intelligence, sample vigilance task performance, and measures of the person's characteristic responses to vigilance task demands. We summarize the literature supporting this approach to predicting individual differences in vigilance in the following paragraphs.

Personality factors. Davies & Parasuraman (1982) summarize the findings for personality dimensions related to vigilance performance; including introversion-extraversion (introverted observers outperform their extraverted cohorts), field dependence-independence (field-independent individuals outperform field-dependent observers), internal-external locus of control (individuals with an internal locus of control outperform those with an external locus of control), and the Type A (coronary-prone) behavior pattern (achievement-oriented Type-A individuals outperform their more relaxed, Type-B counterparts). In addition, Thackray, Bailey, & Touchstone (1977) found that boredom prone individuals may be poorer monitors than those less boredom prone while Robertson, et al. (1997) found that absent-minded individuals, defined by high scores on the Cognitive Failures Questionnaire, did more poorly in than non-absent minded observers and reported higher levels of perceived mental workload than the non-absent minded. Finally, Helton, Dember, Warm, & Matthews (1999) found that optimists perform more effectively on vigilance tasks than do pessimists. Such results indicate that personality profiles should be included as candidates for any approach for developing a vigilance test with reliable predictive features.

Performance sampling as a predictor. A second promising source of predictors of sustained attention ability is the objective measurement of an individual's performance on vigilance tasks themselves. However, traditional laboratory vigilance tasks require a lengthy watch period that would make them impractical as selection tests for large groups of examinees. Recent research, however, shows that brief, highly-demanding, vigilance tasks can be constructed that produce performance that mirrors the vigilance decrements typically observed in long-term vigils (e.g., Matthews, Davies & Lees, 1990; Temple et al., 2000). These tasks show rapid perceptual sensitivity decrements over a period of 10 minutes or less. They also they demonstrate the key diagnostic indicators of being resource-limited: sensitivity decrement, high subjective workload, and sensitivity to stress and arousal factors. Thus, a high level of performance on a short task may be a good indicator of aptitude for longer vigilance tasks.

Differences in subjective responses to vigilance task demands. Finally, recent studies indicate that the perceived workload of vigilance tasks is quite substantial and that workload grows linearly over time (Warm, Dember, & Hancock 1996). Johnson & Proctor (2003) conclude that, rather than being under-stimulating, vigilance tasks place high information-processing demands upon observers. Thus, Resource Theories appear to take precedence over Arousal Theory as models of the factors controlling vigilance performance. However, following Kahneman (1973), Matthews and Davies (2001) argued that Arousal and Resource Theories are not necessarily mutually exclusive, and that they can be integrated by viewing arousal as the agent responsible for resource production. The finding that there seems to be agreement between psychophysiological measures, subjective self-reports, and performance, as predicted by the integrated models, is of considerable significance for selection test development. In addition to workload response differences, Hancock & Warm (1989) found that operators differed in the way they deployed compensatory effort and coping strategies to adapt to demanding performance environments. Short tasks are sometimes insensitive to stressor effects, but as time progresses it becomes increasingly more difficult for the operator to maintain successful coping. Therefore, it may be possible to identify useful predictor measures from an operator's reactions to performing a short vigilance task, which may offer early warning signs of difficulties in coping.

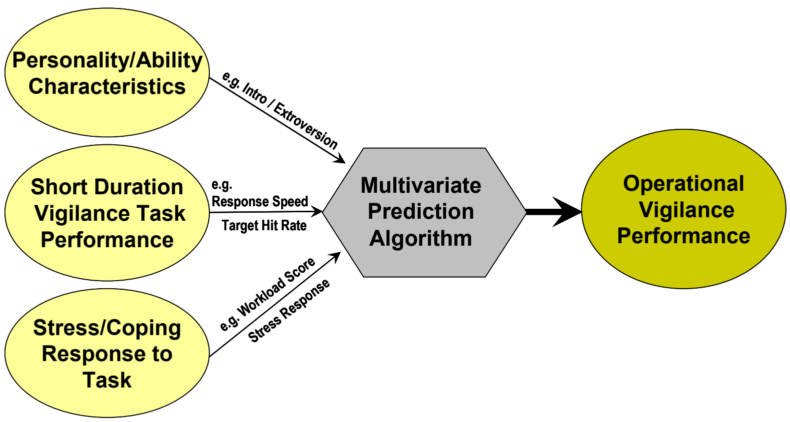


Figure 1. *Vigilant Warrior*™: A model approach to developing a personnel selection tool for sustained attention ability

The proposed model. The challenge presented for developing *Vigilant Warrior*™ was to apply the concepts of vigilance and its measurement discussed above to develop a reliable and valid vigilance prediction toolset. The multidimensional solution to vigilance prediction that was conceived to meet this challenge was to sample key constructs related to (1) personality and analytic skill, (2) objective task performance, and (3) stress, workload and coping responses to vigilance tasks. A primary goal was to extract the optimal measurement instruments from these complimentary approaches and blend them to produce an efficient personnel selection system capable of predicting vigilance performance. A graphic representation of the *Vigilant Warrior*™ personnel selection battery concept is shown in Figure 1.

Preliminary Research and Test Battery Selection

To identify preliminary components for each of the three vigilance prediction dimensions discussed above, we examined the literature addressing the relationship between various personality and analytic skill variables and vigilance performance and documented the limitations and strengths of identified vigilance predictors. Finally, a panel of experts rated the degree of research support and projected utility for each personality dimension. In addition, available brief vigilance tasks were assessed for inclusion in the battery, as well as subjective rating dimensions and scales that could be used to determine an examinee's perceived workload, coping responses, and attitudes associated with performing the vigilance task. Based on the results of these analyses, we developed a candidate vigilance prediction battery composed of personality/analytic skill metrics, brief vigilance-task performance metrics, and resource depletion and allocation metrics. The personality dimensions selected for preliminary research were: Introversion/Extraversion, Intelligence Quotient, Boredom Proneness, Cognitive Failures, Conscientiousness, Trait Sleepiness, Attention Deficit Hyperactivity Disorder (ADHD), Schizotypy, and Propensity to Daydream. Two measures of Analytic skill, Fluid and Crystallized Intelligence, rounded out this group of measures. Two versions of a Short Vigilance Task (SVT) were created for the battery in order to account for the well-known differences in performance and sensitivities to stimulus and environmental variables observed in tasks with (*simultaneous*) and without (*successive*) a comparison stimulus available to classify an event as a signal or a non-signal. The task is a brief (12-minute), paired-symbol vigilance task. Events are presentations of letter pairs in any combination drawn from the letters D, O, and backward D. In the *simultaneous* trials, the signal is any matching pair (e.g., "DD"). In the *successive* version, the signal is defined as the occurrence of the pair "OO." Finally, the Dundee Stress State Questionnaire (DSSQ), the Coping Inventory for Task Situations (CITS), the National Aeronautics and Space Administration Task Load Index (NASA TLX) workload scale, and the Boles Multiple Resource Questionnaire were selected to assess subject attitudes toward, and responses to, performing the SVT. Dimensions assessed by these instruments are Task Engagement, Distress, Worry, Coping (task focused), Coping (avoidance), Coping (emotion focused), Workload, and Multiple Resource Usage.

Refinement Of The Initial Battery

The goals of the main preliminary investigation of the candidate vigilance test battery was to confirm the qualities of the SVT, assess the psychometric properties of the personality, intelligence, and stress/attitude/coping measures to be included in the battery, and to assess their differential abilities to predict vigilance performance on the SVT. The study was conducted with a sample of 210 participants recruited from psychology classes at the University of Cincinnati.

Method. Participants completed a series of questionnaire and performance-based assessments in the following sequence: personality tests, intelligence tests; pre-task stress state, 12-minute SVT; and post-task stress state and coping. During the SVT the character pairs were presented against a masking background at a high event rate. One hundred five (105) participants performed the *simultaneous* version of the task, requiring a comparative judgment to detect the target, while 105 additional participants performed the *successive* version of the task, requiring an absolute judgment to detect the target.

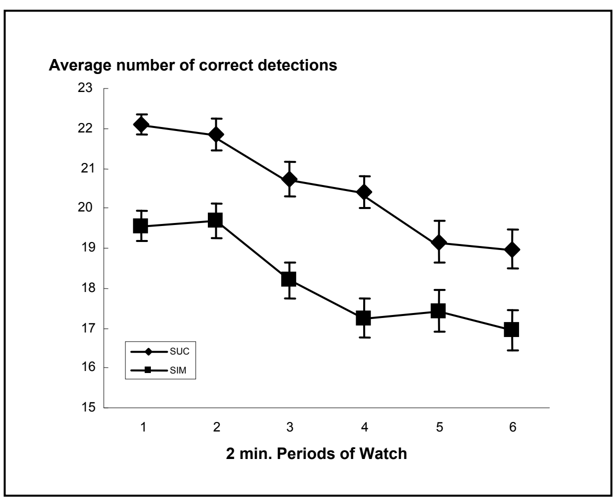


Figure 2. Mean number of correct detections as a function of periods of watch for both *simultaneous* (SIM) and *successive* (SUC) conditions.

Validity of the SVT. One objective of this study was to ensure that the SVT developed for the battery would show the classic performance changes over time that are characteristic of typical longer tasks. Figure 2 shows the average number of correct detections made by subjects performing the *successive* (SUC) and *simultaneous* (SIM) versions of the test over the six continuous 2-min. watch periods. As the graph suggests, the short tasks yielded a common decrement in performance over the 12-min. watch ($F_{(5, 1248)} = 44.74$, $p < .001$.) and a clear difference between the task conditions (F (1, 208) = 19.80, $p < .001$).

Factor analysis of the personality scales. A factor analysis was conducted to test whether the initial set of personality dimensions could be reduced to a smaller number of underlying factors. Analysis of the personality scales showed that these individual difference indicators were intercorrelated. A principal factor analysis was run, followed by an oblique (direct oblimin) rotation. On the basis of the scree test and factor interpretability, a four-factor solution was extracted, explaining 63.7% of the variance. Factor 1 (labeled Cognitive

Disorganization) is defined by various scales linked to disruption of attentional focus, including cognitive failures, mind wandering, and daydreaming, as well as the Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE) Disorganization scale and the Young ADHD Questionnaire-Self-Report (YAQ-S). Factor 2 (Heightened Experience - i.e., enjoyment of events) is defined by O-LIFE unusual experiences and sensation-seeking subscales, and low internal boredom score of the Boredom Proneness Scale. This factor appears to indicate a vivid, excitable mental life. Factor 3 (Sleep quality) brings together the 3 subscales of the Pittsburgh Sleep Quality Index used in the study. Surprisingly, the Epworth Sleepiness Scale (Johns, 1994) fails to load on this factor. Factor 4 (Impulsivity) contrasts the sensation-seeking subscale of the Urgency, Premeditation, Perseverance, and Sensation Seeking (UPPS) scale with the low-premeditation subscale on the I_7 Impulsiveness Questionnaire. The factors were intercorrelated, with the highest correlations found between factors 1 and 4 (r = .51) and between 1 and 3 (r = .44). Factor 2 was largely uncorrelated with the remaining factors.

Correlates of SVT performance. Satisfied that the SVT possesses the fundamental characteristics of a more classical extended-duration task, we examined the Pearson correlations between the SVT and the personality and situational measures. Personality was represented by regression-model factor scores computed on the basis of the factor analysis. Detection frequencies within each 2-min. period were highly intercorrelated (alpha = .93), so average target detection frequency was used as the performance measure for this analysis. Table 1 provides a summary of the correlations of the various scales with performance, for *simultaneous* and *successive* conditions.

Table 1. *Correlations Of Intelligence And Stress Variables With Performance.*

Test Type	Test/Questionnaire	Simultaneous	Successive
Intelligence	Advanced Vocabulary	.084	.294**
	Letter Sets	.274**	.259**
Personality	Cognitive disorganization	-.099	-.089
	Impulsivity	-.170	-.132
	Heightened awareness	.090	.048
	Sleep quality	-.033	.077
Stress (pre)	Engagement	.359**	.122
	Distress	-.135	-.089
	Worry	-.156	-.152
Stress (post)	Engagement	.456**	.402**
	Distress	-.199*	-.180
	Worry	-.120	-.172
Coping	Task-focused	.284**	.402**
	Emotion Focused	-.230*	-.181
	Avoidance	-.429**	-.303**

Note: **Correlation is significant at the .01 level.
*Correlation is significant at the .05 level.

Table 1 shows that the two measures of Analytic skill positively correlate with performance on the SVT. The Educational Testing Service (ETS) Advanced Vocabulary test (Crystallized Intelligence) is a better predictor of performance on the *successive* task, while the ETS Letter Sets test (Fluid Intelligence) correlates with both the *simultaneous* and *successive* tasks. The other correlates with the SVT were the subjective stress states and coping-style measures. Table 1 also suggests that, while *simultaneous* and *successive* tasks have some common correlates, the set of correlates for each type of task may differ somewhat.

Candidate Test Battery for Validation

This preliminary study confirmed that the SVT showed the vigilance decrement characteristic of performance of longer monitoring tasks, qualifying it as the performance sampling component of the battery. The data also replicated findings that personality traits are no more than modest predictors of vigilance**.** However, additional analyses showed that some of the personality factors predict stress and coping during vigilance, which may contribute to their utility in prediction for a longer, sustained monitoring task. In addition, the present data support inclusion of short intelligence tests in the predictive battery. Thirdly, both stress states and coping scales correlated with performance, supporting inclusion of these measures in the battery. Finally, the analyses permitted reductions in both the number of tests and the number of test items in the battery. These reductions allowed construction of a 45-minute automated test battery to be used in the *Vigilant Warrior*™ battery validation study.

Criterion Validation Study

The vigilance criterion task designed to test the predictive capabilities of the *Vigilant Warrior*™ battery employed a simulated, tactical, situation display presented on a computer monitor to provide a two-dimensional plan-view map of a geographical area within which the positions of military combat vehicles were represented. Static components of the display included terrain features and reference grid lines. The dynamic components of the display were moving combat vehicles, the positions of which changed with each display update. The symbolic combat vehicles appeared in three columns that moved from left to right across the screen and returned in the opposite direction with unpredictable directional deviations. The center column of combat vehicles was led by a combat tank with two gun barrels. The display was updated every second, with the gun barrels displayed for 50 msec. Participants were required to report a detection whenever the gun barrels were of different lengths (*simultaneous* condition), or are both were longer than the standard length (*successive* condition). Two additional versions of the *successive* criterion task were created to examine the battery's capacity to predict performance under special task conditions and the concurrent cognitive demands that accompany many real-world vigilance tasks. The target cueing version was intended to simulate vigilance tasks augmented by probabilistic information about potential upcoming signals during screen display updates. The second version of the criterion task represented the common vigilance condition in which the worker is engaged in an additional task; in this case, a secondary auditory task to answer queries about the location of specific vehicles on the map. This additional task was designed to increase the mental resource demands imposed upon the subject to permit testing the ability of the battery to predict vigilance performance under multitasking conditions.

Criterion tasks. Task duration was 60 minutes in all cases, analyzed as 6 successive 10-min. periods of work. Correct detections and false positive responses were recorded for all task versions. The signal detection theory index of perceptual sensitivity, *d*' (Macmillan & Creelman. 2005), was calculated from these response data and was employed as the principal performance index in the validation study.

Participants and procedure. A total of 462 participants were recruited. They were allocated at random to the four criterion task conditions as follows: *Simultaneous* detection task (110), *Successive* detection task (122), *Successive* detection task *with cueing* (122), *Successive* detection task *with auditory competing task* (108). Participants first completed the automated *Vigilant Warrior*™ described above. Then, participants participated in two 2-min. practice sessions for the specific criterion task to be performed followed by the task itself for 60 minutes.

Results. Three sets of predictors were available from the tested battery of measures: (1) The dispositional measures (personality and analytic skill), (2) mean *d*' on the SVT, averaged across the six task periods (Cronbach α = 0.95), and (3) the subjective measures taken following the SVT including three stress state factors (Engagement, Distress, and Worry), three coping scales (Task-focused, Emotion-focused, and Avoidance), and overall workload from the modified NASA-TLX, calculated as an unweighted sum of the 6 rating scales. The performance criterion was mean *d*' on the criterion task, averaged across the six task periods (Cronbach α = 0.97) and was calculated separately for each of the four criterion task versions: *simultaneous*, *successive*, *successive* with cueing, *successive* with secondary task. Bivariate correlations showed that SVT *d*', Analytic skil, post-SVT subjective state, and coping all had some capacity to predict performance on the criterion task while the personality variables were unrelated. We then proceeded to a multiple regression analyis using Analytic skill, SVT *d'*, and the strongest stress/coping/workload measure, the task Engagement stress index.

Table 2. *Summary Statistics For The Regression Of Simultaneous Mean d' Onto The Predictor Sets*.

Step	Predictors	R^2	ΔR^2	df	F
1	Analytic Skill	.086	.086	2, 107	5.01**
2	SVT *d'*	.283	.198	1, 106	29.22**
3	Engagement	.326	.043	1, 105	6.64*

**p*<.05, ***p*<.01

Table 2 shows the summary statistics for predicting criterion mean *d*' on the *Simultaneous* task. The two Analytic skill variables, SVT *d*', and post-SVT engagement all added to the variance explained, explaining about 33% of the variance in the criterion in total. The final equation attained significance ($R = .571$; $F_{(4,105)} = 12.87$, $p < .01$).

Table 3. *Summary statistics for the regression of Successive Mean d' onto the predictor sets*.

Step	Predictors	R^2	ΔR^2	*df*	*F*
1	Task type	.104	.104	2,349	20.23**
2	Analytic Skill	.254	.150	2,347	18.05**
3	SVT *d'*	.374	.120	1,346	46.01**
4	Engagement	.385	.011	1,345	5.25*

* *p* < .05, ***p* < .01

Table 3 shows the summary statistics for the *Successive* criterion tasks. Again, all the predictor sets made a significant contribution, adding 27.1% to the variance explained by task type (the three different *Successive* task versions). The final equation attained significance ($R = .620$; $F_{(6,345)} = 35.95$, $p < .01$).

Conclusions on Assessment of Individual Differences in Vigilance Ability using Vigilant Warrior™

This study validated the *Vigilant Warrior*™ battery against a specific criterion-task simulation in a laboratory setting. While further work will determine the generality of the results, the following conclusions are justified from the large body of data assembled thus far. The results clearly vindicate the multivariate approach to vigilance assessment upon which *Vigilant Warrior*™ was based. Use of multiple objective and questionnaire predictors in *Vigilant Warrior*™ enhances predictive validity. The results also show that the predictor sets are fairly consistent across different versions of the criterion task, implying that the battery has the capacity to predict performance across a range of sustained monitoring tasks and to be practically useful for selecting workers both for superior objective performance on sustained monitoring tasks and for greatest resistance to stress and fatigue.

Acknowledgements

This research was supported by a Phase II Small Business Innovation Research grant from the Army Research Institute under contract number W74V8H-06-C-0049. The views, opinions, and/or findings contained in this chapter are those of the authors and should not be construed as an official Department of the Army position, policy, or decision. The authors also thank Joel Warm, Gerald Matthews, Kyle Behymer, Benjamin Simpkins, Elizabeth Lerner, Victor Finomore, Tyler Shaw, and Mark Crabtree for their contributions to this research.

References

Davies, D.R., & Parasuraman, R. (1982). *The psychology of vigilance*. London: Academic Press.

Hancock, P.A, & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, *31*, 519-537.

Helton, W.S., Dember, W.N., Warm, J.S., & Matthews, G. (1999). Optimism, pessimism, and false failure feedback: Effects on vigilance. *Current Psychology, 18*, 311-325.

Johnson, A. & Proctor, R.W. (2003). *Attention: Theory and practice.* Thousand Oaks, CA: Sage.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Macmillan, N.A. , & Creelman, C.D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Matthews, G. and Davies, D.R. (2001). Individual differences in energetic arousal and sustained attention: A dual-task study. *Personality and Individual Differences, 31,* 575-589.

Matthews, G., Davies, D.R., & Lees, J.L. (1990). Arousal, extraversion, and individual differences in resource availability. *Journal of Personality and Social Psychology, 59*, 150-168.

Robertson, I.H., Manly, T., Andrade, J. Baddeley, B.T., & Yiend, J. (1997). Oops! Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35,* 747-758.

Temple, J. G., Warm, J. S., Dember, W. N., Jones, K. S., LaGrange, C. M., & Matthews, G. (2000). The effects of caffeine and signal salience on performance, workload, and stress in an abbreviated vigilance task. *Human Factors*, 4*2,* 183-194.

Thackray, R.I., Bailey, J.P., & Touchstone, R.M. (1977). Physiological, subjective, and performance correlates of reported boredom and monotony while performing a simulated radar control task. *Vigilance: Theory, operational performance, and physiological correlates*. R.R. Mackie (Ed.). New York, Plenum Publishing Company, pp. 203-215.

Warm, J.S., Dember, W.N., & Hancock, P.A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.). *Automation and human performance: Theory and applications* (pp. 183-200). Mahwah, NJ: Erlbaum.

SELECTION FOR AVIATION RELATED CAREERS: AIR TRAFFIC CONTROL IN THE AIR FORCE AND THE FAA

M. Kathryn Bleckley, Ph.D., Jerry Crutchfield, Ph.D., Raymond E. King, Psy.D., Carol A Manning, Ph.D.
The Civil Aerospace Medical Institute of the U.S. Federal Aviation Administration (FAA)
Oklahoma City, OK, USA

Thomas R. Carretta, Ph.D.
Air Force Research Laboratory, U.S. Air Force (USAF)
Wright-Patterson Air Force Base, OH, USA

This paper discusses selection research and practice, with a focus on air traffic control specialists (ATCSs). In the USAF and FAA, accurate selection of air traffic control (ATC) trainees is essential because of the cost in time and money to train people for this high-consequence occupation. The FAA continues longitudinal validation research for the Air Traffic Selection and Training (AT-SAT) battery. Additionally, validation of the AT-SAT for placement by option would allow the FAA to develop a process for assessing applicants' potential to certify at facilities, providing useful information when determining where placement should occur. Frequently, psychiatric conditions are delineated in medical standards as disqualifying. The value of correctly using psychological testing for screening for psychiatric conditions is addressed. In the future, selection procedures in use today to hire ATCSs who use tactical techniques to separate airplanes might prove to be inappropriate for ATCSs, who will be expected to use strategic ATC methods.

Cognitive ability is the most thoroughly investigated psychological construct in studies of determinants of occupational performance. Accumulated evidence, including several meta-analyses of common selection methods in personnel psychology, has shown that general mental ability (*g*) is the best predictor of training and job performance involving core technical proficiency (Jensen, 1998; Schmidt & Hunter, 1998, 2004). Further, the predictiveness of *g* increases as job complexity increases (Gottfredson, 1997; Hunter, 1983). Gottfredson (1997) concluded that the pervasive utility of *g* in work settings occurs because fundamentally it is the ability to manage cognitive complexity, particularly by complex information processing.

Although *g* is the best predictor of several indicators of occupational performance, its validity can be incremented by other measures. For training, the predictiveness of *g* is incremented by measures of personality, structured interviews, and specialized job knowledge. For job incumbents, it is incremented by personality, job knowledge, and work sample performance. Causal models have shown *g* to exert its influence on job performance both directly and indirectly through the acquisition of job knowledge during training (Ree, Carretta, & Doub, 1998/1999; Ree, Carretta, & Teachout, 1995).

Military Air Traffic Control Specialist (ATCS) Selection

Results from studies of U.S. military ATCSs are consistent with the broader occupational performance literature. Several recent studies have focused on validation of the US military enlistment qualification and training classification test, the Armed Services Vocational Aptitude Battery (ASVAB), and have shown it to be a good predictor of ATC training performance (Carretta & King, 2008; Carretta & Siem, 1999; Held, 2006). Despite the proven validity of the ASVAB, enlisted ATC training and post-training attrition is higher than desirable, contributing to interest in additional selection methods to augment current procedures.

To this end, Carretta and King (2008) examined the utility of the FAA Air Traffic Selection and Training (AT-SAT; King, Manning, & Drechsler, 2006) battery for incrementing the predictiveness of the ASVAB for enlisted US Air Force ATC training. AT-SAT assesses cognitive and perceptual abilities and self-reported workplace characteristics, identified by the Nickels, Bobko, Blair, Sands, and Tartak (1995) job analysis. Air Traffic Scenarios (ATS) is a work sample test that involves the application of complex rules to control air traffic in an interactive, dynamic low-fidelity simulation. ATS requires examinees to learn complex rules and prioritize tasks. The training criteria were the average grade on written tests during an ATC Fundamentals course, the FAA Certified Tower Operator (CTO) test score, and a pass/fail training score. Due to the length of the AT-SAT, students completed one of three overlapping subtest blocks. Sample sizes for the AT-SAT subtest analyses varied from 154 to 326.

All correlations were corrected for range restriction (Lawley, 1943). Those involving the pass/fail training criterion also were corrected for dichotomization (Cohen, 1983). Results confirmed the predictive validity of the ASVAB against all three training criteria. After correction, the correlation between a *g*-loaded composite of the four ASVAB verbal/math subtests and the three criteria were: ATC Fundamentals (.760), CTO test score (.608), and training pass/fail (.630). ATS was the only AT-SAT subtest that demonstrated incremental validity beyond the ASVAB for all three training criteria. The increments in R^2 beyond the ASVAB were small but statistically significant for both the ATC Fundamentals score (.020) and the CTO test score (.016). The R^2 increment for the dichotomous pass/fail training criterion was larger (.156).

Missing from the Carretta and King (2008) study were strong measures of non-cognitive characteristics. A follow-on validation study should include non-cognitive measures, including personality (King, Retzlaff, Detwiler, Schroeder, & Broach, 2003) and improved medical assessment. Almost 25% of the Carretta and King study participants were eliminated for non-academic/non-performance reasons, including anxiety, discipline issues, fear of controlling, and loss of sleep. Neither the ASVAB nor the AT-SAT assess these non-cognitive factors. Finally, it is recommended that follow-on validation studies examine additional training and post-training performance criteria. These include performance in specialized training (control tower and radar approach control operations) and measures of post-training performance (e.g., first-term attrition, supervisor ratings).

FAA ATC Selection

The FAA ATC selection process has multiple stages, some of which are designed to identify candidates who might become ATCSs (select-in), and other stages designed to eliminate those that do not meet medical and/or security requirements (select-out). This section of the paper will include examples of ”select-in” research, as well as “select-out” research in the FAA. When selecting people to train for ATCS positions, two considerations should be kept in mind: 1. does the person have the aptitude to become an ATCS? 2. If so, at which type of position, Tower/Cab, Terminal Radar Approach Control (TRACON), or En Route, would this person most likely succeed? Once selected-in, the candidate must pass medical and security screens. The medical screen includes a psychological assessment, the value of which is discussed at the end of this section.

Between 1981 and 1992, the FAA hired and trained nearly 16,000 new ATCSs to replace those fired during the 1981 strike. This concentrated period of hiring has now led to a concentrated period of retirement as individuals in the replacement workforce achieve 25 years of service. With increasing retirements, the FAA plans to hire approximately 17,000 new ATCSs between 2008 and 2017. As it may take up to 3 years to train a fully certified ATCS, the FAA’s training costs are not trivial. Therefore, to meet the hiring requirements and assure that the right types of individuals are selected for subsequent training, the FAA developed and implemented the AT-SAT battery.

Since its implementation in 2002, nearly 12,000 applicants have taken AT-SAT, including more than 7,000 in the past year. Due to a lack of available ATCS positions until recently, few who were selected by AT-SAT have completed their training and become certified ATCSs. Moreover, former military ATCSs and civilian Department of Defense ATCSs do not take the AT-SAT as part of the hiring process. For selection purposes, ATCS candidates are considered qualified if they score 70-84.999 on AT-SAT and well-qualified if they score 85 or above.

It has only recently become possible to conduct an interim longitudinal validity analysis. In general, selection test validity is judged by its ability to predict job performance. To do this, we would have to wait until enough trainees who were selected based on their AT-SAT have become certified. However, we have access to training performance data that can be used as an interim substitute for job performance. At the end of Initial En Route or Initial Tower/Cab training classes, trainees’ performance is assessed by members of the Air Traffic Organization Training and Development office. Performance verifications (PV) are academic assessments coupled with a skill-based scenario, in which student ATCSs control simulated traffic while a field supervisor observes their performance. If the student’s performance is not satisfactory on day one, they are given additional training followed by a second assessment. Students either pass or fail the PV.

Data were analyzed from 650 students who took AT-SAT as part of the hiring process and completed FAA Academy training as of February 2008. Of these, only 57 failed PV the first time; 593 passed on their first attempt. This substantial inequality creates problems for statistical analysis. To overcome these problems, we randomly selected a subset of 75 of those who passed PV on their first attempt. All who failed PV on first attempt were included in the analyses. The 75 randomly-selected students who passed PV on first attempt (N= 75, $\bar{x}$ = 88.464, std err = .876) scored significantly higher on AT-SAT than did those who failed PV (N= 57, $\bar{x}$ = 85.055, std err =1.002), $t = 2.56$, $p = .012$. As with most selection tests, the range of AT-SAT scores available for analysis is restricted.

Because the PV data are binary (pass/fail), a Logistic Regression (LR) was conducted. LR provides several useful types of information, including an overall classification table. As can been seen in Table 1, AT-SAT correctly predicted who will pass or fail PV for most of the trainees.

Table 1. *Overall classification table from the LR analysis.*

		Predicted PV		
Actual PV		Passed	Failed	Percent correct
	Passed	**61**	14	81.3
	Failed	21	**36**	63.2
			Overall percent correct 73.5	

For the full sample, 93% of those who were well-qualified on AT-SAT passed PV on day 1. Of those who were qualified on AT-SAT, 88% passed on day 1. The difference in pass rates between well-qualified and qualified trainees was significant using Fisher's exact test, $p = .003$.

After making a selection decision, the FAA decides in which ATC option a new hire will be placed. Remember that ATCS options include en route, TRACON, and tower facilities. Currently, the FAA's placement decisions for newly hired ATCS are based only on where and when vacancies occur. Instructors who conduct field training report that 1) some trainees who have aptitude for one type of ATCS option get placed into another option and 2) trainers are sometimes forced to fail a trainee in one option when they believe that he/she would have been better able to perform in a different option. These reports suggest that the FAA needs to develop a process that uses information about a new ATCS's potential to certify at a facility to decide where the individual should be placed. This process will increase the efficiency of placing candidates into jobs and reduce costs associated with training and attrition.

Efforts are being made by the FAA, American Institutes for Research (AIR), and Personnel Decisions Research Institute (PDRI) to validate the AT-SAT test battery for use as a placement tool. Although the development of AT-SAT made extensive use of worker requirements for all three ATCS options, comparison of AT-SAT predictor scores with tower-specific criterion performance measures was not possible in the in the original validation study. As a result, it is not currently known if AT-SAT can be used as a tool to place controllers by option. However, the potential for AT-SAT to be used in this way has been recognized, and a requirement to validate AT-SAT as a tool to inform placement decisions was documented in the FAA's 2005 Controller Workforce Implementation Action Plan.

Four phases must be completed to validate AT-SAT for use as a placement tool. These are: 1) update existing information regarding the activities and sub-activities of the tower cab ATCS; 2) develop criterion performance measures associated with the sub-activities; 3) collect both predictor AT-SAT scores and criterion performance data from incumbent tower ATCSs; and 4) compare and analyze the scores and performance data to determine how AT-SAT subtests should be weighted. Our project is currently completing phase 2. The job performance measure presents simulated air traffic scenarios to incumbent ATCSs then asks them to answer multiple-choice judgment questions about what they observed. Researchers from PDRI worked with FAA ATCS contract instructors from the University of Oklahoma and Raytheon to develop roughly 50 ATCT traffic scenarios and approximately 173 multiple-choice questions that correspond to these scenarios. The scenarios were programmed into a version of the SIGNAL 3D ACTC simulator and recorded for presentation to incumbent ATCSs. ATCSs will see and hear each scenario played on four monitors that represent the out-the-ACTC-window view and one monitor that represents the ATCT radar presentation. Each scenario will be played for a few moments and then paused. When a scenario pauses, a sixth computer display will present relevant multiple-choice questions. The questions were designed to be both standardized across different types of facilities and be challenging enough to differentiate between ATCSs who are good and ATCSs who are exceptional performers.

While determining who has the aptitude for a given career or position within that career is a *select-in function*, determining who is medically fit is a *select-out* function. Medical examinations typically include consideration of the diagnostic categories outlined in the current *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2000). Certain psychiatric diagnoses may be disqualifying if they jeopardize safety or mission completion. Due to the Americans with Disabilities Act (ADA), the determination of fitness (and all medical assessments) can be conducted only after a conditional offer of employment is tendered. In the realm of ATCS selection, all tentatively selected applicants subsequently undergo a medical evaluation, which includes visual, cardiovascular, and psychiatric assessments, as outlined in *FAA Order 3930.3A*. Currently, the FAA administers the Minnesota Multiphasic Personality Inventory-2 (King, Schroeder, Manning, Retzlaff, & Williams, 2008) as a screen. The FAA discontinued use of the 16PF due to a desire for a more thorough psychological screening. For example, of 1,200 ATCS applicants screened with the 16PF in 2006 and 2,101 ATCS applicants screened in 2007, only 3 (.25%) and 1 (.05%), respectively, were determined to be in need of additional assessment. Although further assessment was not mandatory when an applicant was identified with the 16 PF, a psychiatric assessment was typically conducted. Initial psychological testing is only used for *screening*. Candidates that do not clear this screen are referred for additional psychological testing and a clinical interview. A clinical psychologist employed by the FAA and medical personnel review the raw data forwarded by the private practitioners who conduct these follow-up assessments. Applicants are disqualified based on the presence of a personality disorder or other psychiatric conditions (to include substance abuse) that pose a "potential hazard to safety in the Air Traffic Control System" (*p*. 10, FAA Order 3930.3a).

Select-in and select-out processes provide valuable information for hiring authorities; however, practitioners must be careful to not confuse the goals of select-in and select-out testing because tests of psychopathology will not provide useful predictive information about who has the aptitude to succeed in a career field. Conversely, select-in methods will not indicate who is suffering from a psychiatric illness. Optimally, the two approaches should be used in tandem (Carretta & King, 2008) and in the correct order to comply with ADA requirements.

The Future of ATC

The Joint Planning and Development Office (JPDO) proposed a plan for the Next Generation Air Transportation System (NextGen; JPDO, 2007) that is expected to increase airport and airspace capacity to meet future air traffic demands. As a result, considerable changes may be made to the job of the U.S. ATCS. If the ATCS job changes, then procedures used to select ATCSs may also need to change. This section addresses several issues concerning the selection of ATCSs in the NextGen timeframe.

How might NextGen changes affect the ATCS job of the future? NextGen is envisioned to allow pilots to operate with minimal flight interventions. NextGen will provide more data to the cockpit and allow pilots to make more decisions about real-time operations. The likely effect of near-term technology changes, such as Automatic Dependent Surveillance-Broadcast (ADS-B) and DataLink (DL) on the ATCS's job will be to provide more accurate information about aircraft locations, present information in a different format, or make minor changes to the procedures (e.g., standardizing arrival routes used by commercial pilots may reduce the number of manual handoffs required). Other changes, such as airspace redesign and flow efficiencies (FAA, 2007a; FAA, 2007b) might make the job easier (e.g., by reducing the amount of required coordination with other ATCSs), or more difficult (e.g., by increasing the number of runways available, and, thus, the amount of attention required to monitor them). These kinds of changes are minimal and are not likely to affect significantly the ATCS's roles and responsibilities.

However, other proposed NextGen technologies may have a greater impact on the ATCS's job. These include automated conflict resolution (Kirk, Bowen, Heagy, Rozen, & Viets, 2001) and transferring more responsibility for aircraft separation from the ATCS to the pilot (Bilimoria, Sheth, Lee, & Grabbe, 2000) or to automation (FAA, 2007a). Significantly increasing the number of aircraft controlled or reducing separation standards may also result in a dramatic change in the way ATCSs perform their jobs.

Predictions about the job of the future ATC usually involve more monitoring and fewer tactical decisions (Della Rocco, Manning, & Wing, 1990). These predictions produce questions about whether ATCSs can perform tasks effectively if they monitor traffic without controlling it. Can an ATCS quickly resolve a crisis that automation cannot handle? How well can an ATCS be expected to maintain situation awareness while pilots or automation make most of the separation decisions? Moreover, if traffic volume is higher and aircraft are more closely spaced than at present, can ATCSs observe all relevant activity and step in to take appropriate action during an emergency?

What abilities will be required to perform the ATCS job(s) of the future? Making major changes to the ATCS job could affect the knowledge, skills, and abilities needed to perform air traffic control tasks. The abilities required to perform the ATCS's job are measured by selection procedures. The FAA has invested significant effort in developing selection procedures that measure the abilities required to perform today's ATCS job. These are based on a set of 66 "worker requirements" that include communication, computation, memory, meta-cognition, reasoning, information processing, attention, perceptual/ spatial, interpersonal, self-efficacy, work and effort, and stability/adjustment (Morath, Quartetti, Bayless, & Archambault, 2001). The worker requirements were linked with 98 ATCS subactivities associated with ensuring the safe and expeditious flow of traffic and responding to emergencies or special conditions. As long as today's FAA ATCS continues to ensure the safe and expeditious flow of traffic by performing situation monitoring, resolving aircraft conflicts, managing air traffic sequences, routing or planning flights, assessing weather impact, and managing sector and position resources (Ammerman et al., 1987), then the abilities required to perform the job will probably not change much even if ATCS procedures undergo fairly major changes.

Some believe that introducing automation into ATC will not have a big effect on the ability requirements needed to perform the job. For example, Manning and Broach (1992) asked a team of ATCSs who had analyzed operational requirements for a system that provided conflict resolution advisories to assess the effect this automation would have on nine ability requirements. The group believed that the automation would produce some changes in the ATCS's job but predicted that the ATCS of the future would require about the same level of abilities to perform their tasks using the new automation. Moreover, they did not believe that additional abilities would be required to perform the new automated job. However, if more significant changes occurred in job tasks, such as removing responsibility for control decisions, replacing tactical decision making with strategic analysis, and monitoring rather than controlling actions taken by pilots or automation, then the abilities required to perform the job might change. ATCSs may still perform situation monitoring, resolve aircraft conflicts, manage air traffic sequences, route or plan flights, assess weather impact, and manage sector/ position resources but in a much different way (Ammerman et al., 1987) that involves processing information, receiving status updates, choosing automation-identified resolutions, suppressing alerts, and checking conflict violations.

We do not yet know how relevant these abilities will be to performing the future air traffic management job. To obtain a complete answer to this question requires conducting a strategic job/task analysis (SJA) for the new job and using the result to identify the associated future ability requirements. One problem with conducting an SJA is that it is difficult to obtain accurate information about the future job until after important decisions about it have been made. It will be difficult to conduct a reasonable SJA during early developmental stages of a system that has not yet been fielded. However, some methods have been developed to allow describing future tasks based on the limited amount of information available today (Landis, Fogli, & Goldberg, 1998; Schneider & Konz, 1989).

When the job tasks that will be performed by ATCSs in the NextGen timeframe are identified, additional analyses will identify ability requirements associated with the job tasks and tests of new abilities will be obtained or developed. It will be necessary to update information about future job/tasks as revised descriptions of the future job become available.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders*, 4th Edition Text Revision. Washington, DC: APA.

Americans with Disabilities Act of 1990, Pub. L. No. 101-336, §2, 104 Stat. 328 (1991). www.ada.gov/pubs/ada.htm (accessed: 12/15/2008).

Ammerman, H., Becker, E., Claussen, D., Inman, E., Jones, G., Melville, B. & Tobey, W. (1987, July). *FAA air traffic control operation concepts: Volume II: ACF/ACCC terminal and en route controllers.* (DOT/FAA/AP-87/01, Vol. 2). Washington, DC: Federal Aviation Administration.

Ammerman, H., Bergen, L. Davies, D. Hostetler, C. Inman, E., & Jones, G. (1987, November). *FAA Air Traffic Control Operation Concepts: Volume VI: ARTCC/HOST en route controllers* (DOT/FAA/AP-87/01, Vol. 6), Washington, DC: Federal Aviation Administration.

Bilimoria, K. D., Sheth, K. S., Lee, H. Q., & Grabbe, S. R. (2000, August). Performance evaluation of airborne separation assurance for Free Flight. (AIAA-2000-4269). Presented at AIAA Guidance, Navigation, and Control Conference, Denver, CO.

Carretta, T. R., & King, R. E. (2008). Improved military air traffic controller selection methods as measured by subsequent training performance. *Aviation, Space, and Environmental Medicine, 79,* 36-43.

Carretta, T. R., & Siem, F. M. (1999). Determinants of enlisted air traffic controller success. *Aviation, Space, and Environmental Medicine, 70,* 910-918.

Cohen, J. (1983). The cost of dichotomization. Applied Psychological Measurement, 7, 249-253.

Della Rocco, P., Manning, C., & Wing, H. (1990, November). *Selection of air traffic controllers for automated systems: applications from today's research.* (DOT/FAA/AM-90/13). Washington, DC: FAA Office of Aviation Medicine.

Federal Aviation Administration (1980, May 9). *FAA Order 3930.3A - Air traffic control specialist health program.* Washington, DC: U.S. Government Printing Office.

Federal Aviation Administration. (2007a, June). *Executive overview version 1.0 2007-2025 operational evolution partnership*. Available online at: www.faa.gov/about/office_org/headquarters_offices/ato/publications/oep/version1/overview/images/overviewcover.jpg. (accessed: 01/21/2009).

Federal Aviation Administration. (2007b, June). *OEP V1.0 Solution Set Smart Sheets,* Available online at: www.faa.gov/about/office_org/headquarters_offices/ato/publications/oep/version1/solutionsets/, (accessed: 01/21/2009).

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24,* 79-132.

Held, J. D. (2006). *Armed Services Vocational Aptitude Battery (ASVAB) standards: Air traffic control rating, NPRST* Letter Report Ser 3900 PERS-13/00047. Millington, TN: Navy Personnel Research Studies and Technology.

Hunter, J. E. (1983). *Overview of validity generalization for the U. S. Employment Service.* (USES Test Research Report 43.). Washington DC: U. S. Department of Labor, Employment and Training Administration.

Jensen, A. R. (1998). *The g Factor: The science of mental ability.* Westport, CN: Praeger.

Joint Planning & Development Office. (June 2007). *Concept of operations for the next generation air transportation system*, *version 2.0,* Available online at: www.jpdo.gov/library/NextGen_v2.0.pdf, (accessed: 01/21/2009).

King, R. E., Manning, C. A., & Drechsler, G. K. (2006). Operational use of the Air Traffic Selection and Training Battery. *International Journal of Applied Aviation studies, 6,* 207-217.

King, R. E., Retzlaff, P. D., Detwiler, C. A., Schroeder. D. J., Broach, D. (2003). *Use of personality assessment measures in the selection of air traffic control specialists.* Washington, DC: FAA Office of Aerospace Medicine, DTT/FAA/AM-03/20.

King R E, Schroeder DJ, Manning CA, Retzlaff PD, & Williams CA (2008). *Screening air traffic control specialists for psychopathology using the Minnesota Multiphasic Personality Inventory-2.* (DOT/FAA/AM-08/13). Washington, DC: FAA Office of Aerospace Medicine.

Kirk, D., Bowen, K., Heagy, W., Rozen, N., & Vietz, K. (2001, October). *Development and assessment of problem resolution capabilities for the en route sector controller,* AIAA-2001-5255, Available online at: www.mitrecaasd.org/library/documents/aiaa2001_5255.pdf, (accessed: 01/21/2009).

Landis, R. S., Fogli, L., & Goldberg, E. (1998). Future-oriented job analysis: A description of the process and its organizational implications. *International Journal of Selection and Assessment, 6(3)*, 192-198. Available online at: www3.interscience.wiley.com/journal/119131378/abstract?CRETRY=1&SRETRY=0. (accessed: 01/21/2009).

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, Section A, 62 (Part 1),* 28-30.

Manning, C., & Broach, D. (1992, August). *Identifying ability requirements for operators of future automated air traffic control systems*. (DOT/FAA/AM-92/26). Washington, DC: FAA Office of Aviation Medicine.

Morath, R., Quartetti, D., Bayless, A., & Archambault, C. (2001, March) *Chapter 2: Air Traffic Controller Job Analysis*, In R. Ramos, M. Heil, and C. Manning (Eds.) Documentation of Validity for the AT-SAT Computerized Test Battery Volume II. (DOT/FAA/AM-01/6). Washington, DC: FAA Office of Aviation Medicine.

Nickels, B. J., Bobko, P., Blair, M. D., Sands, W. A., and Tartak, E. L. (1995). Selection and Control Hiring Assessment (SACHA): Final job analysis report. (Deliverable Item 007A under FAA Contract DFTA01-91-C-00032). Washington, DC: Federal Aviation Administration, Office of Personnel.

Ree, M. J., Carretta, T. R., Doub, T. W. (1998/1999). A test of three models of the role of *g* and prior job knowledge in the acquisition of subsequent job knowledge in training. *Training Research Journal, 4,* 135-150.

Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior job knowledge in complex training performance. *Journal of Applied Psychology, 80*, 721-730.

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.

Schmidt, F. L., & Hunter, J. E. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86,* 162-173.

Schneider, B, & Konz, A. M. (1989). Strategic job analysis. *Human Resource Management, 28*, 51-63.

THE ROLE OF COMMON METHODS IN PERSONNEL SELECTION

Thomas R. Carretta
Air Force Research Laboratory
Wright-Patterson AFB, OH, USA

Cognitive ability is the most widely researched psychological construct in studies of determinants of occupational performance. Results of meta-analyses of common selection methods in personnel psychology indicate that general mental ability (*g*) is the best predictor of training and job performance involving core technical proficiency. For training, the predictiveness of *g* is incremented by measures of personality and specialized job knowledge. For job incumbents, the predictiveness of *g* is incremented by personality, job knowledge, and work sample performance. In addition to the predictive validity of *g*, personality, and prior job knowledge, their role in the acquisition of additional job knowledge and subsequent job performance has been demonstrated in causal models. These results are consistent with those for diverse military occupations including pilots and several enlisted technical specialties. Several studies are reviewed examining the relations of *g* and other common selection constructs to training performance for military jobs including air traffic controllers.

Personnel selection research provides overwhelming evidence that general mental ability (*g*) is an important determinant of training and job performance (Gottfredson, 1997; Jensen, 1998; Schmidt & Hunter, 1998, 2004). Further, the predictive validity of *g* is directly related to job complexity (Gottfredson, 1997; Hunter, 1983b). Hunter (1983b) demonstrated this in analyses of a US Department of Labor database of 515 diverse jobs. Hunter classified these jobs into categories according to complexity of data handling (low, medium, and high) and complexity of dealing with things (simple feeding/offbearing and complex set-up work). The validity of *g* rose as job complexity increased. The average corrected validities of *g* for the low, medium, and high data complexity jobs were .40, .51, and .58. For the low complexity feeding/offbearing jobs and complex set-up work jobs the corrected validities were .23 and .56. Gottfredson (1997) concluded that the pervasive utility of *g* in work settings occurs because essentially it is the ability to manage cognitive complexity, in particular, complex information processing.

Incrementing the Predictiveness of *g*

Specific Abilities, Knowledge, and Non-Cognitive Characteristics

Several studies have examined the utility of specific abilities and knowledge as well as non-cognitive characteristics for incrementing the predictiveness of *g* versus a wide range of occupational performance criteria. McHenry, Hough, Toquam, Hanson, and Ashworth (1990) investigated the predictiveness of measures of *g*, spatial, perceptual-psychomotor, temperament/personality, vocational interest, and job reward preference for nine US Army jobs. Training criteria were five job performance factors identified by Campbell, McHenry, and Wise (1990): core technical proficiency (job-specific task proficiency), general soldiering proficiency (non-job-specific task proficiency), effort and leadership (demonstrating effort), personal discipline (maintaining personal discipline), and physical fitness and military bearing. General mental ability was predictive of all of the job performance factors and was the best predictor of core technical proficiency and general soldiering proficiency with correlations of .63 and .65 corrected for range restriction. None of the other predictors incremented *g* by more than .02 versus these criteria. For the other job performance factors, temperament/personality was incremental to *g* or superior to *g* for prediction.

Ree, Earles, and Teachout (1994) examined the predictiveness of *g* and specific abilities for job performance in a sample of 1,036 US Air Force enlisted personnel in seven jobs. Job performance measures consisted of hands-on work samples, job knowledge interviews, and a combination of the two called the “Walk Through Performance Test.” Measures of *g* and specific abilities were extracted from a multiple aptitude battery and regressions compared the predictiveness of *g* and specific abilities. Across the seven jobs the average validity of *g* was .40 for the hands-on work sample, .42 for the job knowledge interview, and .44 for the “Walk Through Performance Test.” Adding the specific ability measures increased the validity by an average of only .02. These results are very similar to those of McHenry et al. (1990).

In a large-scale meta-analysis spanning 85 years of published studies, Schmidt and Hunter (1998) examined the utility of measures of *g* and 18 other commonly used personnel selection procedures versus training and job performance. They estimated the predictive validity of *g* to be .56 for training and .51 for job performance. For training, the two combinations of predictors with the highest multivariate validity were *g* plus an integrity test (mean R = .67) and *g* plus a conscientiousness test (mean R = .65). For job performance, the three combinations of predictors with the highest multivariate validity were *g* plus an integrity test (mean R = .65), *g* plus a structured interview (mean R = .63), and *g* plus a work sample test (mean R = .63).

Job Knowledge and Work Sample Tests

Job knowledge and work sample tests are developed around the assumption that examinees already have job-related technical knowledge or know how to do the job. Although they are useful for predicting performance for job incumbents (Schmidt & Hunter, 1998), they generally are not suitable for untrained applicants. Their use for training was unusual enough that Schmidt and Hunter (1998) did not include them in their meta-analyses involving training.

There are some notable exceptions. US military selection and classification batteries such as the Air Force Officer Qualifying Test (AFOQT; Carretta & Ree, 1996) and Armed Services Vocational Aptitude Battery (ASVAB; Segall, 2007) include technical knowledge (non-specific job knowledge) subtests (aviation information, electronics, mechanical, auto/shop) that are used for technical training qualification. These tests measure knowledge that anyone interested in a particular topic might learn from their choice of educational and recreational pursuits. The key concept is that these types of tests are surrogate measures of skill, interest, and motivation in a particular area (Guilford, & Lacey, 1947).

Although work sample tests (e.g., use of flight simulators by commercial air carriers to assess the skill level of experienced pilots) are associated with job incumbents, not all work sample tests are of this type. Beginning in the 1960’s, there have been several efforts to develop work sample tests of trainability suitable for untrained applicants (Robertson & Downs, 1979, 1989). The unique characteristic of work sample trainability tests is that they provide applicants a structured learning experience followed by a test. They also have very good face validity. A disadvantage of these tests is that they typically require long training periods while applicants learn complex rules and procedures. Examples include the Automated Pilot Aptitude Measurement System (Long & Varney, 1975), the Canadian Automated Pilot Selection System (Spinner, 1991), and the FPS 80 (Gress, & Willkomm, 1996) for pilot training and the FAA Air Traffic Scenarios subtest (King, Manning, & Drechsler, 2006) for air traffic controllers. Robertson and Downs (1989) conducted a meta-analysis of work sample tests of trainability and concluded that they provide good prediction of short-term training performance.

Causal Models

In addition to their predictive validity, the causal role of *g*, personality, and prior job knowledge in job performance has been demonstrated. Hunter (1983a) reported causal analyses of meta-analytically derived correlations linking *g*, job knowledge, job performance (work samples), and supervisory ratings from 14 studies with 3,264 participants. Hunter found that *g* (ability) had both a direct and indirect (through job knowledge) influence on job performance. Job knowledge, in turn, had a major causal impact on job performance and supervisory ratings. Ability had no direct effect on supervisory ratings; all effects were moderated. Although job knowledge and work sample performance accounted for all of the relationship between ability and supervisory ratings, the total causal impact of *g* was considerable.

Borman, White, Pulakos, and Oppler (1991) expanded the variables used by Hunter (1983a). Their causal models included measures of cognitive ability, job knowledge, personality (achievement orientation and dependability), task proficiency, problem behavior, and supervisory ratings of performance. Participants were 4,362 US Army personnel in 9 jobs. Cognitive ability, job knowledge, and dependability played strong indirect causal roles on task proficiency and supervisory ratings. Dependability had a modest causal influence on disciplinary actions (problem behavior).

Ree, Carretta, and Teachout (1995) and Ree, Carretta, and Doub (1998/1999) added the construct of prior job knowledge to occupational causal models for US Air Force pilots and enlisted personnel in technical training specialties. Prior job knowledge was defined as job relevant knowledge applicants acquire prior to training. Ree et al. (1995) observed a strong causal influence for *g* on prior job knowledge. No direct path was found for *g* to either of two work sample performance factors derived from check flight grades in early and late jet training; however, its indirect influence moderated through job knowledge was observed. This study also involved a set of three sequential training courses. Most of the influence of *g* was exerted indirectly through the acquisition of job knowledge in the sequential training courses.

Military Air Traffic Controller Selection

The purpose of this section is to evaluate recent US military studies of air traffic controllers in light of the more general findings regarding the determinants of occupational performance.

Over the last decade, the US military has conducted several studies to examine the determinants of enlisted air traffic controller (ATC) performance. Research has focused on validation of the Armed Services Vocational Aptitude Battery (ASVAB; Segall, 2007) and has shown it to be a good predictor of ATC training performance (Carretta & King, 2008; Carretta & Siem, 1999; Held, 2006). Despite the proven validity of the ASVAB, enlisted ATC training and post-training attrition is higher than desirable, contributing to interest in additional selection methods to augment current procedures.

To this end, Carretta and King (2008) examined the utility of the FAA Air Traffic Selection and Training (AT-SAT; King, Manning, & Drechsler, 2006) battery for incrementing the predictive validity of the ASVAB versus enlisted US Air Force ATC training performance. The ASVAB has 9 subtests that measure cognitive ability (verbal, math, and spatial) and technical knowledge. The AT-SAT battery was developed based on results of a job task analysis of the FAA ATC career field. It includes 8 subtests that assess cognitive and perceptual abilities and self-reported life experiences. One of the subtests, Air Traffic Scenarios (AT) is a work sample test that involves the application of complex rules to control air traffic in an interactive, dynamic low-fidelity simulation. The AT subtest requires examinees to learn complex rules and prioritize tasks. The training criteria were the average grade from several written tests during an ATC Fundamentals course, the FAA Certified Tower Operator (CTO) test score, and a dichotomous graduation/elimination training score. The ATC Fundamentals course includes classroom instruction in

ATC fundamentals, control tower operations principles, and ATC radar/non-radar principles. The FAA CTO test assesses job knowledge involving airport traffic control procedures, flight rules, communications operating procedures, flight assistance service, aviation weather, air navigation and aids to air navigation, and en route traffic control procedures. Due to the length of the AT-SAT battery (6 ½ to 8 hours), students were not given all of the subtests. Instead, each student completed one of three overlapping test blocks. Sample sizes for the AT-SAT subtest analyses varied from 154 to 326.

All correlations were corrected for range restriction (Lawley, 1943). Those involving the graduation/elimination training criterion also were corrected for dichotomization (Cohen, 1983). Results confirmed the predictive validity of the ASVAB against all three training criteria. After correction, the correlation between a *g*-loaded composite of three of the four ASVAB verbal/math subtests and the three criteria were: ATC Fundamentals (r = .757), CTO test score (r = .596), and training graduation/ elimination (r = .610). Air Traffic Scenarios was the only AT-SAT subtest that demonstrated incremental validity beyond the ASVAB for all three training criteria. The increments in R^2 beyond the ASVAB were small, but statistically significant for both the ATC Fundamentals score (.034) and the CTO score (.020). The R^2 increment for the dichotomous graduation/ elimination training criterion was larger (.156).

Additional analyses of the Carretta and King (2008) data were conducted to shed light on what is being measured and the sources of predictive validity for the ASVAB and AT Scenarios subtests. After correction for range restriction, the ASVAB verbal/math composite and a composite of the three AT Scenarios subscale scores correlated .695, suggesting that despite its appearance the AT Scenarios test largely measures *g*. Another method to assess what is being measured is to conduct a principal components (PC) analysis of the scores and examine the unrotated component matrix. Results of a PC analysis of the 9 ASVAB and three AT Scenarios scores using data corrected for range restriction yielded two components with Eigenvalues greater than or equal to 1. The first unrotated PC, which for cognitive tests provides a lower-bound estimate of the *g*-saturation of the scores (Ree & Earles, 1991), accounted for 60.3% of the total variance. The average loading for all 12 scores, all 9 ASVAB subtests, the four ASVAB verbal/math subtests, and the three AT Scenarios scores were .771, .796, .820, and .714 respectively. Although the AT Scenarios scores are not as *g*-loaded as the ASVAB subtests, it is clear they have a strong *g* component.

Factor scores were computed using the PC weights and each of the three ATC training criteria were regressed on them. The first principal component score (representing a general factor) was the only one that contributed significantly to the prediction of all three criteria. The PC score that was defined by the three AT scenarios scores also contributed to the prediction of the graduation/elimination training criterion. These results indicate that the AT Scenarios test is predicting unique variance in the graduation/elimination criterion beyond that provided by *g*.

Summary and Recommendations

Accumulated research has shown cognitive ability to be a crucial determinant of occupational performance across a variety of jobs. Further, the predictiveness of *g* is incremented by measures of personality, job-related knowledge, and prior job experience (the later for job incumbents). Results from studies of military ATC training are consistent with the broader occupational performance literature.

A missing component of the Carretta and King (2008) study was the absence of strong measures of non-cognitive characteristics in the test battery. A follow-on validation study should expand the predictors to include non-cognitive measures, including personality (King, Retzlaff, Detwiler, Schroeder, & Broach, 2003) and improved medical assessment. Almost 25% of the eliminations in the Carretta and King study were for non-academic/non-performance reasons, including anxiety, disciplinary, fear of controlling, and loss of sleep. Neither the ASVAB nor the AT-SAT are designed to assess these non-cognitive factors.

Finally, it is recommended that follow-on validation studies examine additional training and post-training performance criteria. These include performance in ATC specialized training tracks (control tower operations, radar approach control operations) and measures of post-training performance (e.g., first-term attrition, supervisor ratings).

References

Borman, W. G., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory job performance ratings. *Journal of Applied Psychology, 76,* 863-872.

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. Special Issue: Project A: The US Army Selection and Classification Project. *Personnel Psychology, 43,* 313-333.

Carretta, T. R., & King, R. E. (2008). Improved military air traffic controller selection methods as measured by subsequent training performance. *Aviation, Space, and Environmental Medicine, 79,* 36-43.

Carretta, T. R., & Ree, M. J. (1996). Factor structure of the Air Force Officer Qualifying Test: Analysis and comparison. *Military Psychology, 8,* 29-42.

Carretta, T. R., & Siem, F. M. (1999). Determinants of enlisted air traffic controller success. *Aviation, Space, and Environmental Medicine, 70,* 910-918.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7,* 249-253.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24,* 79-132.

Gress, W., & Willkomm, B. (1996). Simulator-based test systems as a measure to improve the prognostic value of aircrew selection. *Selection and Training Advances in Aviation: AGARD Conference Proceedings 588* (pp. 15-1-15-4). Prague, Czech Republic: Advisory Group for Aerospace Research & Development.

Guilford, J. P. & Lacey, J. I. (Eds.) (1947). *Army Air Forces Aviation Psychology Program Research Reports: Printed classification tests* (Report No. 5). Washington, DC: U.S. Government Printing Office.

Held, J. D. (2006). *Armed Services Vocational Aptitude Battery (ASVAB) standards: Air traffic control rating,* NPRST Letter Report Ser 3900 PERS-13/00047. Millington, TN: Navy Personnel Research Studies and Technology.

Hunter, J. E. (1983a). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.). *Performance measurement and theory*. Hillsdale, NJ: Lawrence Erlbaum.

Hunter, J. E. (1983b). *Overview of validity generalization for the U. S. Employment Service.* (USES Test Research Report 43). Washington DC: U. S. Department of Labor, Employment and Training Administration.

Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability.* Westport, CN: Praeger.

King, R. E., Manning, C. A., & Drechsler, G. K. (2006). Operational use of the Air Traffic Selection and Training Battery. *International Journal of Applied Aviation studies, 6,* 207-217.

King, R. E., Retzlaff, P. D., Detwiler, C. A., Schroeder. D. J., Broach, D. (2003). *Use of personality assessment measures in the selection of air traffic control specialists.* Washington, DC: FAA Office of Aerospace Medicine, DTT/FAA/AM-03/20.

Long, G. E., & Varney, N. C. (1975). *Automated pilot aptitude measurement system* (AFHRL-TR-75-58). Lackland AFB, TX: Air Force Human Resources Laboratory, Personnel Research Division.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, Section A, 62 (Part 1),* 28-30.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43,* 335-354.

Ree, M. J., Carretta, T. R., Doub, T. W. (1998/1999). A test of three models of the role of *g* and prior job knowledge in the acquisition of subsequent job knowledge in training. *Training Research Journal, 4,* 135-150.

Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior job knowledge in complex training performance. *Journal of Applied Psychology, 80*, 721-730.

Ree, M. J., & Earles, J. A. (1991). The stability of convergent estimates of *g*. *Intelligence, 15,* 271-278.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than *g*. *Journal of Applied Psychology, 79,* 518-524.

Robertson, I. T., & Downs, S. (1979). Learning and the prediction of performance: Development of trainability testing in the United Kingdom. *Journal of Applied Psychology, 64,* 42-50.

Robertson, I. T., & Downs, S. (1989). Work-sample tests of trainability: A meta-analysis. *Journal of Applied Psychology, 74,* 402-410.

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.

Schmidt, F. L., & Hunter, J. E. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86,* 162-173.

Segall, D. O. (2007). CAT-ASVAB form development. *Proceedings of the Defense Advisory Committee on Manpower, Personnel, and Testing.* Monterey, CA: Defense Manpower Data Center.

Spinner, B. (1991). Predicting success in primary flying school from the Canadian Automated Pilot Selection System: Development and cross-validation. *International Journal of Aviation Psychology, 1,* 163-180.

RUNWAY INCURSION PREVENTION USING AN AUDIO INTERVENTION

Nathan Maertens, Christopher DeSalvo, Jessie Chin and J. Michelle Moon

Human Factors Division
University of Illinois at Urbana-Champaign
Savoy, Illinois

From 2004 through 2007 runway incursions (RI), an FAA high priority safety item, have continuously increased (FAA, 2008). The FAA has sought mitigation proposals; here we suggest one such solution. Byrne, Kirlik and their students (2005, 2006, 2007) suggested one possible cause of RIs to be pilots making errors when given counterintuitive taxi instructions (i.e., turns away, as opposed to towards, their ultimate destination). Using this work as a foundation, we identified counterintuitive taxi geometries at Willard Airport (CMI), conducted an experiment with 14 certified flight instructors working at CMI in a simulation of landing and taxiing, and tested potential countermeasures. In addition to replicating Byrne and Kirlik's observations of systematic errors in turns violating experiential and geometrical expectations, we showed that our verbal guidance intervention aided all 7 pilots in correctly navigating the counterintuitive turn whereas only 1 of 7 within the control group did so.

Early in the morning of August 27, 2006 Comair flight 5191 crashed while attempting to take off from Blue Grass Airport in Lexington, Kentucky (NTSB, 2006). In a tragic turn of events, the airplane attempted to take off on runway 26 instead of runway 22. Runway 26 was too short for the aircraft to become airborne and it ran off the end of the runway killing 49 people. This accident illustrates the necessity for pilots to maintain geographical awareness of their assigned taxi instructions and their position on the airfield. These types of failures often result in runway incursions (RI) and, as illustrated in this example, can have tragic results.

Runway incursions are defined as "any occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle or person on the protected area of a surface designated for the landing and takeoff of aircraft" (FAA, 2007, p. 43). RIs increased from 330 to 370 from FY2006 to FY2007. RIs in FY2008 numbered 1009. Likely though, this is due in part to the FAA's adoption of a new, more conservative, definition of runway incursion. (FAA, 2009). An investigation of causes of RIs in a review of 300 Aviation Safety Reporting System reports from FY2004 showed that pilots noted problems with expectations between where they perceived the hold short lines (the lines beyond which clearance is required to proceed) should be and where they anticipated they would hold in about one third of the incidents (FAA, 2005). Inappropriate actions based on these false expectations are known as expectancy violations and often result from the adverse influence of habit patterns. This finding suggests that aiding pilots in maintaining knowledge of position on the airfield may help reduce RIs.

In a study by Byrne, Kirlik and their students (2005, 2006, 2007) it was shown that pilot decision making methods changed across the spectrum from well planned out to guessing as available decision time was reduced in severe low visibility conditions (fog). Further, pilots were more likely to make errors when a counterintuitive taxi instruction (i.e., with turns away, as opposed to towards, their ultimate destination) was given. This characteristic behavior was modeled and validated through comparison with data from a NASA experiment of pilot navigation error in a simulation of Chicago O'Hare International Airport. In fact, every turn-related decision error in the NASA experiment could be explained by the model as owing to a counterintuitive relationship between the direction of the required turn and the turn that most closely corresponded to the direction of the destination gate. Thus, Byrne et al. identified one possible systematic cause of RIs.

Known human factors issues associated with automation are also relevant to the discussion of RIs; any mechanism created to aid pilots in preventing RIs must be designed accordingly. Automation can impair situational awareness (SA) due to changes in vigilance, complacency associated with monitoring, and changes in the quality of feedback provided to the human operator (Endsley & Kiris, 1995). An operator's level of trust in automation matters as well. Over-trust becomes an issue if an operator unduly trusts the system and accepts its counsel without fully understanding the entire situation. This complacency may lead to a difficulty in detecting automation errors, a loss in overall operator SA through failing to fully monitor the situation, and a potential degradation in skill (Wickens & Hollands, 2000). Under-trust, on the other hand, results in the pilot not choosing to use automation when it could increase safety. Despite these concerns, automation can provide novel solutions in addressing RIs when combined with an understanding of these human factors issues.

Based on these tenets, we explored the effectiveness of an automated intervention to prevent RIs due to counterintuitive taxi instructions. This intervention gave pilots progressive taxi instructions to and from their on-field destinations at counterintuitive taxi points and corrected them if they deviated from their prescribed route. We hypothesized that this intervention may help reduce the adverse impact of expectancy violations and habit patterns.

Method

We designed an experiment for experienced flight instructors with the goal of replicating the counterintuitive turn diagnosis of RI identified by Byrne, et al. (2005, 2006, 2007) as well as to test an intervention to prevent these types of errors. With help from a subject matter expert familiar with flight operations at Willard Airport (CMI), we located a point on the airport surface where pilot habit patterns were likely to be previously established and that these habits would be likely to transfer to a simulation. This afforded an opportunity to study expectancy violations due to a counterintuitive turn. To achieve this, an accurate re-creation of CMI airport was developed in an X-Plane® PC-based, flight simulator. To ensure generalizability to CMI, engineering drawings of all taxiway routes and markings were used. Instructor pilots with previous experience flying at CMI were chosen so that issues of unfamiliarity with CMI as well as control of the aircraft itself would not create confounding sources of error in the experimental results. That is, we did not want a pilot's inability to control the airplane to be a factor in the route they chose to taxi.

Our experiment investigated whether an intervention could prevent possible RIs at the point identified on the airfield when taxi instructions for counterintuitive turns were given. To create the setting for a counterintuitive taxi instruction, participants flew multiple approaches to a landing followed by taxiing to a known position (locally called the "orange corral") at CMI over two days. All but the last trial on the second day had the purpose of establishing a habit pattern so that when a counterintuitive taxi instruction was given on the last trial it would be unfamiliar. The pilots were told that they were evaluating the simulator itself and were not aware of the purpose of the experiment until the end of their participation. This was required because knowledge of the experimenters' intent could affect the participants' performance of the task.

On the first day of the experiment all participants were asked to land on runway 32R then taxi to the orange corral on A5, A, and A2 (Figure 1) via prerecorded ATC instructions. This scenario was repeated eight times. On the second day all participants had the same scenario as the first day for the first seven trials. On the eighth, and final, trial of the second day the taxi instruction was altered to taxi to the orange corral via A5, A6, A, A2 (Figure 2). The only difference between the two groups of pilot participants was that a simulation of our intervention was active for the experimental group, but not for the control group. This intervention cued the A6 turn (counterintuitive turn) once the turn off of 32R was completed and gave course corrections in the event the prescribed taxi route was deviated from. The intervention cue for A6 consisted of the statement, "Now approaching alpha six, turn right to zero niner two, onto alpha six."

The researchers took written notes during each session on participant taxi routes and verbalizations. Participants were encouraged to treat the simulated flights as if they were actual flights and they were specifically requested to make read-back calls to ATC as if they were actually using the radio. To better recreate normal taxi conditions, they were also instructed to taxi at speeds they typically used at CMI.

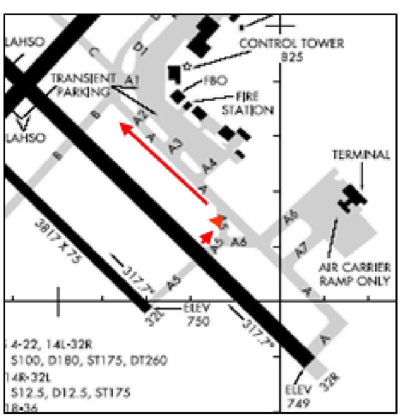


Figure 1. The nominal (experientially & spatially intuitive) taxi route used during all but the final trial of the second day.

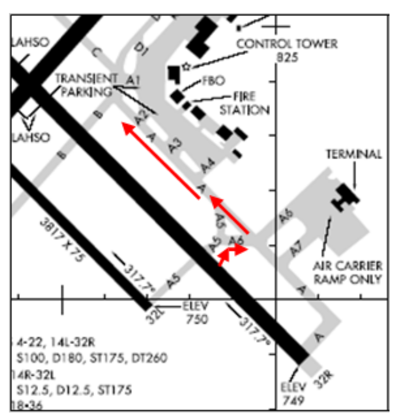


Figure 2. The counterintuitive taxi route used during the final trial of the second day.

In addition to the intervention just described, we tested a different intervention with one additional participant. This intervention was simply to add the phrase to “I say again” to the last trial’s taxi instructions in the attempt to raise the saliency of the counterintuitive turn. For this participant, the procedure was identical to the control group except the taxi instruction was: “Turn right on A5 taxi to Orange Corral via A5, A6, I say again A6, A, A2” in the final trial. The intention was to test if repetition of the counterintuitive turn in the instruction would raise awareness of its counterintuitive nature.

Results

Each participant’s performance was assessed based on his or her taxi instruction read-back and turn to A5 or A6 during the final trial on the second day. Four of the 7 participants in the control group read back the taxi instruction correctly; whereas only 1 in the intervention group did. However, only 1 person in the control group followed the correct taxi route (14% success rate). Despite taking the correct route, the observer noted participant hesitation at the turn’s decision point. Many other control group participants also demonstrated hesitation independent of whether their taxi instruction read-back was correct or not.

In the intervention group, all 7 of the participants made the correct turn onto A6. Hesitation and confusion were noted (e.g., some participants had an inquisitive voice inflection, questioning the taxi instruction’s inclusion of “A6”). The difference in the two groups’ taxi performance was statistically significant ($\chi^2_{Yates} = 7.29$, $p < 0.01$). Finally, for the participant given the taxi “I say again” intervention, the participant made both the correct read-back and correct taxi route. During the post-experiment debriefing the participant stated that hearing A6 stuck out from the routine taxi instructions, but hearing “I say again, A6” made it very salient.

Discussion

This experiment sought to determine if some taxi errors result from expectation failures and if so, test a way of mitigating this problem. Pilot expectations and habits are difficult to overcome in atypical situations; this was demonstrated in that only 1 participant in the control group took the correct route while the 6 remaining participants took the habitual route. This provides converging evidence with Byrne, et. al. (2005, 2006, 2007) that counterintuitive turns are a possible systematic cause of RI errors. In contrast to the control group, all 7 of the participants in the experimental group made the correct turn to A6. These results suggest that our intervention contributed to the experimental group’s success in taking the correct taxi route over that of their habit patterns.

Our anecdotal test of the “I say again” intervention produced quite interesting results. This pilot had distinctly more confidence in the taxi instructions than those in the experimental or control groups. We attribute this to the repetition of the instruction regarding the counterintuitive turn. This repetition provided indication of an awareness of issuing a non-conventional taxi route and also provided confirmation that the taxi instruction was not in error. This type of modification could be accomplished by training each airport’s air traffic to identify non-standard/counterintuitive taxi routes and repeating those sections while issuing taxi instructions.

Conclusion

Runway incursion prevention is crucial to aviation safety. Our intervention was designed to prevent one possible cause of taxi navigation error that has been found, both in our study, and previously by Byrne, et. al. (2005, 2006, 2007). These results provide converging evidence that a systematic and addressable source of taxi error lies in habit patterns and counterintuitive taxi routes. We have demonstrated that interventions that alert and direct pilots at counterintuitive turn points can reduce the prevalence of these errors. With additional research we believe that these types of interventions could be implemented with existing technology in aircraft and on the airfield, and thus could be an effective and timely solution to RI prevention at all sizes of airports. In summary, a possible systematic cause of human error has been identified and an intervention has been proposed and empirically demonstrated to eliminate or reduce these errors.

Acknowledgements

This research was originally submitted to the FAA’s Design Competition for Universities on April 17th, 2008 as a part of the Human Factors Proseminar course at the University of Illinois at Urbana-Champaign under the direction of Professor Alex Kirlik. We appreciate the assistance of Mr. Don Talleur, Prof. Jason McCarley, Prof.

Dan Morrow, and Mr. Jonathan Sivier of the University of Illinois. We also thank Dr. David Foyle of NASA and Prof. Steven Landry of Purdue who were also integral in shaping our research. The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

References

Byrne, M. D., & Kirlik, A. (2005). Using computational cognitive modeling to diagnose possible sources of aviation error. International Journal of Aviation Psychology, 15(2), 135-155.

Byrne, M. D., Kirlik, A. & Fick, C. S. (2006). Kilograms matter: Rational analysis, ecological rationality, and closed-loop modeling of interactive cognition and behavior. In A. Kirlik (Ed.), Adaptive perspectives on human-technology interaction (pp. 267-286). New York: Oxford University Press.

Byrne, M. D., Kirlik, A. & Fleetwood, M. D. (2007). An ACT-R approach to closing the loop on computational cognitive modeling: Describing the dynamics of interactive decision making and attention allocation. In D. C. Foyle & B. L. Hooey (Eds.), Human performance modeling in aviation (pp. 77 – 104). Boca Raton: CRC Press.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. Human Factors, 37(2), 381-394.

Federal Aviation Administration. (2005). FAA Runway Safety Report. Retrieved September 21, 2007, from http://www.faa.gov/runwaysafety/pdf/report5.pdf

Federal Aviation Administration. (2007). FAA Runway Safety Report. Retrieved October 19, 2007, from http://www.faa.gov/runwaysafety/pdf/rireport06.pdf

Federal Aviation Administration. (2008). FAA Runway Safety Report. Retrieved February 13, 2009, from http://www.faa.gov/airports_airtraffic/airports/runway_safety/media/pdf/RSReport08.pdf

Federal Aviation Administration. (2009). Runway Safety Statistics. Retrieved February 12, 2009, from http://www.faa.gov/airports_airtraffic/airports/runway_safety/data/?fy1=2008&fy2=2007&id=fytot

National Transportation Safety Board. (2006). Attempted Takeoff from Wrong Runway Comair Flight 5191 Bombardier CL-600-2B19, N431CA Lexington, Kentucky. Retrieved October 18, 2007 from http://www.ntsb.gov/publictn/2007/AAR0705.pdf

Wickens, C. D., Hollands, J.G., (2000). Engineering psychology and human performance. (3rd Ed.). Upper Saddle River, NJ: Prentice Hall.

SIMULATION RESULTS FOR HIGHLIGHTING RUNWAY SAFETY CRITICAL INFORMATION ON COCKPIT DISPLAYS OF TRAFFIC INFORMATION

Peter Moertl
Kathleen McGarry
James Nickum

MITRE CAASD
7515 Colshire Drive
McLean, VA 22102-7508

This paper describes the results of a human in the loop simulation that evaluated enhancements to a Cockpit Display of Traffic Information (CDTI). Runway safety critical information was highlighted on the CDTI to facilitate flight crew situation awareness and conflict detection for different groups of pilots (General Aviation and commercial) and under different operational settings (crew and single pilot). The evaluated CDTI enhancements are currently being defined by RTCA Special Committee 186 and based on the use of Automatic Dependent Surveillance-Broadcast (ADS-B). The results suggest that highlighting of certain traffic relevant information is promising to increase pilot's hazard detection over a normal CDTI but requires further refinement. This paper describes the primary simulation performance and subjective evaluation results and offers directions for further research and development.

The International Civil Aviation Organization (ICAO) and the Federal Aviation Administration (FAA) both define Runway incursions (RIs) as any occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle or person on the protected area of a surface designated for the landing and takeoff of aircraft. RIs at airports in the United States have been a major area of concern for the National Airspace System (NAS) for the past several years.

To address this problem, extensive human factors research has been performed. This research has generally indicated that human behavior is a root cause for runway incursions (FAA 1998). For example, Adam and Kelly (1996) performed an extensive survey with 1437 pilots from two commercial airlines and interviewed a subgroup of them to identify causal factors for RIs. The study identified factors contributing to runway incursions as related to airport characteristics such as signage, markings, lighting, runway geometry, as well as lack of familiarity of pilots with the airport surface and procedures. Other factors are related to the communication of control clearances via auditory communication channels, which frequently represent an information bottleneck under stress conditions, as well as factors concerning crew and air traffic control operational procedures. Overall, contributing errors may be caused by pilots, controllers (Bales, Gillian, & King, 1989; Steinbacher, 1991), or surface vehicle operators.

Over the last years, multiple approaches have been adopted to reduce runway incursions and collisions. In the United States, enhanced airport surface markings (FAA, 2006a) have been introduced and runway occupancy status lighting systems (FAA 2007) have been developed. Also, controller alerting systems such as the Airport Movement Area Safety Systems (AMASS, FAA 2005), and ASDE-X (FAA 2006b) have been developed. New airport designs, such as end around taxiways that allow aircraft to taxi around a runway instead of crossing it have been developed to reduce the occurrence of runway crossings. A Runway Incursion Information Evaluation Program (RIIEP) was created to learn more about runway safety hazards. In addition, the FAA has been providing guidance to airlines for standardizing ground operations (FAA, 2008, FAA 2003), and the FAA and pilot associations have been providing training and education about runway safety to pilots in various formats including workshops, websites, and DVDs. Flight decks have started to be equipped with moving maps that can display the airport and airport surface movement area. Also auditory systems that are intended to increase pilot awareness about surface hazards (RAAS, Honeywell, 2003) have started to be put on airplanes. Further, standards for the Cockpit Display of Traffic Information (CDTI) have been developed for different platforms on the flight deck such as Electronic Flight Bags (EFBs).

International efforts have included the development of an Advanced Surface Movement Control Guidance System (A-SMCGS) that provides surface traffic management, guidance, and alerting functionality to Air Traffic Control (ATC) and pilots (see IFATCA 2003; Roeder et al., 2008). A-SMCGS has been providing safety functionality for ATC and ongoing development are adding functionality for the flight deck (Vernaleken, Urvoy, Klingauf, 2007).

Specifically, a significant amount of research and development activities has been performed on flight deck-based airport surface safety systems (e.g. Jones 2002; Young & Jones 2001). One aspect of that research that has found its way into aviation standards and cockpits are CDTI capabilities that are targeted to increase flight crew situation awareness (see standards "Airport Surface Situation Awareness" and "Final Approach and Runway Occupancy Awareness", RTCA 2003).

The study that is described in this paper builds on these standards, and is intended to enhance CDTIs to render them more effective for increased runway safety. While traffic displays may increase the situation awareness of pilots under many situations when sufficient time is available to scan the display, pilots may need to find relevant information under safety critical situations more quickly. Specifically, when operating in high traffic volumes on the airport surface, CDTIs may be filled with much traffic that is mostly irrelevant for the flight crew. This would distract from the information that is of critical importance. The research that is described here investigated methods to highlight runway safety critical information on the CDTI that should allow flight crews to better understand the relevant aspects of safety critical situations. In the following sections, first that application is described, then simulation results are presented, and finally conclusions are derived.

DESCRIPTION OF APPLICATION

The described application adds traffic and runway highlighting on a CDTI to help enhance flight crew situation awareness and conflict avoidance. The term "highlighting" is here used to describe the information given to a flight crew to identify traffic and runway status that may become a runway safety hazard even if current conditions are normal (Moertl & Duke, 2008). Such highlighting should not actively attract attention of flight crews and does not represent an alert. It is intended to simplify the pilot task of finding runway safety relevant information on a CDTI before an alerting situation occurs. Highlighting occurs if traffic was either currently on a runway, entered a runway, or was on approach to a runway. Thereby, the flight crew uses the CDTI in combination with other information inside and outside the cockpit to obtain traffic situation awareness and determine the appropriate course of action.

Highlighting as provided in the simulation was context independent and not sensitive to the position, movement, and heading of ownship. The chosen implementation was a preliminary, and relative simple implementation that was intended to collect initial feedback that could then be used to determine if more complex implementations would be required.

A runway could be highlighted as either 'in use' or 'occupied' The "in use" highlighting occurred when an aircraft was currently moving on that runway or was predicted to be moving on that runway at high speed (above 40 knots). A runway was highlighted as "occupied" when an aircraft was stopped or moving on that runway at low speed (at or below 40 knots). For the purposes of this simulation, runway use and occupancy by aircraft was highlighted on all runways ahead of ownship.

Runway and traffic highlighting consisted of both graphical and non-graphical elements. Graphical highlighting included: highlighted runway (A in Figure 1), highlighted traffic (B), and ownship symbol (D). Non-graphical highlighting included a message text box (E), flight ID and groundspeed indications (C), an indicated aircraft data block (F), and a highlighted runway label (H). The graphical runway usage highlighting (G) was not used in the simulation, due to technical limitations. The graphical highlighting was only visible when the zoom setting of the CDTI was such that the traffic and/or runway being highlighted were currently in view. The non-graphical highlighting was visible with all zoom settings. The initial set of highlighting used in this simulation were selected by a panel of pilots and designers, and agreed upon by the working group that developed the application.

Figure 1 displays the graphical and non-graphical highlighting. In that figure, ownship is highlighted as a brown, filled-in triangle in the center of the display, taxiing toward runway 29 behind another aircraft. The highlighted aircraft (UPS 42) has landed and is moving at 118 knots on runway 29.

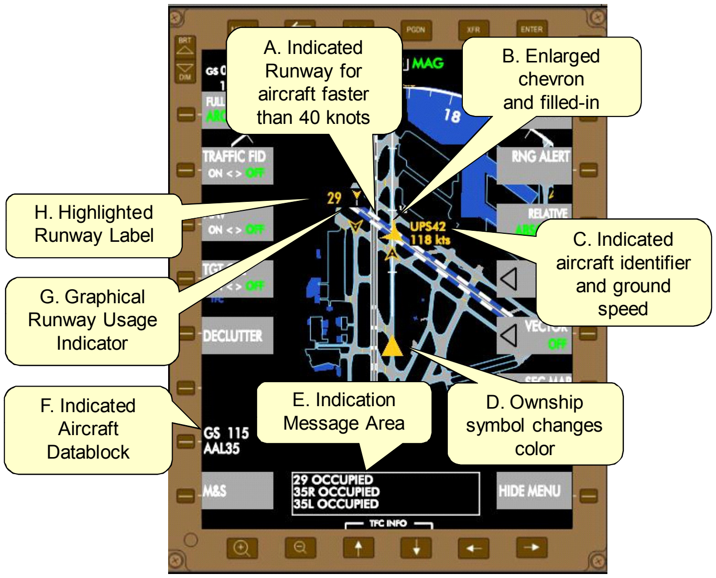


Figure 1 Cockpit Display of Traffic Information and Runway and Traffic Highlighting.

METHOD

The simulation assessed the benefits and limitations of an initial and simple implementation of runway and traffic highlighting on a CDTI under single pilot and crew settings.

Participants

16 pilots participated in this study. Fifteen were male, 1 was female. Eight pilots had private pilot certificates, four had commercial certificates, one was a certified flight instructor, and three had Airline Transport Pilot (ATP) ratings. Their average age was 50 years, ranging from 30 to 73 years, and they had an average flight experience of 3021 hours, ranging from 130 to 20000 hours. A heterogeneous sample of pilots was selected to examine a broad range of human factors related to different pilot experiences, as well as single pilot and crew settings.

Four pilots participated as single pilots, and 12 participated as flight crews. Of the 12 participating as part of a crew, 4 pilots sat in the right seat (pilot monitoring) while a confederate sat in the left seat (pilot flying). The other 8 pilots formed 4 crews of participants.

Scenarios

The simulation used the Louisville Standiford Field International Airport as the setting for all scenarios. The scenarios contained a large amount of traffic to provide an environment where highlighting would be most useful. Participants received clearances from a confederate air traffic controller via radio and heard radio communications with other aircraft that were visible on the CDTI and out the window. All radio communications were on one radio frequency so participants did not have to switch radio frequencies. The pilots monitoring did not perform weight balance or checklists during taxiing to allow them better familiarization with the CDTI. Table 1 shows the eight scenarios with visibility conditions and presence or absence of conflicts. Participants saw scenarios 1, 4, 5 and 8 in the baseline condition (current day CDTI display; no highlighting). They then saw all eight scenarios with highlighting.

Table 1 *Simulation Scenarios*

Scenario	Description	Visibility	Conflict
1	Participant taxies into position, getting a departure clearance while a conflict aircraft is cleared for landing on crossing runway.	High	Yes
2	Participant taxies and crosses runway after holding short for departing traffic.	High	No
3	Participant taxies into position and holds on runway for departing traffic on crossing runway.	Low	No
4	Participant taxies to runway, holds short for landing traffic, then taxies into position and receives a departure clearance while previous landed aircraft has not exited the runway.	Low	Yes

5	Participant is cleared to land while another aircraft is in position and holding for departure on same runway. Aircraft lifts off just prior to participant crossing the threshold.	Low	No
6	Participant is cleared to land while another aircraft receives a land and hold short clearance on an intersecting runway. The participant is told to exit at the end of the runway, causing a conflict with the aircraft that fails to hold short of that runway.	High	Yes
7	Participant lands and is cleared to back-taxi on runway. An aircraft is in position and hold on that runway.	Low	No
8	Participant is cleared to land. After touch down another aircraft lands on the same runway. This requires the participant to initiate an evasive maneuver to avoid collision with the faster moving conflict aircraft approaching from behind.	High	Yes

Simulator

The simulation was performed in a fix-based simulator with a 120 degree out the window view and configured with a primary flight and navigation display. The simulator did not replicate a specific aircraft type and resembled a large, transport category aircraft. The CDTI was shown on an EFB mounted in the left forward field of view for the left seat pilot and in the right forward field of view for the right seat pilot.

Procedure

Participants were briefed about the purpose of the study and then received an introductory briefing on the CDTI and the highlighting. They were given a short training manual to read through, followed by an introduction to the use of the flight simulator. Participants were then shown three scenarios to practice using the simulator and using the CDTI.

After the practice scenarios, participants were fitted with an eyetracker that was used to collect information about pilot's attention allocation during the scenarios. Eyetracking data results are not presented here.

For data collection, participants saw scenarios in two conditions: a baseline condition with a current day CDTI and no highlighting, and the experimental condition with the highlighting of traffic and runways. Participants saw all four baseline scenarios first, followed by the eight experimental scenarios. Scenario order was randomized within both conditions. The experimental condition was presented second to focus feedback on the highlighting after gaining sufficient familiarity with a baseline CDTI.

After each scenario, participants completed a questionnaire. After the last scenario, pilots also completed a post simulation survey and then participated in a debriefing session with the experimenter.

RESULTS

This report presents preliminary simulation results. Analysis of the remaining data is ongoing and will be presented at a future time.

Pilots were asked if they found the highlighting helpful in determining critical runway safety information. A majority of pilots (14 of 16) agreed or strongly agreed that highlighting were helpful. This is a statistically significant finding using chi-square ($\chi^2 = 9$, $p < 0.01$, df = 1). Pilots were also asked if the highlighting provided them with additional information as compared to the baseline CDTI information. Pilots reported that the highlighting did provide additional information ($\chi^2 = 15$, $p < 0.01$, df = 1). Pilots also reported that the highlighting helped them locate relevant traffic ($\chi^2 = 12.25$, $p < 0.01$, df = 1).

Pilots were asked to rank the seven functioning CDTI highlighting features according to usefulness (excluding the runway usage indication). The pilots' rankings were averaged, and are shown for each of the 7 highlighting features by operation type (arrival, departure, taxi) in Figure 2. Lower numbers equate to better ranking. The four best ranked features were: A. runway highlighting; C, highlighting of FID and ground speed; B, enlarged target; and D. Runway Message Area (see also Figure 1). The rankings confirmed pilot comments and observations made during the simulation by the experiment observers. Specifically, the graphical runway and traffic highlighting seemed to provide pilots a fast and intuitive indication of runway occupancy that stood out against other traffic. Participants stated they found the indicated ground speed useful in determining traffic movement on the runway which was especially relevant during approach scenarios. In contrast, the highlighted runway label (H) was ranked as less useful. Finally, runway usage information was also available to pilots via the runway status text box independent of the current range setting of the CDTI. Several pilots reported that the runway status text box could be used as a procedural trigger to initiate a sequence of actions, such as setting a CDTI range allowing them to determine the traffic situation. Though some pilots commented on the highlighting of ownship (D) as a useful general indicator about runway occupancy, pilots overall did not seem to find ownship highlighting overly helpful as that information was available in other places. Finally, pilots reported they did not find the indicated data block (F) useful. That data block indicated FID and groundspeed of the aircraft that was closest to ownship. This low perceived usefulness may have been because the non-graphical nature of this highlighting made it hard for pilots to relate the traffic to ownship, or because it was difficult to determine which traffic it was related to.

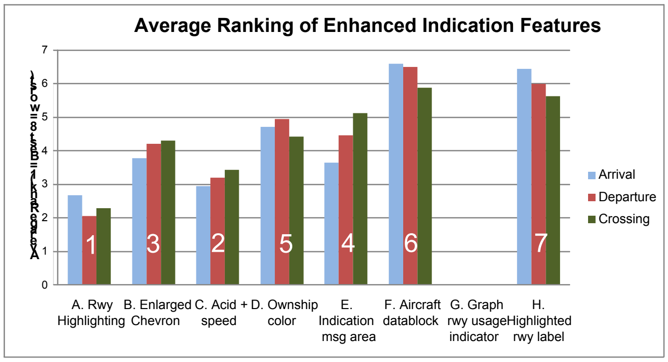


Figure 2 *Pilots' Averaged Rankings of Highlighting Feature Usefulness*

Participants indicated that zoom usage depended on whether highlighting was presented or not. There was a difference in agreement to the statements on zoom usage when pilots used a baseline CDTI versus when they used highlighting (5 response levels, $\chi^2 = 24.67$, $p < 0.001$, $df = 3$). A majority of pilots expressed that zoom usage increased workload in the baseline condition ($\chi^2 = 4.8$, $p < 0.05$, $df = 1$) but not so in the highlighting condition ($\chi^2 = 1.8$, $p > 0.05$, $df = 1$). This perceived difference in workload for adjusting the zoom level between the conditions was rather small compared to the overall scenario workload, where no difference in workload between conditions could be detected.

In the baseline condition participants did not reach agreement on the statement of whether the CDTI distracted them from surface operations or not. In the highlighting condition, pilots reported that the CDTI did not distract them from operating the aircraft ($\chi^2 = 12$, $p < 0.01$, $df = 1$).

While participants generally agreed on the usefulness of the enhanced display, participants also commented on display limitations. Specifically, participants did not reach agreement on the question if the right amount of traffic was highlighted. However, five pilots commented that too much traffic was highlighted and that this could be confusing. Pilots also asked that only safety information that was relevant to ownship should be highlighted.

To estimate the effectiveness of highlighting and how it might increase pilots' conflict detection, participants were exposed to conflict scenarios. Participants saw three of these conflict scenarios in both the baseline and the experimental condition. Reported conflict detection was compared between the two conditions to determine if highlighting traffic and runways increased conflict detection performance. Results are shown in Figure 3.

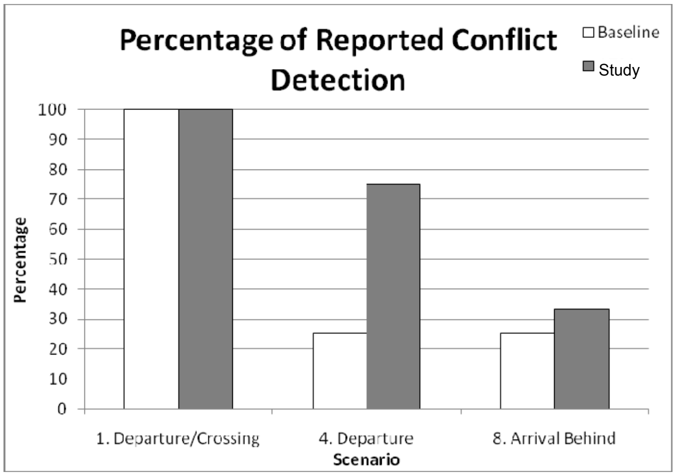


Figure 3 *Conflict Detection Performance with Highlighting (Study) and Without (Baseline)*

Results showed that highlighting seemed to increase reported conflict detection in two scenarios though it did not reach statistical significance. In the departure scenario (4) participants were cleared for takeoff while an aircraft was ahead on the runway. In that scenario, the percentage of reported conflict detection went from 25% to 75% and fewer pilots initiated a takeoff than in the baseline condition, however this was not a statistically significant finding. The increase in reported conflict detection in the experimental condition was likely due to the fact that pilots had to set the CDTI at the appropriate range setting to detect traffic on the departure runway in the baseline condition, but were provided with safety information in the experimental condition independent of zoom setting. In the arrival behind scenario (8), an aircraft landed behind and converged with ownship while the pilots taxied on a runway. In the baseline condition, pilots would have needed to have the zoom set to a level that included the approach area to see the conflict aircraft, while in the highlighting condition, pilots were provided with highlighting regardless of the zoom setting. However, there was no statistically significant difference in reported conflict detection between the two conditions. Four pilots who reported not detecting the conflict, did state that they had noticed highlighting on the CDTI. This finding suggests limitations of highlighting in cases where pilots cannot see the conflict aircraft out the window.

In the departure/crossing scenario (1), while conflict detection was relatively easy, pilots reported using different information for initial conflict detection in the baseline versus the highlighting condition. More pilots used the CDTI to detect the conflict in the highlighting condition (80 % of trials) than in the baseline condition (50 % of trials, marginally significant difference between conditions, $\chi^2 = 3.6$, $p < 0.06$, $df = 1$).

SUMMARY

Pilots with varying degrees of experience evaluated a set of display features for highlighting traffic and runways on a CDTI. Highlighting was based on a simple algorithm that graphically indicated runway usage and occupancy. Pilots also evaluated the highlighting features as a way to improve detection of critical runway safety information and hazards by using the CDTI both with and without highlighting. Preliminary simulation results demonstrated that participants generally reported runway and traffic highlighting was helpful in locating relevant traffic information on the CDTI though no statistically significant performance differences could be found. Pilots rated the usefulness of the various aspects of highlighting features. They rated runway highlighting, enlarged chevron, aircraft flight ID and groundspeed, and text message box as more useful than the other highlighting features. Pilots also identified shortcomings and areas for improvement with the highlighting. Pilot comments related to areas for improvement will be reported at a later time.

Several pilots expressed their preferences for fewer highlighting, and, specifically only if traffic was relevant for ownship. We take this to mean runway and traffic highlighting may be improved by making it sensitive to ownship's context of operation.

We also identified limitations of the experimental methodology. Because we assessed the CDTI usage under a broad range of pilot experience and cockpit settings, the generalizability of results for a specific environment is limited. Also, because we did not always task pilots to the levels of workload that they would experience under real world operations, we may have allowed our participants more time to dedicate to the CDTI than they would likely have had available in their real work place. On the other hand, under real operational conditions, pilots would have had more time to become familiar with the system and likely had more training. Therefore, we consider the results of our study as somewhere in-between maximal and minimal expected benefits.

In terms of hazard detection, highlighting was not associated with deterioration in pilot reported conflict detection performance. Highlighting seemed to be less effective in supporting conflict detection when the traffic was behind ownship and not visible in the forward field of view. In the departure scenario, while there seemed to be improvement in reported conflict detection, there was no significant difference between conditions. This suggests the need for further research into the benefits of highlighting over a baseline CDTI, looking specifically at objective performance data.

The findings of this study point toward important considerations for the usage of CDTIs. The amount of information on a CDTI can be considerable. It can include runway and taxiway layout information, buildings, hold lines, centerlines, traffic, traffic movement and interrelations thereof. This information is shown in a relatively small display area that needs to be continuously adjusted to ensure needed information is in view. For example, during taxiing, a flight crew may select a close-in zoom setting to confirm their intended taxi- and runways. However, while crossing a runway, they may use a different zoom setting that allows them to check for traffic on the entire runway. Extracting the needed information requires a planned effort by the flight crew and needs to be considered as part of the flight crew activities. Based on pilot comments on the amount of highlighted display features on the CDTI, future research will attempt to reduce the amount of highlighted safety information. The displayed information should match a minimum set of information that allows pilots to determine hazards in few, quick glances while minimizing the amount of interactions with the CDTI. “Only give me the information that I need” describes what we frequently heard from pilots.

These simulation results will be used to refine the display and triggering rules of traffic and runway highlighting. A subset of the seven initial highlighting features will be used and more context dependent triggering rules will be implemented. In addition, we think to have determined a need for alerting because pilots had difficulty detecting a conflict either when their attention was not dedicated to the CDTI or when the traffic was outside their field of view. An auditory and visual alert message should help pilots to resolve such situations. We expect that alerting will play a significant role in increasing conflict detection and resolution because alerts will actively attract the attention of flight crews toward critical safety information and thereby facilitate faster responses. We are continuing our work by designing and evaluating an added alerting capability and will report this work in the future.

DISCLAIMER

REFERENCES

Adam, G., D. Kelley, J. G. Steinbacher, 1994, Reports by Airline Pilots on Airport Surface Operations: Part 1. Identified Problems and Proposed Solutions for Surface Navigation and Communications, Mclean, VA, The MITRE Corporation, MTR 94W0000060.

Adam, G., D. Kelley, 1996, Reports by Airline Pilots on Airport Surface Operations: Part 2. Identified Problems and Proposed Solutions for Surface Operational Procedures and Factors Affecting Pilot Performance, Mclean, VA, The MITRE Corporation, MTR 94W0000060, v2.

Bales, R. A., M. R. Gilligan, & S. G. King, 1989, An Analysis of ATC-Related Runway Incursions, with Some Potential Technological Solutions, Mclean, VA, The MITRE Corporation, MTR-89W00021.

Federal Aviation Administration's Runway Incursion Program Office (ATO-102) (1998), Airport Surface Operations Safety Action Plan to Prevent Runway Incursions and Improve Operations, Washington, DC, Department of Transportation.

Federal Aviation Administration (2003), Advisory Circular 120-74A Guidelines to standardize Flight Crew Procedures during Taxi Operations, Washington, DC, Department of Transportation, Federal Aviation Administration.

Federal Aviation Administration (2003), Advisory Circular 91-73 Single Pilot Procedures During Taxi Operations, Washington, DC, Department of Transportation, Federal Aviation Administration.

Federal Aviation Administration (2005), AMASS Specifications, Washington, DC, FAA-E-2069a.

Federal Aviation Administration (2006a), Advisory Circular – AC 150/5340-1J, Washington, DC, Federal Aviation Administration.

Federal Aviation Administration (2006b), ASDE-X Interface Control Document, Washington, DC, NAS-IC-24032410-xx.

Federal Aviation Administration (2007), Runway Status Light Engineering Briefing, Washington, DC, 64a.

Federal Aviation Administration (2007) Automatic Dependent Surveillance Broadcast (ADS-B) Out Performance Requirements to Support Air Traffic Control (ATC) Service, Notice of Proposed Rulemaking, Washington, DC, Department of Transportation, RIN 2120-AI92.

Honeywell (2003). Product Description Runway Awareness and Advisory System. Unpublished document.

IFATCA (2003), A-SMGCS Implementation in Europe, Publication of the International Federation of Air Traffic Controllers' Associations.

Jones, D. (2002). Runway Incursion Prevention System Simulation Evaluation. Presented at the 21st Digital Avionics Systems Conference.

Moertl, P. M. & J. Duke (2008) Enhanced Traffic Situational Awareness on the Airport Surface with Indications and Alerts (ATSA SURF IA) Application Description, Version 1.2.1. (Working document, contact pmoertl@mitre.org for a copy.)

Roeder, M., Jakobi, J., Nuzzo, A., Teotino, D. (2008), EMMA2 Airport Surface Management, Presentation at ASAS TN, April 14-15, 2008, Paris, France.

RTCA (2003) Minimum Aviation System Performance Standards for Aircraft Surveillance Applications (ASA), DO-289, Washington, DC, RTCA.

Steinbacher J. G., 1991, An Analysis of ATC-Related Runway Incursions in the National Airspace, Mclean, VA, The MITRE Corporation, WP 91W00234.

Vernaleken, C., Urvoy, C., Klingauf, U. (2007). International Symposium Enhanced Solutions for Aircraft and Vehicle Surveillance. March 6-7, 2007, Bonn, Germany.

Young, S. D.; and Jones, D. R. (2001): Runway Incursion Prevention: A Technology Solution. Proceedings of the Flight Safety Foundation 54th Annual International Air Safety Seminar, Athens, Greece, Nov. 5 – 8, 2001.

NEAR-TERM NEXTGEN AND CLASS 2 EFBs

Thomas L. Seamster, Ph.D.
Cognitive & Human Factors
Santa Fe, New Mexico
Barbara G. Kanki, Ph.D.
NASA Ames Research Center
Moffett Field, California

This study is based on data collected at the Electronic Flight Bag (EFB) Advanced Software and Authorization Workshop for US operators currently involved in EFB software evaluation or implementation for their own fleets. With most US operators not taking delivery of new, larger aircraft in the next few years, they are considering ways of displaying near-term NextGen data on board existing aircraft through systems such as the EFB. The workshop collected operator near-term needs in the areas of EFB user interface and standardization and EFB advanced software applications. The analysis of the data collected during the workshop provided a prioritized list of operator needs over the next few years with an emphasis on runway safety and related NextGen systems. The study reports on those needs in the context of near-term NextGen systems and Class 2 EFBs.

The NASA/FAA Operating Documents Group held the Electronic Flight Bag (EFB) Advanced Software and Authorization Workshop jointly sponsored by NASA Ames Human Systems Integration Division and FAA ATO-P Research and Development during the last quarter of 2008. The primary audience for this workshop was North American operators currently involved in EFB advanced software evaluation or implementation. Topics for the workshop included implementation of EFB software applications such as moving maps, satellite weather, and data overlays. This workshop emphasized operator needs rather than manufacturer or vendor capabilities and provided operators with an opportunity to identify key EFB issues with a focus on EFB advanced software applications. Operators had an opportunity to hear about and discuss their EFB challenges, lessons learned, and how the EFB authorization process should be streamlined. The researchers, who have focused on the effects of EFB on crew performance (Kanki & Seamster, 2007) took the opportunity to collect data on operator EFB advanced software needs and issues and then had the operators rate each of those items with regard to how important they were in the context of their operations. The most important EFB issues identified through this workshop point the way to several near-term safety and efficiency improvements especially in surface operations that can be developed and implemented while the aviation industry is working toward full NextGen implementation.

Background

NextGen incorporates several significant advancements to air traffic control to meet the substantial increase in traffic anticipated between now and 2025. NextGen concentrates on the main technological shifts from ground based to satellite navigation, from voice communication to digital data, from disparate to centralized weather with the ability to operate in a fuller range of adverse weather and terrain conditions. NextGen is being planned and designed top down and its full implementation will require the implementation of a number of operational improvements that will not be available until the longer term. Looking at the near-term, NextGen is conceived from the bottom up starting with specific research and development activities, some of them leading to enabling technologies which in turn combine to provide more accurate navigation, weather and real-time broadcasting of related information necessary for the more accurate and tightly spaced management of air traffic. The research and development activities cover many areas including trajectory and performance-based operations, safety management, security, weather information services and a net-centric infrastructure.

Although these research and development areas are interrelated, it helps to focus on one area, in this case Trajectory-Based Operations (TBO). A full implementation of TBO requires near real-time and highly accurate navigation, surveillance and weather information that is accessed over a secure national integrated network. Prior to that full implementation, there are several enabling technologies that will, by themselves, improve operational performance, with an emphasis on crew performance on the flight deck. Starting with the research and development, TBO will require the technical development of critical data exchange of flight clearances, algorithms for real-time trajectory management that incorporate multiple user preferences, separation standards and automated en-route flight plan negotiation that accommodates changing weather and other operational conditions (JPDO, 2008). There are also several research and development areas that look at pilots and the allocation of roles, responsibilities and tasks between controllers and flight crews as well as between computers and their operators. Although crew performance using the EFB (Seamster & Kanki, 2007) is not a driving force across NextGen research, it was a key concern for the workshop participants who represented the operators and who, in most cases, were active pilots.

Near-Term NextGen

The timeframe being addressed in this study is from 2009 through 2012 which coincides primarily with the near-term NextGen work plan. One of the near-term operational enhancements for TBO is improved surface traffic management. This operational enhancement is based on a set of interrelated enhancement with an emphasis of controller data and decision aids. These enhancements are designed to increase both safety and efficiency of the surface movements of not only aircraft but in the long run, also of other ground vehicles. Specifically, it will improve the safety of active runway crossings and reduce aircraft departure wait times (JPDO, 2008). From a top-down perspective, improved surface traffic management requires advanced surface management systems to reduce the time aircraft spend on the surface as well as to optimize the use of gates, taxiways, and runways under a full range of operating conditions. NextGen plans to improve surface movement through the combination of automation, transmission of data instead of just voice communications as well as improved surveillance and displays. The full implementation of improved surface traffic management will require systems integration between surface and aircraft automation. The plan is also to include a runway incursion alerting system that provides controllers and pilots notification of potential incursions. This has been identified as an area needing additional research to determine key alert characteristics including the form, context and other human factors issues (JPDO, 2008). A related area for technology that will extend these surface capabilities will provide aircraft with the ability to taxi in near-zero visibility through a combination of Automatic Dependent Surveillance-Broadcast (ADS-B) OUT along with airport moving map and flight deck traffic displays. From a top-down view of NextGen, improved surface traffic management requires a complex of research and technology developments to achieve full implementation. By shifting the perspective away from a top-down, controller-centric view to a set of near-term operator and pilot needs, it is possible to obtain a clearer view of some less complex innovations that can lead to improved surface safety and efficiency in the next few year.

The NextGen work and implementation plans emphasize the Air Navigation Service Provider (ANSP) as it tracks the delegation of separation responsibilities ensuring that the responsibility is clearly communicated. The long-term plan leads to what NextGen calls, cooperative surveillance, based on ADS-B IN and ADS-B OUT where data is available about all aircraft in the area. Devices and displays will be needed for both the controller and flight deck side of operations to support receiving and understanding flight and traffic information. The air or pilot side can be enhanced through the use of flight deck displays in the graphic representation of surface clearances, conditions and changes. Some of the enabling technologies that will require flight deck display of airport and surface data include electronic maps and charts with own-ship position on airport ramps, taxiways, and runways with the eventual representation of other surface vehicles. Additionally, there will be the cockpit display of nearby surface traffic. This will be followed by a more advanced display of traffic information that includes both surface and airborne aircraft. A further capability will be a device to allow aircraft to expedite the crossings of active runway and perform delegated separation procedures at high-density airports as well as under low-visibility conditions. This complex of technologies may take a decade or more to develop, but there is an important tool that is being implemented by a number of operators that can provide pilots access to some of the capabilities that will improve surface safety and efficiency.

NextGen and the EFB

The EFB has the potential to display near-term NextGen capabilities in a cost effective manner on existing aircraft. This coincides with NextGen implementation plans to leverage existing aircraft systems and capabilities throughout the near-term. The EFB is being used by an increasing number of operators to display charts, manuals, and weather data. Recently, the FAA has allowed portable EFBs to display own-ship position on airport moving map displays. The FAA further authorizes installed EFBs that are certified to be integrated with other avionics, such as the Flight Management System (FMS), to support some of the implementations of the advanced NextGen capabilities. The EFB could play a significant role in the NextGen scenario where pilots will receive the final flight plan data, which could be in both a text and graphic format. Own-ship position would also be displayed on the flight deck showing it as it taxies along with the position of other aircraft in the vicinity and other surface vehicles. Rather than having a number of separate devices and displays, the EFB could also be considered as a way to provide runway incursion alerts integrated with the moving map and own-ship position.

EFBs have different certification requirements depending on their classification. Class 1 EFBs are portable computing devices that are not mounted to the aircraft. Class 2 EFBs are computing devices that are attached to the aircraft during normal operations while Class 3 EFBs are installed on the aircraft allowing for a wider range of applications. The name, electronic flight bag, describes the initial concept of the device which was to replace the pilot’s bag of operational charts and documents with a computer and display that would provide full access to that information in a more usable form. As the pilot’s EFB has evolved and has been networked not only with the other pilot’s EFB but also with other flight deck systems, it is being viewed by pilots and operators as an innovative display and control device that can be used well beyond its initial intent providing a number of NextGen functions.

A candidate control and display of near-term NextGen data on the flight deck for US operators is the Class 2 EFB. This is due in part because major carriers will not be taking delivery of substantial numbers of new aircraft in the near-term with the overall estimates of new aircraft deliveries to the domestic operators being revised downward. This will affect the availability of Class 3 EFBs that are generally obtained through new aircraft purchases by the major operators. With approximately 17% of aircraft being stored by the major US operators, and three of those operators with more than a quarter of their aircraft stored (see Figure 1), the demand for near-term deliveries of new aircraft with Class 3 EFBs is has been reduced.

Methods

The workshop was designed to provide participants with an opportunity to present, discuss and rate EFB advanced software experiences and needs. There were 25 participants at the workshop involved in identifying the key EFB advanced software issues. They included representatives from the main operators evaluating or implementing EFBs as well as other industry members including regulators. All participants were given the opportunity to specify EFB issues, and then 16 of the participants performed the actual rating of those issues. The raters had an average of eight years of EFB experience and an average of 4,300 hours of total flight time. The range of total flight time was between 0 hours for the three engineers and 15,000 hours, with the raters having substantial operational experience.

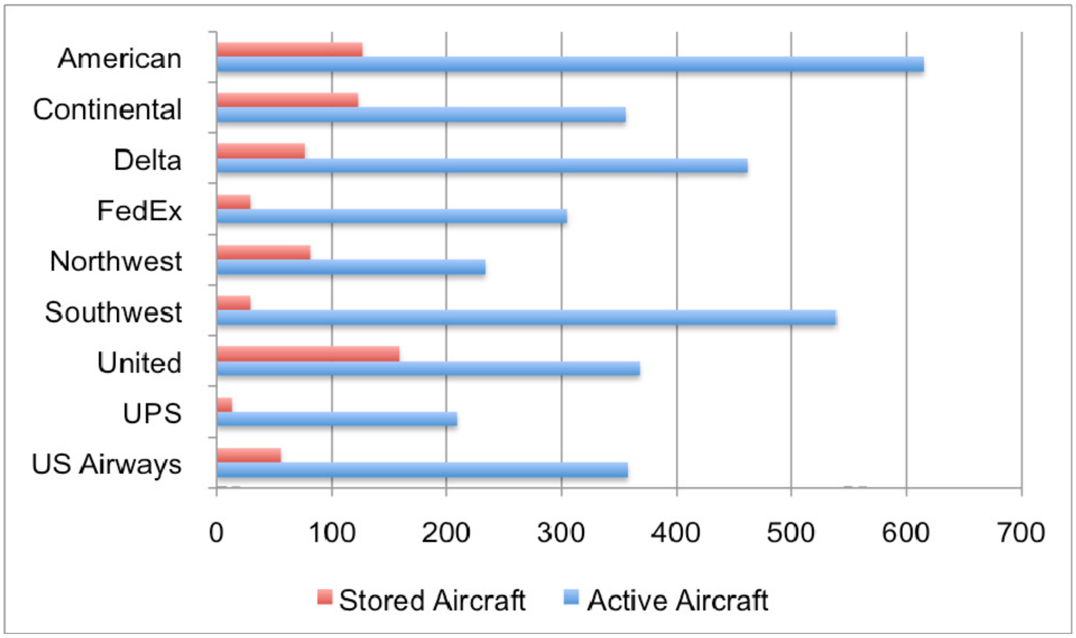


Figure 1. Major US operators with approximate numbers of active and stored aircraft (data source Planesregister.com).

Leading up to the workshop, participants were asked to submit topics that they wanted to present and also those they were interested in hearing about. During the workshop, participants, working as a group were encouraged to identify EFB issues in the following four areas plus any additional EFB issues:

1. EFB User Interface and Standardization to include Multi-Tasking, Color Coding and Symbols
2. EFB Advanced Software Applications including MET/WX, Charts Graphical Overlays.
3. Integrating EFB with SOP, Training, Best Practices and Flows
4. Improving Crew Performance with EFB to Include Situation Awareness, Workload Management and Runway Safety.

After all the EFB issues were identified and discussed, participants were provided with a ratings form listing the 25 issues organized by the above areas. They were asked to rate each issue as to its importance using a six-point scale where 6 represented 'Extremely Important,' and 1 represented 'Extremely Unimportant."

Results and Discussion

Although some of the issues proposed by the participants pertained to more than one category, the issue was placed in the area where it was first identified. The participants specified six issues related to the EFB touch screen functions, standardization of information organization, high level EFB functions, lower level chart and map details as well as standards applied to key features of the ground-based and flight deck EFBs. They also identified issues specifically related to advanced software including the display of own-ship position, airport moving maps, other traffic, and weather. In the area of EFB SOP and training, the group specified issues of crew coordination, company procedures and best practices, and integration with existing training and crew assessment. In the area of improving crew performance, participants were concerned with managing multiple applications on the EFB, integrating applications and standard usage of some of the advanced applications.

Table 1. *Top 10 rated EFB issues based on 16 raters with 6 representing Extremely Important.*

EFB Advanced Software Issues	**Mean Rating**
Top Five Most Important Issues	
Display of aircraft position	5.47
Runway and taxiway safety	5.43
Airport moving map plus traffic and advisories	5.20
How many and which applications degrade crew performance	5.19
Managing multiple applications	5.13
Next Five Most Important Issues	
Coordination across pilots	5.06
Multi-tasking issues, minimize button pushes	4.93
Select Function: touch/select/drag/scroll interface design and training terminology	4.87
Company-specific standard callouts, EFB use	4.81
Integration with training, qualification standards	4.81

The top ten EFB issues rated as most important are shown in Table 1 along with their mean ratings. The top five important issues, with an average rating between Extremely to Very Important group into an integrated set of EFB research and development activities that should be considered as a way to make available some NextGen data in the near-term. Airport moving map with own-ship position is just now being approved for operational use on Class 2 EFBs. Operators see the importance of extending that functionality to further enhance safety by determining ways to add traffic and advisories plus other available NextGen data. This combines with the issues of integrating, what are currently, separate applications into a form that will improve pilot information management without degrading crew performance. The next five important issues group into a set related to crew coordination, SOP, training and the EFB input interface research activities. Based on these two groupings of issues, operators are most concerned with the integration of additional surface data and advisories into an easy to use EFB display. They have a secondary concern on how to ensure that this advanced technology can be used to improve crew coordination through procedures and training as well as how to improve the EFB interface, with an emphasis on inputs via the touch screen.

Class 2 EFBs provide operators with an economical way of displaying and controlling some of the important near-term NextGen data on existing aircraft. Interpreting the ratings data, one of the research and development challenges is to provide that data in ways that will improve, rather than potentially degrade, crew performance. From a research perspective, there are several key challenges for providing the display and control of this NextGen data in an integrated manner, especially on Class 2 EFBs. One research area involves evaluating the different user interface metaphors as the EFB transitions from being just a flight bag that displays documents and charts to becoming a flexible display of both static and dynamic information to improve decision making and situational awareness while reducing crew workload. The industry is working with a number of distinct metaphors that are either under development or that have been implemented (see Figure 2 for some examples). Some of the metaphors are based on the FMS controls with either hard or soft buttons around the edges of the display for user input. Other metaphors have been derived from paper document trip books or clips that pilots have used traditionally to organize their charts before and during flights. Still other metaphors under development have utilized a browser for accessing and displaying information. The browser metaphor shows potential for transitioning the EFB from a flight bag to a more display and control device, but developers will have to address the challenges and limitations of using a browser interface for critical applications on a flight deck rather than at a desktop.

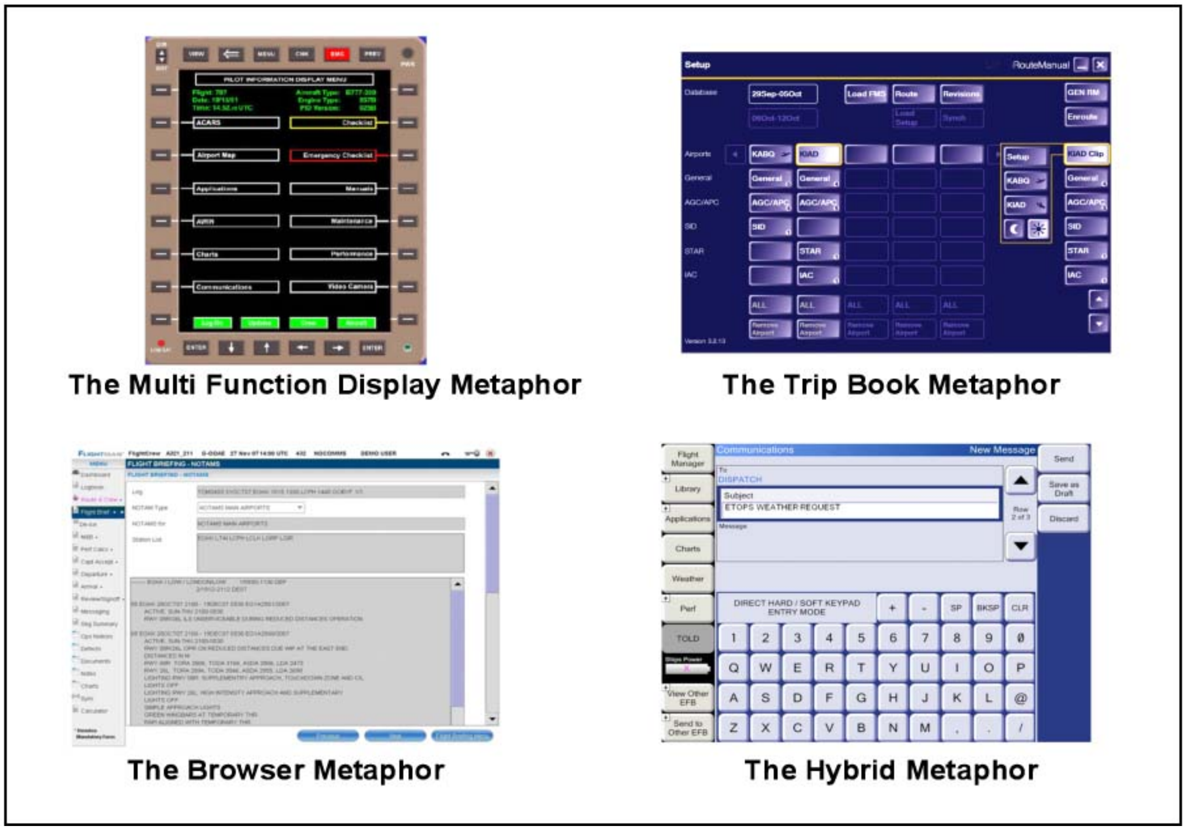


Figure 2. Examples of EFB User Interface Metaphors under development or in current use.

In order to incorporate NextGen capability, the current trip book metaphor needs to be extended and the browser interface would have to be refined before it can be used as a way to access information on the flight deck. In addition, the EFB Class 2 small screen size presents substantial limitations for data display. On most flight decks, the EFB screen size cannot be increased substantially in part because of the limited space and potential for blocking existing displays and controls. Even with these limitations, the Class 2 EFB should be evaluated as a way of graphically displaying additional airport data such as traffic, taxi clearances, closed runways, construction and other temporary obstacles normally made available to pilots through text messages. With operationally relevant research and development, EFB constraints can be overcome allowing the display of safety critical near-term NextGen data.

References

JPDO (2008). *Next Generation Air Transportation System Integrated Work Plan: A Functional Outline*. Washington, DC: Joint Planning and Development Office.

Kanki, B. G., & Seamster, T. L. (2007). Optimizing EFB use through training, standards, and best practices. *Proceedings of the Fourteenth International Symposium on Aviation Psychology*. Dayton, OH: The Wright State University.

Seamster, T. L., & Kanki, B. G. (2007). Beyond Electronic Flight Bag (EFB) approval: Improving crew performance. *Proceedings of the Fourteenth International Symposium on Aviation Psychology*. Dayton, OH: The Wright State University.

ANALYSIS OF RAMP DAMAGE INCIDENTS AND IMPLICATIONS FOR FUTURE COMPOSITE AIRCRAFT STRUCTURE

Barbara G. Kanki, Ph.D.
NASA Ames Research Center
Moffett Field, CA 94035-1000

Connie L. Brasil
San Jose State University Research Foundation
NASA Ames Research Center
Moffett Field, CA 94035-1000

As aircraft manufacturers use increasing amounts of composite materials in primary aircraft structures, an understanding of how composite damage may occur is crucial. One likely setting for composite damage events is the ramp and gate areas where "ramp rash" is a common occurrence. Costly consequences to airlines and the potential to jeopardize safety are an everyday hazard. In order to better understand how such events unfold in today's operations, 104 ramp damage reports that were voluntarily submitted to the NASA Aviation Safety Reporting System (ASRS) were analyzed. Factors including environmental conditions, aircraft state, aircraft damage locations and types of ramp vehicles or equipment involved were examined in order to describe the scenarios in which damage occurs. Results provide a starting point for identifying and characterizing possible operational risks for tomorrow's advanced composite aircraft.

The manufacturers of the next generation of commercial transport airplanes are making a major shift in airframe technology. Primary components that have typically been constructed of metal are now being designed with composite materials. In the 1980s and 90s composites were widely used but never exceeded more than 12-14% of the airframe by weight. Now, in service for less than two years, the Airbus A380 is constructed with about 25% composite materials by airframe weight. Scheduled to be in service by next year, the Boeing 787 makes the most significant shift as it is produced with 50% composite structure by weight, with almost 100% of the aircraft skin/fuselage being composite materials (Boeing, 2007). It is undeniable that the use of composites provides great benefits. Weight savings alone will result in significant fuel savings, and resistance to corrosion and fatigue is expected to lengthen maintenance intervals thereby decreasing maintenance costs over the lifespan of the aircraft.

The present-day, predominately metallic, aircraft have accumulated a long service history and a knowledge base of standards and best practices from which manufacturers, regulators, and operators draw upon with confidence. In contrast, the introduction of advanced composite aircraft with very little comparable service history, present new unknowns. Huang and Lin (2005) note:

> Past reliability and structural risk studies have focused on fatigue of aging aircraft, which is mainly an issue unique to metal structures. Composite structures are fatigue and corrosion resistant, but are much more sensitive to damage threats such as hail, bird strikes and ground vehicle collisions because of brittle behavior during failure. Furthermore, there may be no visible evidence of damage to composite structures, even though significant internal damage has been sustained. (p. 2)

Such concerns provide an impetus to researchers and industry groups to investigate some of these anticipated risks. For example, the Commercial Aircraft Composite Repair Committee (CACRC) is an international industry group that shares regulatory and research updates, and develops standards for composite maintenance processes and materials. Operators in this group raise many issues, ranging from damage detection and characterization to specific repair problems. In addition, they offer valuable insights into the nature of damage threats to composite structures from their own operational experiences. Blohm (2007) gives detailed examples of damage that involves ground service vehicles, towing and docking equipment and passenger jet bridges, as well as the more typical cases of runway debris and tire separation. Figure 1 below illustrates a simplified view of the ramp environment and many of the potential everyday threats to composite aircraft structures.

Figure 1. Typical ramp and gate area with a variety of service carts, vehicles, and passenger boarding equipment.

Kim (2008) is conducting a detailed investigation of wide area blunt impact damage to composite fuselage areas that are associated with ramp activity.

> With new all-composite fuselage transport aircraft coming into service, significantly more composite skin surface area is exposed to ground vehicles and equipment. To address the difficulties that exist in being able to predict and detect the damage resulting from blunt impact, and to aid in assessing its effect on structural performance, the development of basic tools to characterize blunt impacts is needed. Of particular interest is damage that can be difficult to visually detect from the exterior, but could be extensive below the skin's outer surface. (Kim, 2008)

Another means of identifying and characterizing the current potential operational threats associated with ramp damage is through a query of the NASA Aviation Safety Reporting System (ASRS) database. Since 1976, the ASRS has been a repository of voluntary, confidential safety information provided by aviation personnel. The largest percentage of reports comes from pilots, but there is small but steady input from controllers, mechanics, ramp workers, flight attendants, dispatchers and others. Since ASRS reports are submitted voluntarily, they are subject to self-reporting biases and the data cannot be used to infer the prevalence of the specified problems within the entire National Airspace System. Nevertheless, the database provides a useful means of acquiring a large number of firsthand reports on a variety of aviation safety issues. Of particular value are the reporters' narratives that often provide great detail on the context, conditions and other personnel and organizations involved in the incident. Pertinent to the topic of ramp damage, the ASRS introduced a specialized maintenance reporting form in 1996 and began to actively encourage the reporting of ground incidents. Thus, in this study, we were able to investigate the problem from the perspective of ground personnel as well as pilots. Detailed information on the ASRS can be found on the following website. http://asrs.arc.nasa.gov/overview/summary.html

Method

Search Criteria

In order to select records that would best fit our research interest, we used the following search criteria in the ASRS Online Database query conducted in August 2008 at
http://akama.arc.nasa.gov/ASRSDBOnline/QueryWizard_Begin.aspx

- Operator: Air Carrier
- Federal Aviation Regulations (FAR) Part: 121
- Flight Phase: Ground: Parked, Preflight, Pushback and Maintenance
- Event Type: Ground Encounters; Vehicles and People and Ground Excursions; Ramp
- Text: Damage (Narrative and Synopsis)

Of the 140 reports found in the initial ASRS search criteria, 36 reports were removed due to lack of damage and/or irrelevance to the ramp damage issue. The remaining 104 incidents were reported between 1999 and 2007. Aircraft type is included in each report description but there were too many different types to be able to partition the data with sufficient numbers of aircraft in each category.

Factors of Interest

Environmental Conditions. Time of day and environmental conditions were two simple factors considered to have possible impact on ramp incidents. The ASRS report form gave the options of daylight, night, dawn or dusk for time of day. Environmental conditions included options such as ice, snow, rain, fog, thunderstorm, and other.

Aircraft State. In order to better understand when aircraft damage occurred, we defined a variable called Aircraft State to capture the operational phase and whether the aircraft was parked or moving at the time of the event. This was further complicated by incidents in which aircraft that were supposed to be parked, moved due to a malfunction or error. Thus the five Aircraft State values include the following:

1. Aircraft moving: AC parked but brakes or chocks malfunction or error made
2. Aircraft moving: AC during pushback
3. Aircraft moving: AC taxiing
4. Aircraft not moving: AC parked with brake set
5. Aircraft state unknown

Ramp Vehicle/Equipment Type. A large variety of vehicles and ground equipment operate in the vicinity of aircraft on the ground, particularly during turnarounds at the gate. Collisions involving catering vehicles, baggage carts, passenger-boarding bridges, and other servicing equipment can cause significant damage. In addition, equipment employed for cargo and passenger loading and unloading are obviously an integral part of every arrival and departure. We wanted to see what types of equipment or service vehicles were involved in aircraft damage. After discovering more than 17 different types of vehicles, we collapsed them into six categories.

1. Belt or cargo loaders
2. Carts: including baggage carts, maintenance carts and oxygen carts
3. Passenger boarding equipment; jet bridge and passenger stair trucks
4. Service vehicles: including catering, fuel, lavatory, vans and various unknown ground vehicle types
5. Service/maintenance equipment: including deicing truck, lift equipment and other unknown ground equipment types
6. Tugs or tow bar

Note that when ‘unknown’ is applied, it meant that the reporter did not know the exact type of service vehicle or equipment was involved, or they did not actually observe the vehicle or equipment firsthand.

Aircraft Damage Location. We also wanted to learn more about where on the aircraft the damage was located. As manufacturers make the shift in airframe technology from metals to composites it will be important to know where potential danger zones may exist. The report narratives usually indicated where the main aircraft damage was located. After coding and consolidating some of the categories, the following seven damage locations were defined:

1. Doors: including aircraft door, food service doors, cargo door, avionics door and main gear doors
2. Engines: external damage including both left and right engine cowlings and thrust reverse fairings
3. Fuselage: including left, right and aft
4. Nose: including nose cone and nose gear
5. Tail: including elevator, horizontal stabilizer, rudder, tail cone, APU
6. Wings: including left and right wings, ailerons and flaps
7. Unknown: location of damage not stated

While damage location was pretty reliably reported, type of damage (e.g., composite, metallic) could not be inferred from the damage location, nor was it reported consistently in the narratives. Therefore, Environmental Factors, Aircraft State, Ramp Vehicle/Equipment Type and Aircraft Damage Location comprised our initial factors of interest. Numbers of subcategories were somewhat constrained by the total number of reports.

Results and Discussion

Analysis of Factors of Interest

Environmental Conditions. The breakdown of 'Time of Day' for the 104 reports resulted in: Daylight (52, 50%), Night (24, 23%), Dawn (4, 4%), Dusk (4, 4%) and Unknown (20, 19%). The unknown category means that damage was discovered but the time of day was not reported. From the whole set of 104 reports there were only 15 cases that cited environmental weather factors (snow, rain, fog, and ice) as a possible contributing factor. While the greater percentage of events occurred during daylight (50%) compared to non-daylight conditions (31%), this factor could easily be confounded by volume of activity in daylight versus non-daylight hours. In addition, reporters could have interpreted Time of Day as the time when they observed the damage event versus when they detected the damage versus when they filed the report. Thus, we did not try to analyze this factor any further.

Aircraft State. As shown in Figure 2, the 104 reports were also broken down by Aircraft State including the five categories described earlier. The unknown category means it was difficult to determine the aircraft state from the report. A simple comparison of Moving versus Not Moving yielded almost equal numbers of Moving (49, 47%) and Not Moving (48, 46%).

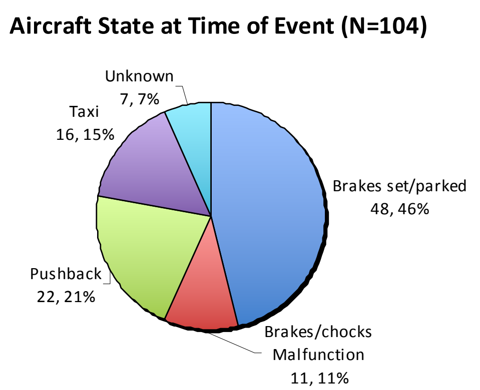


Figure 2. 104 ASRS Reports broken down by Aircraft State.

It should be noted that when an aircraft is parked with brakes set, it might seem that the cause of damage can be attributed to the ramp worker using or driving the service equipment. However, 11% of the reports describe complex situations in which flight and ground crew experienced an unexpected malfunction or error that resulted in an aircraft impact. During pushback and taxi is no simpler in terms of damage cause or initiator. For example, an impact of aircraft and tug can be due to a flight or ground crew problem, or both, and does not automatically imply one or the other is at fault. And more than one instance was cited in which a moving aircraft was "hit by a cart or object" because of a jet blast from a nearby taxiing aircraft. In short, the breakdowns identify factors to consider because they occur in actual operations, but a simple analysis of "what hit what" or "what was moving versus what was not" does not tell the whole story of how and why events occur.

Ramp Vehicle/Equipment Type. Figure 3 depicts the breakdown of 104 ASRS reports by type of ramp vehicle and equipment involved. While the large variety of types is not surprising given the ramp environment and the number of activities that must be accomplished for each flight, it immediately leads to the question of how to narrow the research focus to the most critical types of damage. In this analysis, damage can occur in many ways and we need to know which events warrant more detailed investigation.

Ramp Vehicle/Equipment Type (N=104)

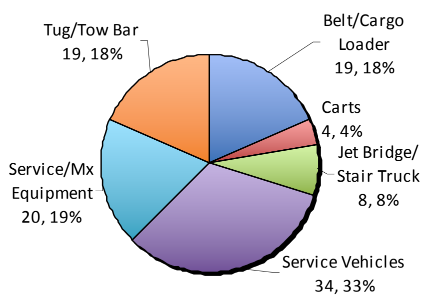


Figure 3. 104 ASRS reports broken down by Ramp Vehicle/Equipment Type.

In addition to the breakdown in Figure 3, we also considered whether Ramp Vehicle/Equipment Type was related to Aircraft State. For example, in the 49 reports where the Aircraft State was Moving, the most frequent impacts were with Service Vehicles and Tugs/Tow Bars (62%). In the 48 reports where the aircraft was Not Moving (parked), most impacts involved Belt/Cargo Loaders, Service Vehicles and Service/Mx Equipment (78%). However when considering each Ramp Vehicle/Equipment Type the following relationships emerged:

- Belt/Cargo Loaders more often involved Not Moving Aircraft (76%)
- Service/Mx Equipment more often involved Not Moving Aircraft (72%)
- Tug/Tow Bars more often involved Moving Aircraft (63%)

While Service Vehicles make up a sizeable proportion of the aircraft damage events (33%) it doesn't seem to matter if the aircraft is moving or not.

Aircraft Damage Location. Figure 4 shows the general breakdown of the 104 ASRS reports by Aircraft Damage Location. The only standout in terms of overall percentages is the 39% of incidents that result in wing damage. Given the large wing area that is exposed to ramp vehicles and objects, it is probably not surprising. Although the fuselage also covers a large expanse, targeted areas such as doors and access panels possibly constrain the hit area somewhat.

Aircraft Damage Location (N=104)

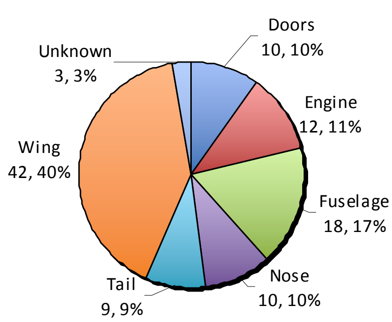


Figure 4. 104 ASRS Reports broken down by Aircraft Damage Location.

In addition to the general breakdown, we also considered whether relationships could be found between Aircraft Damage Location and Ramp Vehicle/Equipment Types. While the numbers are too small to indicate more than tendencies, the following relationships emerged when looking at each Ramp Vehicle Type:

- Service Vehicles more often impacted Wings and Fuselage (71%)
- Service/Mx Equipment more often impacted Wings and Tails (80%)

Considering each Aircraft Damage Location:

- Fuselage damage more often involved Belt/Cargo Loaders and Service Vehicles (56%)
- Nose damage more often involved Tugs/Tow Bars (80%)
- Tail damage more often involved Service/Mx Equipment (56%)
- Wing damage more often involved Service Vehicles and Service/Mx Equipment (71%)

While we can easily imagine the scenarios in which the relationships above could take place, it is important to note that very few relationships are one-to-one. Aircraft damage can be caused in multiple ways so it is important to consider this when developing awareness training and other mitigation strategies.

Conclusion

The objective of this study was to gain an understanding of what types of ramp damage scenarios occur today and how they unfold. The results are not meant to directly generalize to future advanced composite aircraft since damage risk depends on aircraft configuration and specific damage consequences. Still, the analysis of 104 ramp incidents clearly showed that ramp damage occurs in a wide variety of ways involving many different ramp vehicles and service equipment, during all times of the day and affecting nearly every part of the aircraft. Damage events can occur when the aircraft is parked and when it is moving during pushback and taxi. While some relationships among factors emerged, none were strictly one-to-one. Even when ramp activities suggest certain groupings of factors (such as a tug in the nose area during pushback, or belt loader close to the fuselage while the aircraft is parked) these relationships were only somewhat validated by the data. Partly this is because anomalies, malfunctions and errors occur thus increasing the chance of damage at unexpected times. In addition, damage events can be the outcome of a chain of events rather than a single cause. Finally, there were a number of cases where damage was reported but the actual impact event was not reported. When damage is visibly obvious, this may not pose a major problem, but non-reporting of impact events with advanced composite aircraft may have serious consequences. As Hall (2008) states, "Composite airframe structures may not visibly show damage as readily as traditional metallic structures. . . Awareness and reporting of significant impact events is essential". The data from these incident reports further underscore the need to support awareness training and reporting for all personnel who work in the ramp and gate areas; flight crews, maintenance, inspection, drivers of ramp vehicles and others.

References

Aviation Safety Reporting System (ASRS). (2007). *Program Briefing.* on website at http://asrs.arc.nasa.gov/overview/summary.html

Blohm, H. (2007). *Lufthansa Perspectives on Safe Composite Maintenance Practices*. Presented at the FAA Damage Tolerance and Maintenance Workshop, May 9th – 11th, 2007, Amsterdam.

Boeing Commercial Airplanes (2007). *787 Airplane Characteristics for Airport Planning*. D6-58333, Rev A. Preliminary Information, September, 2007. http://www.boeing.com/commercial/airports/787.html

Hall, H. (2008). *Awareness and Reporting of Significant Impact Incidents involving Aircraft Structure*. Presented at the Air Transport Association 51st Annual NDT Forum, September 23-25, 2008, Seattle, WA.

Huang, C.K. & Lin, K.Y. (2005). *A Method for Reliability Assessment of Aircraft Structures Subject to Accidental Damage*. Presented at the 46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference, April 18-21, 2005, AIAA-2009-1830, Austin, TX.

Kim, H. (2008). *Impact Damage Formation on Composite Aircraft Structures*. Presented at the FAA Joint Advanced Materials and Structures (JAMS) Center of Excellence 4th Annual Technical Review Meeting, June 17-18, 2008, Mukilteo, WA.

A METHODOLOGY AND TOOLS FOR THE PROSPECTIVE IDENTIFICATION OF NEXTGEN HUMAN FACTORS ISSUES

Ken Funk
Oregon State University
Corvallis, Oregon, USA

The Human-Machine Systems Engineering Methodology (HMSEM) is a systematic method to prospectively identify relevant human fallibilities, potential errors, and general human factors issues in a complex, high-risk system, then develop design recommendations for remediations to counteract the fallibilities, avoid or mitigate the errors, and resolve the issues. HMSEM uses IDEF0 functional modeling, task analysis, human fallibilities analysis, and Failure Modes and Effects Analysis, organizing the information for and from the analyses in a workbook. The results of its application to several tasks on the NextGen flight deck suggest that it can be a valuable complement to other means to anticipate and resolve human factors issues in NextGen development.

The problem of human performance in complex, high risk systems was described concisely, accurately, and usefully by Wiener in the phrase, "fallible humans and vulnerable systems" (Wiener, 1987), and the Next Generation air transportation system (NextGen) threatens to be a system highly vulnerable to the errors of its fallible human operators. From the documentation available at this time (e.g., JPDO, 2007), NextGen appears to be a technology-driven system, not a human-centered system, and we know from past experience that technology-driven systems can be particularly vulnerable to human error. Already, some NextGen human factors issues have been identified (e.g., Sheridan et al, 2006; Funk et al, 2009), but much remains to be done. The aviation human factors/psychology community can make a valuable contribution to the development and implementation of NextGen through the thorough and systematic identification of human factors issues. Those issues must be identified, organized, and presented in such a way as to be understandable by and useful to NextGen system architects and engineers.

Objectives

The objectives of this project were to develop a systematic, analytical methodology to prospectively identify human factors issues and recommend remediations, then apply the methodology to the NextGen flight deck.

Human-Machine Systems Engineering Methodology, Tools, and Application to NextGen

The result of the development, the Human-Machine Systems Engineering Methodology (HMSEM), is a formal, systematic methodology to identify important human fallibilities relevant to a system, identify specific errors likely to arise from the interactions of those fallibilities with characteristics of the system, identify general human factors issues arising from the potential errors, identify remediations, and organize the findings in a way useful to analysts, Subject Matter Experts (SMEs), system architects, and system engineers. HMSEM analysts, supported by SMEs, work through the following stages: 1) formal functional modeling using IDEF0, 2) task analysis, 3) human fallibilities identification, 4) Failure Modes and Effects Analysis, 5) issue identification, and 6) requirements development. HMSEM was applied, in a test case, to the NextGen flight deck, and HMSEM and the application are described and discussed in the remainder of this paper.

IDEF0 Modeling

Many human factors methodologies begin with some form of hierarchical task analysis (HTA), but HMSEM requires a richer and more detailed representation of system processes (activities, functions, tasks) than HTA typically provides. This requirement is met by modeling the system with IDEF0, a graphical language for modeling system functions. The Oregon NextGen Flight Deck Functional Model (ONFDFM) is an IDEF0 model of a generic NextGen commercial flight deck based on NextGen literature available at this time (e.g., JPDO, 2007) and knowledge of present-day commercial flight deck operations. Figure 1 shows ONFDFM's top-level diagram, its most general representation of flight deck functions.

In IDEF0, a function is a process, performed by mechanisms (humans, devices), that transforms inputs (matter, energy, information, system states) to outputs (matter, energy, information, system states), subject to controls (information, factors) that guide, facilitate, or constrain the process. IDEF0 uses boxes labeled with verb phrases to represent functions and arrows labeled with noun phrases to represent mechanisms, inputs, outputs, and controls. So, omitting some details, Figure 1 represents that the human flight crew [h FC] and flight deck systems (devices) [d FD systems] perform flight deck tasks [Perform flight deck tasks] to transform the aircraft system [s Acft] to a managed and controlled aircraft system [s Acft, managed & controlled]. The performance of flight deck tasks is guided (controlled) by information in flight deck procedures [i FD procedures] and Federal Aviation Regulations [i FARs] and influenced (controlled) by performance shaping factors [f Performance shaping factors], like the aircraft's performance limitations and the flight crew's decision biases. To perform flight deck tasks also transforms the flight crew's mental model [i FC MM] to an updated mental model [i FC MM, updated], utilizes NextGen systems [s NG systems] and the Air Navigation Service Provider [h ANSP], and is controlled by information received from the NextGen system [i NG info] and the ANSP [i Comm from ANSP].

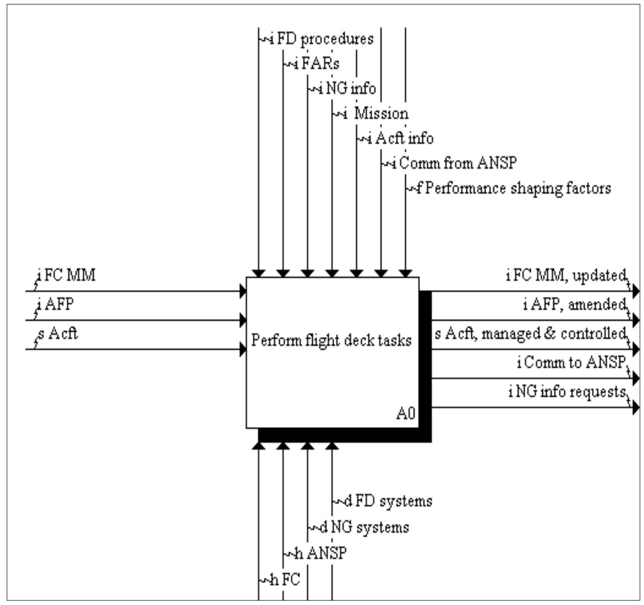


Figure 1. Top-level IDEF0 diagram of the Oregon NextGen Flight Deck Functional Model.

In IDEF0, general functions are detailed or decomposed into more specific functions, those functions are further detailed, and the modeling process continues until a representation sufficiently detailed for further analysis is produced. For example, in the ONFDFM, the function [Perform flight deck tasks] is detailed into [Collaboratively manage FP (flight plan)], [Manage 4DT (4-dimensional trajectory)], [Manage acft (aircraft) systems], and [Control acft]. Those are in turn detailed, and so on. Table 1 shows a portion of the function hierarchy of the ONFDFM, elaborating part of the [Manage 4DT] branch. A-numbers (A#s) define a function's place in the hierarchy ("A" for "Activity" being inherited from IDEF0's precursor, SADT). The hierarchy is, effectively, the task hierarchy resulting from a typical HTA, but the detailed IDEF0 diagrams underlying the hierarchy bear much more information than does the typical HTA. As shown in Table 1, the detailing of [Manage 4DT] ultimately yields [Get traffic info using HSI/CDTI (Horizontal Situation Indicator/Cockpit Display of Traffic Information)], part of whose IDEF0 diagram is shown in Figure 2.

Table 1. A portion of the ONFDFM function hierarchy, elaborating the [Manage 4DT] branch.

A# Function
A0: Perform flight deck tasks
A1: Collaboratively manage FP
A2: Manage 4DT
A21: Receive ANSP clearances
A22: Assess 4DT WRT AFP & clearances
A23: Assess 4DT WRT terrain
A24: Assess 4DT WRT obstacles
A25: Assess 4DT WRT traffic
A251: Get traffic info from ANSP advisories
A252: Get traffic info from FD alerts
A253: Get traffic info using HSI/CDTI
A2531: Configure HSI to display traffic
A2532: Locate traffic symbols on CDTI
A2533: Select traffic for detailed info
A2534: Determine traffic IDs, bearings, ..., from CDTI
A2535: Estimate traffic trajectories from CDTI info
A254: Get traffic info visually
A255: Integrate traffic info
A256: Assess integrated traffic picture
A26: Adjust 4DT
A3: Manage acft systems
A4: Control acft

ONFDFM was developed using KBSI Inc.'s AI0Win IDEF0 modeling software. An HTML version of the full model, generated by AI0Win, is accessible at http://flightdeck.ie.orst.edu/NextGen/Models/ONFDFM1.0/.

IDEF0 diagrams and the glossary of model elements underlying them provide a very rich representation of the functions performed in and by a complex system. An important benefit over HTA is that IDEF0 explicitly models not only functions (or tasks), but relationships among functions via mechanisms, inputs, outputs, and controls. Those relationships can be identified in the IDEF0 model by examining related diagrams and tracing

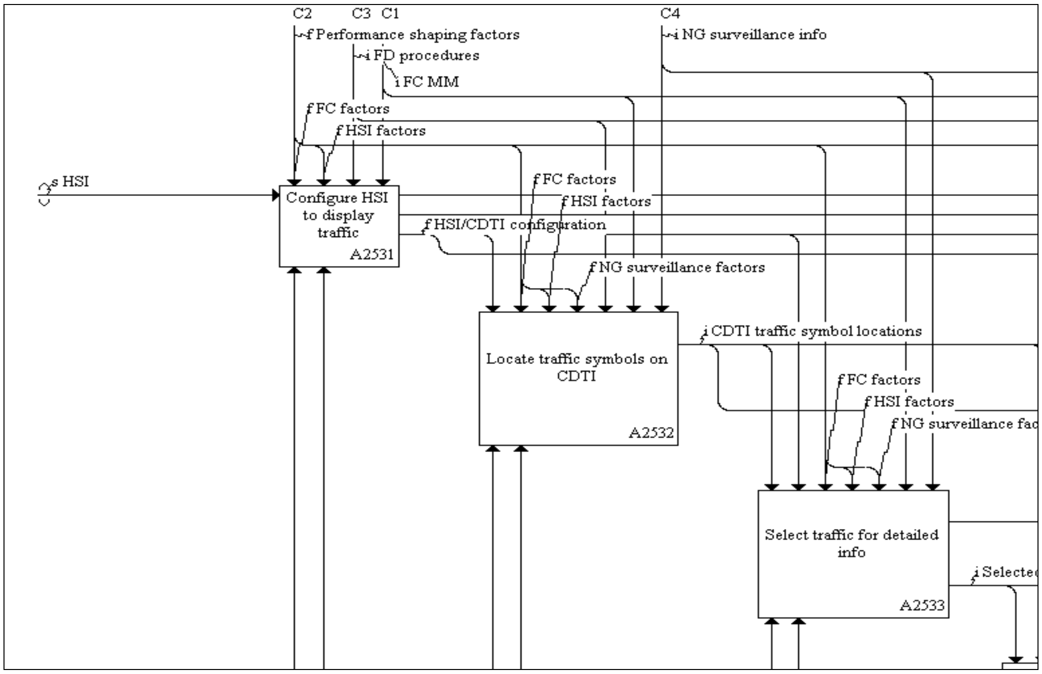


Figure 2: Detail from the IDEF0 diagram of the [Get traffic info using HSI/CDTI] function from the ONFDFM.

arrows. However, a complex IDEF0 model may have many diagrams, and navigating them to identify relationships, although in principle straightforward, is in practice difficult and prone to error. As in any reductionist method, it is tempting for the analysts to focus on a small part of the IDEF0 model and ignore its context, thus to "lose the big picture" or "miss the forest for the trees". HMSEM uses the prototype IDEF0 Navigator (INav) to avoid that. INav operates on an IDEF0 model providing an alternative representation to the IDEF0 diagrams. An arrow entering an IDEF0 diagram can come from another part of the model outside the immediate diagram or from outside the system itself. The INav representation abstracts out some of the details of the IDEF0 diagrams to show from where each arrow (or each group of related arrows) comes or where it goes, allowing the analyst to explore details in the context of the entire model in a single view.

Task Analysis

In HMSEM, task analysis is used to further analyze the most detailed IDEF0 functions – referred to as tasks – to compile, from the model and elsewhere, information needed for human fallibilities identification. The analysts enter, for example, task location and timing information into the HMSEM workbook. Table 2 shows the results of task analysis of [A2534: Determine traffic IDs, bearings, ranges, & relative altitudes from CDTI].

Table 2. Results of the task analysis of [A2534: Determine traffic IDs, bearings, ranges, & relative altitudes from CDTI].

Task Analysis Attribute	Value
Purpose / Value Added	Necessary to detect conflicts and determine if separation and spacing is appropriate.
Location	Flight deck
Frequency & Timing	Continuous, intermittent
Environmental Conditions	Darkness (red illum) to direct sunlight, glare, etc.; Noise; Vibration (low, high freq.)
Information Requirements	i Selected traffic; i CDTI traffic symbol locations; f HSI/CDTI configuration; i FD procedures; i FC MM; i NG surveillance info
Sensory/Cognitive/Motor Actions	View CDTI; Identify traffic; Estimate bearings, ranges, relative altitudes, and probable trajectories

Human Fallibilities Identification

Human factors analysis sometimes employs a Human Error Identification (HEI) technique like SHERPA (Embrey, 1986) to identify errors that could occur in a system. HEI techniques typically start with a functional representation of the system (often from HTA) and analysts and SMEs, referring to that representation, use their knowledge and experience to hypothesize potential errors that could occur in specific tasks. HEI techniques rely heavily on analyst and SME memory and judgment (and, one could say, serendipity) to compile a comprehensive list of likely errors and are, therefore, subject to the same kinds of limitations that affect human performance in systems like the one they are studying. Rather than to attempt to identify errors directly, HMSEM first identifies the human fallibilities likely to be significant in each task and, from system and task information from the IDEF0 model and task analysis, proceeds to project errors that could occur as a result of those fallibilities interacting with system and task characteristics.

HMSEM uses the Human Fallibilities Identification and Remediation Database (HFIRDB) for fallibilities identification. The HFIRDB is a database consisting of human fallibilities and remediations for them compiled from Wickens' and Hollands' *Engineering Psychology and Human Performance* (Wickens & Hollands, 2000). The user interface leads the analysts through a series of questions about each task to be analyzed for fallibilities and errors and the analysts refer to the IDEF0 model and the task analysis to answer them. The HFIRDB first asks the analysts to select from among seven information processing stages (i.e., sensory registration, perception, attention allocation, working memory, long-term memory, decision-making, and response control) those employed in the task under consideration. Next the analysts are asked to choose general human fallibility categories (e.g., visual display processing or working memory limitations) that apply to the selected information processing stages. Then HFIRDB asks the analysts to choose from a list of possibilities just those conditions that exist in the task under consideration. For example, that operators must appropriately allocate attention to concurrently process or selectively attend to visual stimuli presented in displays is a condition necessary for visual display processing fallibilities to be relevant. The HFIRDB uses a series of queries to produce a list of human fallibilities that may manifest themselves in performance of the task, such as the sensitivity-related vigilance decrement, the tendency for operator performance to degrade during vigilance tasks as a result of a decrease in sensitivity level. The HFIRDB then asks the analysts to confirm task conditions that enable manifestation of the fallibilities and a complete list of relevant fallibilities is generated, which may be copied into the HMSEM workbook for the next analysis stage.

Failure Modes and Effects Analysis

Failure Modes and Effects Analysis (FMEA) is an analytic technique used to prospectively identify the ways in which a system can fail. FMEA begins with a process or functional description of the system to be analyzed. For each function, the analysts use knowledge of the function to identify failure modes, that is, ways in which it could fail to achieve its intended outcome. For each failure mode, the analysts identify the causes of or contributing factors to the failure mode, and try to predict its consequences. To prioritize the failure modes for further study or remediation, the analysts assign numeric ratings as to the severity of the consequences of the failure mode, the probability or expected frequency of its occurrence, and the likelihood that it would not be detected in time to avoid the consequences. These three ratings are multiplied to give a Risk Priority Number (RPN) for each failure mode and the RPNs are used to prioritize the failure modes for further analysis or remediation.

In HMSEM, FMEA is used to identify potential operator errors as failure modes. The analysts use the IDEF0 model, operator fallibilities identified with the help of the HFIRDB, and general domain and human factors knowledge to identify specific failure modes – i.e., operator errors – that could occur in performing the task as a result of the interaction of system and task characteristics with those fallibilities. These are entered into the HMSEM workbook. Table 3 presents some results from FMEA applied to the task [A2534: Determine traffic IDs, bearings, ranges, & relative altitudes from CDTI].

Issue Identification

To identify issues, the HMSEM analysts collect similar failure modes and those related by common fallibilities and task characteristics. For each such collection, the analysts compose a statement which, if it is or should become true in the implementation and operation of the system, describes a condition or situation related to system operations where natural human characteristics, capabilities, limitations, and tendencies are very likely to lead to significant problems with system effectiveness, efficiency, or safety. These issues are added to the HMSEM workbook. Table 4 presents some NextGen flight deck failure modes and general issues arising from them.

Requirements Development

Perhaps hundreds of human factors issues related to the NextGen flight deck may be identified in this and

Table 3. Excerpts from the Failure Modes and Effects Analysis of the task, [A2534: Determine traffic IDs, bearings, ranges, & relative altitudes from CDTI].

Human_Fallibility	Other Contributing Factor(s)	Potential Failure Mode	Potential Effects of Failure Mode	Severity	Probability	Nondetect.	RPN
Perceptual competition	High symbol density on HSI/CDTI	MM error: FC confuses two CDTI traffic symbols, mis-estimates bearing/range/altitude/trajectory of one or both.	Inaccurate perception and projection of traffic bearing/range/altitude/trajectory, loss of separation/spacing.	5	4	5	100
Negative skill transfer	CDTI display format, symbology differ from those of similar equipment.	MM error: FC misinterprets CDTI traffic info, mis-estimates bearing/range/altitude/trajectory.	Inaccurate perception and projection of traffic bearing/range/altitude/trajectory, loss of separation/spacing.	5	4	4	80
Strategic task-management bias	Other high-priority, concurrent tasks/stimuli.	TM error: FC fixates on CDTI, fails to perform other high-priority tasks.	Other tasks ignored or performed poorly.	4	5	4	80

Table 4. Some general issues identified by analysis of the NextGen flight deck.

Related Failure Modes	In Task(s)	Resulting General Issue
Miss: FC misses traffic on CDTI.	A2532	The flight crew's CDTI traffic detection performance decreases over long periods of self-separation authority.
Delay: CDTI scan is prolonged. Miss: FC fixates on one region of CDTI, misses other traffic.	A2532	The effectiveness and efficiency of the flight crew's CDTI traffic scan is very susceptible to stress and other performance-shaping factors and performance can suffer as a result.
mistake: FC chooses and sets HSI/CDTI to inappropriate config.	A2531	Complex device configuration procedures induce pilots to select suboptimal configurations, leading to diminished performance when the devices are used.
slip: FC sets HSI/CDTI to unintended config. lapse: FC omits step to properly configure HSI/CDTI. MM error: FC misinterprets CDTI traffic info, mis-estimates bearing/range/altitude/trajectory.	A2531, A2534	Attempting to perform two or more tasks that require the same mental resources concurrently causes the performance of at least one of them to be diminished,

other ways, but unless guidance is given to avert the potential effectiveness, efficiency, and safety problems they raise, merely citing them is of little value. Here is an opportunity for aviation human factors scientists and practitioners to go the next step toward solution. In addition to human fallibilities information, the HFIRDB contains general guidance information for remediations to reduce the likelihood that human fallibilities will interact with system and task characteristics to manifest themselves as errors. With fallibility, failure mode, error, and issue information from the HMSEM workbook, the analysts may turn again to the HFIRDB to retrieve countermeasures it suggests to counteract the fallibilities. Table 5 presents some suggested requirements for the NextGen flight deck. Following requirements engineering convention, terms and phrases enclosed in asterisks (* ... *) are, for the time being, ambiguous and unverifiable. Further analysis, and possibly research, will be required to refine them.

Discussion

HMSEM is prospective, systematic, and is based on validated human factors knowledge. Moreover, its use of a rich functional modeling formalism provides a framework to organize human fallibilities, potential errors, human factors issues, and recommendations or requirements in a way compatible with the functional models used by system architects and engineers. It thus offers a natural way for human factors scientists and engineers to collaborate with system designers in the critical early stages of system development. But HMSEM has important limitations. In its present form, it is a time-consuming process. Most HMSEM tools are presently in the prototype stage. Despite its

Table 5. Some preliminary NextGen flight deck requirements to address issues identified in HMSEM analysis. Asterisks (* ... *) denote as-yet unverifiable terms.

A#	Requirement	Type
A0	NextGen flight crews shall receive concurrent task management training, including *topics TBD*.	Training
A253	CDTI traffic symbol visual coding, for whatever purpose, shall *manifest* exactly three levels of salience corresponding to the three levels of traffic priority: low for the symbols of normal priority traffic, medium for symbols of intermediate priority traffic, and high for symbols of high priority traffic.	Equipment
A2532	CDTI procedures shall *recommend or specify* a *systematic* display scan pattern that covers the entire display each cycle and which cycle is completed in no more than *C* seconds.	Procedures

attempt to be systematic, its application is still subject to analyst biases and analyst knowledge and cognitive limitations. Its application to NextGen, described in this paper, is limited in scope to a few tasks related to CDTI-based traffic awareness. The functional model itself is limited in scope and based on as-yet very limited documentation on the envisioned NextGen flight deck.

Recommendations

Therefore, the knowledge base of the HFIRDB should be expanded to address more dimensions of human performance and the HMSEM workbook should be converted to a more robust software tool that integrates the other tools, provides a repository for findings, and generates publishable reports. A team of human factors analysts, SMEs, and engineers should be assembled to continue applying HMSEM to NextGen. They should refine and expand the ONFDFM to incorporate the most recent plans for NextGen implementation, modeling, in detail, the full scope of flight deck functions. They should use the model and refined tools to identify human fallibilities, potential errors, and human factors issues, and make recommendations for engineering requirements to guide NextGen system design. Throughout this process, the team should work with NextGen system architects and engineers to make the ONFDFM consistent with functional models used for NextGen development, to utilize the latest NextGen plans in their analyses, and to organize and present their findings in a way compatible with NextGen design documents. In this way, human factors analysis and recommendations will be more likely to have greater impact on NextGen.

Acknowledgments

The author is grateful to Oregon State University for a term of sabbatical to finish the development of HMSEM and apply it to the NextGen flight deck. Melissa Hastings and Leslie Thompson developed INav and HFIRDB, respectively, for their industrial engineering MS thesis research projects.

References

Embrey, D.E., (1986). SHERPA: a systematic human error reduction and prediction approach, In *Proceedings of the International Topical Meeting on Advances in Human Factors in Nuclear Power Systems*. La Grange Park: American Nuclear Society.

Funk, K., R. Mauro, & I. Barshi (2009). NextGen flight deck human factors, to be published in *Proceedings of the 15th International Symposium on Aviation Psychology*, Dayton, Ohio, 27 – 30 April 2009.

Joint Planning and Development Office [JPDO]. (2007). *Concept of Operations for the Next Generation Air Transportation System, Version 2.0*. Washington, DC: author.

Sheridan, T.B., K.M. Corker, & E.D. Nadler (2006b). *Final Report and Recommendations on Human-Automation Interaction in the Next Generation Air Transportation System*. Cambridge, MA: US Department of Transportation, John A. Volpe National Transportation Systems Center.

Wickens, C. D. and Hollands, J. G. (2000). *Engineering Psychology and Human Performance*, Third Edition. New Jersey: Prentice Hall.

Wiener, E.L. (1987). Fallible humans and vulnerable systems: lessons learned from aviation, in J.A.Wise and A. Debons (Eds.), *Information Systems: Failure Analysis*, NATO ASI Series, Vol. F32, Heidelberg: Springer-Verlag, 165-181.

AVIATOR 2030 - ABILITY REQUIREMENTS IN FUTURE ATM SYSTEMS

Hinnerk Eißfeldt
Deutsches Zentrum für Luft- und Raumfahrt DLR
(German Aerospace Center)
Institute of aviation and space medicine
Department of aviation and space psychology
Hamburg, Germany

‘Aviator 2030’ is a project at DLR on ability requirements for operators in future ATM systems. Several workshops have been conducted with pilots and air traffic controllers to learn how today’s aviation professionals see their jobs develop in future. Using separated workshops first, pilots and air traffic controllers were introduced to current developments within the context of Single European Sky SES, a large-scale program comparable to NextGen in the United States. Following the ‘future-workshop’ concept participants developed scenarios of future ATM from their professional background and experience. In a third workshop pilots and controllers met to exchange and discuss their concepts. Together they developed a shared view of future ATM systems, using role-plays and other forms of presentation. They also used the Fleishman Job Analysis Survey F-JAS in a special version to express their view on future ability requirements.

Improvements in air traffic management (ATM) and aircraft systems as well as organisational structures have become one of the key challenges of aviation in 21st century. This is especially important with regard to the considerable increase in air traffic. To allow maximum capacity and safety as well as minimum impact on environment and cost, Single European Sky (SES) will be implemented to coordinate the traffic in Europe.

The key question of the project ‘Aviator 2030’ deals with changes that will concern pilots and air traffic controllers introducing SES. Which modifications of operators’ tasks, roles and responsibilities can be expected? Will pilot or air traffic control trainees selected today ever work in the ATM system reflected in the current job analysis? If not, what ability requirements will change, what will remain?

Aviator 2030

Based on domain experts’ point of view, Aviator 2030 develops future scenarios of ATM. Key aspects of these scenarios are tested with human operators in low-fidelity simulators which combine on-board and ATC systems. Thus, potential changes in ability requirements for pilots and air traffic controllers will be identified prospectively and allow for timely adjustment of selection profiles (figure 1).

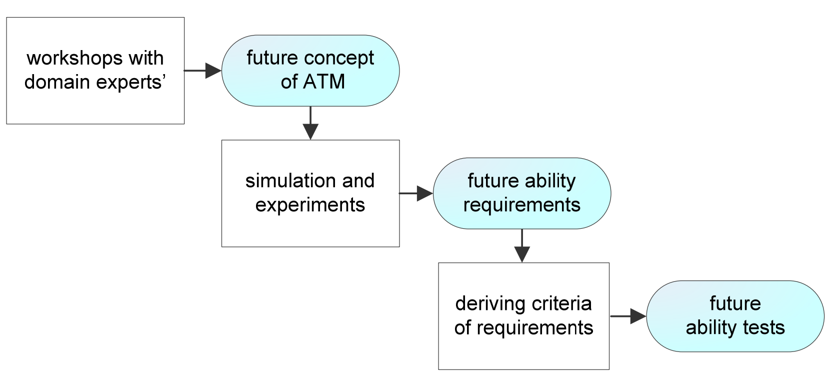


Figure 1. Flowchart of the project Aviator 2030

Workshops with experienced air traffic controllers and pilots have been conducted separately to obtain job incumbents expectations regarding their future tasks, roles and responsibilities. The first two-day workshop was conducted with nine air traffic controllers of the Deutsche Flugsicherung GmbH (DFS GmbH) and the second involved ten pilots of Deutsche Lufthansa AG (DLH AG). Both workshops were designed correspondently using the *future workshop* concept. This technique developed by Jungk in the 1970s (Jungk & Muellert, 1987) enables a group of people to generate new ideas or solutions of mainly social or organizational issues. It has been used for the first time in a technical aviation context with good success.

Each future workshop started with an information session: Participants were informed about the general idea of the project, the goals of the 'Vision 2020' for European aeronautics and the Concept of Operations for the Single European Sky (SESAR CONOPS, Sesar 2007).

Participants and controllers were then asked for their criticisms about 'Vision 2020' and SESAR CONOPS. Both ATC and pilots emphasised the risk of single workplace replacing teamwork, shift of competencies or incapacitation and inappropriate system design. Upon collecting risk about future aviation, participants were asked for their ideas about future aviation. Visionary scenarios dealt with the process of negotiation of 4D-trajectory, tactical planning and operating of flights, improvements of human resource planning, first draft of a virtual workspace and a new approach to line and recurrent training. All scenarios consist of innovative approaches for handling possibilities and changes in the future. Finally, participants checked their scenarios with regard to further steps, workplace design and potential obstacles.

About four months later an *integrative workshop* with the same pilots and air traffic controllers was conducted to exchange the ideas and concepts. First, results of the future workshops were presented and discussed. Controller and pilots enjoyed sharing their future scenarios. Second, mixed groups consisting of controller and pilots elaborated several ideas: a concept of trajectory negotiation, procedures for operating flights in the future and an integrated training system for pilots and air traffic controllers. In general, participants developed future scenarios including ATC' and pilots' perspectives. Finally, participants derived future scenarios which should according to their background be simulated and tested in the ongoing project. A detailed description of the layout and the outcome of the workshops is provided by Bruder, Jörn & Eißfeldt (2008).

To receive a first impression on potential changes in ability requirements in a more standardised way, participants of the workshops were finally asked to rate the ability requirements for the future ATM system. To do so participants teamed up in pairs with always one of each background to enable a mutual understanding of scales to be rated and to support the exchange of views. Each participant then gave his rating for his professional role in the light of his understanding of the future ATM system.

Method

The Fleishman Job Analysis Survey F-JAS (F-JAS; Fleishman 1992) was used to depict ability requirements for the future ATM system. With the F-JAS job incumbents are asked to use a 1 to 7 scale to "rate the task on the level of the ability required, not the difficulty, time spent or importance of the ability" (Fleishman 1992b, p.7). The F-JAS has been used at DLR in a number of studies with good success, for instance in a simulator study at the DFS Research & Development Centre on the effects of ATM systems comprising datalink (Eißfeldt, Deuchert & Bierwagen, 1999).

The F-JAS Fleishman Job Analysis Survey (Fleishman 1992) is a survey measuring human abilities, providing detailed definitions and anchored rating scales for 72 scales covering the domains of cognitive, psychomotor, physical and sensory abilities as well as interactive/social and knowledge/skills scales, the latter being still under research. It comes together with a detailed 'Administrators Guide' (Fleishman & Reilly 1992a) and the 'Handbook of Human Abilities' (Fleishman & Reilly 1992b) providing some theoretical background and lists of validated tests measuring a certain abilities including reference data of test providers. In 1995 the F-JAS was republished with 52 scales covering cognitive, psychomotor, physical and sensory/perceptual abilities. In 1996 the F-JAS Kit Part 2 was published covering 21 social/interpersonal abilities (MRI 1996).

With the Aviator 2030 project a special version of the F-JAS was developed including not only the original scale material but anchors representing requirements of current pilots and air traffic controller jobs in addition. These mean ratings reflect the results of prior studies with air traffic controllers of

Deutsche Flugsicherung GmbH (N = 88; Eißfeldt & Heintz, 2002) and pilots of Deutsche Lufthansa AG (N = 141; Goeters, Maschke & Eißfeldt, 2004). In this special F-JAS aviator version the mean rating for air traffic controllers of DFS is depicted in a blue box on the left, the mean rating for pilots of DLH in a yellow box on the right side of the central scale. Figure 2 shows an example for a scale as used in the project Aviator 2030 with integrated anchors for air traffic controllers and pilots.

1. Oral Comprehension	This is the ability to listen and understand spoken words and sentences.

How Oral Comprehension Is Different From Other Abilities		
Oral Comprehension: Involves **listening** to and **understanding** words and sentences spoken by others.	vs.	*Written Comprehension*: Involves **reading** and **understanding written** words and sentences.
		Oral Expression and *Written Expression*: Involve **speaking** or **writing** words and sentences so others will understand.

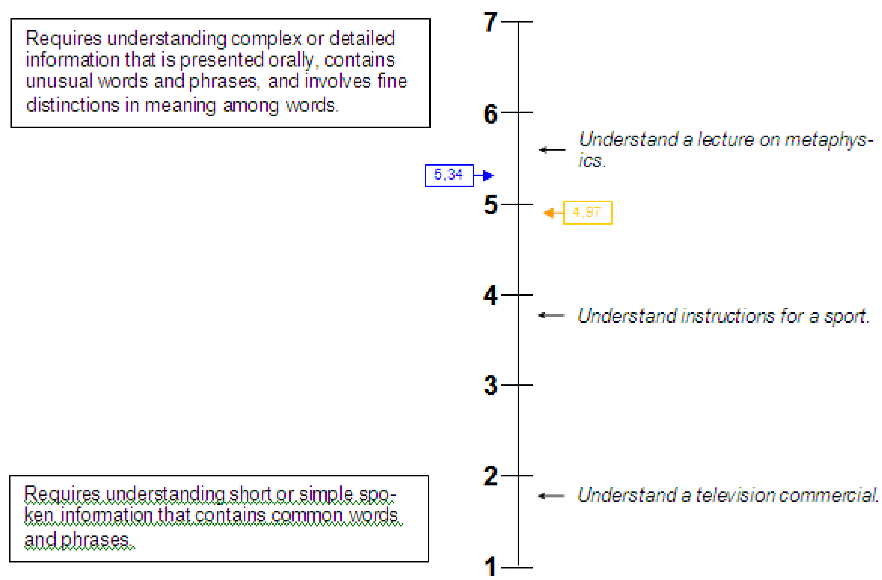


Figure 2. Example scale F-JAS aviator: Oral comprehension with added anchor scales for air traffic controllers and pilots. Adapted from Fleishman (1992), with permission.

To integrate these anchors graphically on the scale better allows interpreting results as increasing or decreasing requirements compared to today. In an earlier study this was achieved by working the F-JAS twice: First to obtain the ratings for the everyday job experience as air traffic controller and second, after days of training and simulation in a new datalink environment to collect the ratings for the new system (Eißfeldt 1999). Due to time constraints this approach was not possible for the Aviator project; however the special experience of this unique group of aviation professionals after 4 days of dealing with issues of future ATM demanded the use of scientific standardized material. After some first trials it was decided to integrate the anchors as numerical values in coloured coding directly on the scales. The F-JAS aviator version proved to be easy to work with, a total of 15 sets of ratings (8 pilots, 7 air traffic controllers) were collected. Although this sample does not reach a size allowing for strong interpretation, the combination with larger existing data sets (141 pilots, 88 air traffic controllers) should enable interpretation of ratings obtained from workshop participants. However, it has to be considered that these results are preliminary.

Results

In the following only the results for the cognitive abilities of the F-JAS aviator will be discussed. As Figure 3 shows many of the scales in the cognitive domain were rated very similar for the future ATM system as for the current job requirements. For air traffic controllers, strong increase was found with 'problem sensitivity' and 'speed of closure'; strong decrease was rated for 'originality', memorization' and 'spatial orientation'. For pilots a strong increase was indicated for 'deductive reasoning' and a strong decrease in 'number facility'. Given that 'Abilities with mean ratings of four or greater are generally considered to be important for the job (Fleishman & Reilly 1992, p.10)' the impression is that the profile of cognitive ability requirements will not change essentially with future ATM concepts for both professions, with some minor adjustments being proposed.

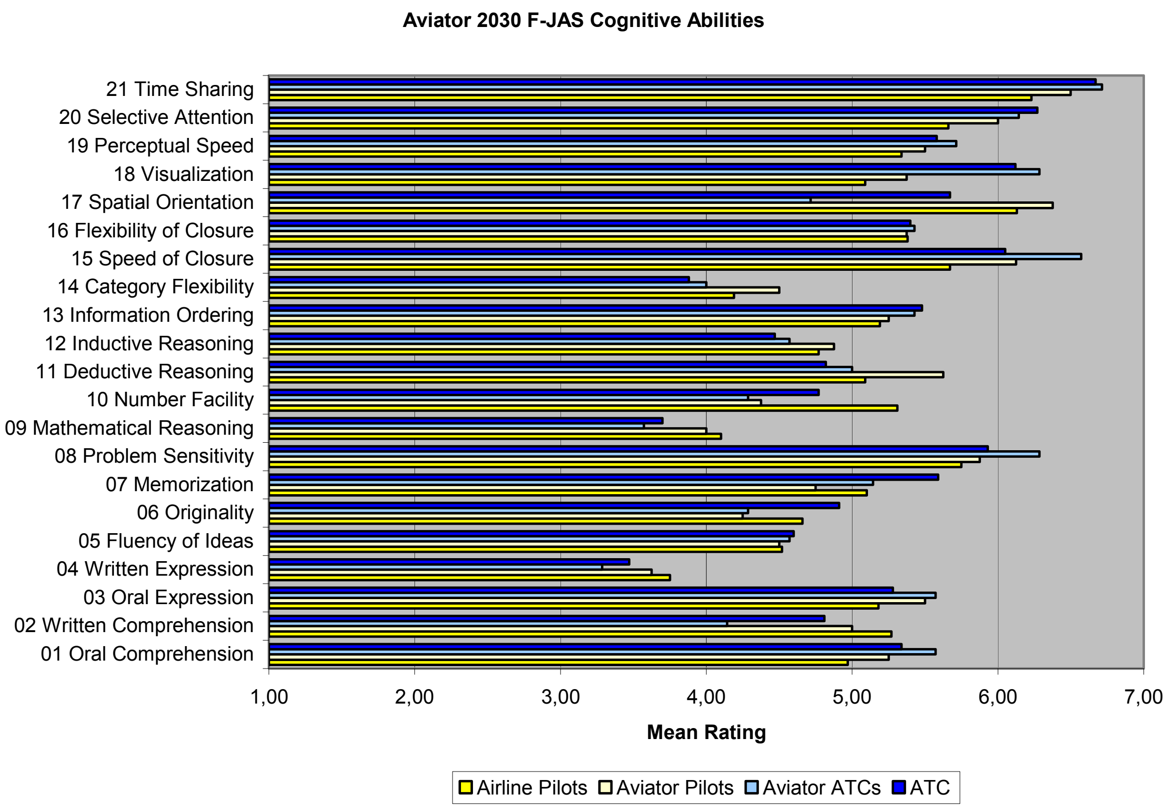


Figure 3. F-JAS Cognitive Abilities for pilots and air traffic controllers in Aviator 2030

A second look concerns the similarity of ratings for pilots and controllers: in the domain of cognitive abilities most of the ratings are not much different for the two groups. Only two of the cognitive scales showed significant differences between pilots and air traffic controllers: 'spatial orientation' and 'visualization'.

Looking at the pattern of results for 'visualization' in both groups there was a slight increase with the future ATM concepts, as was seen with a lot of the cognitive abilities. Also 'oral comprehension', 'oral expression', 'problem sensitivity', 'deductive reasoning', 'inductive reasoning', 'category flexibility', 'speed of closure', 'perceptual speed' and 'time sharing' all showed a slight increase with the future ATM concepts for both professional groups.

With 'spatial orientation' it was different; there was an increase in relevance for the pilots and a decrease for the air traffic controller group. A similar but only slight tendency was found in the ratings for 'selective attention' and 'information ordering'. There was not a single cognitive ability showing the opposite pattern of decrease of relevance with pilots and increase with future ATM concepts for air traffic controllers.

In a third pattern of results the relevance of abilities decreased with the future ATM concepts for both professional groups. 'Written comprehension', 'written expression', 'originality', 'memorization', 'problem sensitivity', 'mathematical reasoning', and 'number facility' all showed decreasing relevance with future ATM concepts as discussed in the Aviator 2030 workshops.

Discussion

To follow up the changes in ability requirements of core aviation professions remains a never ending task for those dealing with aviator selection. Especially the introduction of new automation has to be controlled for effects on tasks, roles and responsibilities, and in consequence on selection profiles (Eißfeldt 1991). However, when cognitive abilities are focussed, there seems neither relief nor much intensification of ability requirements to be stated. What can be foreseen are pilot and air traffic controller profiles assimilating with regard to cognitive abilities mostly linked to the tasks of airborne separation issues.

Acknowledgements

The author wishes to express his thanks to all air traffic controllers and pilots participating in the Aviator 2030 project and to Deutsche Flugsicherung GmbH and Deutsche Lufthansa AG for supporting this research. Special thanks go to Edwin A. Fleishman for commenting an earlier version of this paper.

References

Eißfeldt, H. (1991). Automation in ATC: How does it affect the Selection of Controllers? In J. A. Wise, V. D. Hopkin & M. L. Smith (Eds.), Automation and System Issues in Air Traffic Control (Vol. NATO ASI Series F, Vol F 73, pp. 461-465). Berlin: Springer.

Eißfeldt, H. (1999). Ability requirements for Air Traffic Controllers in future ATM systems. In R. S. Jensen, B. Cox, J. D. Callister & R. Lavis (Eds.), Proceedings of the Tenth International Symposium on Aviation Psychology, Columbus, Ohio USA (Vol. 1, pp. 592-597). Columbus: The Ohio State University.

Eißfeldt, H., Deuchert, I. & Bierwagen, T. (1999). Ability requirements for future ATC systems - a simulation study using research facilities of DFS Deutsche Flugsicherung GmbH. In DLR-Forschungsbericht 99-15. Hamburg: DLR.

Eißfeldt, H. & Heintz, A. (2002). Ability requirements for DFS controllers - current and future. In H. Eißfeldt, M. C. Heil & D. Broach (Eds.), Staffing the ATM system - the selection of air traffic controllers (pp. 13-24). Aldershot: Ashgate.

Fleishman, E. A. (1992). Fleishman Job Analysis Survey (F-JAS) Rating Scale Booklet. Potomac, Maryland: Management Research Institute, Inc.

Fleishman, E. A. & Reilly, M. E. (1992a). Fleishman Job Analysis Survey (F-JAS) Administrator´s Guide. Potomac, Maryland: Management Research Institute, Inc

Fleishman, E. A. & Reilly, M. E. (1992b). Handbook of Human Abilities. Definitions, Measurements, and Job Task Requirements. Potomac, Maryland: Management Research Institute, Inc.

Fleishman Job Analysis Survey F-JAS (Part 2) Rating Scale Booklet (1996). Potomac, Maryland: Management Research Institute, Inc.

Goeters, K.-M., Maschke, P. & Eißfeldt, H. (2004). Ability requirements in core aviation professions: Job analysis of airline pilots and air traffic controllers. In K.- M. Goeters (Ed.), Aviation Psychology: Practice and Research (pp. 99-119). Aldershot: Ashgate.

Jungk, R. and Muellert, N. (1987). Future workshops: How to create desirable futures. London: Institute for Social Inventions.

SESAR: The ATM Target Concept (2007). In SESAR Definition Phase Project (Ed.), DLM-0612-001-02-00a. Brussels: SESAR Consortium.

DETERMINING JOB REQUIREMENTS FOR THE NEXT AVIATOR GENERATION

Hans-Juergen Hoermann
German Aerospace Center (DLR)
Hamburg, Germany

Dirk Schulze-Kissing & Oliver Zierke
German Aerospace Center (DLR)
Hamburg, Germany

The aviation industry is envisioning a tremendous growth of air traffic within the next two decades. New technologies and operational concepts will be the key enablers to accommodate the increasing amount of movements in a safe, efficient and environment friendly manner. Current working concepts reach from improved interoperability of national ATM systems, via satellite based navigation, collaborative decision making, and self separation of aircraft up to fully automated air-ground-space systems. It can be expected that the introduction of such concepts will have a significant impact on the working conditions and job requirements of future air traffic controllers and pilots, who were selected on traditional job profiles reflecting the current and past operational settings. Our paper is presenting elements of a prospective job analysis of future aviators assigned to specific operational tasks within the future air transport system. Results will be based on reviews of available international concept papers, conducted future workshops with present job holders and low fidelity simulation runs of collaborative air traffic control and aircraft separation tasks. Relevant en-route and arrival scenarios will be discussed and presented at the symposium with some preliminary data of the initial tryout studies.

One important basis of a fair and efficient selection system is the adequate identification of job requirements. However, in times of rapidly developing working conditions, job profiles of knowledge, skills, abilities and other characteristics have an increasing short half-life period. The classical selection rationale that job requirements should match a person's individual pattern of stable aptitudes and interests will become perishable because of significant environmental influences. The drivers of that change are economical, technological and societal in nature (Anderson & Herriot, 1997). Such changes will lead to altered sets of typical job tasks, procedures and resources, which may or may not be congruent with the selection methods used to predict the job holders' success at the time of hiring. In some cases, staff members might even have to be retrained or reassigned to different tasks. This will bear equal challenges for experts in selection as well as in training.

The aviation system is facing tremendous challenges in the coming decades due to the economic needs to expand transportation capacities by a factor 2x or beyond while maintaining the same or better safety levels (Ky & Miaillier, 2006; Krois, McCloy & Piccione, 2007). Such growth will be enabled by new technologies and operational concepts, which will significantly affect the work roles and tasks of all human actors in the future air transport system. Work roles may become more interchangeable, flexible, and proactive. For example, human operators could control aircraft from the ground or air traffic controllers might give instructions to aircraft clusters instead of single aircraft while spacing and separation could be assured by pilots in the cockpit. A new job profile might develop, for which we use the term *Aviator*. This paper is part of the project called *AVIATOR 2030* (Eissfeldt, 2006), which intends to elaborate tools and methods for a prospective analysis of job requirements and work roles in future commercial aviation. Future workshops and simulation are the main approaches applied in this project. While the paper of Eissfeldt et al. (see Symposium Proceedings) describes the results of the future workshops, this paper will provide an outline of the derived scenarios, which will be implemented in a simulation environment called *AviaSim* in order to investigate potential new work roles of pilots and controllers.

New Concepts for the Air Transport System within NextGen and SESAR

Current developments in the aviation system are driven in the US and in Europe by two large-scale industry-government programs called NextGen and SESAR. SESAR is Europe's *Single European Sky Air traffic Research* system. NextGen is the US' *Next Generation Air Transport System*. Both programs are aiming to prepare the future air transport system for the increased demands in the years 2020 and beyond. The common vision is to integrate and implement new technologies and operational concepts that will boost performance of the air traffic management system (ATM) on a sustainable basis. Both, SESAR and NextGen combine increased automation with new procedures to achieve safety, economic, capacity, environmental, and security benefits. The programs will be aligned with each other to establish common standards for technical equipment and interoperability (JPDO, 2007; SESAR, 2007).

A key component is the cooperative ATM-model (C-ATM), where aircraft are constantly sharing their position data (from navigational satellites), flight path intent, and other relevant aircraft parameters with each other and with ATC. Automatic Dependent Surveillance Broadcast, known as ADS-B is one of the technological preconditions to determine navigational data at a much higher degree of precision. This system can be used to transmit with high accuracy the same traffic information to pilots and air traffic controllers (ATCOs) and hence assure safe aircraft separation even if minima are reduced in high density airspace or at the airport. The new paradigm for planning and executing system operations will be the aircraft’s 4D-trajectories: a 4D-trajectory is the aircraft path in three spatial dimensions related to time, from gate-to-gate. SESAR’s ATM target concept is based on a further number of key features (SESAR, 2007):

- Trajectory management with minimized constraints by airspace design or pre-defined routes
- Collaborative planning continuously reflected in the Networks Operations Plan to ensure strategic de-conflicting even where resources are constrained
- Capacity gains by integration of airport operations and greater coordination between the stakeholders
- New separation modes supported by ATCOs and airborne separation systems will minimize potential conflicts and interventions
- System wide information management (SWIM), which integrates all ATM operational relevant data and links all relevant users into collaborative decision making (CDM) processes

- Humans will be central as managers and decision-makers even though advanced levels of automation support will be required to exploit the complexity.

The nature of roles and tasks for human actors within the future system will necessarily change. This will affect equipment design, staff selection, training (especially for unusual situations and degraded mode of operations), competence requirements and relevant regulations. For example, SWIM will cause a shift from mutually exchanging information to publishing, broadcasting, and goal-directed retrieval and usage of information.

AviaSim – A New Simulation Platform with Multiple Actors

Future workshops were conducted with a number of experienced ATCOs of the Deutsche Flugsicherung (DFS) and airline pilots of Deutsche Lufthansa (DLH). As described in the paper by Eissfeldt (2009), the workshop participants generated several future scenario elements such as trajectory negotiation, tactical flight planning, self-separation, working in distributed teams, or teaming with automation. On the basis of these workshop results and the review of NextGen's and SESAR's future operational concepts, a simulation platform called *AviaSim* has been developed by the authors, which should allow to investigate processes of the tactical decision making, task allocation, attention, monitoring, and information management of human actors working together collaboratively in a distributed team environment.

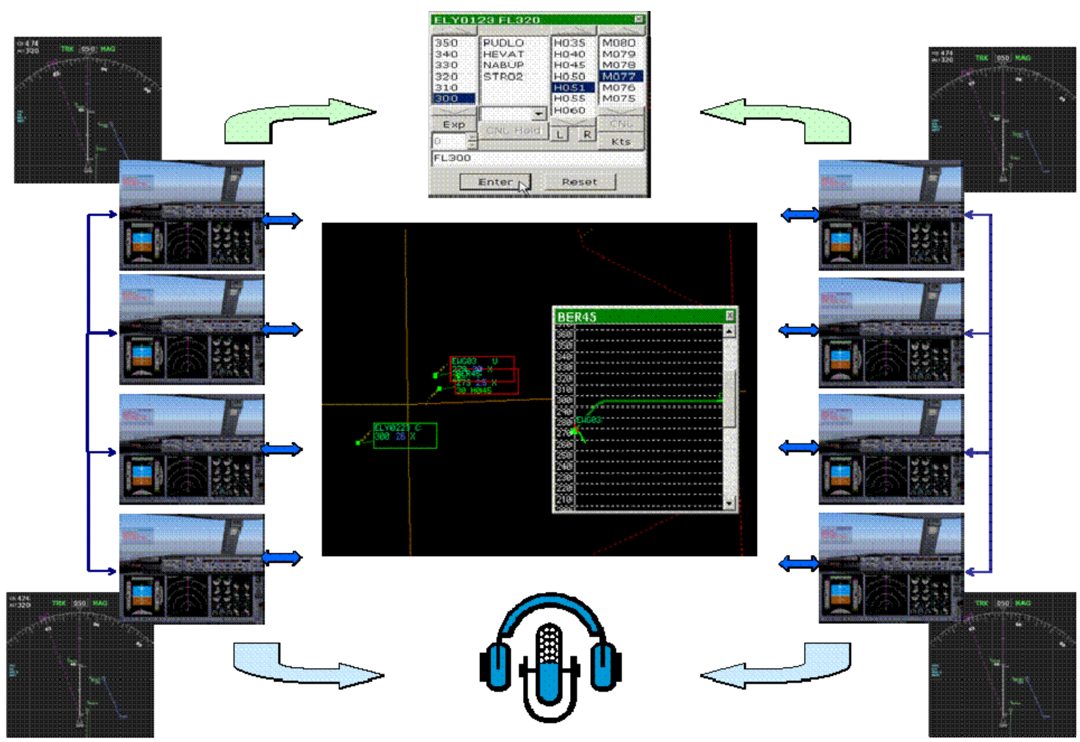


Figure 1. *AviaSim simulation platform with a networked configuration of eight flight simulators and one air traffic control simulator. Workplaces are equipped with standard technology and additional decision support systems. Communication is via data link and VOIP (Hoermann, Schulze-Kissing & Zierke, 2008)*

With open local area network architecture, AviaSim is currently configured for up to nine aviator workplaces: one for an air traffic controller and eight for pilots. Additional traffic can be generated with pre-determined flight plans per experimental script files. Each workplace has the standard equipment with additional automatic assistance functionality to support tactical decision making and continuous monitoring tasks. Figure 1 displays a configuration with traffic information displays and collision warning functionality. Communication processes are facilitated through VOIP and advanced by data link channels. This configuration serves primarily the simulation of en-route scenarios. However, with different support systems such as airport moving maps or arrival/departure managers we can also simulate with AviaSim traffic situations on ground or during departure and arrival. The type of aircraft also permits to introduce military traffic and uninhabited aerial systems.

Development of a Potential Future Scenario

The choice of a potential future scenario was guided by a number of project-specific criteria and constraints:

- Realistic simulation of the working environment
- Reflecting results of the future workshops
- Consideration of required expertise of the subjects
- Air-ground simulation of the collaboration between multiple actors with distributed roles
- Facilitating control of experimental factors and measurement of a variety of relevant dimensions, including hard data, observation, subjective rating and eye-point of gaze
- Low fidelity simulation platform with open architecture.

The main purpose of the scenario development is to create an environment, in which it is possible to investigate work processes of aviators in the future air transport system. However, the technology development has not yet progressed so far that specific descriptions would be publicly available so that operational procedures could be elaborated. Therefore, our scenarios just have a certain probability of being realistic. In order to maximize this probability, it was essential to review current proposals of SESAR and NextGen for concepts of operation as well as to conduct the workshops with present jobholders. As a result of this, the focus of our initial scenario scripts is on how to define the functionality of future inborard/onground human to human communication interfaces as well as how to integrate new automation systems in the future work processes. The collaboration between distributed human and automatic team players during operational decision making processes from gate-to-gate is a main facet of the future air transport system, which we intend to investigate (see Figure 2).

The general task is to plan and execute effective separation of traffic by complying with the needs of the user while assuring separation minima. The authority for separation control should be transferred between ATCOs and pilots during the scenario. The different human actors will cooperate with particular assistance systems which can be attached or detached to the workplace (*Concept of Control Sharing*). They can choose to communicate with each other per data link or per voice transmission in a dyadic or in a partyline manner. The airspace is sectored into managed and unmanaged areas separated by transition layers. Following specific handover procedures, separation authority will be transferred from ground to air or back from air to ground upon transitions between managed and unmanaged sectors (*Concept of Control Transfer*). When an aircraft is in self-separation mode, it will have to follow a certain set of rules to prevent the risk of loss of separation. This en-route scenario challenges the crews' abilities of planning ahead, situation awareness, communication, information management and decision making as well as their attitudes towards *Compliance to Rules* and *Trust in Automation.*

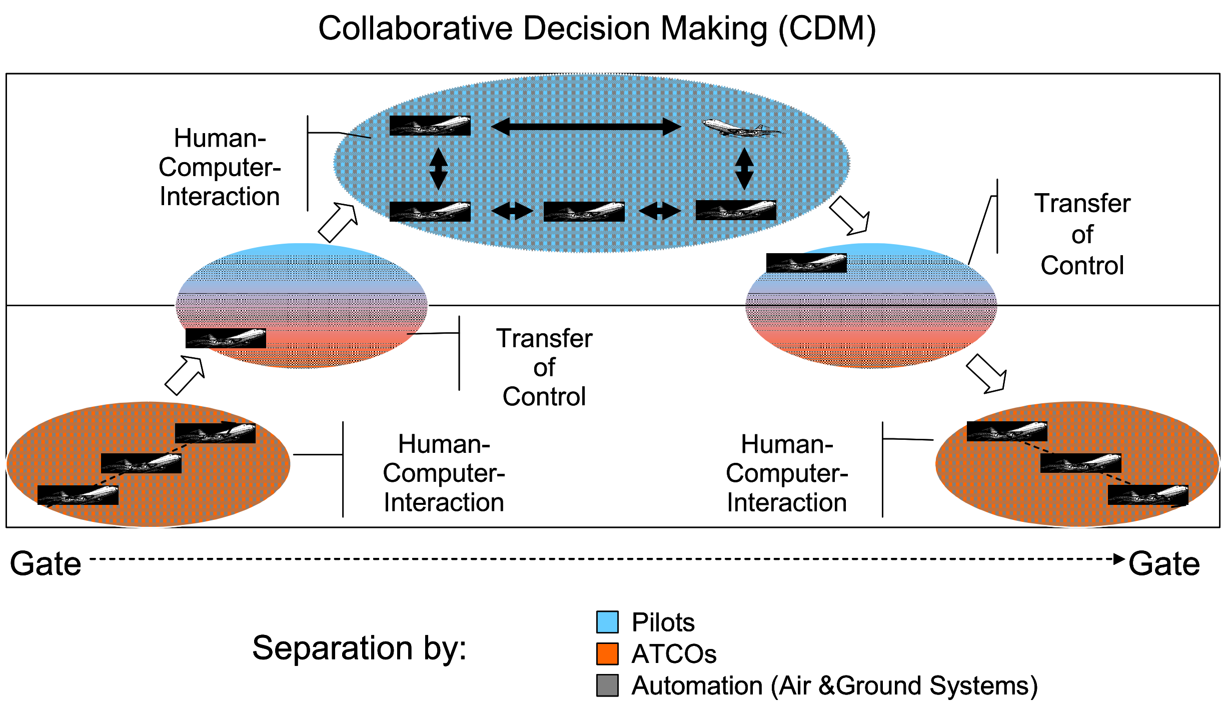


Figure 2. *Control Sharing and Control Transfer for separation tasks in future gate-to-gate operations (Hoermann, Schulze-Kissing & Zierke, 2008)*

It can be expected that such handover situations of authority carry a slightly increased risk of misunderstanding. Therefore, we expect flexibility of attention, communication, foresight and shared situation awareness of all actors to become critical factors of system performance. During self-separation, the ATCO can shift certain degrees of attention to secondary tasks. Eye-point-of-gaze measurement will be applied to record monitoring behavior. Being displayed on a different screen, secondary cognitive and perceptual tasks can also be inserted into the cockpit environment to gain some information on the workload, attention control, planning and monitoring behavior of the pilots.

In summary, task performance in the described en-route scenario should be relatively independent of the degree of subjects' technical knowledge and expertise. However, they will of course have to be current license holders. It is further intended to create a normal operations scenario without a significant amount of technical failures. The focus will be the behavior of the human actors. The determination of basic job requirements will not be linked to emergency situations in this phase of the project.

Outlook

At the time of writing this paper, the project is preparing for first tryout studies and data collection phases. The AviaSim platform is already equipped and checked with all technical features described above. Up to 20 subjects will be recruited from DFS and DLH to activate the system and to participate in the real-time simulation of authority and control sharing and transfer in the future aviation system. It is envisioned to apply a customized version of Fleishman's Job Analysis Survey as described by Eissfeldt (2009) to collect information about the cognitive task requirements. In addition, a number of performance indicators will be collected. In future, we intend to use the AviaSim platform for cognitive task analyses of aviators beyond the en-route scenario. Arrival and departure scenarios with respective assistance systems have already been drafted and will be followed by surface movements at airfields. Assistance systems with higher levels of intelligence are also being designed. The open system architecture offers plenty degrees of freedom for expanding the equipment parallel to definition and implementation phases of NextGen and SESAR.

References

Anderson, N. & Herriot, P. (1997). *International handbook of selection and assessment*. West Sussex/UK: Wiley.

Eissfeldt, H. (2006). *Projektplan „Aviator 2030"*. Hamburg/Germany: DLR ME-PS.

Eissfeldt, H. (2009). Aviator 2030 – Determining ability requirements in future ATM systems. *International Symposium on Aviation Psychology*. April 27-30, Dayton/OH.

Hoermann, H.J., Schulze-Kissing, D. & Zierke, O. (2008). Kooperation und Verantwortung im zukuenftigen ATM-System. Paper presented at the 4th AVIATOR2030 Progress Meeting. DLR: Hamburg/Germany: August, 18th, 2008.

JPDO (2007). *Concepts of operations for the next generation air transportation system* (Version 2.0). Joint Planning and Development Office. http://www.jpdo.gov/ .

Ky & Miaillier (2006). SESAR: towards the new generation of air traffic management systems in Europe. *Quarterly Issue of the ATCA Journal of Air Traffic Control*.

Krois, McCloy & Piccione (2007). Research portfolio for the Next Generation Air Transportation System (NextGen). *International Symposium on Aviation Psychology*. April 23-26, Dayton/OH.

SESAR (2007). *The ATM target concept - D3*. SESAR Definition Phase – Deliverable 3. http://www.sesar-consortium.aero/ .

MODELING PILOT COGNITIVE BEHAVIOR FOR PREDICTING PERFORMANCE AND WORKLOAD EFFECTS OF COCKPIT AUTOMATION

Guk-ho Gil, David Kaber, Sang-Hwan Kim, Karl Kaufmann & Theo Veil
Industrial & Systems Engineering, North Carolina State University
Raleigh, NC
Paul Picciano
Aptima Corporation
Boston, MA

The objective of this research was to develop a model of pilot cognitive behavior to predict performance and workload while using varying degrees of cockpit automation to serve as a basis for future systems design. A cognitive task analysis (CTA) was conducted on expert pilot performance a flight control panel (FCP), control-display unit (CDU) and flight management system, and an enhanced CDU (CDU+) providing pre-programmed arrivals from air traffic control in a simulated landing and approach task. Cognitive models were developed from the CTA using an enhanced form of the GOMS language, including a set of additional task operators, to represent pilot actions on cockpit interfaces. Pilot performance and workload data from a parallel empirical study of the same flight tasks were used as a basis for validating the cognitive model output. Indices of automation complexity were formulated based on counts of task methods and steps, required chunks of information, and information transactions coded in the enhanced GOMS models. These indices revealed high complexity for the FCP mode and low complexity for the prototype CDU+ mode. The automation index values were positively and significantly correlated with pilot heart rate (as an objective measure of workload) and vertical path deviation error from the experimental data set. The computational cognitive models of pilot behavior in using forms of cockpit automation were demonstrated to be a viable tool for predicting pilot workload and flight performance under high workload flight conditions.

Early research on cockpit automation (e.g., Wiener & Curry, 1980) identified potential human performance consequences resulting from a technology-centered approach to automation design implementing automation wherever and whenever possible, while leaving unanticipated and unstructured tasks to the pilot. These consequences include pilot complacency, vigilance decrements, loss of situation awareness and decision making problems. A number of empirical studies subsequently demonstrated such negative effects of technology centered automation design (e.g., Parasuraman et al., 1992; Endsley & Kiris, 1995) both in the aviation context and other domains. On this basis, human-centered approaches to cockpit automation (e.g., Billings, 1997) were proposed. This includes considering the information processing and performance capabilities of pilots as well as how pilots interact with cockpit interfaces. Empirical studies were conducted to determine the impact of various levels of automation on human performance, workload and situation awareness in aviation-related tasks (e.g., Endsley & Kaber, 1999), which led to guidelines for the use of intermediate modes of automation (between manual control and full automation). Beyond this, qualitative models for selecting the types and levels of automation applicable to human-machine systems (Parasuraman et al., 2000) were developed.

The main issue with the existing approaches to cockpit automation design is that they require empirical data as a basis for design alternative selection or they are based on collections of design guidelines with limited theoretical explanation of why such guidelines might be effective. Experimental studies to obtain necessary data are time consuming and costly. Also, the lack of a cognitive explanation for why certain design principles may be useful limits understanding of when and how guidelines should be applied. With this in mind, there is a need to develop computational models of pilot behavior in interacting with cockpit automation as a basis for reducing experimentation to assess or validate specific forms of automation. Such models can also provide a basis for explaining the effects of automation design guidelines in terms of perceptual processing, memory transactions, decision rule use, and response execution; thereby providing a more theoretical foundation of human-centered design of automation. Based on the prior research, the objective for the present study was to develop a computational (computer-based) model of pilot cognition interacting with various forms of cockpit automation as a basis for future system design.

Method

Flight Simulator and Flight Scenario

A PC-based flight simulator was setup for cockpit automation prototyping and to collect data on pilot performance for use in the cognitive model validation step. The simulator setup consisted of two PCs and flight deck controls, including a yoke, a throttle quadrant, and rudder pedals (see Figure 1 (a) for the simulator setup and displays) integrated with the X-Plane simulator software. Two LCD monitors were arranged vertically with the lower display presenting the instrument panel of the Boeing 767-300, including the primary flight display (PFD), flight control panel (FCP), and control display unit (CDU) (or flight management system (FMS)) interface. The upper display showed an out-of-cockpit view of the dynamic flight situation rendered by X-Plane. The display contents of the two monitors were synchronized using a TCP/IP network supported by the X-plane software.

(a) (b)

Figure 1. *Simulator setup (a) and image of X-Plane displays (b).*

A realistic arrival and landing scenario was created to support the objectives of conducting a CTA on pilot interaction with cockpit automation and the experimental study of the performance effects of automation in addressing normal events during a high workload phase of flight. Reno-Tahoe International Airport (KRNO) was chosen for its proximity to significant terrain and selection of instrument approaches and arrivals. There were three critical events pilots encountered in the flight scenario. The first critical event was a re-clearance from the northern standard terminal arrival (STAR) to the southern STAR due to a runway changing. This occurred 5 NM from the first waypoint, which served both STARs, and pilots had a very short period of time to interpret the clearance and command the aircraft to turn onto the new STAR. The second critical event was a northbound leg of the STAR to align the aircraft with the ILS final approach. This leg was defined as the backcourse of the ILS serving the opposite runway. Backcourse procedures are familiar to all instrument rated pilots, but they are not often encountered in normal service. This required extra effort from pilots to recall and carryout the correct procedures at the proper times. The last critical event was a clearance to descend from the initial altitude. If there was any delay in beginning the descent or if the rate of descent was too low, intercepting the glideslope became very difficult.

Three Interfaces Representing Different Forms of Cockpit Automation

There were three different modes of cockpit automation (MOAs) that were simulated through the X-Plane software. Each MOA had four types of information processing functions (TOF) including perception of flight status (TPF-P), flight information analysis (TOF-IA), decision making on flight path (TOF-DM), and pilot action implementation (TOF-AI). In the FCP mode, X-Plane presented the B-757/767 flight control panel. Pilots used the FCP display for tracking altitude and speed (TOF-P) and they dialed-in flight path targets (TOF-AI) during the experiment. Because, X-Plane does not provide the B-757/767 CDU, a new realistic CDU interface was developed using the X-Plane SDK. This was then employed for the CTA and pilot performance study. With respect to the CDU+ mode, the main difference from the CDU mode was that the system was capable of presenting to the pilot (TOF-P) ATC suggested routes including vertical path, when changing or deciding on other routes (TOF-DM) under inclement weather conditions, etc. With these pre-planned routes, pilot control actions (TOF-AI) were dramatically reduced, as the CDU+ required no pilot interaction during the STAR, once the desired runway for landing was selected.

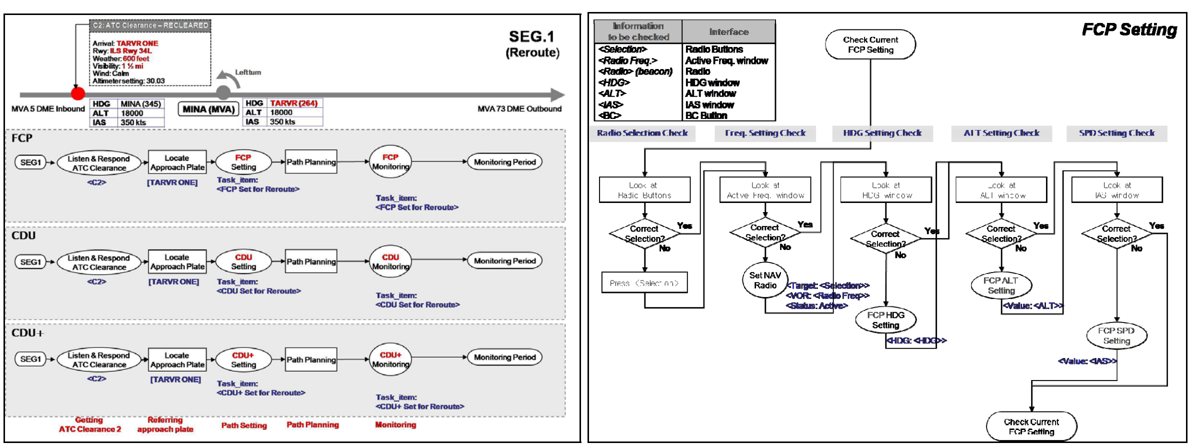
Cognitive Task Analysis

There was a need to develop an understanding of the commercial transport pilot's working context as a starting point for the cognitive modeling effort. Kieras (1997) suggested that cognitive modeling starts with a CTA. The purpose of this step in the research was to identify expert pilot behaviors in flying the high workload landing approach scenario using the different forms of cockpit automation simulated through the enhanced X-Plane setup. Specifically, the CTA was expected to reveal pilot goals, decisions, information requirements, and tasks in achieving goals at various stages in the approach. Information from verbal protocols and goal-directed task analyses (Endsley, 1993) was used to develop the computational cognitive models of pilot behavior with the FCP, CDU and CDU+ modes of control.

The CTA required several steps, including: (1) videotaping expert pilot performance with the X-Plane simulation in the test flight scenario; (2) recording pilot verbal protocols and transcribing them; (3) formulating pilot task lists for each MOA. Table 1 shows an example of task items for the FCP mode at a specific location (73 DME from the MINA VOR (MVA) outbound) after receiving a clearance from ATC according to the flight scenario; (4) developing pilot action flow diagrams (AFDs) of overt and cognitive behaviors as the basis for cognitive model coding. Figure 2 shows example AFDs for the use of the three different MOAs in the rerouting task (Figure 2(a)) and a sub-task flow to check FCP settings and the required information for the task (Figure 2 (b)); and (5) expert pilot verification of the AFDs for accuracy in describing behaviors with the automation in the various phases of the approach. For the first, second and fifth steps of this procedure, a highly experienced former USAF transport pilot (C-130) with ATP certification served as the expert pilot.

Table 1. *Example of task items for FCP use.*

Location	Current Status (Expected)		ATC Clearance		FCP	
					Tasks	Objects
MVA 73 DME outbound	NAV1 NAV2 Source Altitude IAS HDG	MVA/I-RNO FMG NAV1 18000 350 283	Altitude Speed Altimeter	16000 250 30.03	Descending (to 16000) Speed down Switch Radio HDG Setting (344) BC toggle on Altimeter Setting	V/S mode IAS knob NAV1 Radio HDG knob BC button Altimeter knob



(a) (b)

Figure 2. *Example of AFDs for general flow of rerouting task (a) and sub-task for checking FCP setting (b).*

Development of GOMS Models

Enhanced GOMS (E-GOMS) models were created based on the results of the CTA, specifically the AFDs. The general structure and flow of the models was similar to NGOMSL (Kireas, 1997); however, E-GOMS included an expansion on theGOMSL available operator set to more accurately represent pilot actions on cockpit interfaces (e.g., dialing knobs). The E-GOMS models included a main (task) goal, sub-methods and operators for each sub-method, as well as task item representation. Two major features of each model were the description of the action flow and the information object set. Models not only represented pilot behaviors, but also the information to be manipulated during flight tasks (e.g., from external ATC clearances or internal path planning). All information objects were coded as audio objects with their own variables and values. For example, the information object for the CDU SPD/ALT setting had two variables, a SPD value and an ALT value. Internal path plans were represented as task-items.

Empirical Study

A lab experiment was conducted to assess the effects of the FCP, CDU and CDU+ modes of automation on pilot performance, and subjective and objective workload responses (NASA-TLX and heart rate, respectively). The experiment used the same scenario as used for the cognitive model development (high workload landing approach with a "last minute" reroute, steep descent and speed reduction). The main objective was to test hypotheses on the potential for pilot flight control errors in response to critical events based on the nature of the automation interfaces and functionality (e.g., the CDU MOA was expected to produce greater waypoint over shoot errors upon the reroute due to the complexity of flight path reprogramming). The experiment also served to generate a data set for preliminary validation of cognitive model output.

Results

Experiment

Pilot performance results revealed highly significant effects of MOA among data segments including vertical and lateral path deviations ($p<0.0001$). Pilot objective workload (heart rate) revealed significant effects of MOA and there was an interaction of MOA and flight segment across test trials ($p=0.0487$) when trial order was considered in the statistical model. Pilot subjective workload ratings (NASA-TLX) revealed a marginally significant effect of MOA ($p=0.0949$) when trial order was considered in the model. In general, these results indicated an influence of the FCP, CDU, and CDU+ modes of control on pilot behavior and motivated the cognitive model development effort.

Cognitive Model Outputs

As previously mentioned, the cognitive models were analyzed manually for pilot performance predictions with the various forms of cockpit automation. Since the flight scenario was divided into three segments for analyzing the actual pilot performance data from the lab experiment, the cognitive model outputs were also determined and analyzed according to the same three segments (rerouting, turning, and final approach). In general, the outputs from the E-GOMS models can be characterized as task complexity indices for each MOA and flight segment. Four indices were determined for this study, including: (1) the number of sub-methods to perform tasks during a flight segment; (2) the total number of steps in the model, including those as part of required sub-methods; (3) the required number of information elements to complete a task during a segment (including the sub-methods); and (4) the number of information transactions between WM and LTM or external memory (e.g., pilot notes on an approach plate).

Table 2 shows the values for the task complexity indices for each MOA and flight segment, as determined from the E-GOMS models. It should be noted that the indices for the final flight segment are the same across MOAs because only the FCP mode was used in this segment. In general, the FCP mode produced larger index values than the CDU and CDU+ modes. The CDU+ mode generated the smallest index values among all modes. Therefore, the CDU+ mode was considered to pose the lowest level of task complexity and use of the FCP mode yielded the highest level of task complexity.

Table 2. *Calculated task complexity indices for each MOA and flight segment.*

		Mode of Automation		
		FCP	**CDU**	**CDU+**
Seg. 1	# of Sub-methods	7	6	6
	# of Steps	74	77	69
	# of Information	32	24	24
	# of Transactions	4	3	3
Seg. 2	# of Sub-methods	16	12	10
	# of Steps	169	151	124
	# of Information	63	46	42
	# of Transactions	12	11	11
Seg. 3	# of Sub-methods	7	7	7
	# of Steps	75	75	75
	# of Information	31	31	31
	# of Transactions	7	7	7

On the basis of these index values, the potential for flight errors can be predicted. Kieras (1997) noted that, if more than five (5) chucks of information must be maintained in WM at any given time, this lead to cognitive overload and, consequently, induce errors in performance. Figure 3 shows a plot of the number of chunks of information required by a pilot during the second flight segment (turning) under each MOA. It can be seen from the plot that the number of chunks for setting the FCP control to turn the aircraft at TARVR is 16, while the other modes of control (CDU and CDU+) required less than two (2) chunks of information. Even though the task of setting the FCP for turning can be further decomposed into heading setting, altitude setting, radio setting and air speed setting, the amount of information that must be manipulated by a pilot at a given time exceeds the criteria suggested by Kieras (1997) and the "magic number" of working memory capacity identified by Miller (1956). Thus, it can be predicted based on the cognitive model output that a pilot may make flight errors in setting the FCP for turning descent of the aircraft under high workload conditions. Based on the results of the experiment with actual pilots, it was observed that some participants did not set the FCP appropriately at this point in the flight and this produced greater path deviation than for the CDU or CDU+ modes.

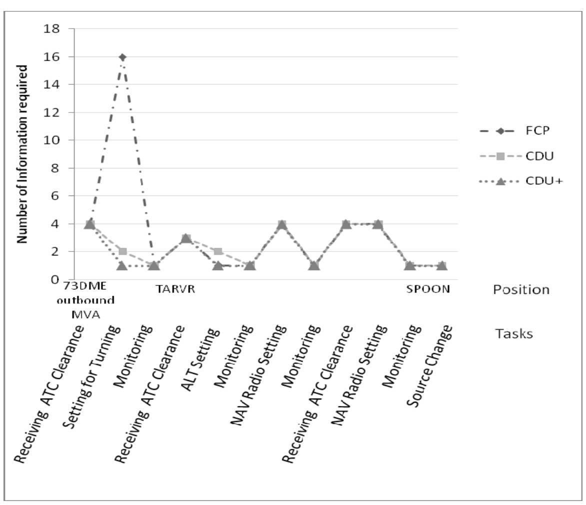


Figure 3. *Number of chunks of information required during the second flight segment.*

Comparison of Model Outcomes with Experiment Data

Non-parametric correlation (Spearman) analyses were conducted on the task complexity index data and observations on the workload and performance response measures from the experiment. Since only the FCP mode of control was used in the final segment of the flight scenario, data for the first and second segments were used for comparison of model outputs with the pilot HR and path deviation responses. In addition, a composite task difficulty index was determined based on the E-GOMS models for all three segments of flight (across all pilots) for correlation with the NASA-TLX scores, determined at the close of trials.

Results revealed the pilot HR responses were highly correlated with all four model-based task complexity indices (r= 0.928, 0.829, 0.928, 0.883; number of sub-methods, number of steps, required chunks of information, and information transactions, accordingly) with a significance level of p=0.05. Additional correlation results revealed NASA-TLX scores to be positively correlated with the number of sub-methods, number of method steps, and number of required chunks of information. Unfortunately, there were too few data points for the significance levels to be considered reliable. Related to this, the number of information transactions was not significantly correlated with the subjective workload data. In addition, there were positive linear relations between vertical path deviation and model outcomes including: number of sub-methods (r=0.978, p=0.008); number of steps (r=0.886, p=0.019); number of required information elements (r=0.978, p=0.008); and number of information transactions (r=0.971, p=0.001). However, there was no significant correlation between the lateral path deviation data and model outcomes. These results suggested that for the specific flight scenario, vertical path control performance may be most sensitive for revealing differences in cognitive processing due to modes of cockpit automation.

Conclusion

The computational cognitive models of pilot behavior in using the various forms of cockpit automation were demonstrated to be a viable tool for predicting pilot workload and flight performance under high workload flight conditions. The new cognitive modeling approach may support the development of a general models of pilot cognition, which may facilitate future automated cockpit design.

Acknowledgement

This research was supported by NASA Ames Research Center under Grant No. NNH06ZNH001. Mike Feary was the technical monitor. The opinions and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of NASA.

References

Billings, C. E. (1997). *Aviation automation: The search for a human-centered approach*. Mahwah, NJ: Lawrence Erlbaum Assoc.

Endsley, M. R. and Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462-492.

Endsley, M. R. and Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37, 381–394.

Endsley, M. R. (1993). A survey of situation awareness requirements in air-to-air combat fighters. The International Journal of Aviation Psychology, 3(2), 157-168.

Kieras, D. E. (1997). A guide to GOMS model usability evaluation using NGOMSL. In M. Helander, T. Landauer & P. Prabhu. (Eds.), *Handbook of Human-Computer Interaction*, 2nd Ed. completely revision ed., 733-766.

Miller, G.A. (1956). *The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information*, V.63, N.2, pp.81-97.

Parasuraman, R. , Bahri , T. , Deaton, J. , Morri son, J. & Barnes, M. (1992). *Theory and design of adaptive automation in aviation systems* (Progress Rep. No. NAWCADWAR-92033-60). Warmister, PA: Naval Air Warfare Center, Aircraft Division.

Parasuraman, R. , Sheridan, T. B. and Wickens, C. D. (2000). A model of types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics*, 30, 286–297.

Wiener, E. L. & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, 23, 95–1011.

USAGE DATA FROM USERS OF TWO SYNTHETIC VISION SYSTEMS

Dennis B. Beringer
FAA Civil Aerospace Medical Institute
Oklahoma City, OK

Although much research has been conducted regarding display design and formatting criteria for terrain-depicting or synthetic-vision displays, little data have been collected concerning how General Aviation pilots use fielded displays. Structured interviews were conducted with a small sample (10; 33% response rate) of users of two fielded synthetic-vision (SV) displays, one with selectable Highway-In-the-Sky (HITS) guidance and one without. Questions were asked concerning pilots' experience (both general and specific with display systems) and use of the SV systems by phase of flight. Use rates for the first system (with a selectable HITS) were high, with "always used" being reported for 57% or more of the sample during cruise, descent, and approach. Including "often used" increased this to over 71%. Patterns were slightly different for the second SV system users, and were likely attributable to the smaller proportion of sampled users and to format and content differences; all found the displays extremely useful.

Forward-looking perspective pictorial displays (synthetic vision, SV) are becoming more available in general aviation (GA) and experimental aircraft. A significant number of research efforts have been initiated to determine design guidelines based upon both pilot performance and pilot preference**.** These include studies examining display design characteristics (Schnell et al., 2003), guidance symbology (Beringer, 2000), applications to specific flight tasks (Bartalone et al., 2004), and the characteristics of the terrain representations (Lemos et al., 2003). An Advisory Circular on these systems in Part 23 aircraft has been published by the FAA's Small Airplane Directorate (FAA, 2005). Other design guidance has been published (SAE Aerospace, 2005), and Minimum Aviation System Performance standards for a number of synthetic-vision-related systems has just been completed (Radio Technical Commission for Aeronautics, Special Committee 213). However, the focus has largely been on defining design parameters for synthetic vision systems (SVS) and the minimum performance acceptable in fielded systems. It was also of interest to see how those few systems that had already been approved and fielded (two in particular) were being used by pilots. As is often seen in the introduction of new systems, users often find new and sometimes unanticipated ways of using them. A structured interview was prepared for use with pilots having some experience flying these two display systems to determine (1) how frequently and under what conditions the displays and certain features were used (phase of flight, weather), (2) what the pilots perceived as the most and least useful features and (3) what additional features or functions were found desirable but lacking.

METHOD

Interview instrument

An interview form was constructed to assess a number of demographic, experiential, and system-use items. *Demographic information* included age, sex, year that the pilot began licensed flying, certificates held, ratings held, restrictions on the medical certificate (e.g., eyeglasses required), and date of last recurrency check, proficiency check, or biennial flight review. *Pilot experience questions* included summaries of categorized flight hours (VFR, IFR, simulator, etc.), type of aircraft flown most frequently, experience with head-up displays (HUD), electronic primary flight displays (PFD), PFDs with terrain representations, enhanced vision systems with/without flight guidance, night-vision goggles, forward-looking infrared (FLIR) displays, and any training related to HUDs or FLIR displays.

Questions regarding *synthetic vision system features, usage, and evaluations/ratings* included (1) type of hardware used (manufacturer/model), (2) terrain flown over while using (6 categories reported by percentage), (3) type of operations in which used (day/night, VMC/IMC), (4) altitudes at which most flying was done using the system, (5) frequency of use of SVS by flight phase, (6) frequency of use of pathway (highway-in-the-sky, HITS) guidance if available by flight phase, (7) the 4 most useful functions of the system, (8) the 4 least useful functions of the system, (9) features desired that were not available, (10) training provided for use of the system and format of that training, (11) strengths and weaknesses of training when provided, (12) operations made possible by the SVS that could not previously been accomplished or that would have been difficult without it, (13) operations that could be performed using the system but were not yet allowed under operational rules, (14) ratings of reliability/accuracy and safety of both the SVS and flight guidance information, and (15) location of the primary flight display and the standby instruments on the panel of the aircraft most frequently flown with the SVS.

Participants

Of the 30 certified pilots contacted, 10 agreed to participate (33% response rate). Names of potential interviewees were provided courtesy of 2 manufacturers of Electronic Flight Instrumentation Systems (EFIS) currently approved for use in Part 21 airplanes. Demographics for the 10 pilots who chose to participate are as shown in Table 1. Of these 10 pilots, 5 had primarily piston-engine time, 2 had turboprop time, and 3 had turbojet time.

Table 1. Descriptive statistics for participants' ages and flight experience as Pilot in Command (PIC).

	Age	Years flying experience	Total hours PIC	PIC hours last 90 days
Mean	55.7	28.9	6,859	69
Median	61.0	26.0	4,397	54
SD	11.2	11.2	6,358	42
Maximum	67.0	44.0	17,800	145
Minimum	33.0	13.0	325	25

All of the participants were users of 1 of 2 SVSs certified at the time of the interviews. One of the SVSs (to be referred to as System 1, n = 7) had an egocentric point-of-view forward-looking terrain display (monochrome terrain) on the Primary Flight Display (PFD) and flight guidance provided by a selectable HITS. The other system (System 2, n = 3) did not have the HITS but did have color-coded terrain. Although both systems now have egocentric perspective-view depictions of the forward view of terrain on the PFD, this had only been in certified status for System 2 for about 2 years at the time of the interviews. As such, 2 of the 3 interviewed users for System 2 had experience largely with the exocentric-view version on the multi-function display (MFD). In this version, the viewing point for the terrain depiction, ownship, and perspective courseline was above, behind, and slightly to the right of ownship. Depictions of the terrain, however, were similar except for coloration.

Procedure

Potential participants were contacted by telephone, and the intent of the proposed interview, to occur at a later date, was explained. Participants were informed that they would be compensated for their time, although the majority who participated indicated a willingness to do so whether they were or were not compensated. For those who agreed to participate, a copy of the proposed interview questions was sent to them via e-mail so that they could formulate complete and accurate responses (particularly regarding flight experience and logged hours) in advance of the interview. A date and time was then determined for the interview, and the participant was telephoned at the appointed time. Responses were recorded by the interviewer for each of the various questions, as well as the documentation of unsolicited commentary not fitting within the context of one of the specific questions.

RESULTS

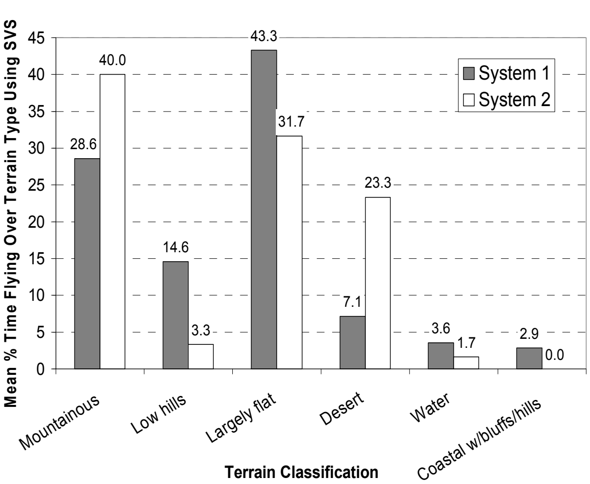


Figure 1. Mean percentage of time using display by terrain type being flown over for Systems 1 and 2.

Terrain

Interviewees were asked about what types of terrain were flown over, by percentage, when using the SVS. Figure 1 shows the distribution of flight time by category of over-flown terrain for each system. Interestingly, the smaller sample using System 2 spent more time flying over desert and mountainous terrain than did those using System 1, and the System 1 users spent more time flying over low hills than did the System 2 users. Both had a large proportion of time spent flying over largely flat terrain where the benefits of the terrain depiction would be minimal.

Illumination/Weather

The participants were asked what percentage of time they flew in various illumination and lighting conditions. Figure 2 summarizes the data for each system. In each case, the system was used predominantly in day Visual Meteorological Conditions (VMC). System 1 operators used their system more in night VMC than did System 2 operators (6.7% as compared with 1.7%), and they also used their system more in day Instrument Meteorological Conditions (IMC) (8.6% versus 1%), but less in night IMC (1.6% versus 4%). One should keep in mind that System 1 provided the heading-oriented out-the-window analog and could provide HITS guidance, with this combination likely to influence use. However, the small sample size makes it impossible to say too much concerning the differences.

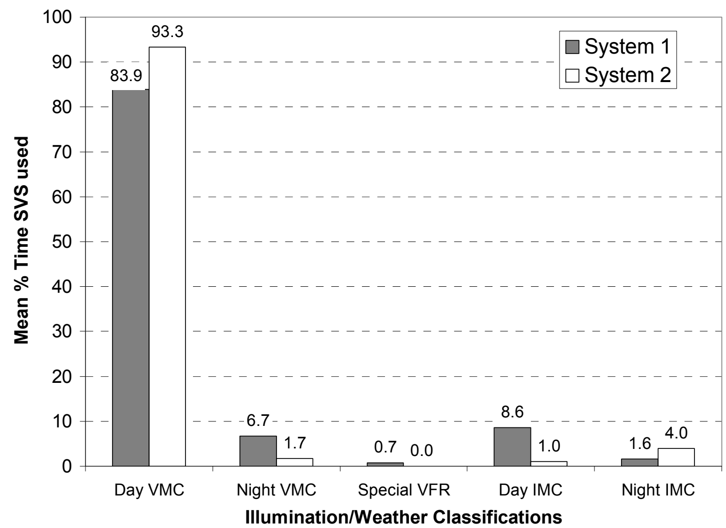


Figure 2. Percentage of time spent in various weather/illumination conditions for System 1 and System 2.

Altitude Brackets

The participants were also asked at what altitudes they flew most during cruise flight with the SVS. Figure 3 presents the data for each group. It is interesting that one sees a dichotomous distribution of altitudes with about a quarter of the flights using each system occurring between 300 and 3000 feet AGL. The rest, however, were at or above 10,000 feet MSL (57% for System 1 and 76% for System 2). These display systems are most useful close to the terrain or near significant terrain features, so much of the cruise flight indicated by these users would not be in close proximity to terrain, with the exception of high mountain peaks.

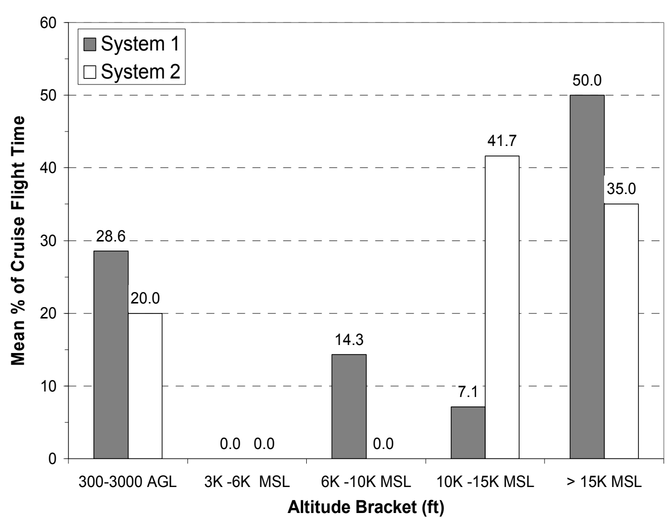


Figure 3. Altitudes most often used for cruise flight for System 1 and System 2 users.

Phases of Flight

General system use. The next series of questions asked system users how often they used the SVS during various phases of flight. Figures 4 and 5 depict the percentage of each group of system users that categorized their use of the display system during the listed phases of flight by each of the listed frequencies ("always," "often," "sometimes," or "never"). Combining two categories, "always" and "often", to serve as an index of frequent use, 71% of the System 1 users indicated that they used the SV display frequently during climb, 86% used it frequently during cruise/enroute, 86% used it frequently during descent phases, and 100% "always" used it during approach. While there were smaller values (Figure 4) for "sometimes" use, it should be noted that no System 1 pilot reported "never" using it. If we combine the same two categories for System 2 as a frequent-use indication, the values are slightly different in that frequent use during climb was reported by 67%, frequent use during cruise/enroute was reported by 33%, frequent use during descent was reported by 67%, and frequent use ("always") during approach was reported by 67%. It is clear for both systems that the most frequent use in any phase of flight is during approach. It should be pointed out that some of the differences between the uses of the two systems are likely due to the majority of System 2 users having experience with the exocentric-view version of that system on the MFD as opposed to the egocentric-view presentation on the PFD in System 1.

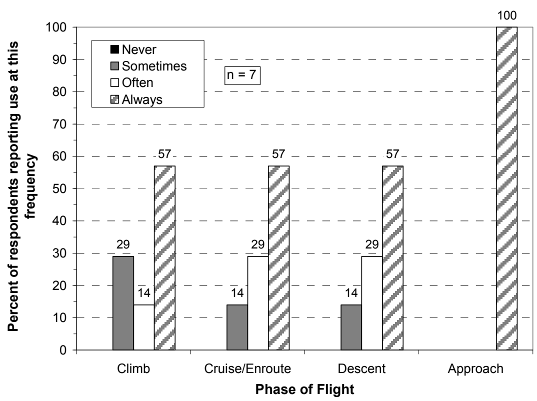


Figure 4. Frequency of use of System 1 by phase of flight.

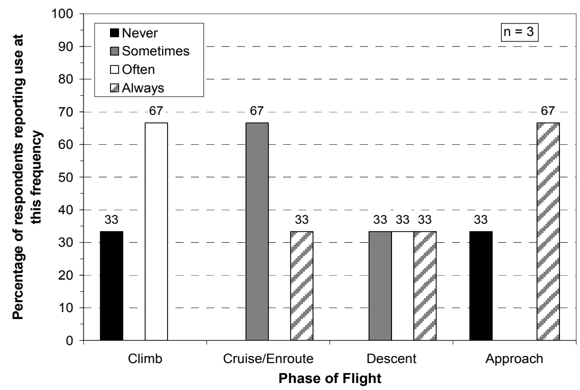


Figure 5. Frequency of use of System 2 by phase of flights.

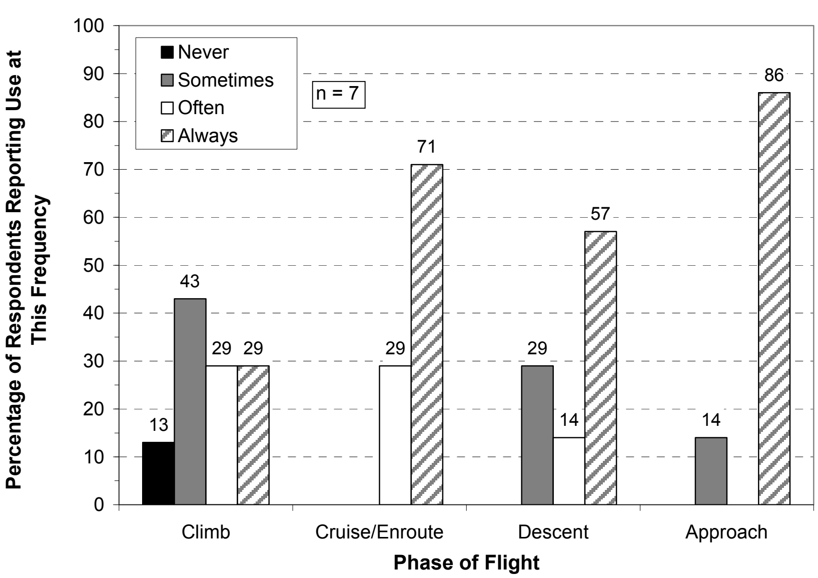


Figure 6. Frequency of use of System 1 HITS by phase of flight.

HITS use. Only System 1 specifically used the HITS guidance representation at the time of the interviews. Combining the "often" and "always" categories to represent frequent use, as done previously, 58% of the pilots frequently selected the HITS guidance for climb, 100% frequently selected it during cruise/enroute (71% "always"), 71% frequently selected it during descent, and 86% frequently ("always" in this case) selected the HITS on approach (see Figure 6). Thus, HITS guidance was more likely than not to be enabled for all phases of flight using this system.

Most Useful Functions

Pilots were asked to indicate which 4 system/display features they considered to be the most useful. This was followed by the inverse question, which were the 4 least useful or distracting features. Finally, the pilots were asked what features they would like to see implemented that were not available at that time. Table 2 presents the items, along with the frequency of mention. Items mentioned with a high frequency are paired with an "n" indicating the size of the sample to which the item has relevance.

Table 2. Frequency of mention of (1) most useful features, (2) least useful features, and (3) desired features.

Most useful features		**Least useful features**		**Desired features**	
Feature	Mentions	Feature	Mentions	Features	Mentions
Terrain depiction / warning / color coding	10 (of 10)	Terrain turns off at extreme bank	1	Egocentric view	1
Highway in the sky	6 (of 7)	Too many button presses	1	Turn coordinator	1
Off-level bank indication	1	HITS chasing in IFR	1	Match Garmin database	1
Velocity vector	1	Small symbology	1	Runway	1
Altitude/Airspeed on PFD	1	Difficulty loading approaches	1	TCAS targets	1
Nearby airports	1			More terrain realism	1
Winds aloft	1			Combine with FLIR	1
Descent rate	1			Sensor inset	1
Easy visual scan	1			Remaining runway indication.	1
Grid on terrain	1				
Flight path on MFD	1				
Digital pitch readout	1				
Radar altimeter	1				
Runway depiction	1				
Airport map	1				

It is clear that the terrain-related features and the HITS were the most universally valuable to users of these systems. Other items were less universally useful but received mention by one or another individual pilot. Two pilots mentioned sensor-image (e.g., FLIR) insets as a desirable feature. This approach is being incorporated into other systems recently fielded or introduced for certification.

Available Training

Training used. Interviewees were asked what training was available for the system they were using and which types of training they had used. In a follow-up question, they were asked about particular strengths and weaknesses of the training they used. For System 1, all of the respondents indicated they had used the handbook, 5 of the 7 said that they used DVD or videotaped instructions, 2 used embedded (in the device) simulation, and 1 each used in-aircraft training, computer-based instruction, and classroom instruction (the latter indicated a preference for an independently offered course that he felt was superior to the one offered by the manufacturer; it was noted that classroom instruction had to be paid for). For System 2, 2 of the 3 pilots said they used the handbook, 1 used computer-based instruction, and 1 also received classroom instruction. Although the pilots using System 2 indicated that recorded-media training was available, none of them had used it. Across these 2 systems, then, the most widely used training aid was the pilot's operating handbook (90%), followed closely by recorded media (50%). In a tie for third place were embedded simulation training (20%) and computer-base instruction (20%).

Training strengths and weaknesses. Regarding *perceived strengths* of the available training, 2 System 1 pilots rated the DVD-based instruction highly. One favored an independently authored short book on using the system. Of the System 2 users who commented, 1 thought the 1-hour course did a good job of covering system operation, and the other favored the handbook. *Perceived weaknesses* in the System 1 training mentioned included 2 references to the need for a software simulator independent of the device or simulator training, 2 references to a need for interactive training using the device (interactive tools), and two references to the handbook (too lengthy; too difficult to understand in isolation from the actual hardware).

Operations Now Possible with System

The participants were also asked what operations they believed they could now perform legally with the SVS that they could not perform before and, additionally, what might be possible technically but was not approved operationally. In the first instance, there were multiple references to low-level, low-visibility terrain avoidance and approaches and Category II - and even Category III - approaches/landings. For "all things possible" but not as yet approved, the pilots mentioned all-weather operations, low-level IMC in mountainous terrain, credit for IFR approaches into airports without published approaches, lowered approach minimums, approaches using HITS without an ILS on site, and Categories II and III approaches/landings.

Reliability / Accuracy / Safety

Finally, the pilots were asked to rate their SVS for reliability/accuracy and overall safety (poor = 1, below average = 2, above average = 3, and excellent = 4). The mean ratings were: System 1 – reliability/accuracy = 3.83, overall safety = 4; System 2 – reliability/accuracy = 4, overall safety = 3.67. Those individuals using System 1 were also asked to rate the HITS in the same way: reliability/accuracy = 4, overall safety = 4. Thus, the users of these systems perceived them as providing high levels of reliability and accuracy, as well as being very safe.

CONCLUSIONS

It is clear that the features making these systems unique, as compared with other flight displays, namely a perspective view depicting terrain and some form of pictorial guidance, are the ones that users of the systems found most useful and appealing. While it may seem discouraging that the systems are being used predominantly in day VMC when they could be used beneficially in other situations, the frequency-of-use data should be tempered by the proportion of time pilots are exposed to actual IMC. However, the systems were being used extensively in descent and on approach, phases of flight where they can make significant contributions to flight safety. Additionally, it should be viewed as positive that the systems are being used, with what many of the interviewees reported as reductions in workload when compared with more conventional instrumentation systems.

ACKNOWLEDGMENTS

The author extends his sincere thanks to his contact points at the two avionics manufacturer's sites who were willing to provide candidates for interviews, to the pilots who took the time to answer the interview questions and have in-depth conversations concerning their evaluations of the systems that they fly, and to Mr. Lowell Foster, FAA Small Airplane Directorate, for his work in preparation of the structured interview and review of the manuscript.

REFERENCES

Bartalone, A.P., Hughes, M.F. and Wong, D.T. (2004). Symbology development for general aviation synthetic vision primary flight displays for the approach and missed-approach modes of flight. In *Proceedings of the Human Factors and Ergonomics Society 48th annual meeting*, 208-212.

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

FAA (2005). Synthetic Vision and Pathway Depictions on the Primary Flight Display. Advisory Circular AC 23-26.

Lemos, K. Schnell, T., Etherington, T., Vogl, T., and Postikov, A. (2003). Synthetic vision systems: human performance assessment of the influence of terrain density and texture. In *Proceedings of the 22nd Digital Avionics Systems Conference,* vol.2, pp 9.E.3-91-10.

Radio Technical Commission for Aeronautics (RTCA). Special Committee 213 (in review). Minimum Aviation System Performance Standards (MASPS) for Enhanced Vision Systems, Synthetic Vision Systems, Combined Vision Systems, and Enhanced Flight Vision Systems for current operational applications. DO-135. Washington, D.C.: RTCA, Inc.

SAE Aerospace (2005). Human engineering considerations for design and implementation of perspective flight guidance displays. ARP (Aerospace Recommended Practice) 5589. Society of Automotive Engineers – Aerospace.

Schnell, T., Lemos, K., Keller, M., Yang, S. and Guiterres, F. (2003). Synthetic vision displays: Optimal display characteristics. University of Iowa: Operator Performance Laboratory, Final Report to NASA Langley Research Center.

THE EFFECT OF VIDEO WEATHER TRAINING PRODUCTS ON GENERAL AVIATION PILOTS' FLIGHT BEHAVIOR

William Knecht, Jerry Ball
Civil Aerospace Medical Institute, Federal Aviation Administration, Oklahoma City, OK, USA
Michael Lenz
FAA Headquarters, Washington, DC, USA

This study examined the effect of video weather training products on general aviation (GA) pilot flight behavior.

Fifty pilots participated. Training products were two popular GA weather training videos, plus a non-weather video as control. Pilots watched one video. Then, in the CAMI flight simulator, they flew a challenging 1.5-h visual flight rules (VFR) mission. Along the route, terrain rose slowly, with cloud bases squeezing pilots between ground and clouds.

The control group penetrated significantly farther into the deteriorating weather. Otherwise, no significant safety differences were observed for time spent in instrument meteorological conditions (IMC), time scud running, or time below 500' ground clearance. Neither instrument rating nor locality of pilot residence appeared to affect these safety variables.

Flight behavior—complete penetration of the weather versus diverting—could be predicted for 80% of pilots, using a model with training product + initial takeoff hesitation + pilot age.

The term "adverse weather" involves multiple factors such as restricted visibility due to low cloud ceilings, fog, rain, snow, thunderstorms, or airframe icing. Adverse weather is a perennial concern to GA. Analyses of GA accidents from the 1970s-2000s show that, despite a relatively low incidence rate for weather-related accidents (4-5%, depending on data source and classification scheme), their fatality rate is 3-4 times higher than for other GA accident causes (Bazargan, 2005; Bud, Mengert, Ransom, & Stearns, 1997; NTSB, 1989; NTSB, 2005). This is largely because weather accidents often involve flight into terrain or loss of control, which typically kills all onboard.

Civil Aerospace Medical Institute (CAMI) researchers were tasked to explore whether video weather training products significantly affect pilot flight behavior in instrument meteorological conditions (IMC). The research was conducted in two phases. Phase 1 examined data collected from January to July, 2008.

Method

Weather training products/control materials.

Learning theories fall into 3 categories: behaviorism, cognitivism, and constructivism. Because behaviorist training methods arguably apply best to procedural tasks, one cognitivist and one constructivist training product were selected.

Cognitivism focus on brain functions, particularly memory and information processing, suggesting that brains may functionally resemble computers, processing inputs to produce outputs (Waltz & Feldman, 1988). Whereas, constructivism expands on the computer metaphor, elevating cognition from a straightforward "information vector-transformation" role to a somewhat richer "construction of an inner world" (von Glasersfeld, 1995). This "inner world" involves an organized set of mental representations of external objects, relations, and interactions.

Two well-known video weather training products were selected. Product 1 constituted the "constructivist product." It focused mainly on the aeronautical decision making aspects of weather flight. It offered systematic, mnemonic risk factor checklists applicable to specific factors such as the weather in question, internal pilot factors affecting performance (e.g., skill, health, fatigue), and factors external to the pilot that could affect risk-taking (e.g., passengers needing to arrive at their destination by a certain time). After each video lecture session, it presented hypothetical flight scenarios for students to evaluate, based on the lecture content presented so far.

Product 2 constituted the "cognitive product." This focused largely on the recognition of different cloud types, visibility conditions, horizon recognition, and terrain clearance. Exercises showed still pictures of weather situations as seen aloft, asking what recognition factors were problematic, and for go/no-go decisions on VFR flight.

The third video group—the Control group—received an FAA-produced video on aviation physiology, having nothing whatsoever to do with weather.

Research design

Table 1 depicts the basic design. Training product, instrument rating, and pilot's state of residence were primary independent variables; age and flight hours were secondary. This gave a 3x2x2 between-groups design with 12 treatment cells, ≥4 Ss per cell. Cells were equilibrated for age and flight hours during pilot assignment to treatments.

Table 1. Experimental structure.

Independent variables (IV)					Dependent variables (DV)					
Training product	Instrument rating	Pilot's state of residence	Age	Flight hr	Flight duration	Minimum distance to ABQ	Minutes spent in IMC	Minutes scud running	Minutes spent < 500' AGL	Takeoff decision (Y/N)
1	VFR	OK	12 treatment cells were equilibrated for age and flight hr.							
1	VFR	Non-OK								
1	IFR	OK								
1	IFR	Non-OK								
2	VFR	OK								
2	VFR	Non-OK								
2	IFR	OK								
2	IFR	Non-OK								
Control	VFR	OK								
Control	VFR	Non-OK								
Control	IFR	OK								
Control	IFR	Non-OK								

Participants

Following IRB approval, 50 GA pilot volunteers participated, providing informed consent. Mean age was 41.0 (median = 39, SD = 17.5), mean flight hours was 1314 (median = 268, SD=2709).

Local pilots (those currently living in Oklahoma) were recruited from a list of pilots having participated in previous studies and by placing fliers in local flight schools. Non-local pilots were recruited through an advertisement in *Flying* magazine.

Advanced General Aviation Research Simulator (AGARS)

AGARS is a real-time, fixed-based GA flight simulator configured as a Piper Malibu for this study. A high-resolution visual system and a 150° field of view allow precise presentation of meteorological conditions. AGARS captures up to 150 variables continuously at 30Hz for a four-hour mission and includes up to 85 programmable non-routine events. It utilizes an experimenter operating station (EOS) and an ATC workstation. During a flight scenario, EOS allows the experimenter to visually monitor the cockpit and simulation environment. A digital camera records the cockpit, as well as pilot, ATC, and experimenter communications onto a stand-alone digital video recorder.

Flight mission

Pilots planned a VFR flight from Amarillo, TX (AMA), to Albuquerque, NM (ABQ)—approximately 90 minutes in the Malibu at high-speed cruise. They were instructed to plan with two cockpit VORs (VHF OmniRange Navigation System), an ADF (Automatic Direction Finder), with access to in-flight Automated Weather Observing System (AWOS) weather updates. A data-collecting Web-based weather emulation of *www.aviationweather.gov* was written by the experimenters and made available for preflight planning (Figure 1). After preflight, a 15-minute break was given to each pilot. Subsequently, a 30-40-minute training session with AGARS was given, including autopilot, horizontal situation indicator (HSI), and Malibu flight parameters and characteristics (e.g., maximum/stall speeds, associated power settings).

The assigned route consisted of gradually rising terrain during the first two thirds of the flight, followed by a dramatic elevation change during the last one third. During the flight, pilots experienced deteriorating VFR weather conditions. Initially, visibility was set at 8 nm and was gradually decreased to 5 nm two thirds of the way along the route. Concomitantly, cloud ceilings were lowered from 4500 feet AGL to 3500 AGL across the same terrain. As a result, ceilings gradually squeezed the pilots closer to the ground, resulting in a potentially dangerous flying situation with hazardous encounters throughout the course of the flight.

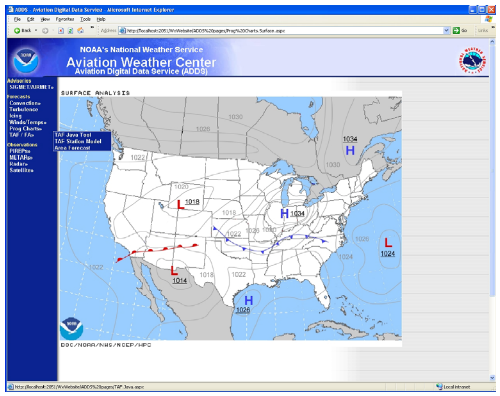


Figure 1. Sample screenshot from the Web-based emulation of *www.aviationweather.gov*.

Results

Flight duration was the only DV to satisfy the Kolmogorov-Smirnov normality test. Therefore, most analyses were done using non-parametric statistics. Table 2 shows key correlations (2-tailed *p*-values are in parentheses).

Table 2. Correlations between key variables.

Variable 1	Variable 2: Instrument rating (1=instrument rated)[1]	Locality of residence (1=Local)[1]	Pilot age[2]	Pilot flight hours[2]	Flight duration[2]	Minimum distance to ABQ[2]	Minutes scud running (<500' cloud base clearance)[2]	Minutes in IMC[2]
Instrument Rating	1.0							
State of Residence	[3]	1.0						
Pilot Age	.523[1] (.0001)	[3]	1.0					
Pilot Flight Hours	.401[1] (.004)	[3]	.757 (<.001)	1.0				
Flight Duration	-.039	.042	**-.423** (.002)	-.270	1.0			
Minimum Dist to ABQ	.013	.013	**.422** (.002)	.303 (.032)	**-**.936 (<.001)	1.0		
Minutes scud running	-.013	-.012	.051	.107	.013	-.042	1.0	
Minutes in IMC	-.020	-.005	-.089	-.084	.028	-.035	.676 (<.001)	1.0
Minutes < 500' AGL	-.281 (.048)	.144	-.167	-.289 (.041)	**.379** (.007)	**-.384** (.006)	-.095	-.174

[1]r_{pb} = Point-biserial correlation; [2]r_s = Spearman rho correlation; Low *p*-values are in parentheses (all others are non-significant (NS)); [3] No correlation run because sample had been partitioned for these factors. All *p*-values are 2-tailed.

Trivial correlations are discussed first (highlighted light gray). Older pilots were more likely to be instrument rated and to have more flight hours. Pilots with high flight hours were more likely to be instrument rated. Flight duration x minimum distance to ABQ (r_s = -.936) merely shows that the longer pilots flew, the more likely they were to get close to ABQ. Minutes scud running x minutes in IMC (r_s = .676) turned out to be partially a complex-but-trivial side effect of the way scud running was defined.

Other correlations (medium gray) are non-trivial. Instrument-rated pilots spent slightly less time too close to the ground (<500' AGL, r_{pb}= -.281)—one indicator of potential hazard. Higher flight-hour pilots also spent less time too close to the ground (r_s = -.289) and tended to stay farther away from ABQ (r_s = .303), reflecting an inclination to divert before completing the flight.

Finally, four significant *and* meaningful correlations (dark gray, boldface) show older pilots tending to have shorter flights (r_s = -.423, .422, respectively). Effect size (r^2) was about 18%. Ground clearance was also better-maintained on shorter flights (r_s = .379, -.3 84, respectively), capturing the flight scenario's tendency to "squeeze" pilots between clouds and terrain near ABQ.

Takeoff hesitancy

Pilots were told that the best way to give good flight data was to treat this mission as if it were a real flight. Given those instructions, 12 of the 50 pilots initially stated that, having to fly this mission VFR, they would choose to not even take off. This was perhaps predictable, given the weather plus the fact of being scrutinized by FAA officials at an FAA facility. Therefore, to overcome any reservations they might have about being scrutinized, pilots who hesitated taking off were explicitly asked to do so and fly at least briefly. All complied.

Locality of pilot residence had no significant effect on takeoff hesitancy—18% of local (Oklahoma) pilots hesitated versus 32% of non-local (non-Oklahoma) pilots (2-tailed p_X2 = .251, NS). Neither did instrument rating predict hesitancy (15% for instrument rated v. 33% for non-instrument rated, 2-tailed p_x^2 = .138, NS). Finally, despite the confidence often associated with experience, neither age nor flight hours seemed to affect hesitancy (2-tailed Mann-Whitney U, p_U = .146, .625 respectively, NS). Overall, the cause of takeoff hesitancy appeared mysterious.

Effect of the weather training products on takeoff hesitancy. Hesitancy could have been caused by the weather training products. Table 3 shows the numbers of pilots who initially hesitated versus values expected by chance (in parentheses). The Yates-corrected p_x^2 is .034, implying that the training groups differ. However, a statistical caveat clouded the results: Half the cells had expected frequencies < 5, violating the 20% convention. Given that caveat, *if* this were indeed a reliable effect, pairwise tests of odds-ratios implied that the unusual group was the Control, where 17 of 18 pilots showed no hesitancy to take off.

Table 3. Takeoff hesitancy.

		Trg Prod 1	Trg Prod 2	Control
Initial takeoff decision	Yes	11 (11.3)	9 (12.1)	17 (13.6)
	No	4 (3.7)	7 (3.9)	1 (4.4)
Pairwise odds-ratios, 1-tailed *p*		← .152 → (Trg Prod 1 vs Trg Prod 2)		
			← .004 → (Trg Prod 2 vs Control)	
		← .037 → (Trg Prod 1 vs Control)		

In other words, studying a weather training product may have made pilots more hesitant to take off into deteriorating weather. However, cognitive priming is an alternate hypothesis, and will be revisited in the Discussion section.

Effect of takeoff hesitancy on subsequent flight safety. The 12 hesitators did not fly demonstrably safer than the remaining 38 pilots. There were no significant differences between hesitators and non-hesitators for minutes spent in IMC, minutes scud running, or minutes < 500' AGL (2-tailed Mann-Whitney p_U = .102, .147, .498 respectively, all NS). However, hesitators did seem to continue their conservatism into their flight, making significantly briefer flights (p_U = .002), with consequently less penetration into the marginal weather close to ABQ (p_U < .001).

Net effect of the weather training products on subsequent flight safety. Did viewing a weather training product affect flight safety? Some signs point to yes, some to no.

The Control group showed significantly less takeoff hesitancy. It also displayed greater flight duration and, consequently, lower minimum distance to ABQ (Kruskal-Wallis p_{KW} = .007, .005 respectively). Follow-up pairwise Mann-Whitney U tests implied that the Control group was significantly different from both weather training products ($p_{U\text{-}TRG1\ x\ CONTROL}$ = .011, .004 respectively and $p_{U\text{-}TRG2\ x\ CONTROL}$ = .004, .005 respectively), although the two weather products themselves did not significantly differ (p_U = .867, 1.0 respectively, NS).

Now—because the maximum hazard of this flight lay near the destination—we might be tempted to conclude that the longer flights of the Control group should predict greater risk exposure. However, no significant overall differences were found between the three training groups for subsequent minutes spent in IMC, minutes scud running, or minutes < 500' AGL (p_{KW} = .245, .158, .812 respectively, all NS). Even though the Control group showed less hesitancy and longer flight duration, and even though longer flight duration correlated significantly with minutes < 500' AGL, the net effect of the weather training videos on subsequent flight safety seemed nonsignificant.

So, how can there be no significant differences in flight safety between the three training groups? If seeing the weather training video related to takeoff hesitancy, and takeoff hesitancy related to flight duration, and flight duration related to minutes spent < 500' AGL—how could weather video not relate to minutes spent < 500' AGL?

The answer lies in the nature of causation versus correlation. If each factor perfectly *caused* the next factor in the chain, then the first factor would perfectly predict the last. In symbolic logic, A $\Rightarrow$ B (A implies B), and so on, so

A ⇒ B ⇒ C ⇒ D, therefore A ⇒ D. This is easy to see in a Venn diagram (Figure 2a). But, if each factor only *partially predicts* the next factor, then the overall relational strength between the first and last factors can theoretically be zero (Figure 2).

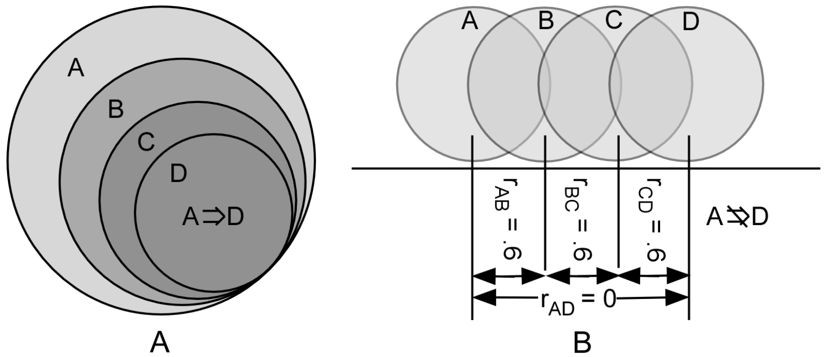


Figure 2. a) Venn diagram embodying causation A ⇒ B ⇒ C ⇒ D; b) Venn diagram embodying correlation A r_{AB} B r_{BC} C r_{CD} D.

Modeling the effect of training videos on flight behavior

When simple correlational models fail to explain effects, we turn to multivariate modeling. Here, cluster analysis and binomial logistic regression were used to generate models capable of explaining these pilots' flight behavior. Specifically, we wanted to predict if pilots would risk flying completely through the deteriorating weather (DV = *To ABQ* = 1/Yes or 0/No). Table 4 summarizes the smallest set of variables capable of doing that reliably.

Table 4. Binary logistic regression for To ABQ

	B	$p_{if\ term\ removed}$
Age	- 0.081	.002
TO decision	-21.20	.016
Control		.006
Trg Prod 1	- 3.08	
Trg Prod 2	- 2.53	
Constant	4.64	
Nagelkerke R^2 = .594		

Here, *Takeoff Decision* reflects "takeoff hesitancy," as discussed earlier. The training product is broken out into its three groups. Negative B-weights mean that a *positive* value for the independent variable subsequently related to a *reduced* groupwise tendency to fly all the way to ABQ. For example, pilots hesitant to take off (TO Decision = 1) subsequently showed a reduced tendency to fly all the way to ABQ. Similarly, pilots receiving either weather training product subsequently showed reduced tendency to fly all the way to ABQ, compared to the Control group.

In practical terms, this is a moderately strong model, accounting for 64.0% of the explainable variance in the data. It implies that pilot experience (flight hours) may work *in combination with* an instinctive reaction to a weather situation and a training video to affect ultimate continuation into adverse weather. This elaborates somewhat on the conclusion reached earlier about training product, so we will revisit that theme in the Discussion section.

Table 5 compares the prediction success rate for completed flight to ABQ made by logistic regression (boldface) versus cluster analysis (italics, in parentheses). Grey cells represent successful predictions.

Table 5. Success rate for **binary logistic regression** versus *(cluster analysis)*

Observed *To ABQ*	Predicted *To ABQ*		
	Did not make it to ABQ	Made it to ABQ	% correct
Did not make it to ABQ	**26** *(27)*	**4** *(4)*	86.7 *(87.1)*
Made it to ABQ	**4** *(0)*	**14** *(16)*	77.8 *(81.3)*
Overall % correct	Base logistic prediction rate = 62.5%		83.3 *(91.5)*

This shows that a simplified logistic model containing only pilot age, initial takeoff decision, and training product correctly predicted 83% of these pilots' overall decisions whether or not to fly through the deteriorating weather all the way to ABQ.

Overall, this three-variable model produced a gain of about 21% from the base rate predicted by a constant only (62.5%). Compare this to the eight-variable cluster model (not shown) correct predictions of 91.5%, versus a "complete" 15-variable logistic regression (not shown) where 100% of all cases were predicted correctly. However, note that the "complete" model was vastly overfitted, meaning it contained too many predictors, given the number of cases.

Discussion

Assessing the influence of a video weather training product on GA flight behavior turns out to be a subtle task. If we try to demonstrate statistically significant *direct* training product effects on hazardous-flight variables, we can claim none. However, if we examine *takeoff hesitancy* in the face of marginal cloud ceiling and visibility at the destination, we see greater average hesitancy in the two training product groups than in the control group. Pilots who hesitate tend to *continue* this conservativism into the flight, showing a greater tendency to divert to an alternate, rather than continuing on into deteriorating weather. So, there is a temptation to think that training product→takeoff hesitancy→shorter flight→lower risk exposure.

However, things are not quite that simple. First, training product does not directly correlate highly with hazardous flight variables. Second, takeoff hesitancy could reflect nothing more than artificial conservatism induced by the presence of FAA experimenters at an FAA testing facility. Pilots receiving the training video may merely have been primed to know that the experiment was about weather and may have simply given the experimenters the initial behavior they thought the experimenters wanted, namely, a conservative response to a weather situation.

The situation brightens when we model pilots who made it all the way through the weather versus those who diverted to an alternate. In that case, we can predict how about 80% of pilots will behave, based on nothing more than whether they received a training product, whether they hesitated to take off, and their total flight hours. This may imply that video weather training products "bring out the conservative" in some pilots, but less so in others.

However, we should stop short of making either extreme claim—that these products have either no effect, or some definite positive effect. In fact, the entire question is analogous to building a house. A single brick, no matter how well-crafted, will not suffice to build an entire house. In other words, weather is a complicated subject. No matter how good any given chapter is, we need to read the entire book.

Phase 2 of this study revisited the flight behavior of these same pilots after of a time lapse of several months. The data are currently being analyzed, to be reported shortly. If flight behavior of the experimental pilots regresses to the mean, then we can more strongly assume that cognitive priming was operating and that measuring hazardous weather flight *in simulo* is an even more challenging task than we already know it is.

Acknowledgments

The authors gratefully acknowledge Tammy Harris (FAA), and Janine King, Sally Glasgow, and Suzanne Thomas (Xyant Technology). Special thanks go to the pilots, without whom this study would have been impossible.

References

Bazargan, M. (2005). Analysis of national top 10 first causes of fatal accidents of general aviation between 1983 and 2002. Part 1-national-all aircraft types. (Technical Report 61022). Daytona Beach, FL: Embry-Riddle Aeronautical University, Center for General Aviation Research.

Bud, M.J., Mengert, P., Ransom, S.,& Stearns, M.D. (1997). General aviation accidents, 1983-1994: Identification of factors related to controlled flight into terrain (CFIT) accidents. (Technical Report DOT/FAA/AAR 100-97-2). Washington, DC: Federal Aviation Administration, Office of Aviation Research.

NTSB. (1989). Safety report: General aviation accidents involving visual flight rules flight into instrument meteorological conditions (Technical Report NTSB/SR-89/01). Washington, DC: National Transportation Safety Board.

NTSB. (2005). Risk factors associated with weather-related general aviation accidents (Technical Report NTSB/SS-05/01). Washington, DC: National Transportation Safety Board.

von Glasersfeld, E. (1995). Radical constructivism: A way of knowing and learning. London: Falmer Press.

Waltz, D. & Feldman, J.A. (1988). Connectionist models and their implications. Readings from cognitive science. Norwood, NJ: Ablex.

CONSTRUCTING ACCURATE AND PRECISE TIMELINES FOR MAJOR AVIATION ACCIDENT INVESTIGATIONS

Aaron S. Dietz
John J. O'Callaghan
Bruce G. Coury
Joseph M. Kolly
National Transportation Safety Board
Washington, D.C.

A clear, precise, and accepted description of what happened in an accident is a necessary first step in understanding why an accident happened. Although timelines are routinely used in accident investigations, constructing an accurate and precise one can be difficult. Large volumes of information must be correlated to a common time base, and the significance of events can change as the investigation develops. This paper describes the development of a timeline application to help overcome the difficulties associated with accident timelines. Development has emphasized interactive capabilities that allow users to manage the content and format how evidence related to the accident sequence is presented. The paper concludes with a discussion about how accident timelines can enhance communication and information access.

Timelines are routinely used in accident investigations to establish *what* happened in the accident, a necessary first step in determining *why* the accident happened. Their value lies in the identification of critical events, issues, and relevant evidence, especially in the early stages of the investigation. As the investigation develops and additional information is uncovered, more detail about the events and underlying conditions can be included on the timeline.

In addition, a timeline can be used to show the juxtaposition of events and underlying conditions that explain what happened in the accident. Recognizing the relevant relationships from events and information may point to causal and contributing factors and shape the direction of the investigation. In this way, accident timelines help bridge the gap between what happened in an accident and why it happened.

Despite such added value, constructing an accurate and precise depiction of critical events in a major aviation accident can be difficult. Investigators must correlate large volumes of information from numerous sources to a common time base, and the significance of particular events often changes as the investigation develops and new information becomes available. As a result, the selection of critical events from the complete set of available information and a meaningful presentation of those events can be a challenge. To overcome these challenges, the National Transportation Safety Board (NTSB) developed the Accident Critical Events Sequence (ACES) timeline application as a user-centered timeline application to support major aviation accident investigations.

The purpose of this paper is to describe the development and implementation of ACES, and show how it displays the sequence of events leading to an accident and gives investigators rapid access to related information. The paper begins by pointing to aspects of investigative activity that make constructing the sequence of events leading to an accident difficult. The discussion then turns to the motivation to develop ACES to overcome these challenges. The paper concludes with a discussion of effective areas of ACES implementation in ongoing investigations.

ACES is being developed as part of a larger NTSB effort to evaluate ways to improve the management of a major accident investigation. The Principal Issues Management Model (PIMM) being used by the NTSB focuses on managing principal issues, which are defined as significant aspects of an accident that directly relate to the factors underlying events and actions that occurred (Coury, et. al., 2008). Briefly, principal issues comprise the hypotheses or questions that the investigation must answer. Principal issues arise as the investigation progresses, and may require intensive efforts by multiple, interdependent investigative groups to gather evidence to answer questions raised by these issues. Because many of the questions associated with principal issues concern the chronological sequence of accident events, an accident timeline is essential. ACES is being developed to display critical events and related information, to provide a way to manage information from specific investigative tasks, and to communicate important time-related information to the entire investigative team.

The initial development of ACES drew upon other efforts to develop accident timelines. For instance, Events and Causal Factors Charting is employed by the United States Department of Energy to represent the multiple events and underlying conditions that contribute to the occurrence of an accident (DOE, 1999). The Transportation Safety Board of Canada uses a similar method—a Sequence of Events and Underlying Factors Diagram—to document the sequence of events leading to an accident (Ayeko, 2002). Finally, Sequential Timed Events Plotting (STEP) is an investigative methodology based on a multi-linear display that shows how events interact to produce an accident (Hendrick and Benner, 1987). Although ACES has some of the same characteristics as these other types of timelines, it is unique in its ability to depict, integrate, and display events and time-related data from multiple sources. The specific investigative challenges considered during the development of ACES are discussed in the next section.

ACES

NTSB has developed the ACES timeline application to help investigators depict and describe the sequence of events leading to an accident. Currently, ACES is a prototype built on Microsoft Excel 2003. Early development centered on establishing the functional requirements of the application based on the needs of the individual investigator and the investigative team. Updates and modifications to ACES relied on data collected through interviews with NTSB investigators, through usability testing, and through the observation of ongoing investigations to identify the specific investigative challenges that ACES should address, as described below.

First, NTSB investigators spend a significant amount of time and energy identifying what happened in an accident. This understanding forms the basis for determining the causal and contributing factors that explain why the accident happened and the actions necessary to prevent its recurrence. However, the management and analysis of information available to reach these conclusions can be overwhelming and presenting it in a way that is digestible can be difficult.

Second, because accurate and reliable timing is fundamental to a useful depiction of the sequence of events in an accident, careful correlation of all of the times used for data derived from event recorders [for example, the cockpit voice recorder (CVR) and flight data recorder (FDR)] is necessary. This correlation requires specialized knowledge and understanding of the timing involved in relevant systems.

Third, NTSB investigators focus on collecting evidence related to their fields of expertise, and there is a need for a centralized repository where diverse event-related evidence from each of the investigative groups can be displayed. Such an integrated depiction would help investigators identify issues requiring further investigation and help establish the relationship of events from different functional areas. In addition, the precision and relevance of time-stamped data can change over the course of the investigation, and these changes must be verified and communicated to the entire investigative team to ensure a shared understanding of accident events.

To overcome these challenges, ACES development has emphasized interactive capabilities that allow users to easily add, remove, and modify information to generate timelines that meet both individual and group needs:

- Users can customize how much detail is presented.
- Presentation options allow users to view events from different information sources that overlap, interact, or occur at the same time.
- Events and parameter data are color-coded so that different types of information can be easily distinguished from each other.
- External files from documents, pictures, and records can be linked to timeline events to provide access to more detail without cluttering the display.
- Finally, the synchronization of different time sources can be easily defined and updated.

The ways in which ACES manages the content and format of information presented to the user is expected to help overcome the challenges to constructing accident timelines and provide a mechanism to enhance communication and information access among interdependent investigative groups. ACES is also a repository where diverse, event-related evidence can be displayed in one place. The section below describes how accident data are used to generate an accident timeline on the ACES Graphical Display.

ACES Graphical Display

ACES works with text-based event data and numeric parameter data. Users enter these data types on individual worksheets within an Excel workbook that have been designated for that information source. Ultimately, these data are integrated into the accident timeline on the ACES Graphical Display as vertical text boxes and time-history plots, respectively. An example of the ACES Graphical Display is presented in Figure 1.

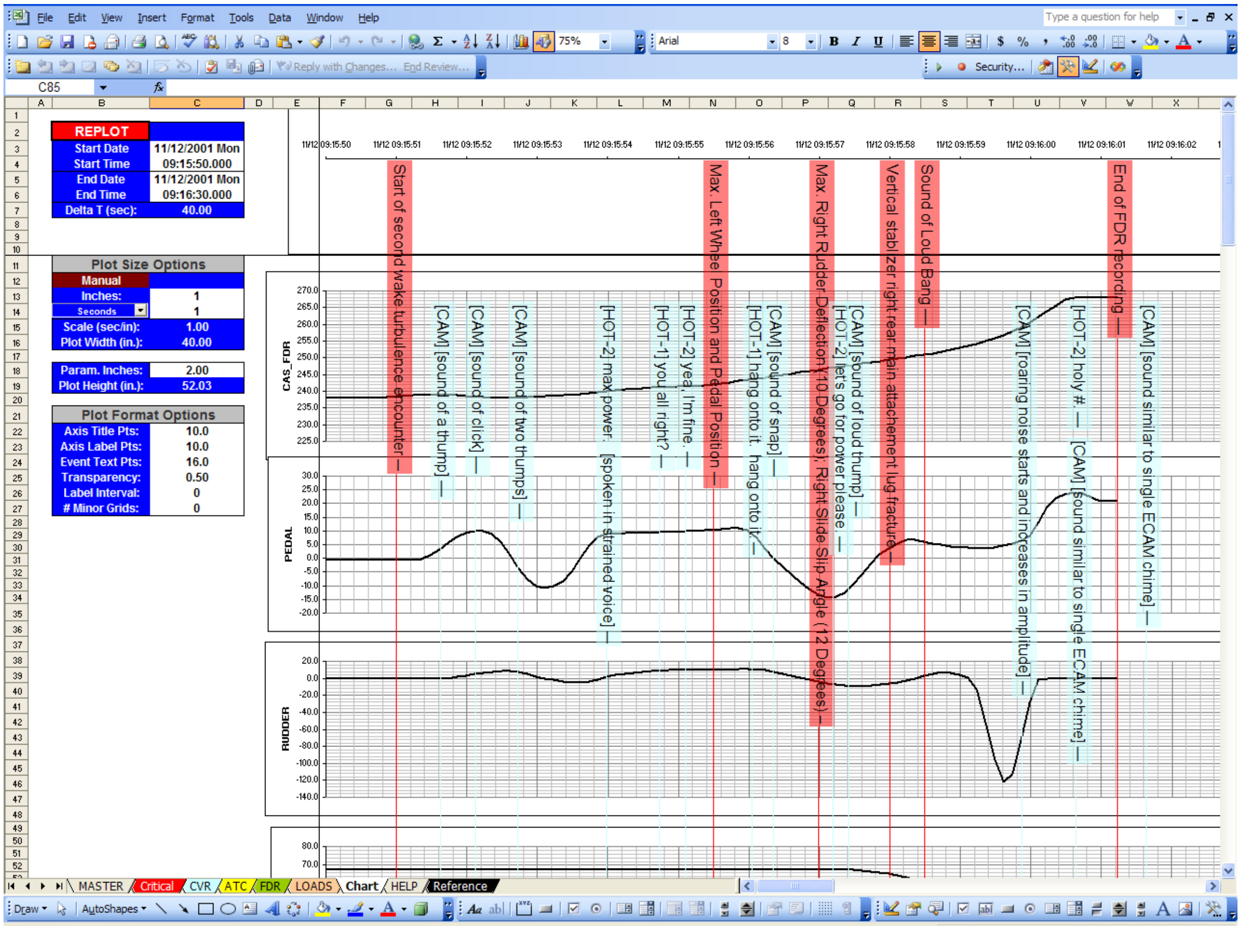


Figure 1. Example of ACES Graphical Display illustrating events from the NTSB investigation of American Airlines Flight 587 (NTSB, 2004).

The horizontal axis represents the master time for the investigation. Time runs from left to right and text-based events are printed vertically underneath the times when they occurred. Optional time-history plots of parameter data appear on the vertical axis and are intersected by lines dropping from the times corresponding to the text-based events. To the left of the timeline is the user interface, where investigators can manipulate the displayed time range, the scale of the horizontal axis relative to physical screen space, and other display settings. Finally, the sheet tabs located at the bottom of the screen allow users to navigate between each information source, a master sheet that amalgamates all of this information in a tabular format, and the ACES Graphical Display.

The information used to generate the accident timeline presented above was derived from the air traffic control (ATC) transcript, CVR, and FDR and was correlated to a common time-base. Additional information from weather reporting facilities, pre-flight maintenance logs, dispatch logs from emergency responders, training records, witness interviews, etc., can be incorporated on the accident timeline as well. ACES' ability to manage this diverse event-related evidence is described below.

Effective Areas

The evaluation of ACES during several ongoing aviation accident investigations indicates it is an effective investigative resource. ACES has been found to be most useful in three areas:

1. Documentation and illustration of *what* happened in an accident
2. Support of collaborative investigative decision-making and problem-solving
3. Resolution of time discrepancies from multiple time-stamped data sources

The first area is critical for any accident investigation. A clear, accurate, and accepted description of *what* happened in an accident is a necessary first step in understanding *why* the accident happened. For example, the identification of an event may prompt accident investigators to recognize the relationships among other events in the accident sequence, support conclusions made about other issues, or ask new questions that otherwise may have been delayed or overlooked. ACES effectively documents, catalogs, and illustrates what happened in an accident.

The second area results from the complexity of an accident investigation and the need for input from many individuals, representing different areas of expertise, to find solutions to problems and make sound decisions. For instance, determining the configuration of an aircraft during landing may require evidence from the Operations Group to determine if the aircrew configured the airplane properly, evidence from the Vehicle Performance Group to determine the airplane's behavior, and evidence from the Human Performance Group to determine the effect of task complexity on crew resource management. This example highlights the interdependencies between investigative groups and underscores the importance of providing investigators with rapid access to evidence related to critical events at any point in the investigative process. ACES provides the capability for diverse event-related evidence to be displayed in one place and manipulated so that investigators can see the relevant relationships.

The third area relates to the synchronization of time-stamped data sources. The time bases underlying information from event recorders, radar data, witness statements, and other sources of time-related data are generally not synchronized and can vary in accuracy. However, building a precise depiction about what happened in an accident depends on the accurate placement of events in relation to one another. Consequently, synchronizing time-stamped data from multiple sources is of paramount importance. An accident timeline provides a mechanism for merging all the "clocks" from different information sources and synchronizing them to a master time. ACES performs this synchronization and presents an integrated timeline of events referenced to a common master time.

It is also worth mentioning that the initial development and implementation of ACES assumed that the application would be centrally managed by a single individual, with investigators working with that person to obtain necessary data plots and timelines. Development of ACES has changed course as a result of the ongoing evaluation to move the application in the direction of a stand-alone tool that can be used by investigators to create their own data plots and timelines. For instance, a user's manual and training modules were developed to accompany and provide guidance in the use of the application.

Conclusion

ACES was developed to help organize, present, and communicate factual information relating to the accident sequence to the entire investigative team. ACES clearly conveys the sequence of events leading to an accident and enables investigators to customize the content and format of information to meet both individual and group needs. Currently, ACES allows users to select subsets of accident data and synchronize time-stamped data from different information sources.

ACES is a new approach for constructing accident timelines and its potential to support investigative activity as part of PIMM will continue to be evaluated. This paper has addressed specific investigative needs that must be considered when constructing an accident timeline and discussed the ways in which ACES has demonstrated its value as an investigative resource. Future research is planned to determine the steps necessary to fully integrate ACES into the accident investigation process.

Acknowledgements

The views expressed in this paper are not those of the National Transportation Safety Board and are solely the views of the authors.

References

Ayeko, M. (2002). Integrated Safety Investigation Methodology (ISIM) – Investigating for Risk Management. In the *Proceedings of the Workshop on the Investigation and Reporting of Incidents and Accidents*, Glasgow, Scotland, July 17-20, 2002.

Coury, B.G., Kolly, J.M., Gormley, E., and Dietz, A. (2008). The Central Role of Principal Issues in Aviation Accident Investigation. In the *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (Santa Monica, CA: The Human Factors and Ergonomics Society), 99-103.

Hendrick, K. and Benner, L., Jr. (1987). Investigating Accidents with STEP. Marcel Dekker, 1987.

National Transportation Safety Board Aircraft Accident Report Number AAR/04-04, 2004, *In-Flight Separation of Vertical Stabilizer American Airlines Flight 587, Airbus Industrie A300-605R, N14053, Belle Harbor, New York, November 12, 2001.*

United States Department of Energy (1999). DOE Workbook: Conducting Accident Investigations, Revision 2. May 1, 1999. U.S. Department of Energy, Washington D.C.

LOGISTIC REGRESSION ANALYSIS OF OPERATIONAL ERRORS AND ROUTINE OPERATIONS

Elaine M. Pfleiderer
Cheryl L. Scroggins
Carol A. Manning
FAA Civil Aerospace Medical Institute
Oklahoma City, Oklahoma

Two separate logistic regression analyses were conducted for low- and high-altitude sectors to determine whether a set of dynamic sector characteristics variables could reliably discriminate between operational error (OE) and routine operation (RO) traffic samples. Dynamic sector characteristics submitted as predictors were: Average Control Duration, Number of Handoffs, Number of Heading Changes, Number of Intersecting Flight Paths, Number of Point Outs, and Number of Transitioning Aircraft. In the low-altitude sector model, the Number of Intersecting Flight Paths, the Number of Point Outs, and the Number of Handoffs produced a 75% overall classification accuracy. In the high-altitude sector model, the Number of Intersecting Flight Paths, the Number of Heading Changes, the Number of Transitioning Aircraft, and Average Control Duration produced a 79% overall classification accuracy. Classification rates achieved through the use of the selected sector characteristics support the assumption that elements of the sector environment contribute to the occurrence of OEs.

A considerable amount of research has focused on the relationship between sector characteristics and controller workload or perceived complexity. However, relatively few studies have examined the relationship between sector characteristics and the occurrence of OEs. In many early studies of OE causal factors, examinations of sector characteristics were limited to purely theoretical relationships (e.g., Arad, 1964) or to traffic counts and altitude transitions of the involved aircraft (e.g., Schroeder, 1982; Spahn, 1977). Grossberg (1989) expanded on this by collecting ratings from 97 controllers and supervisors regarding various aspects of the sector environment in the Chicago Air Route Traffic Control Center (ARTCC). Rodgers, Mogford, and Mogford (1998) evaluated the relationship between sector characteristics and the incidence of OEs at the Atlanta ARTCC. In both the Grossberg (1989) and Rodgers, Mogford, and Mogford (1998) studies, sector characteristics were evaluated without comparison with routine operations (ROs). Yet, for every OE that occurs in a sector, there are hundreds (possibly thousands) of hours in which an OE did not occur. Variables that correlate with sector OE frequency do not describe what was different about the sector at the time of the OE. To truly understand the environmental and contextual factors that contribute to OEs, it is necessary to identify what was different about the sector environment at the time the OE occurred.

Pfleiderer and Manning (2007) conducted an investigation to determine whether logistic regression analysis of objective sector characteristics could discriminate between OE and RO traffic samples. Two separate logistic regression analyses were performed for low- and high-altitude sector samples at the Indianapolis ARTCC (ZID). In the low-altitude sector sample, variables included in the final model were the Number of Point Outs, the Number of Handoffs, and the Number of Heading Changes. This model was able to accurately classify 79% of the low-altitude OE and RO traffic samples. In the high-altitude sector sample, a logistic regression model comprising the Number of Heading Changes, the Number of Transitioning Aircraft, and Average Control Duration was able to accurately classify 80% of the OE and RO traffic samples. Unfortunately, the study was flawed. Available traffic data consisted of OEs from 9/17/2001 to 12/10/2003 and ROs from 2/25/2005 to 3/3/2005. Clearly, the time differential between the OE and RO traffic samples was a confounding influence because it represented an uncontrolled, systematic difference between the two groups. A second problem with the design involved pairing OE and RO traffic samples (by sector, day of week, and time of day). Logistic regression analysis assumes that all cases are independent of one another. Only random selection of RO traffic samples would have guaranteed independence.

In the present study, OE and RO traffic samples are again compared using logistic regression analysis, but some important modifications were made to the design. OEs occurring in ZID airspace between 2001 and 2003 were compared with RO traffic samples from 2003, thereby reducing the time differential between the OE and RO groups. No attempt was made to match the RO traffic samples to the OE samples, thus meeting the assumption of independence.

Separate logistic regression analyses were conducted for the low- and high-altitude sector samples because there was reason to suspect they represent heterogeneous sub-samples (Pfleiderer & Manning, 2007). Therefore, combining sector strata would probably produce a model that fit the high-altitude sectors poorly and the low-altitude sectors not at all. Predictor variables were restricted to dynamic sector characteristics. The variance of static variables would be seriously limited because multiple OEs occurred in many of the same sectors in the sample. Consequently, even if static sector characteristics were related to OEs, it is unlikely this relationship would be detected. The dynamic sector characteristics variables Average Control Duration, Number of Handoffs, Number of Heading Changes, Number of Intersecting Flight Paths, Number of Point Outs, and Number of Transitioning Aircraft (described in detail in the Method section) were submitted to logistic regression analysis to determine the degree to which they could discriminate between OE and RO traffic samples.

Method

Traffic Samples

All traffic samples were initially derived from System Analysis Recordings (SAR) generated by en route Host Computer Systems. The Host features data reduction programs that generate text reports of selected subsets of SAR data. The information used to calculate the predictor variables was extracted from reports produced by one such program, the Data Analysis and Reduction Tool (DART).

OE traffic samples were derived from reconfigured DART information from Systematic Air Traffic Operations Research Initiative (SATORI) files. SAR data require a prohibitive amount of storage space. SATORI re-creations require less space and so these files are often the only traffic data saved after an OE. Therefore, the primary constraint on the size and range of the data set was the availability of SATORI re-creations. SATORI data meeting processing criteria (i.e., five minutes prior to the initial loss of separation) were only available for 119 OEs occurring in the ZID airspace from 9/17/2001 through 12/10/2003. Of these, 40 occurred in low-altitude sectors and 79 occurred in the high-altitude sectors.

RO traffic samples were derived from ZID SAR data recorded on 5/8/2003 (15:55 to 17:05, 18:55 to 20:10, and 20:50 to 22:15 ZULU), 5/9/2003 (0:00 to 1:10 ZULU) and 5/10/2003 (11:20 to 12:40 ZULU). DART text reports were first encoded into database files and then processed in 5-minute intervals using custom software designed to calculate objective measures from routinely recorded NAS data. This produced a total of 2644 RO traffic samples. Of these, 992 occurred in low-altitude sectors and 1652 occurred in the high-altitude sectors.

The 40 low-altitude OE traffic samples were combined with 40 randomly-selected low-altitude RO traffic samples to produce a total of 80 traffic samples for the low-altitude sector analysis. The 79 high-altitude OE traffic samples were combined with 79 randomly-selected high-altitude RO traffic samples to produce a total of 158 traffic samples for the high-altitude sector analysis. The number of traffic samples in the RO and OE groups was kept equal because widely disparate group size produces logistic regression models that favor the largest group. Equal group size also ensures that classification accuracy in excess of 50% represents improvement over chance.

Predictor Variables

Average Control Duration. Aircraft control duration is influenced by a number of factors, including aircraft performance characteristics, Traffic Management Initiatives (TMI), and sector size – all of which have been associated with sector workload or complexity (Grossberg, 1989; Mogford, Murphy, & Guttman, 1994; Pfleiderer, Manning, & Goldman, 2007). Average Control Duration is the mean of the durations (in seconds) of all aircraft controlled by the sector within a processing interval. Control time occurring before or after the interval was not included in the calculations.

Number of Handoffs. Although traffic count remains the best single predictor of the number of OEs on a national level, previous research suggests that it is not an effective predictor of OEs at the sector level (Schroeder, 1982; Schroeder, Bailey, Pounds, & Manning, 2006; Spahn, 1977). Perhaps the biggest drawback to traffic count is that it tends to be highly correlated with other traffic-related measures, thereby creating redundancies that may overshadow more effective predictors. Handoffs are correlated with the number of aircraft in the sector, but may also capture elements of communication workload, coordination, and required procedures. The Number of Handoffs is the total number of handoff initiates and handoff accepts occurring within the 5-minute processing interval.

Number of Heading Changes. Heading changes have demonstrated a relationship with controller ratings of activity (e.g., Laudeman et al., 1998), workload (e.g., Stein, 1985), and complexity (e.g., Kopardekar & Magyarits, 2003). Heading changes are involved with a number of procedures such as merging and spacing, Standard Terminal Arrival Routes (STARs), Standard Instrument Departure Routes (SIDs), and holding. The Number of Heading Changes is a count of all turns in excess of 10° per 12-second radar update that continue in the same direction for at least three updates. Heading changes made in an attempt to avoid an imminent OE were excluded.

Number of Intersecting Flight Paths. This was one of the highest rated complexity factors in the high-altitude and super high-altitude sectors in the Pfleiderer, Manning, and Goldman (2007) study. A similar factor (several traffic flows converging at the same point) was highly rated in an investigation of Maastricht airspace conducted by Eurocontrol (2006). The Number of Intersecting Flight Paths is the maximum number of flight paths that might be expected to intersect, irrespective of altitude, within a 10-minute projected time frame given the aircraft's current speed and trajectory. Projections were calculated for every 12-second radar update within each minute of data. The length and slope of the projected paths were based on the distance and angle of the current and previous radar position coordinates.

Number of Point Outs. Point out entries represent one of the few instances in which coordination between sectors is recorded. The Number of Point Outs is the total number of point out entries made by the Radar and Radar Associate controllers during the 5-minute processing interval.

Number of Transitioning Aircraft. The amount of climbing and descending traffic has long been recognized as a contributor to the difficulty of working a sector (e.g., Arad, 1964; Grossberg, 1989; Kopardekar & Magyarits, 2003). The Number of Transitioning Aircraft represents the number of aircraft making one or more altitude changes during the 5-minute processing interval. To be counted as a change, altitude must increase or decrease by a minimum of 200 feet per 12-second radar update and must continue to change in the same direction for at least three updates. Altitude changes resulting from last-minute clearances made in an attempt to avoid an OE were excluded.

Results

Stepwise elimination was employed for the logistic regression analyses because such methods are extremely valuable in exploratory research. Backward elimination was used because it is less prone to omit useful variables, since all variables are in the model at the beginning of the process. The likelihood-ratio test, which compares the fit of the model with and without each predictor at every step, was the selection criterion because it is more rigorous than other methods ((Menard, 1995). A criterion level of .10 was used to ensure that all relevant variables are included in the logistic regression model.

Low-Altitude Sector Sample

Tolerance values were >.45 for all predictors, far in excess of the <.20 that would indicate multicollinearity in the low-altitude sector sample. The logistic regression model for the low-altitude sample generated a Model $X^2(3, N{=}80){=}23.82, p{<}.01$, indicating significantly improved prediction over the model with the constant only. The non-significant Hosmer-Lemeshow $X^2(8, N{=}80){=}1.61$, $p{=}.99$ suggests that the model fit the data well. Logistic regression coefficients (*B*), standard errors (*S.E.*), estimated odds ratios (Odds), and significance values for the likelihood-ratio tests for the low-altitude sector sample are provided in Table 1. Note that neither the logistic regression coefficients nor standard errors are inflated, indicating a sufficient ratio of cases to predictors.

Table 1. *Logistic Regression Summary: Low-Altitude Sector Sample (N = 80).*

	B	*S.E.*	Odds	*p*
Number of Intersecting Flight Paths	1.06	.36	2.89	.00
Number of Point Outs	.45	.24	1.57	.04
Number of Handoffs	.17	.11	1.19	.10
Constant	-1.94	.57	.14	

In the low-altitude sample model, the Number of Intersecting Flight Paths had the highest odds ratio (2.89), followed by the Number of Point Outs (1.57), and the Number of Handoffs (1.19). In other words, each intersecting flight path increased the likelihood that the traffic sample was an OE by 189%, each point out increased the likelihood by 57%, and each handoff increased OE likelihood by 19%. Classification accuracy in the low-altitude sample is shown in Table 2. Of the 40 ROs in the low-altitude sample, 32 (80%) were correctly classified and 8 (20%) were misclassified as OEs. Of the 40 OEs in the sample, 28 (70%) were correctly classified and 12 (30%) were misclassified as ROs. Overall, the low-altitude model had a 75% classification accuracy. This represents 25% improvement over prior probabilities (i.e., the number that would be correctly classified by chance).

Table 2. *Classification: Low-Altitude Sector Sample (N = 80).*

Observed	Predicted		
	Routine Operations	Operational Errors	Total
Routine Operations	32 (80%)	8 (20%)	40
Operational Errors	12 (30%)	28 (70%)	40

High-Altitude Sector Sample

As with the low-altitude sample, Tolerance values were high (.56 and above) for all predictors. The logistic regression model for the high-altitude sample generated a Model $X^2(4, N=158) = 73.01$, $p<.01$, indicating significantly improved prediction over the model with the constant only. The non-significant Hosmer-Lemeshow X^2 $(8, N=158) = 3.33$, $p=.91$ for the high-altitude sample verified that the model fit the data. Logistic regression coefficients (*B*), standard errors (*S.E.*), estimated odds ratios (Odds), and significance values for the likelihood-ratio tests for the high-altitude sector sample are provided in Table 3. Note that neither the logistic regression coefficients nor standard errors are inordinately large, indicating a sufficient ratio of cases to predictors.

Table 3. *Logistic Regression Summary: High-Altitude Sector Sample (N = 158).*

	B	*S.E.*	Odds	*p*
Number of Intersecting Flight Paths	.69	.28	2.00	.01
Number of Heading Changes	.31	.15	1.36	.03
Number of Transitioning Aircraft	.24	.11	1.27	.02
Average Control Duration	.01	.01	1.01	.01
Constant	-4.43	1.11	.01	

In the high-altitude sample model, the Number of Intersecting Flight Paths had the highest odds ratio (2.00), followed by the Number of Heading Changes (1.36), the Number of Transitioning Aircraft (1.27), and Average Control Duration (1.01). In other words, each one-unit increase in the Number of Intersecting Flight Paths increased the likelihood that a traffic sample was an OE by 100%, each one-unit increase in the Number of Heading Changes increased the likelihood by 36%, every Transitioning Aircraft increased the likelihood by 27%, and each one-second increase in Average Control Duration increased the likelihood by 1%. Classification accuracy in the high-altitude sample, shown in Table 4, was slightly better than that of the low-altitude sample. Of the 79 ROs in the high-altitude sample, 64 (81%) were correctly classified and 15 (19%) were misclassified as OEs. Of the 79 OEs in the sample, 60 (76%) were correctly classified and 19 (24%) were misclassified as ROs. Overall, the high-altitude model had 79% classification accuracy. This represents 29% improvement over prior probabilities (i.e., the number that would be correctly classified by chance).

Table 4. *Classification: High-Altitude Sector Sample (N = 158).*

Observed	Predicted: Routine Operations	Predicted: Operational Errors	Total
Routine Operations	64 (81%)	15 (19%)	79
Operational Errors	19 (24%)	60 (76%)	79

Discussion

The results of the logistic regression analyses indicate that a sufficient model may be constructed from sector characteristics variables. Overall classification accuracy between 75-79% is remarkable for models constructed solely of environmental and contextual factors. After all, other factors (e.g., human elements, organizational influences) also contribute to the occurrence of OEs. Unfortunately, all the logistic regression models were better at classifying ROs than OEs. Classification of OEs ranged from as low as 70% in the low-altitude sector sample, to 76% in the high-altitude sample.

Low-Altitude Sector Model

The most influential variable in the low-altitude sector model was the Number of Intersecting Flight Paths (Odds=2.89), followed by the Number of Point Outs (Odds=1.57), and the Number of Handoffs (Odds = 1.19). In Pfleiderer and Manning (2007), the most influential predictor was the Number of Point Outs (Odds=3.30), followed by the Number of Handoffs (Odds=1.54), and the Number of Heading Changes (Odds=1.49). The predictive strength of the Number of Point Outs and the Number of Handoffs in the Pfleiderer and Manning (2007) results suggested that coordination played a primary role in the development of OEs in the ZID low-altitude sectors. This impression was bolstered by the Pfleiderer et al. (2007) data, in which controllers and supervisors at ZID rated coordination as one of the primary sources of complexity in low-altitude sectors. Consequently, the emergence of the Number of Intersecting Flight Paths as the most influential predictor in the current low-altitude logistic regression model was surprising, because ratings for this complexity factor were moderate in the low-altitude sectors. The results of the logistic regression analysis suggest that coordination may be a contributing factor, but converging traffic patterns are of greater consequence.

High-Altitude Sector Model

The Number of Intersecting Flight Paths was the most influential predictor in the high-altitude sector model, followed by the Number of Heading Changes. This is consistent with previous research. Controllers and supervisors rated the Number of Intersecting Flight Paths as one of the most influential complexity factors in the high- and super high-altitude sectors (Pfleiderer et al., 2007), and the Number of Heading Changes received the highest beta weight in a linear multiple regression analysis of controller ratings of activity in four sectors at the Denver ARTCC. Laudeman et al. (1998) attributed the influence of heading changes to the “significant arrival traffic in all the sectors that were observed” (p. 7). Arrival and departure traffic complexity is generally considered to be a low-altitude phenomenon, but this perception may be inaccurate. In the present study, the Number of Heading Changes was only influential in the high-altitude sector model. The third most influential factor in the high-altitude logistic regression analysis was the Number of Transitioning Aircraft, which has long been recognized as a contributor to the difficulty of working a sector (e.g., Arad, 1964; Grossberg, 1989; Kopardekar & Magyarits, 2003). This finding is also consistent with Pfleiderer et al. (2007) in which the complexity factor Climbing and Descending Traffic received the highest complexity rating for the high-and super high-altitude sectors.

Future Research

Logistic regression cannot be used to directly identify causal factors (i.e., prediction is not the same as causation), but elements of the models reveal aspects of the sector environment that might be altered to reduce the number of OEs. For example, the combination of the Number of Point Outs and the Number of Handoffs in the low-altitude sector model may indicate that the location of sector boundaries increases coordination workload and complexity. On the other hand, the combination of the Number of Point Outs and the Number of Intersecting Flight Paths may point to problems with the orientation of traffic paths relative to those boundaries.

Because of the research that remains to be accomplished, these results must be viewed as preliminary. Multiple studies must be conducted at a number of facilities before such models might be viable for practical applications. Nevertheless, the methodology of comparing OE and RO traffic samples is promising. Continued investigations along these lines may highlight complexity factors that should be addressed to ensure that safety is maintained.

References

Arad, B.A. (1964, May). The control load and sector design. *Journal of Air Traffic Control,* 5, 12-31.

Eurocontrol (2006). *A complexity study of the Maastricht upper airspace centre.* (EEC Report No. 403). Brétigny-sur-Orge, France: Eurocontrol Experimental Centre.

Grossberg, M. (1989, April). Relation of sector complexity to operational errors. *Quarterly report of the Federal Aviation Administration's Office of Air Traffic Evaluations and Analysis.* Washington, DC: Federal Aviation Administration.

Kopardekar, P., & Magyarits, S. (2003, June), Measurement and prediction of dynamic density. *Proceedings of the 5th Eurocontrol/FAA Air Traffic Management Research and Development Conference*, Budapest, Hungary. Available online at: http://atm2003.eurocontrol.fr.

Laudeman, I.V., Shelden, S.G., Branstrom, R., & Brasil, C.L. (1998). *Dynamic density: An air traffic management metric.* (NAS/TM-1998-112226). Moffett Field, CA: NASA Ames.

Menard, S. (1995), *Applied logistic regression,* Sage University Papers Series on Quantitative Applications in the Social Sciences, No. 106. Thousand Oaks, CA: Sage.

Mogford, R.H., Murphy, E.D., & Guttman, J.A. (1994). Using knowledge exploration tools to study airspace complexity in air traffic control. *International Journal of Aviation Psychology, 4*(1), 29-45.

Pfleiderer, E.M., & Manning, C.A. (2007). *Prediction and classification of operational errors and routine operations using sector characteristics variables.* (Report No. DOT/FAA/AM-07/18). Washington, DC: FAA Office of Aerospace Medicine.

Pfleiderer, E.M., Manning, C.A., & Goldman, S.M. (2007). *Relationship of complexity factor ratings with operational errors*. (Report No. DOT/FAA/AM-07/11). Washington, DC: FAA Office of Aerospace Medicine.

Rodgers, M., Mogford, R., & Mogford, L. (1998). *The relationship of sector characteristics to operational errors* (Report No. DOT/FAA/AM-98/14). Washington, DC: FAA Office of Aviation Medicine.

Schroeder, D.J. (1982, October). The loss of prescribed separation between aircraft: How does it occur? *Proceedings of the Behavioral Objectives in Aviation Automated Systems Symposium, Society of Automotive Engineers (SAE), P-114*, 257-269.

Schroeder, D., Bailey, L., Pounds, J., & Manning, C. (2006). *A human factors review of the operational error literature.* (Report No. DOT/FAA/AM-06/21). Washington, DC: FAA Office of Aerospace Medicine.

Spahn, M.J. (1977). *The human element in air traffic control: Observations and analyses of the performance of controllers and supervisors in providing ATC separation services. Supplement 2: Analysis of the System Effectiveness Information System (SEIS) data base.* (Report No. MTR-7655, Supplement 2). McLean, VA: METREK.

Stein, E. (1985), *Air traffic controller workload: An examination of workload probe.* (Report No. DOT/FAA/CT-TN84/24). Atlantic City, NJ: FAA William J. Hughes Technical Center.

PROFESSIONALISM IN AIRLINE OPERATIONS?...AND ACCIDENT INVESTIGATION?

Captain Kevin M. Smith
United Air Lines (Ret)
Mesquite, NV
Ronald John Lofaro, PhD
Embry-Riddle Aeronautical University
Orange Beach, AL

The views and opinions expressed in this paper are solely those of the authors. They do not necessarily represent the positions or policies of any private, public or governmental organizations.

When we read the findings of NTSB report AAR-07/06, Southwest Airlines flight #1248, we felt transported to a parallel universe whose occupants seem to be lacking any ability to reason analytically. These "findings' seemed to turn logic on its head, were insufficient in scope and incorrect as to causation. This paper will analyze the SW accident using the ODM model; will show the deficiencies in the NTSB report and finally, a. Indicate how to design line oriented flight training (LOFT) scenarios that reflect actual operating conditions and are aircraft type-specific. b. Show where/how a separate DM crew training module should be placed in flight crew training. The result would be training that provides both instruction and simulator practice for all Captains in such a way as to make timely and accurate decisions, thus avoiding the very accident we are discussing and, truly meet the requirements of an Advanced Qualification Program (AQP) of pilot training and certification.

Complex systems behave counter—intuitively:
That is the plausible tends to be wrong.
____ *J. W. Forrester*

Purpose

We will be questioning and critical. However, we will go beyond somewhat facile critiques and raising questions that seem not to have been asked. We will offer solutions, some of which we believe should have been in place already. Before we begin: the 6 page limit for papers has resulted in some condensing of our original paper. We believe that this version of the paper still "fills the bill."

Why Now?

We have, as do all in aviation, a deep concern for safety. To that end, we have been heavily involved for over 15 years with early, middle and later CRM, LOFT and other flight crew human factors training and evaluation. In the early through mid-2000's, we were pleased to see the recognition of what we had said many times, beginning in 1993: *The pilot's main function and responsibility is that of risk manager* and that the pilot and crew's main functions were risk identification, assessment and mitigation. This recognition is true, at least, in military aviation with its Operational Risk Management (ORM); a checklist completed prior to launch, which can result in mission planning changes and even aborting the mission. However, there is one problem with this proactive approach: it does not provide for changing conditions and factors aloft that can result in a rising risk after launch. The entire purpose of risk identification, assessment and mitigation is to enable the pilot to make the most timely, and accurate decision, in real time, in a time-compressed and unforgiving environment. More later on this.

The Tipping Point

When we read or observe something so outrageous and devoid of logic, we think we have somehow been transported to a parallel universe whose occupants seem to be lacking any ability to think and reason in a coherent way. Such was the case when we read the bizarre findings of the NTSB report AAR-07/06. We remain completely amazed and puzzled that pilots and other aviation experts have not risen as a group and demanded this report be revised…or, better, almost completely re-done in order to have congruence with the reality and facts of the accident. This, as yet, not having occurred, we now feel constrained to explicate our objections and recommendations to prevent another accident of this type…as well as other accidents where risk identification/assessment and the decision-making needed to deal with high risk are involved. In aviation, accurate operational decisions must be made, often with incomplete *or conflicting* information, in a time-compressed environment that is unforgiving of error. (Smith, Hastie, 1992). In this case, all the information needed was available and still incorrect decisions were made.

The Accident

On December 8, 2005, a Southwest Airlines flight #1248, attempted a landing at Midway International airport in adverse conditions, rolled through a blast fence, an airport perimeter fence, and onto an adjacent roadway striking a passenger automobile. One innocent bystander's (outside the airport perimeter) life was lost, people on-board seriously injured, and property destroyed. The weather at the time of the accident was such that only the most carefully flown aircraft had even the slightest chance of landing safely at this airport and that prospect was rapidly fading when the plane was well out of the approach phase to Midway.

The NTSB inexplicably determined that the probable cause was the pilot's failure to stop the airplane on the runway (under conditions that would almost guarantee that this could not even be possible). So, the accident happened because the pilot did not complete the landing within the confines of the runway. What an enlightening piece of information; one supposes that, under this logic, when a CFIT happens, the cause is that the plane hit the ground. While superficially plausible to an uninformed observer, this NTSB finding is manifestly wrong. As an oblique afterthought, the NTSB made some reference to the fact that perhaps a diversion to another airport was in order. These findings turn logic on its head; they imply that it is somehow perfectly acceptable if an airline sends flights from both coasts to the Midwest in wintertime, into a snowstorm, and then "hope for the best." Thus, the NTSB tosses the accumulated of knowledge of more than 50 years of flying military and civil aircraft out the window.

A Re-Look at the Conditions and Events

a. Weather and Adverse Conditions.

Low visibility and falling snow were compounded by the fact that the runway was slippery and **braking action advisories were in effect.** Since there are only two types of advisories issued to pilots and dispatchers, (wind shear and braking action) we ask this question: what are these good for? Should they be seriously considered in the Mission Planning phase, or since it is not "illegal" to operate under these conditions, do we relegate them to the dustbin? Further, at the time of the accident, an 8 knot tailwind did exist…beyond acceptable limits…and, when coupled with the poor braking, a recipe for disaster. Were these devastating facts beyond the science of weather prediction? We think not. Many of the unsung heroes of aviation are the meteorologists, and their predictive accuracy in our 40 years of aviation experience is, to say the least, exceptional. Armed with the latest forecast of wind and weather, we wonder why this flight was even attempted…and, even more telling: why was there not an in-flight diversion to a designated, alternate field, or, as a last resort, a rejected landing?

b. Dispatch and Mission Planning.
Part 121 Carriers are not permitted to launch airplanes into the wild blue yonder any time they feel like it, but must comply with Dispatch protocol and constraints. Some conditions make successful completion of some flights so improbable that they cannot be attempted. For example, if the forecast weather at the intended destination is below landing minimums, the flight should not and cannot be flown… period! A meaningful conference call between Dispatch, the Captain, and the weather expert would have resulted in the decision that *attempting to land in low visibility conditions, with breaking action advisories in effect, with a tail wind on a short runway and with no real overrun is inadvisable.* Why did not the captain abandon the approach, reject the landing and proceed to the alternate?

c. The Decisions of Both the Pilot and Dispatch.
In 1993, we proposed the concept of "Pilot as Risk Manager", and further proposed that this is the quintessential Captain's activity, super ordinate to all others. We developed an operational decision-making model, to be used in flight. The end-result of using our ODM paradigm is taking the actions needed to keep the risk level low: 1. If the risk is low, continue with the mission plan. 2. If the risk is moderate, modify the mission plan in order to prevent the risk from rising. 3. If the risk is high, abandon the mission plan and/or cancel the mission (Smith, Lofaro 2003, Smith, Lofaro 2001; Lofaro, Smith 2008, 2000, 1999, 1998, and 1993). The question remains: why did the Captain continue?

d. A Brief Look at ODM
In order to deal effectively with the challenges of AQP, we developed an Operational Decision Making (ODM) paradigm. It deals with risk: Its identification, quantification and management as the basis for making decisions in the operational environment…decisions based on a precise and accurate understanding of risk. All flying can be visualized as operating in a four-sided figure, the operational envelope, where the sides consist of the critical factors to safe flight. The actual envelope is 3-D as the plane can fly in any direction as well as climb or descend. The continuous task of the pilot is risk identification and location by using situational knowledge. Situation(al) knowledge is that part of the ODM consisting of the continually changing set of elements (knowledge bits) that comprise the Captain's awareness of (a) the area of the ops envelope where the captain perceives the aircraft is located and (b) which of the critical factors of the ops envelope boundaries are in play. In this way, the pilot can ascertain what is the cumulative effect of these factors and thus, **re-locate** the aircraft's actual position in the ops envelope. The pilot can then use the rising risk scale as a decision for action responses. See Figure 1

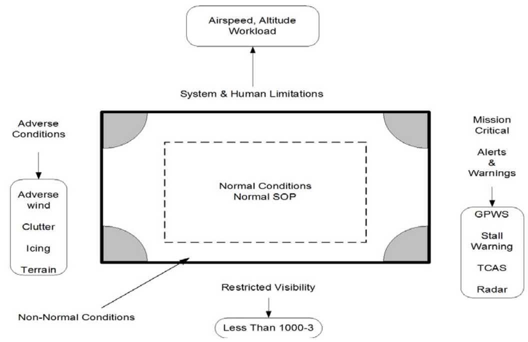


Figure 1.
The Operational Envelope
Note: Some examples of critical factors that are parts of the sides are shown.

To return to the accident: The Captain of flight 1248 was clearly faced with a rapidly rising risk he either did not understand or was constrained in his decision by other, non-safety of flight, issues. But, rising risk is non-linear. While we tend to think of one thing at a time, like wind or visibility, the reality is a cumulative effect can, and often does, occur where the real impact of the conditions taken together result in a much higher risk than the conditions taken as discrete events. Take, for, example low visibility operations. The Captain was faced with about a 200 foot ceiling and ½ mile visibility, close to CAT 1 minimums. But he also had to deal with unfavorable wind conditions, and runway contamination reducing breaking effectiveness *on a short runway*. When the factors are combined, the risk trajectory moved from moderate to high risk. Clearly, the Captain had an aircraft that was outside normal conditions and into non-normal operating conditions (again, see *Figure 1)*…and, therefore at high risk. Using the concepts of Operational Decision Making, we assert here that the Captain should have abandoned the approach and proceed to the assigned alternate. Let us emphasize this point: A landing should have never been attempted.

e. The Landing.
As we said before, the conditions were such that only the most perfectly executed landing could possibly have brought the aircraft to a safe stop on the existing runway. That this was not done is obvious. While the NTSB report highlighted this fact, it failed to acknowledge one of the most important rules of aviation: " Never ever get yourself in a situation where extraordinary piloting skills are required." Had the crew done so, a completely different result and report would have emerged.

What Do We Posit As The Real Cause?

The accurate "most probable cause" of this tragedy was the failure of the Captain to make a timely and accurate decision to abandon the approach and landing and proceed to the assigned alternate, given the deteriorating weather conditions, marginal breaking action and adverse wind conditions. This set of conditions and their cumulative effect, show the plane to be in a rapidly rising risk spiral, where the risk had gone to the highest level. In aviation systems, if decisive actions are not taken concerning critical events, risk will continue to rise to a point beyond which catastrophic mission failure results. Such a point is called the Critical Event Horizon (CEH); the Captain and his plane had passed through their CEH.

What of the NTSB finding that the fault was in the SW training that was either not received or not tested as to engine thrust reversers and auto brake systems? Not the real culprits as with the time/distances/speed needed before they would/could be deployed, they would have had little effect. However, we do not mean to say that SW did not (does not?) have training deficiencies. The Captain did lack some needed training, both in winter ops, in his 737 model…and, in risk identification, management and, most importantly, decision-making. These should be both SW and FAA concerns. More importantly, such training applies to all situations.

But, the real deficiency is somewhere else. Some history: In the early '90s a national task force was formed, which we were a part of, to develop the next generation Airline Pilot Training program. Originally promulgated by SFAR 58, it is called AQP. Guidelines for an AQP program were developed and all carriers were invited to design and implement such a state of the art curriculum. The major feature of AQP is to provide "mission realistic" training and evaluation, concentrating not only on flight maneuvers, but on higher order skills like decision-making needed in actual line operations. Indeed, Captain's authority, workload management, and decision-making are the three underpinnings of any successful AQP. (Captains Kevin Smith and Bill Hamman of United were key players in much of the United AQP R&D). These higher order skills can and must be taught and evaluated by any airline that wants to produce and maintain quality, trained pilots. As said, accurate risk location is the key, when in a (rising) risk situation, to making the optimal selection of a course of action, i.e., an action response that is an alternative to the original mission plan

Recommendations and Questions

The first one is that SW (and all carriers) develop and use LOFT's that reflect actual operating conditions and are aircraft type-specific. Event sets can be obtained from the carrier's accident/incident reports as well as using input from line pilots. A template for the development, and crew evaluation, of such LOFT scenarios can be found in the 14th Chapter ("Flight Simulators and Training"; Lofaro, R.J. and Smith, K.M., 2008) of *Human Factors in Training and Simulation*. This chapter also includes the Mission Performance Model (MPM) that is the template for specifying the critical components of flightcrew "effectiveness" (effective performance). Secondly, since the apex of crew responsibilities is decision-making, it should be made a separate crew training module, not a part of CRM. Plainly put: The requirements of an AQP must be properly implemented.

Finally, was the penultimate cause of this accident the failure, at the most senior levels of management, to develop, install and implement a quality, state of the art pilot training program: AQP? Such a program would provide both instruction and simulator practice for this Captain (and all other Captains) in how to exercise Captain's Authority in such a way as to make timely and accurate decisions and avoid the very accident we are discussing. The January, 2008 B-777 accident at Heathrow is another example of a decision-making process that boggles the mind and gives rise to the question: What is the type, quality and evaluation of training being received by airline aircrews?

References

Lofaro, R.J. and Smith, K.M. (1993). The Role of LOFT in CRM Integration in the *Proceedings of the Seventh International Symposium of Aviation Psychologists*. Columbus, OH: ISAP.

Lofaro, R.J. and Smith, K.M., (1998). Rising Risk? Rising Safety? The Millennium and Air Travel in *Transportation Law Journal, Vol.25, No.2 (entire issue).* Denver CO: University of Denver Press.

Lofaro, R.J. and Smith, K.M., (1999). Operational Decision-Making (ODM) and Risk Management (RM): Rising Risk, the Critical Mission Factors and Training in *Proceedings of Tenth International Symposium on Aviation Psychology.* Columbus, OH: ISAP.

Lofaro, R.J. and Smith, K.M., (2000) Operational Decision-Making: Integrating New Concepts into the Paradigm in *Proceedings of Eleventh International Symposium on Aviation Psychology*. Columbus, OH:

Lofaro, R.J. and Smith, K.M., (2008). Flight Simulators and Training. In D. A. Vincenzi, M. Mouloua, P. A. Hancock, and J. A. Wise (Eds.), *Human Factors in Training and Simulation*, (pp. 257-286). Philadelphia, PA: Taylor and Francis.

Smith, K.M. and Hastie, Reed, (1992). Airworthiness as a Design Strategy in *Proceedings of the FSI Air Safety Symposium.* San Diego, CA: FSI.

Smith, K.M. and Lofaro, (2001). A Paradigm for Developing Operational Decision-making (ODM) in *Proceedings of 2001 SAE World Aviation Congress (WAC) Conference*. Seattle, WA: SAE WAC

Smith, K.M. and Lofaro, R.J, (2003). The Finalized Paradigm for Operational Decision-Making (ODM) paradigm: Components and Placement in Proceedings *of the 12th International Symposium on Aviation psychology.* Dayton, OH: ISAP

A HUMAN-MACHINE INTERFACE FOR REPLANNING OF 4D TRAJECTORIES

Bart J. A. van Marwijk
Mark Mulder
Max Mulder
M. M. (René) van Paassen
Clark Borst
Department of Aerospace Engineering, Delft University of Technology
Delft, The Netherlands

To accomplish air traffic growth in a safe and efficient way, future air traffic management concepts require aircraft to accurately plan and execute 4D trajectories. A trajectory planned prior to takeoff, may, however, require in-flight revision. To support the flight crew in their task of accurately re-planning a flight plan up to a meter fix, in four dimensions, a dedicated planning interface has been designed. The interface allows direct manipulation of the ground track and the descent profile. Constraints on trajectory planning are mapped onto candidate waypoint locations, highlighting the possibilities for acceptable ground track geometry in the horizontal situation display. In the vertical situation display, these constraints are mapped onto candidate top and bottom of descent locations. It is hypothesized that the designed interface enables pilots to efficiently plan suitable 4D trajectories, while allowing for adaptive behavior and supporting situation awareness, even under high workload conditions.

To increase airspace capacity, future air traffic management (ATM) environments will not only require greater diversity and flexibility in the routes that can be flown, but also greater accuracy and timeliness with which aircraft adhere to these routes. This has major consequences for both ground-based ATM and airborne navigation planning. Indeed, in most of the proposed new ATM systems, the capability to accurately plan, implement, and execute a flight plan in four dimensions (4D), that is, in space and time, is a central ingredient (Swenson, Barhydt & Landis, 2006). The planning, guidance and navigation tasks of the flight crew will change when adhering to strict time constraints becomes of key importance.

Currently, airborne planning, implementation, and execution of a flight plan is automated with the help of the flight management system (FMS). Although the FMS has evolved at an exceptional rate in available features and functionality (Lidén, 1994), programming a flight plan still is a cumbersome task. The specification of the sequence of waypoints, flight levels, speed and time constraints, etc., needs to be entered alpha-numerically through the keypad of the command and control display unit (CDU).

The need for pilots to exploit the powerful functionality of the FMS quickly and accurately, in accordance with future ATM concepts, calls for a re-design of the navigation planning interface. This paper proposes a flight deck interface to the FMS, which allows for direct manipulation of the flight plan by the flight crew, during the task of airborne trajectory revision. In other words, the interface allows the crew to directly manipulate their flight plan in space and time (see also Kaber et. al, 2002; Winterberg, 2002; Vandenbussche, 2005; and Mulder, Winterberg, van Paassen & Mulder, 2009).

The design goals were threefold: 1. To find a suitable representation of constraints; 2. To support adaptive behavior of expert workers; and 3. To lower the required level of cognitive behavior for the trajectory revision task to skill and rule based behavior, allowing pilots to perform revisions under high workload conditions.

The proposed interface is designed in accordance with the principles of Ecological Interface Design (EID), see, for example, Rasmussen (1999), Borst, Suijkerbuijk, Mulder & van Paassen (2006), and van Dam, Mulder & van Paassen (2007). To find suitable representations of the work domain and task constraints, a cognitive work analysis of the airborne trajectory revision task was performed as part of the preliminary design phase, see, for example, Vicente (1999). Experienced pilots with backgrounds in commercial aviation and research were asked to provide input on the design work.

Cognitive Work Analysis of the Airborne Trajectory Revision Task

The process of cognitive work analysis (CWA) consists of five steps, which involve quite a diverse set of modeling methods. Vicente (1999) formulated these five steps as follows:

1. Work Domain Analysis – What are we working with? With what purpose?
2. Control Task Analysis – What must be done?
3. Strategies Analysis – How can it be done?
4. Social Organization and Cooperation Analysis – Who can best perform each (sub)task?
5. Worker Competencies Analysis – How can human actors be supported in their task?

Work Domain Analysis

System Boundaries. A detailed scenario of an aircraft requesting a trajectory modification would involve numerous interacting systems. However, this research focuses on supporting the crew during the (re-)planning task, not on subsequent interactions with for instance the air navigation service provider. The system considered for work domain analysis will therefore be limited to an aircraft flying in an airspace with obstructive elements, for example, adverse weather cells or restricted airspace.

Abstraction Hierarchy. The abstraction hierarchy (AH) uses different levels, from abstract to concrete, to describe the same system in terms of means and ends. The highest level of abstraction provides insight in the system's overall goals. The lower levels provide a more detailed representation of means and sub-goals. Four levels of abstraction were considered: 1. Functional purpose, i.e., what is the purpose of the work domain? 2. Abstract function, i.e., what are the underlying laws and principles? 3. Generalized function, i.e., what specific processes are involved? 4. Physical function. i.e., what tools are available to influence these processes?

Control Task Analysis

Decision ladder. The control task can be mapped as a sequence of subtasks (Vicente, 1999). The four dimensions of the trajectory to be defined are interdependent. For example, once a spatial trajectory has been defined, the time constraint at the meter fix, and the aircraft performance capabilities, limit the possibilities for temporal planning considerably. A distinction of the task in terms of temporal and spatial planning is therefore considered useful. Two decision ladders, with interactions, are shown in Figure 1.

Internal and external constraints. Assuming that the aircraft has the required navigation capability to execute user preferred routes in future ATM operations, the remaining internal constraints are: the flight envelope, the aircraft dynamics, and the fuel available. The constraints imposed on the aircraft are: first, obstructions of the flight path; second, operational regulations, and third, arrival requirements at a meter fix, which is typically near the destination and may designate the transition from user preferred routing airspace into airspace managed by air traffic control (ATC).

Strategies Analysis

Three experienced professional pilots, with backgrounds in civil aviation and research (see Table 1), were consulted for input on the interface design. After an introductory discussion of future ATM concepts and the implications on in-flight trajectory modification, the subjects were questioned on their preferences with regards to trajectory generation and interface content regarding a new FMS planning interface design.

Regarding trajectory alternatives in case of obstructions, three preferences were expressed: 1. To plan descents which are performed at constant throttle setting, 2. To separate speed changes from flight level changes, 3. To minimize fuel consumption resulting from the revision. Concerning the interaction with the automation, a preference for decoupled planning of the ground track and vertical profile was expressed. All pilots were in favor of direct manipulation of the flight plan geometry through a cursor control device. When asked to name display content that would be useful in performing a trajectory revision, the following answers were given: a visualization of time constraints at the meter fix, a preview of throttle settings and speeds per planned trajectory segment, a preview of the maximum rate of descent, an estimate of the fuel consumption corresponding to the modified trajectory, and an outline of the original trajectory during the editing process.

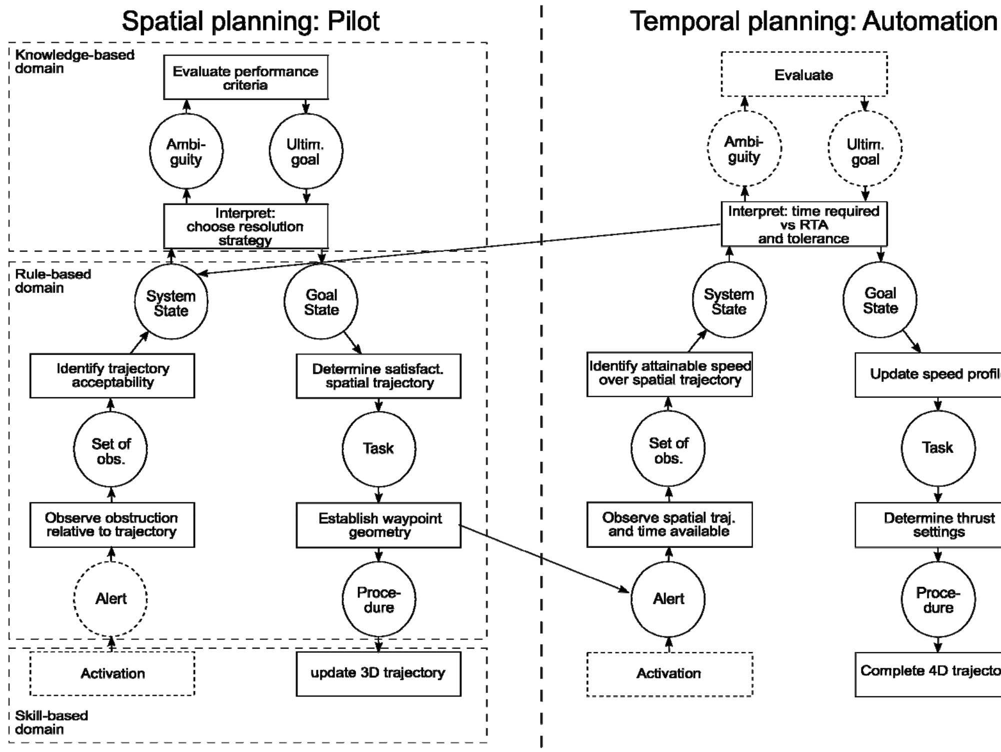


Figure 1. *Decision ladder of the 4D planning task, separated into spatial planning, which is allocated to the pilot, and temporal planning, which is allocated to the automation.*

Table 1. *Age, gender and experience of interview subjects.*

Pilot	Gender	Age	Flight Hours	Aircraft Types
A	Male	32	2,000	Cessna Citation II, Piper PA-31
B	Male	64	9,000	Boeing 737-200/300, 757-200, 767-300 ER
C	Male	69	14,000	Boeing 747-300/400

Social Organization and Cooperation Analysis

The spatial trajectory planning task is left to the pilot. The temporal planning of the trajectory is allocated to the automation. Once the pilot has defined the new spatial trajectory, the automation completes it by suggesting speed and altitude profiles that satisfy the 4D constraints at the meter fix, that, additionally, optimize fuel efficiency. The interactions between the automation and pilot decision ladders in Figure 1 illustrate how the outcome of spatial planning affects temporal planning, and vice versa. To allow the pilot to quickly make a well informed decision on the spatial resolution, the effect of his actions on adherence to constraints can be previewed in a spatial affordance zone, which is realized through automated pre-evaluation of numerous spatial trajectories (Mulder, Winterberg, van Paassen, & Mulder, accepted). Upon definition of a new spatial trajectory by the pilot, the automation adds a corresponding temporal plan, so as to ultimately obtain a complete 4D trajectory.

Worker Competencies Analysis

The purpose of the worker competencies analysis is to identify the level of cognitive behavior required to perform the tasks allocated to the human. Using the skills, rules, knowledge taxonomy (Rasmussen, 1983) as a qualitative framework for assessment, the hypothetical benefits of the proposed planning interface can be explained. Through the proposed automation support, the pilot's task of choosing and implementing a satisfactory resolution strategy is reduced to applying expertise to the signs on the display. His cognitive behavior is thus supported on the rule based level. Since establishment of the waypoint geometry is achieved through direct manipulation with a cursor control device, it is now best categorized as skill based behavior. If the complexity of a situation requires knowledge based behavior, the automated representation of constraints supports the pilot in his tasks of interpretation and decision making. The resulting demands on pilot cognitive behavior are hypothesized to allow effective in-flight re-planning even under high workload.

Interface Design

The interface design is a combination of conventional and novel display elements. There are several reasons for building on conventional displays, rather than designing 'from scratch'. First of all, the existing representations used for aircraft navigation, that is, the Horizontal and Vertical Situation Displays (HSD and VSD respectively), have already proven their value both in experiments (for example, Prevôt & Palmer, 2000) and practice. Second, it was not so much the navigation display that needed re-designing, but rather the way pilots could interact with it. Third, a practical advantage of extending conventional display functionality is that it may facilitate speedy implementation of the proposed re-planning interface in future systems. The proposed horizontal and vertical planning displays are shown in Figure 2.

Novel Display Elements.

Horizontal Situation Display. In the HSD, control action is equivalent to modification of the location of the selected waypoint. The constraints that bound the affordance zone, are the time available in which to reach waypoint FIX and the achievable ground speed range (Mulder *et. al.*, accepted). Modification of the trajectory by means of waypoint relocation will generally result in a different distance-to-fly to waypoint FIX. By adjustment of the speed profile to the trajectory length resulting from waypoint relocation, the estimated time of arrival (ETA) of the new trajectory can be made close or equal to the required time of arrival (RTA) of the contracted trajectory. Since there is a tolerance window around the RTA, candidate trajectories may be classified in three types, according to the smallest possible difference between ETA and RTA at waypoint Fix: 1. The ETA equals the RTA. Candidate locations that would result in this type of a trajectory are represented by the light shade of the affordance zone (see (1) in Figure 2). 2. The ETA lies within tolerances, but the RTA itself cannot be achieved. Waypoint locations corresponding to such trajectories are represented by the dark shade of the affordance zone (2) 3. The ETA is outside RTA tolerances. Corresponding waypoint locations are not part of the affordance zone, as the resulting trajectory would require renegotiation of a slot in the landing queue. The HSD additionally includes a representation of the speed profile selected by the automation (3), and the location of the top and bottom of descent (4).

Vertical Situation Display. Since manipulation of waypoints in the HSD only results in a (re)definition of the ground track, the VSD is used to facilitate modification of the vertical profile. After manipulations to the ground track, the automation will present the corresponding optimal vertical profile on the VSD by default. Analogous to the ground track, the vertical profile can be modified by manipulation of its nodes, which are the top and bottom of descent. The affordance zone in the VSD consists of a horizontal band (5), which appears when either the top or bottom of descent is selected, and highlights the alternative locations for the selected waypoint that would result in a feasible vertical profile. Placing the top or bottom of descent within this band ensures that the resulting trajectory is not too steep to allow for descent with constant ground speed. The second element of the vertical affordance zone is an outline of the descent envelope (6), which is bounded by the steepest descent from the earliest and latest top of descent location, and of the initial and final altitudes. To assist the pilot in evaluating the vertical profile, a numeric representation of the maximum vertical speed is included (7). Finally, since it is important for pilots to form an accurate mental picture of the 4D flight plan during evaluation and re-planning, the trajectory edit display reveals the time dimension of the trajectory means of ground speed targets, along with predicted throttle settings (8).

Showing this information is hypothesized to increase situation awareness and reduce the risk of the crew being surprised by, for example, automatically executed speed changes.

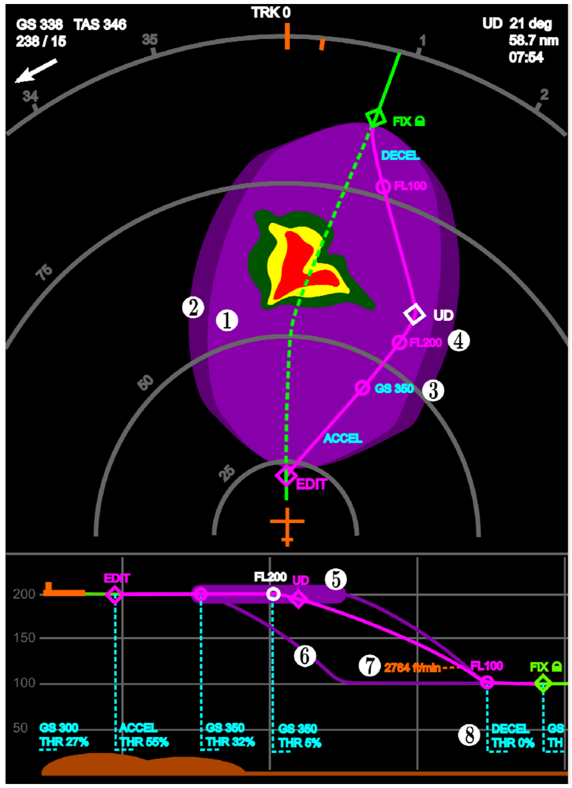


Figure 2. *The proposed horizontal and vertical navigation displays.*

Concluding Remarks

A new interface for modifying a 4D flight trajectory was introduced, which was designed using the principles of ecological interface design. The interface visualizes the constraints for the re-planning task to the pilot in a manner that is consistent with the constraints pilots have to take into consideration during this task (Mulder *et. al.*, accepted). It is expected that this interface will enable pilots to quickly generate alternative 4D trajectories when faced with the necessity of making route changes, allowing them to make efficient use of the powerful capabilities of the FMS. Furthermore, as was shown in the design rationale, the interface supports cognition on all levels of the skills, rules, and knowledge taxonomy of Rasmussen (1983). This display will be evaluated in a flight simulator in the near future.

References

Borst, C., Suijkerbuijk, H. C. H., Mulder, M., & van Paassen, M. M. (2006). Ecological Interface Design for Terrain Awareness. *The International Journal of Aviation Psychology*, 16(4), pp. 375-400.

van Dam, S. B. J., Mulder, M., & van Paassen, M. M. (2007). Ecological Interface Design of a Tactical Airborne Separation Assistance Tool. *IEEE Transactions on Systems, Man, Cybernetics, Part A*, 38(6), pp. 1221-1233.

Kaber, D. B., Riley, J. M., & Tan, K. W. (2002). Improved Usability of Aviation Automation Through Direct Manipulation and Graphical User Interface Design. *The International Journal of Aviation Psychology*, 12(2), pp. 153-178.

Lidén, S. (1994). The Evolution of Flight Management Systems. *Proceedings of the Digital Avionics Systems Conference*, AIAA/IEEE, Phoenix, AZ, USA, pp. 157-169.

Mulder, M., Winterberg, R. van Paassen, M. M., & Mulder, M. Direct Manipulation Interfaces for In-flight Four-dimensional Navigation Planning. *International Journal of Aviation Psychology* (accepted December 2008).

Vandenbussche, J. (2005). Direct Manipulation Interface for Airborne 4D Trajectory Planning. Unpublished MSc. Thesis, Department of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands.

Vicente, K. J. (1990). *Cognitive Work Analysis: Towards Safe, Productive, and Healthy Computer-Based Work*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Winterberg, R. (2002). Design and Evaluation of an Airborne Human-Machine Interface for Planning Four-Dimensional Trajectories. Unpublished MSc. Thesis, Department of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands.

Prevôt, T. and Palmer, E. (2000). Staying Ahead of the Automation: A Vertical Situation Display can Help. *SAE Technical paper 2000-01-5614*, 2000 World Aviation Conference, San Diego, CA.

Rasmussen, J. (1999). Ecological Interface Design for Reliable Human-Machine Systems. *The International Journal of Aviation Psychology*, 9(3), pp. 203-223.

Swenson, H., Barhydt, R., & Landis, M. (2006). Next Generation Air Transportation System (NGATS) – Air Traffic Management (ATM) – Airspace project. *Technical report*, Washington D.C.: National Aeronautics and Space Administration

Joint Planning and Development Office (2007). Concept of Operations for the Next Generation Air Transportation System (NextGen). *Technical Report*, Joint Planning and Development Office.

Performance Review Commission Eurocontrol (2008). An Assessment of Air Traffic Management in Europe during the Calendar Year 2007. *Technical Report PRR 2007*, Eurocontrol.

PERFORMANCE VISUALIZATION METHOD OF AIR TRAFFIC CONTROL TASKS FOR EDUCATIONAL PURPOSE WITH UTILIZING COGNITIVE SYSTEM SIMULATION

Daisuke Karikawa
Tohoku University
Sendai, JAPAN
Hisae Aoyama
Electronic Navigation Research Institute
Tokyo, JAPAN
Makoto Takahashi
Tohoku University
Sendai, JAPAN
Kazuo Furuta
The University of Tokyo
Tokyo, JAPAN

In the dynamic and multi-task condition of air traffic control, an Air Traffic Controller (ATCO) must utilize effective strategies to control traffic to prevent potential collision of aircraft and also to reduce his cognitive workload. Therefore, the strategy building skill of an ATCO is quite important for aviation safety and efficiency. In the present research, for supporting education of strategy building, a function which can visualize the difference of task performance has been implemented into the Air Traffic Controller Cognitive Simulation (ATCCS). Using this function, the effect of ATCO's control on air traffic flow has been successfully visualized, which helps trainee to understand the differences of the consequences of the different strategy. This result has strongly implied that ATCCS equipped with performance visualization function can be utilized as a supporting tool for education of ATCO trainees.

It is strongly required to achieve higher level of safety in aviation along with the rapid increase of the demand in air traffic recent years. The human factors in Air Traffic Control (ATC) area are one of the most important issues to be tackled for enhancing aviation safety.

The ATC tasks are characterized by multiple tasks under the time-pressure condition. ATCOs are sometimes required to control over 10 aircrafts which have multiple performance and different demands at the same time. It means that the task environment of ATC essentially involves potential causes of human errors such as cognitive overload and inappropriate attention allocation. However, our previous research of cognitive task analysis for ATCOs (Inoue, K. et.al. 2006, Inoue, S. et.al. 2005) has revealed that they have typical skill to develop effective air traffic strategies which can prevent potential conflict of aircraft well in advance and also can reduce their cognitive workload. Such strategy building skill of

ATCOs is definitely important for enhancing safety in the heavy traffic condition.

Based on such recognition, our research group has explored possible application of computer simulation as a supporting tool for acquiring strategy building skill in the basic training process of ATCO by visualizing the possible consequences of various task plans. In the ATC area, Fast Time Simulations (FTS) have already been utilized as an effective supporting tool for prediction of ATCO's workload and also for evaluation of airspace design. However, as conventional FTSs have mainly focused on generating discrete ATC events, their problem solving strategy tends to be somewhat different from that of human controllers in a specific situation. That is because the cognitive processes concerning the decision making by ATCOs have not been modeled properly in those conventional FTSs. Therefore, conventional FTSs have not been capable of being utilized for educational purposes. In our opinion, further elaboration of FTSs is definitely required in order to realize the realistic computer based simulation for the initial education of ATCOs.

In the present study, the cognitive system simulation including the detailed cognitive model of ATCO called Air Traffic Controller Cognitive Simulation (ATCCS) has been developed, which has been designed based on the results of cognitive task analyses of an ATCO performed with the help of ATCOs working regularly. The implementation of prototype supporting function for educational purpose and results of its preliminary evaluation are described in this paper.

Air Traffic Controller Cognitive Simulation (ATCCS)

Major Characteristics of ATCCS

The major characteristics of the proposed simulation framework are described in the following:

Uncertainty. According to the interviews with ATCOs conducted by our research group, uncertainty of an air situation is an important factor affecting ATCO's cognitive strategy and workload. In

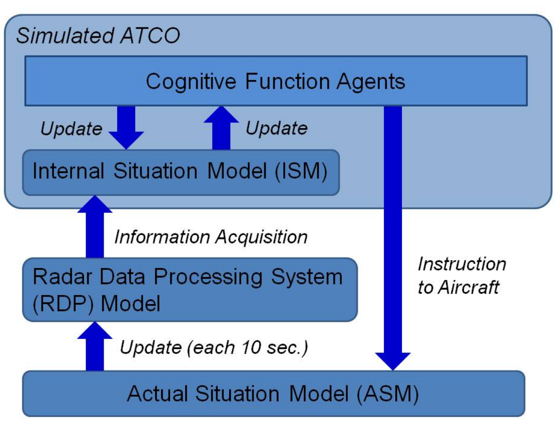


Fig.1 Basic Structure of ATCCS

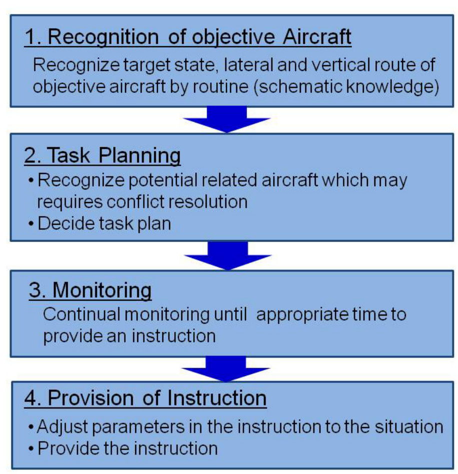


Fig.2 Cognitive Process of Simulated ATCO

our ATCCS, the uncertainly concerning the future situation of aircraft (e.g. future trajectory and flight path of aircrafts, time delay of pilot's reaction to ATCO's instruction) has been taken into consideration. The proposed ATCCS can simulate ATCOs' behavior when future situation cannot be determined exactly, which requires the extensive monitoring and the adjustment of strategies by the simulated ATCO according to the ongoing situation.

Bounded rationality. Cognitive activity of ATCO in the ATCCS is limited by multiple cognitive resources based on Wickens's theory (Wickens, C. D. et.al. 1984); they are visual, auditory, cognitive and motor resources. In addition, ATCO in the ATCCS has the Internal Situation Model (ISM), which is separated from the Actual Situation Model (ASM). The ISM represents ATCO's situation awareness, which may differ from the actual situation. In other words, it is ATCO's mental representation concerning task environment involving temporal and spatial aspects. On the other hands, the ASM represents situation of the actual world. ATCO's actions are determined based on the ISM which needs to be updated by information acquisition from the Radar Data Processing Unit (RDP) Model or predictions based on obtained external information and ATCO's inherent knowledge. This architecture realizes a simulation taking the model of bounded rationality into consideration. It also enables ATCCS to simulate the situation in which chain of errors occurs resulting from the discrepancy between ISM and ASM caused by erroneous recognition of a parameter and inappropriate attention allocation.

Schematic knowledge. Our previous research has revealed that ATCOs have schematic knowledge defined as "routine" involving dynamic descriptions of typical situations which can serve as a significant basis for comprehension and prediction of situations (Inoue, S. et.al. 2005). It has also indicated that routines involve the packages of heuristics to handle the situations effectively. The routines represented in the developed ATCCS provide necessary knowledge for developing three-dimensional flight image in objective sector based on the destination and route of the focused aircraft. The routine is also utilized to detect and recognize related aircrafts among the number of aircrafts in the sector.

Basic Cognitive Process of Simulated ATCO

Cognitive Process of Simulated ATCO has been designed based on a cognitive process model of ATCOs constructed with the help of ATCO in our group working regularly. Fig.2 shows simple overview of the cognitive process of the simulated ATCO. In the ATCCS, ATCO's cognitive functions are implemented as an assembly of various agents. Each agent has a specific cognitive function such as information acquisition from a radar screen, execution of communication with a pilot, storing a schematic knowledge, and so on. Those agents activate each other, and the activation levels of agents determine the overall behavior of the simulated ATCO.

Table 1 Definition of Task Levels (proposed by Aoyama, H. et al.)

Task Level	Situation	Display Color in ATCCS
1	No necessity of providing instruction without any requests from a pilot	Green
2	Typical tasks such as setting intrail, changing altitude are existed	Yellow
3	Typical tasks and conflict resolution tasks are existed	Orange
4	Task level 2 or 3 + Time pressured situation	Red

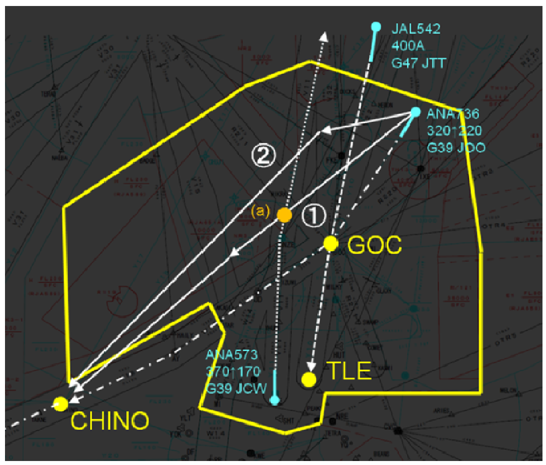


Fig.3 Scenario of Simulation Experiment

Visualization of Task Levels

For educational purposes, ATCCS has a visualization function of task levels which are defined by actual ATCO in our research group. The definition of task levels is shown in Table 1. The transition of task levels depends on the effectiveness of the applied strategy.

Implementation & Evaluation

The proposed ATCCS has been installed on the PC with using C++ language. For preliminary evaluation of proposed simulation, a numerical experiment has conducted based on scenarios in which simulated ATCO is required to provide descent clearance to specific aircraft with resolving conflicts among them so that each aircraft can accomplish its altitude target. In the simulation experiment, ATCCS could successfully simulate the typical behavior of human controllers in the similar situations (Karikawa et.al. 2008).

For evaluation of visualization function of task levels, additional simulation experiment based on the scenario described in Fig. 3 has been conducted. In this scenario, the simulated ATCO must accomplish the requirement of altitude target of JAL542, which is 13000 feet at TLE. Two departure aircrafts, ANA573 and ANA736 must be also controlled so that they can reach their cruises altitudes within this sector. The original flight planed route of ANA 736 is shown by dashed-dotted line in Fig 3. However, in this case, it is ineffective to follow the original planed route because it can lead to confliction between descending JAL 542 and climbing ANA 736 near GOC. Therefore, human ATCOs often reroute aircraft in order to resolve the conflict effectively in such a situation. In this simulation experiment, two possible control strategies were given by an actual ATCO instructor. The strategy 1 is making ANA736 shortcut to the prior fix directly. The strategy 2 is to lead ANA736 to west by radar vector and then providing instruction to direct to the prior fix after crossing. Both strategies are for resolving the confliction between JAL 542 and ANA736 by crossing both aircraft at earlier stage. The consequences of these strategies have been visualized and compared by using ATCCS (Fig.5).

The result of the simulation has shown in Fig. 4 ~ Fig. 7. In the Fig. 4 and Fig. 5, the task levels are overlaid on the trail of each aircraft with color code. The Fig. 6 and Fig. 7 are time series graphs of task levels estimated by ATCCS. The result of the simulation has indicated that the strategy 1 has lead to continuous higher task levels due to another confliction between ANA 736 and ANA 573. On the other hands, strategy 2 could successfully resolve not only the confliction between ANA 736 and JAL 542 but

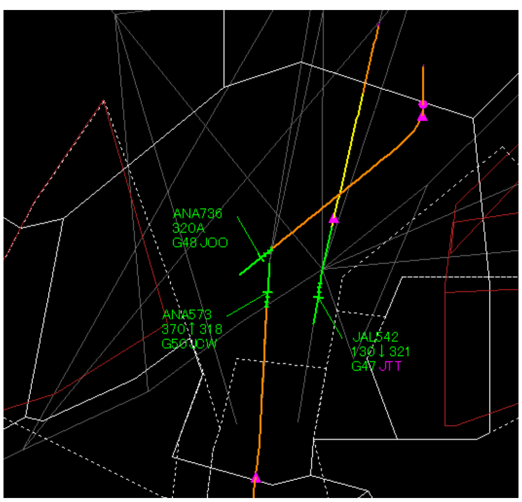


Fig.4 Result of Simulation Experiment (Strategy 1)

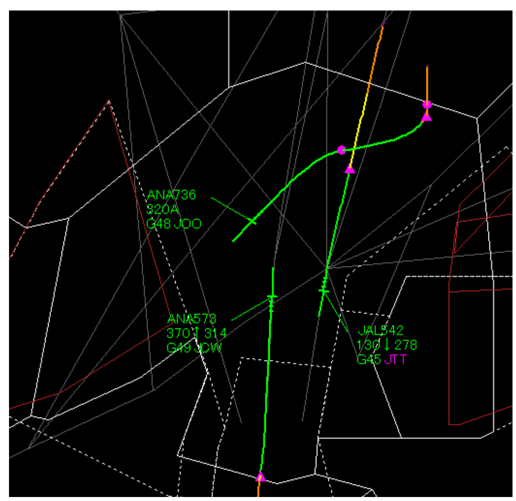


Fig.5 Result of Simulation Experiment (Strategy 2)

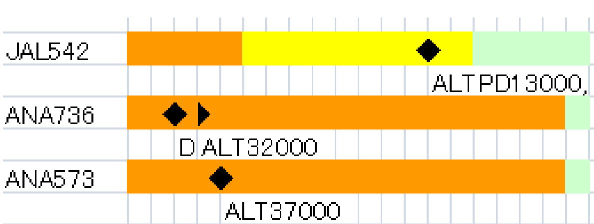


Fig.6 Result of Simulation Experiment
(Strategy 1, Time Series Graph)

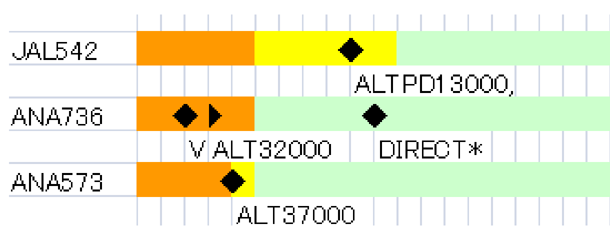


Fig.7 Result of Simulation Experiment
(Strategy 2, Time Series Graph)

(Task Level 1: Green, Level 2: Yellow, Level 3: Orange, Level 4: Red)

also the confliction between ANA 736 and ANA 573 by displacing crossing point to north where ANA 573 is certainly expected to reach enough high altitude to maintain vertical separation with ANA 736. The task level has been reduced in the earlier time frame when strategy 2 has been adopted. This result indicates that the strategy 2 has an advantage in terms of reducing possible risk of confliction although it requires one more instruction for radar vector. It can also contribute to reduce cognitive load of an ATCO to monitor and resolve conflicts.

Through the simulation experiment, the developed ATCCS could successfully visualize the effect of ATCO's control on air traffic flow for different strategies which is consistent with the opinions of actual ATCOs. This function of the ATCCS can helps trainee to understand the differences of the consequences of the different strategy more effectively.

Conclusion

In the present study, a cognitive system simulation of an air traffic controller called the Air Traffic Controller Cognitive Simulation (ATCCS) has been developed based on the results of cognitive task analyses of an ATCO. The function visualizing the difference of task performance has been implemented into the ATCCS. Using this visualization function, the effect of ATCO's control on air traffic flow has been successfully visualized. Although the development of this simulation framework is still underway, the result of the simulation experiment has strongly implied that ATCCS equipped with performance visualization function can be utilized as a supporting tool for education of ATCO trainees.

Reference

Inoue, K., Ando, H., Aoyama, H., Yamato, H. (2006). A Research on Task Analysis System for En-route Air Traffic Control: *In Proc. 8th Int. Conf. Probabilistic Safety Assessment and Management (PSAM8)*, 0254.

Inoue, S., Aoyama, H., Kageyama, K., and Furuta, K. (2005). Task Analysis for Safety Assessment in En-Route Air Traffic Control: *In Proc. 13th Int. Symp. on Aviation Psychology*, Oklahoma, USA, 253-258.

Wickens, C. D., and Holland, J. G. (1984). Engineering Psychology and Human Performance, E. Merrill Publishing

Karikawa, D., Aoyama, H., Takahashi, M., Furuta, K. (2008). Study on Air Traffic Control System based on Cognitive Systems Engineering - Cognitive System Simulation of Air Traffic Controller -: *In Proc. of Annual Conference of European Association of Aviation Psychology (EAAP) 2008*

SOLUTION SPACE-BASED COMPLEXITY ANALYSIS OF ATC AIRCRAFT MERGING TASKS

Patrick Hermes, Max Mulder, M. M. (René) van Paassen
Faculty of Aerospace Engineering – Delft University of Technology
Delft, The Netherlands

Hans Huisman
Air Traffic Control the Netherlands
Schiphol Oost, The Netherlands

J.H.L. (Harjan) Boering
To70 Aviation Consulting
The Hague, The Netherlands

Air traffic controller workload is considered to be a limiting factor in the growth of air traffic. In this paper a new method of assessing controller task demand load will be developed and tested. Based on the hypothesis that workload is primarily caused by the complexity of the task to be conducted, the concept of the "solution space" is described. For any particular air traffic control problem, the solution space describes the constraints in the environment that limit (and therefore, guide) air traffic controller decisions and actions. The complexity of that particular control problem can then be analyzed by considering the properties of the solution space. The task of merging an aircraft into a stream of other aircraft that fly along a fixed route is considered. An experiment has been conducted in which subjects were instructed to solve merging problem scenarios of varying complexity. After completing each scenario, subjects were asked to rate the task complexity. High correlations are found between several solution space properties and reported complexity.

Air traffic controller (ATCo) workload is considered to be one of the main constraints in air traffic growth (Hilburn, 2004). It is important to be able to predict the effect of developments in the air traffic management system on the ATCo. Currently, these effects are mainly assessed using expert judgment. For reasons of cost and time, it is preferred to perform this analysis already during the initial fast-time simulation (FTS) phase.

The analysis of ATCo workload in FTS has been the subject of a large number of studies (Phillips & Marsh, 2000; Crutchfield & Rosenberg, 2007). The immediate flaw that is found in workload assessment using FTS programs is that it is impossible to assess the mental workload, i.e., the workload as *experienced* by the operator, as here subjective elements such as training, equipment, and stress level play an important role. Instead, developers of FTS programs aim to analyze the ATCo's task demand load (Stassen, Johanssen, & Moray, 1989), which is considered to be an objective measure of the complexity of the task to be performed by the controller. ATCo workload is hypothesized to be composed of a number of factors, such as level of training, type of equipment and sector complexity (Stein, 1985; Kirchner & Laurig, 1971). Sector complexity is often used as the means to describe ATCo task demand load. The underlying hypothesis is that – as in the current research – ATCo workload is coupled to task demand load (i.e., an increase in task demand load leads to an increase in workload), and that task demand load, in turn, is coupled to sector complexity.

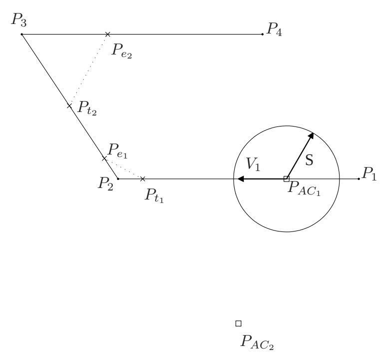
For the current generation of FTS programs, task demand load metrics are constructed using a weighted combination of scenario properties. Examples are the number of aircraft involved, the sector size, the ratio of climbing and descending aircraft, or the count of weighted controller events (Kopardekar & Magyarits, 2002; Majumdar, Ochieng, Bentham, & Richards, 2005). The properties that are relevant to the task demand load analysis, and what weighing factors need to be used, are determined through expert judgment and regression analyses. The validity of this method is questionable, however, since the scenarios that are being analyzed might differ heavily from the baseline scenarios used for the regression analysis. ATCo task demand load has proven to show non-linear behavior, and can vary greatly due to slight changes in the situation being controlled. Therefore, another, more objective and also more widely applicable method of task demand load analysis is required.

This paper aims to demonstrate how a new method of complexity analysis can be used to perform a task demand load analysis for air traffic control related problems, in a more accurate and objective manner than current techniques. In this method, the complexity of a particular controller task is analyzed by examining the – what we refer to as the – "solution space of the problem". The solution space can be defined as the subset of all possible vector (combined heading and velocity) commands that can be issued by ATC, that satisfy constraints of safety, productivity, and efficiency. These constraints are imposed by the situation at hand. To evaluate the validity of this method, only the task of merging aircraft is considered in this paper, in the horizontal two-dimensional plane.

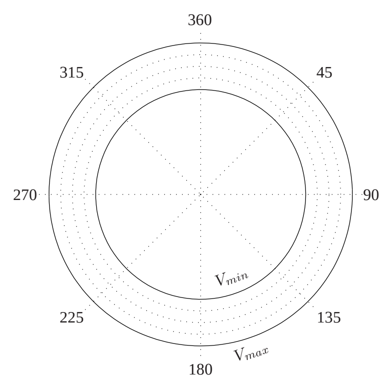
The solution space of aircraft separation problems has been researched by Van Dam et al. (Van Dam, Mulder, & Van Paassen, 2008) from the pilot's perspective, and it was hypothesized that their systematic approach might also be applicable to the ATC problem. Basically, the solution space is a measure of the set of possible solutions that are at an operator's disposal to deal with a particular problem. In the present context, for any particular air traffic control problem the solution space describes the constraints in the environment that limit – and therefore, *guide* – the air traffic controller's decisions and actions.

As a first step in the development of the solution space method, the problem of merging aircraft onto a single fixed route is analyzed. Merging situation occur, for instance, as aircraft approach an airport and need to be lined up for landing. The solution space analysis is performed for aircraft that are not on the route and aims to find out what combinations of heading and velocity – the ATCo *vectors* – lead to a successful merge.

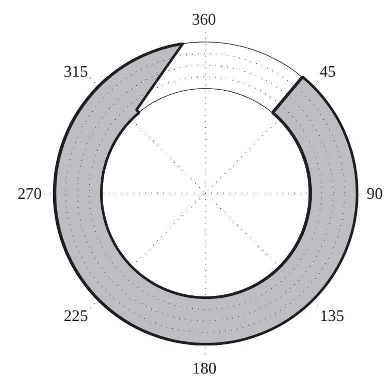
The solution space is defined as the state space that represents possible vector commands issued by ATC that satisfy particular well-defined constraints. For the current analysis these are: (1) *Productivity*: the vector must be such that the free aircraft flies toward the route; (2) *Safety*: the vector may not lead to loss of separation at any point in time; and (3) *Efficiency*: the vector must allow for direct route interception, no additional commands shall be necessary. The solution space computations are conducted in a number of steps, discussed in full detail in (Hermes, Mulder, Van Paassen, Boering, & Huisman, 2009).



(a) Overview of example problem



(b) Initial solution space



(c) Computed solution space

Figure 1: Example problem overview, initial and computed solution space.

An example problem is illustrated in Figure 1(a), showing a route that runs from fixed route points P_1 to P_4, via P_2 and P_3. An aircraft – referred to as the 'route' aircraft – is defined by its position P_{AC_1}, velocity V_1 and separation minimum S, flies along the route. It is assumed that the aircraft travels along the route with a constant velocity V_1, and has a fixed turning angular rate of three degrees per second. When more aircraft are flying along the route, the solution space calculations need to be repeated for all route aircraft.

Another aircraft – the 'free' aircraft – intends to intercept the route, and is initially located at P_{AC_2}. The goal of the analysis is to determine which combinations of heading and velocity commands can be given to the free aircraft, in such a way that the vector command satisfies the productivity, efficiency, and safety criteria. The initial solution space can be drawn like Figure 1(b), in which all possible combinations of heading and velocity are present. Note that the velocity possibilities are limited by the minimum and maximum velocity, V_{min} and V_{max}, respectively, of the free aircraft.

Certain combinations of velocity and heading commands meet the criteria of productivity, safety and efficiency, others don't. Areas in the solution space that contain these vectors are labeled *safe areas*. Areas containing vectors that do not satisfy one of the constraints are called *unsafe areas*. Figure 1(c) shows the solution space for this particular example, with the unsafe area indicated in grey (Hermes et al., 2009).

Using the method of solution space analysis on a merging problem, all properties of the scenario are systematically combined into a single metric. A solution space-based metric is therefore hypothesized to be a more objective and also scenario-independent metric than a weighted combination of scenario properties, in which the weights highly depend on the baseline scenarios considered.

In order to investigate if and how the solution space analysis can be used to assess the aircraft merging problem complexity, a validation experiment was performed. In this experiment, subjects were confronted with scenarios in which they were required to merge a free aircraft onto a route using heading and velocity instructions. Correlations were computed between properties of the initial solution space (the static solution space at the start of the scenario), such as size of safe areas, and the scenario complexity as experienced by the test subjects.

Method

Experiment set-up

Apparatus A stand-alone simulator was developed using MATLAB^{TM}. An interface was presented that consisted of two parts. The left part was a conventional Plan-View Display (PVD), where the route and the aircraft, represented as a square with a label, can be found. The right part was the command window, allowing subjects to give commands to the aircraft. Subjects could send heading and velocity commands (either one by one, or simultaneously), and the command to intercept the route.

Subjects and Instructions Nineteen subjects participated. Based on experience, they were divided in two groups. The first group, six subjects, aged 33 to 50 ($\mu = 39.3$, $\sigma = 6.4$), had operational ATC experience. The second group consisted of thirteen inexperienced subjects, aged 23 to 51 ($\mu = 29.2$, $\sigma = 8.2$).

Subjects were instructed to maneuver the free aircraft onto the route. They were free to choose any point on the route for interception, but were not allowed to merge in front of the first route aircraft, or behind the last route aircraft, because the stream of aircraft was finite. Their subgoal was to use as few commands as possible.

In practice most, if not all, subjects merged the free aircraft on the route segment that lied in-between the initial heading bandwidth BW_{head}. And surprisingly, although subjects were told that they could also command the motions of the route aircraft, they all worked only with the free aircraft.

Procedure Subjects first got familiarized with the interface using an interactive, explanatory tutorial. Then, a minimum of nine training scenarios, hypothesized to be ascending in complexity, were presented. Subjects were introduced with the questionnaire. Then, fifteen measurement scenarios were done, in randomized order (random in hypothesized complexity and random per subject).

Questionnaire The questionnaire consisted of the following six questions, constructed to find out how complex the subjects perceived the scenarios to be, and why they assessed it as such: (1) How *complex was the scenario* to solve? (2) Did you feel that *time pressure* influenced the complexity of solving the scenario? (3) Did you feel that *aircraft limits* influenced the complexity of solving the scenario? (4) Did you feel that *route design* influenced the complexity of solving the scenario? (5) Did you feel that *traffic* influenced the complexity of solving the scenario? (6) Did you feel that *initial conditions* influenced the complexity of solving the scenario? Each of these questions were answered using an 11-point (0-10) Likert scale.

Experiment scenarios

Aircraft All aircraft moved at a certain heading with a certain velocity (200 knots). They turned with a rate of three degrees per second, and accelerated/decelerated with three knots per second square. The simulation was run four times as fast as real-time, due to the relative simplicity of the task. The simulation was two-dimensional, altitude was not taken into account. All aircraft had a fixed V_{min} and V_{max} of 175 and 225 knots, respectively.

Airspace and routes The Dutch airspace was used as a background to increase the fidelity of the simulation. Subjects did not have to take sector boundaries into account when performing the task, however. Routes were constructed in such shapes and lengths as to create certain solution space diagrams and merging problems.

Traffic In each scenario, traffic was placed such that (in combination with route design) certain initial solution space properties were achieved. All traffic was present at the start of the scenario; aircraft disappeared from the PVD as they reached the end of the route. Route aircraft that were not on the actual route yet, traveled toward the first route point: no aircraft intercepted the route at any point aside from the first one.

Initial scenario properties Initial scenario properties that were hypothesized to be indicators of scenario complexity, were identified prior to the experiment, using results from literature and through expert judgment. The following properties were considered: (1) Number of route aircraft (N_{AC}); (2) Number of approaching aircraft (N_{AC_a}); (3) Distance to the route (d_{route}); (4) Turns in the route (N_{turns}); (5) Length of the route (l_{route}); and (6) Bunching (B), a measure of aircraft being in close proximity to each other. For every two aircraft that have intersecting or touching separation circles at scenario initialization, B is increased by one.

Initial solution space properties Several initial solution space properties were hypothesized to be possible complexity indicators. The following variables were examined: (1) Heading band range (BW_{head}); (2) Number of safe areas (N_{safe}); (3) Number of relevant aircraft ($N_{AC_{rel}}$); (4) Total solution space size (A_{safe_t}); (5) Size of largest safe area (A_{safe_l}); (6) Average safe area size (A_{safe_a}); and (7) Safe area size deviation (σ_{safe}).

Dependent measures

The questionnaire results consisted of the subjects' answers to the six questions, and additional comments. As different subjects exhibit different rating behavior, all quantitative data were first corrected for inter-subject differences. This correction was performed by calculating the Z-scores for every test subject.

Hypothesis

Our main hypothesis was that, when using the initial solution space properties, a metric can be constructed that has a stronger (i.e., higher) correlation to the subjectively-reported complexity than the other metrics based on either the initial scenario properties or logged properties such as number of commands or separation violations.

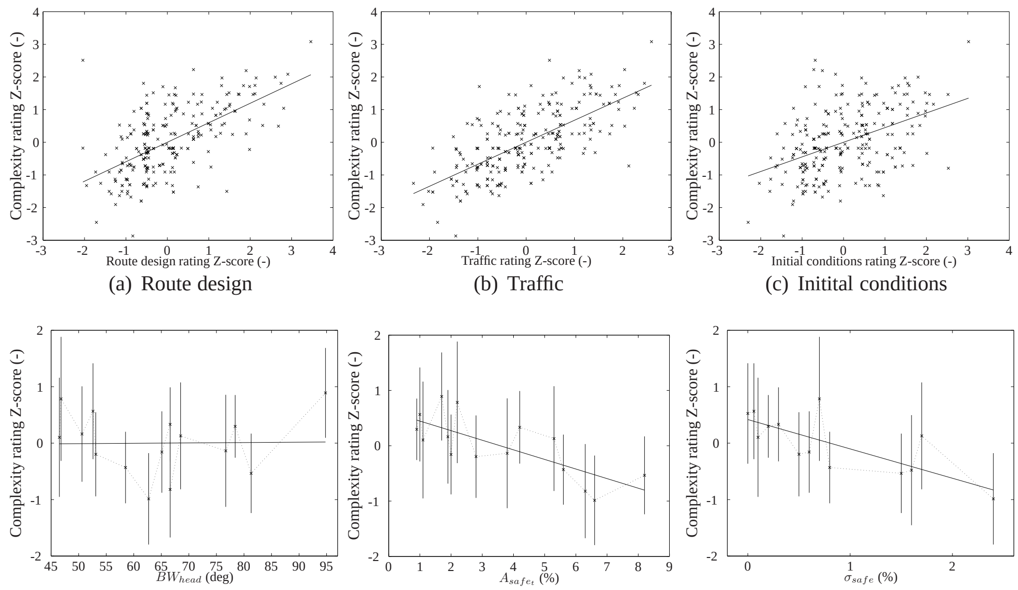
Results

A total of 285 experiment runs were performed using nineteen test subjects and a total of fifteen measurement scenarios. Using analysis of variance it was shown that no training effect was present in the data. Results from an outlier analysis and a group correlation analysis showed that the most illustrative results would be obtained if the full data set was used, and if all subjects were considered to be members of a single group (Hermes et al., 2009).

Correlation analyses

One-way analyses of variance were conducted, with the subjective complexity rating the dependent measure. Since all but one of the possible complexity predictors showed significant p-values ($p < 0.05$) in these ANOVAs, correlation analyses were performed in order to determine how well the possible predictors correlated with the test subject complexity ratings. For each possible predictor, a Pearson's R value of linear correlation was calculated. This was done using data from all experiment runs individually *and* using the means of experiment data per scenario, that is, averaged over all subjects.

Correlation between complexity and other questionnaire variables Figure 2 shows the Z-score complexity rating plotted against some of the questionnaire variables (also Z-scores), together with a best-fit linear relationship. Although all questionnaire variables showed statistically significant correlation, subjects linked complexity most strongly to the "Traffic" involved in the scenarios, i.e., the presence of the other aircraft flying on or towards the route (means: R=0.9740, $p < 0.001$; all: R=0.6743, $p < 0.001$).



(a) Route design (b) Traffic (c) Initital conditions

(d) BW_{head} (e) A_{safe_t} (f) σ_{safe}

Figure 2: Complexity rating versus other questionnaire ratings (top) and initial solution space properties (bottom).

Correlation between complexity and initial scenario properties The number of aircraft N_{AC} (means: R=0.4454, $p = 0.0962$; all: R=0.2434, $p < 0.001$), bunching B (means: R=0.4237, $p = 0.1156$; all: R=0.2316, $p < 0.001$), and the number of approaching aircraft N_{AC_a} (means: R=0.3580, $p = 0.1901$; all: R=0.1957, $p < 0.001$), correlate to complexity strongest. Since these are all properties related to traffic, this finding supports the hypothesis that subjects linked complexity most strongly to traffic properties. It corresponds well with the questionnaire findings.

Correlation between complexity and logged properties Statistically significant correlation was found for the number of commands N_{com} (means: R=0.8244, $p < 0.001$; all: R=0.2361, $p < 0.001$) and the number of separation violations N_{SV} (means: R=0.7220, $p = 0.0024$; all: R=0.1978, $p < 0.001$). Hence, the correlations for the initial scenario traffic-related properties and the statistically relevant logged properties are in the same order of magnitude, with R values of approximately 0.2. Especially the fact that the number of aircraft and the number of commands correlate to complexity approximately equally strong is interesting, since they are both currently used as preliminary indicators of workload in FTS. This provides confidence regarding the validity of the present experiment.

Correlation between complexity and initial solution space properties In Figure 2, the Z-score complexity rating is plotted against some of the initial solution space properties, including the best-fit linear relationship. In the initial solution space properties correlation analysis, the safe area percentages A_{safe_t} (means: R=-0.7423, $p = 0.0015$; all: R=-0.4047, $p < 0.001$), A_{safe_l} (means: R=-0.7224, $p = 0.0024$; all: R=-0.3949, $p < 0.001$), and σ_{safe} (means: R=-0.7284, $p = 0.0021$; all: R=-0.3981, $p < 0.001$), correlated to complexity most strongly. The absolute R value for these three initial solution space properties is approximately twice as high as the absolute R values of the best predictors from the initial scenario properties and the logged properties. This leads to the conclusion that solution space properties, and specifically those that link to *solution space size*, were the best predictors of complexity in this experiment. This supports our main hypothesis, namely that a more accurate complexity predictor could be constructed using the solution space concept.

Regression analysis

In the regression analysis, initial solution space properties were combined in metrics in an attempt to obtain stronger correlations, and thus more accurate complexity predictors. The regression was performed using “means” calculations. By combining all seven initial solution space properties, an absolute R value of 0.839 could be achieved. Furthermore, it was observed that the total safe area size A_{safe_t} was present in the best 36 metrics, suggesting that this is the most important solution space property and the best complexity predictor. This finding is supported by the fact that a metric that contains only A_{safe_t} already has an absolute R value of 0.742, only 0.097 lower than the absolute R value for the best metric, the one including all descriptors. The relatively small increase in correlation in the regression analysis also suggests, however, that the initial solution space properties that were analyzed in this experiment are coupled. Whether this means that solution space *size* is the most relevant of all solution space properties, or that another property that has not been analyzed can add significant additional predictive capability, should be determined in a more elaborate study.

Overall, the results suggest that the solution space does indeed lead to more accurate complexity predictors. However, it is important to realize that the two-dimensional merging problem that was analyzed in this research is not the only, or main, task that an air traffic controller performs in a normal work setting. Yet, although the results should be treated with care, they certainly provide a solid basis for further research into the development of a complexity metric which is based on the solution space concept.

Conclusions and Recommendations

This study investigated whether the solution space of a two-dimensional air traffic merging problem can be used to assess the complexity of that problem more accurately and objectively than current metrics. An experiment was conducted which showed that the initial solution space properties, in particular those related to solution space size, are indeed more accurate complexity assessors than traditional metrics, while being at least as objective. This result provides a solid basis for expanding the solution space research. Possible future research paths include dynamic solution space analysis, three-dimensional air traffic control problems and the development of solution space based interfaces to support air traffic controller decision making and situation awareness.

References

Crutchfield, J. M., & Rosenberg, C. W. (2007). *Predicting subjective workload ratings: A comparison and synthesis of operational and theoretical models* (Tech. Rep. No. DOT/FAA/AM-07/6). FAA.

Hermes, P., Mulder, M., Van Paassen, M. M., Boering, J. H. L., & Huisman, H. (2009). Solution space based complexity analysis of the atc task of merging aircraft. *Proceedings of the AIAA Guidance, Navigation and Control Conference, Chicago (IL), USA, August 10-13*(AIAA 2009-).

Hilburn, B. G. (2004). *Cognitive complexity in air trafic control - a literature review* (EEC Note No. No. 04/04). Bretigny-sur-Orge, France: EUROCONTROL.

Kirchner, J.-H., & Laurig, W. (1971). The human operator in air traffic control systems. *Ergonomics*, *14*(5), 549-556.

Kopardekar, P., & Magyarits, S. (2002, 27-31 October). Dynamic density: Measuring and predicting sector complexity. In *Proceedings of the 21st digital avionics systems conference* (pp. 2C4-1–2C4-9).

Majumdar, A., Ochieng, W. Y., Bentham, J., & Richards, M. (2005). En-route sector capacity estimation methodologies: An international survey. *Journal of Air Transport Management*, *11*(6), 375-387.

Phillips, M. R., & Marsh, D. T. (2000). The validation of fast-time air traffic simulations in practice. *Journal of the Operational Research Society*, *51*, 457-464.

Stassen, H. G., Johanssen, G., & Moray, N. (1989). Internal representation, internal model, human performance model and mental workload. *Automatica*, *26*(14), 811–820.

Stein, E. S. (1985). *Air traffic controller workload: An examination of workload* (Technical Note No. DOT/FAA/TC TN06/19). Atlantic City, NJ, USA: FAA.

Van Dam, S. B. J., Mulder, M., & Van Paassen, M. M. (2008). Ecological Interface Design of a Tactical Airborne Separation Assistance Tool. *IEEE Transactions on Systems, Man & Cybernetics, Part A*, *38*(6), 1221-1233.

DESIGN AND EVALUATION OF A COGNITIVELY ENGINEERED SYSTEMS MONITORING DISPLAY

Michael J. Findler
Wright State Research Institute
Dayton, OH USA
Dr. S. Narayanan,
Wright State University
Dayton, OH USA
Jeff S. Collier
Sytronics, Inc
Beavercreek, OH USA
Dr. William Marshak
Sytronics, Inc
Beavercreek, OH USA

A new format was derived from a Visual Thinking cognitive psychology paradigm and permits easy understanding of multiple system parameters with different directions and unit scaling. This new "Sprocket" format allows rapid cross check, characterizing multiple failure thresholds, and easy detection of out-of-tolerance conditions and a gestalt state awareness. The format was evaluated in a dual task, aviation-oriented experiment.

The advent of the all glass cockpit in aircraft makes new display formats possible beyond the traditional discrete gauge display that dominated the first 100 years of aviation. Graphics allow greater flexibility in information display, but is not without its dangers. A British B737 accident was attributed to an aircrew unfamiliar with a new engine display format shutting down the wrong engine in response to an engine fire. New format designs must be cognitively compatible with both the aircrew's mental model of aircraft function, at least consistent with existing displays affecting reading through transfer of training, and provide significant advantages in reading accuracy, speed, and reduced display space.

Visual Thinking

The Visual Thinking phrase was first coined by Arnheim (1969) to describe the relationship of perception and cognition. Inherent in vision is the ability to preprocess data and recognize visual patterns.

> Vision is not perception and perception is not thinking. The mind gathers information and processes it. Note that I said information, information is data plus meaning. Before the mind conveys the information your eyes must observe it, and some preprocessing needs to be done to turn this data into information. (Arnheim 2004)

The key to Arnheim's thesis is that vision and thinking are not necessarily disjoint concepts. When a person perceives an object with their eyes, before deep thoughts about the object can be conceived, the simple sight of that thing at least causes classification (placing the object in the context of other objects like it). For instance, if you see a cat, before any separate thinking is performed about the cat, it has already been placed in the category of "cat". This is a particularly useful cognitive trait to have when that cat is a dangerous one that needs to be fled from, such as a tiger.

Arnheim contends the idea of visual thinking is an old one, going back to the ancient Greek philosophers: Plato, Socrates and Aristotle. These philosophers were the first to make a distinction between perceiving and reasoning, mainly because perception from direct senses could not always be trusted. (We have all experienced "our eyes playing tricks on us", or heard tales of mirages in the desert.) Reasoning was considered to be the "correction of the senses" and the "establishment of truth".

It can be reasonably argued that Arnheim's Visual Thinking is an almost instantaneous pattern classification. It is not the perception of the object that classifies the object, but rather the very well travelled mental pathways that react with almost lightening quick classification. The perceiving of the object (cat) does not require new neurons to fire off and create new paths; the existing short pathways have always succeeded previously.

Design of the Visual Thinking Sprocket

Physiologically, the eyeball as an information-gathering instrument scans the world under the guidance of cognitive attention centers. The eyeball fixates on a region of interest. An image is buffered and scanned, like a massively parallel computer, to find objects within the image through feature extraction. Once extracted, these objects are serially scanned at about 25 items per second. Since the eye scans quickly, reacquiring a new image about 10 times a second, only four to twelve objects are recognized before the eye jumps to another fixation. These physical boundaries must drive the design of cognitively sensitive displays.

Furthermore, when designing a display, two attributes must be balanced: the **overview** of the situation and the **details** within the situation. The overview is a qualitative "aspect of data preferably acquired rapidly and even better, pre-attentively; that is, without cognitive effort" (Spence 2007). A well designed overview display uses visual cues that are acknowledged to be pattern classifier aids so information "pops out" at the operator. On the other hand, details are quantitative and should only be presented to the operator on an as needed basis, i.e., when the operator requests more in-depth information, presumably because of the overview display observations.

Within the design of the Visual Thinking Sprocket display, primary attention is devoted to the overview pattern classifier aids. A design that presents an overview of a situation must be designed simply and stress those features that can be pre-attentively processed. According to Ware (2004), features that can be pre-attentively processed can be organized in the following categories:

- Form: Line orientation, line length, **line width**, line collinearity, **size**, **curvature**, **special grouping**, blur, added marks, numerosity
- Color: **Hue**, Intensity
- Motion: Flicker, **Direction of Motion**
- Spatial Position: 2D position, Stereoscopic depth, convex/concave shape from shading

The features in **bold** were the pre-attentive cues the Visual Thinking Sprocket attempted to model. Examining ***Figure 1***, one can see the intentional feature implementation on the initial single-threshold Visual Thinking Sprocket design. This Visual Thinking Sprocket was intended to be a decision support aid within a larger flight simulator.

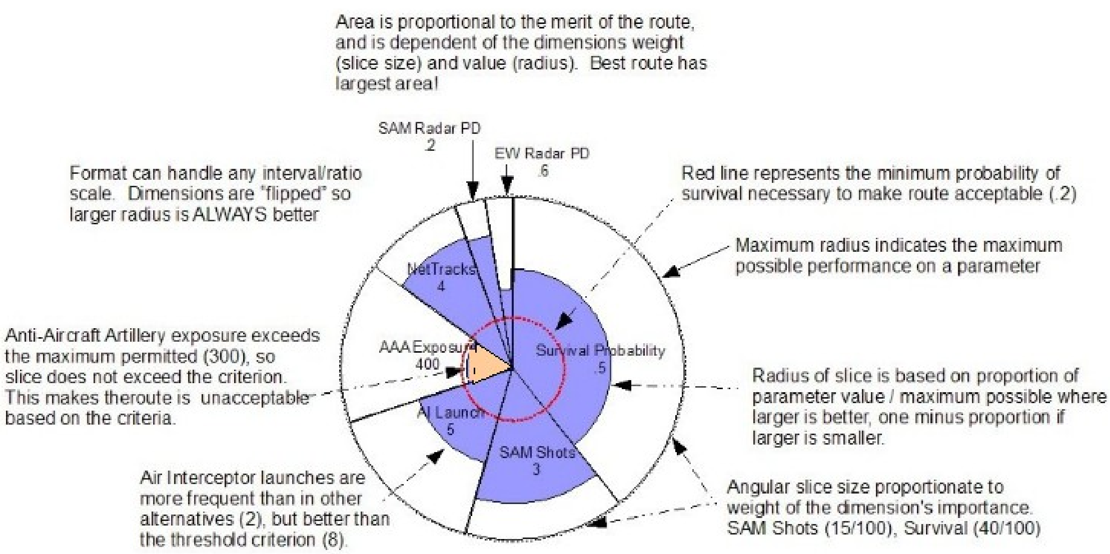


Figure 1. Early drawing of a multi-dimensional, multiple scaled decision support display. The raw detail data display is visible by mouse roll-over of the slice.

Encoded into this initial Visual Thinking Sprocket were (1) angular slices proportional to the weighting of the dimension; (2) acceptability of specific dimensions (pink – unacceptable, blue – acceptable); (3) individual dimension "health" or "preference" (larger colored area is always better); (4) slices nearer the red tolerance line are less optimal, those nearer the maximum radius are deemed near optimal; (5) labels naming individual dimensions and their associated current values; (6) a normalized rescaling of the dimensions; and (7) the global preference of the decision – bigger sprockets are better than smaller sprockets. Finally, if the operator wanted more information about a specific dimension, a simple "mouse-over" displays the detailed raw data behind the image.

From Figure 1, one can see why the resulting circular figure is called a sprocket, with geared teeth of varying length, resembling the tooth embellished wheel that drives a chain, or in this case, cognitive understanding. Experiments were performed to examine the viability of using Visual Thinking as the cornerstone in designing displays. Figure 2 illustrates two versions of a Visual Thinking Sprocket, each of which can be instantiated through the same software library.

Figure 2.A The decision support aid is a static single-threshold two-color sprocket – it compares three alternate routes that an operator may select. The operator has to choose which route alternative was best based on the presented dimensions, called the Figures of Merit (FOM). For this display, bigger is always better, so Route 2 is the best route. This is obvious without placing exact values on the display. The sprockets were instantaneous snapshots of the route alternatives. Finally, dimensional weights are displayed in the slice angular subtend so dimensional contribution to total area is clearly expressed.

Figure 2.B The dynamic double-threshold three-color system monitoring display. The operator is monitoring four unmanned aerial vehicles (UAVs). In this example each of the UAVs has serious health issues, for example UAV_1 and UAV_2 have Fuel Temperatures that are below the minimum threshold, while UAV_0 has an RPM that exceeds the upper threshold. The amount that UAV_1 and UAV_2 fail to reach the minimum threshold is illustrated by the size of the wedge – UAV_2 "barely" fails to reach the minimum threshold, indicated by the numeric value and less white space in between the wedge and red lower threshold ring. UAV_3 is "healthy", since it is all gray. Best is not the biggest or smallest area; but rather, the most circular gray pie. Optimal is indicated by the light blue ring. Color coding is meant to mimic American water faucets, e.g., blue (cold/low) and red (hot/high).

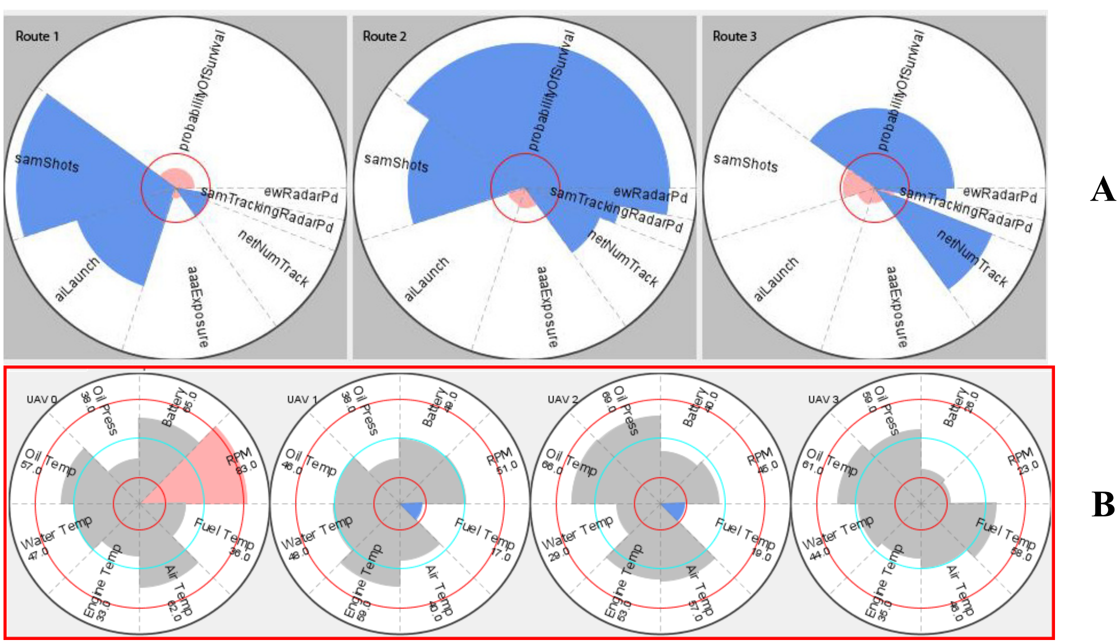


Figure 2. Static single-threshold (A) and dynamic double-threshold (B) Visual Thinking Sprocket designs

The new sprocket format is designed to achieve greater situation awareness (SA) as a basis for the decision. Endsley and Kiris (1995) defined three different levels of SA. Level 1 (SA1) deals with the "perception" of elements of the environment. Level 2 (SA2) describes the comprehension of those elements and indicates a deeper level of cognitive comprehension. Level 3 (SA3) refers to the ability to use that comprehension to make predictions based on them. When operators choose between generated routes, they necessarily make an SA3 assessment to predict which route will best accomplish the mission. The Visual Thinking design paradigm encourages greater comprehension because it reveals not only which route is superior, but why it is superior. It facilitates integration of the various dimensions via the rescaling of the dimensions and translation into the area representation.

The Visual Thinking Sprocket was compared to a text table (similar to the current UAV display technology (circa 2007) used in a Predator Ground Control Station (GCS) Variable Information Tables (VIT) and bar charts (considered chosen as a first naive attempt at a "graphical user interface"). It is understood that these are discrete interfaces, not integral interfaces, which is because there exist no integral interfaces for multiple UAVs.

Experimental Design

The experiment employed the Sprocket as a decision aid to choose which of three alternative missions were best based on weighted criteria. The criteria included dimensions like probability of survival, number of missile shots, minutes of radar exposure, etc.; measures differed in their direction and scales. This trio of routes was presented in a two factor repeated-measures design with full-model partitioning. Each image was generated from the same data, i.e., a bar chart, text table and sprocket image were generated from datasets 1, 2, …, 12. Each subject was shown a series of generated images of the Figures of Merit (FOM) for three alternative paths' and a question. The subject responded to the first question (rank order the three routes from best to worst), and then the second question was displayed, and so on – the presented image did not change and remained visible during the questioning (SPAM – Situation Present Assessment Method). The four questions and levels of situation awareness are listed in Table 1. The questions were designed to explore specific decisions considered typical within a multi-UAV mission.

Table 1. Questions asked for each set of generated images.

	SA Level	Question	Possible Answers
Q1	SA2	Rank order the routes [Best to worst]	1-2-3, 1-3-2 2-1-3, 2-3-1 3-1-2, 3-2-1
Q2	SA1	Do any of the routes meet the all minimum criteria?	Yes, No
Q3	SA3	Which route is Best if Dimension X is dropped? [Where X was chosen from among the 4 top weighted available dimensions]	1, 2, 3
Q4	SA1 & SA2	Which route has Best Dimension Y? [Where Y was chosen from all available dimensions]	1, 2, 3

Results and Conclusions

The accuracy and response time for each question must be examined. To choose this new display, the results of the experiment must show that it is better than the alternatives (current text based display or naive graphic display) in accuracy and/or response time – preferably both. Furthermore, the subjects should find the new display "intuitive".

Looking at the "correctness of answer" per question data first, the generated data was examined prior to presentation to the subjects to determine the "correct answer" to each question with a weighted combination of the parameters. The operator's "correctness" response was then defined as whether the operator responded with the pre-calculated correct answer. A correct response was assigned a value of 1 and a wrong answer was assigned a value of 0. The sum of the correct answers was then used as a measure to determine the cognitive ease of use for each display type – each question had a maximum score of 12 points for correctness.

For example, the Minimum Criteria question (Q2) asks whether all of the mission alternatives presented to the operator meet the minimum criteria on all dimensions: (1) the possible answers are `Yes` or `No`, and (2) the correct answer is `Yes` (the three mission statements are all valid).[1] Then for each operator that answered `Yes`, a

[1] The Text Table and the Sprocket color coded a failure of each dimension, i.e., if the "Probability of Success" dimension failed to reach the minimum value, then the text (pie piece) would be colored red. For the Bar Chart, no color coding was attempted.

counter would be incremented by 1. The maximum value the counter could reach is 12, so if the final counter value was 12 (out of 12), then 100% of the subjects responded with the correct answer. If ½ of the subjects responded correctly, then the final value would be 6 (out of 12).

Table 2 summarizes the statistical results for each question. Significant statistical differences in accuracy were found in questions 1, 2 and 3.

Table 2. Analysis of accuracy results

Question	Statistic	Tukey *post hoc* analysis	Conclusion
Q1 Rank Order	[F(2, 46)=11.70, p<0.0001]	Tukey Grouping / Mean / N / Block A / 7.17 / 24 / Sprocket B / 5.38 / 24 / Bar Chart B / 4.96 / 24 / Text Table	Sprocket significantly better
Q2 Meets Criteria	[F(2, 46)= 7.82, p=0.0012]	Tukey Grouping / Mean / N / Display A / 11.29 / 24 / Sprocket A / 10.96 / 24 / Text Table B / 9.92 / 24 / Bar Chart	Bar chart signicantly worse
Q3 Drop One Dimension	[F(2, 46)= 14.03, p<0.0001]	Tukey Grouping / Mean / N / Display A / 9.17 / 24 / Sprocket B / 7.75 / 24 / Bar Chart B / 7.17 / 24 / Text Table	Sprocket significantly better
Q4 Best Dimension	[F(2, 46)= 2.03, p<0.1432]		No significant difference

Now that the statistical results of the accuracy among the questions have been examined, we looked at the response times for each question. If the main effects of Display were found to be significant, there is a statistical difference in the response times among the displays, a Tukey *post hoc* test was performed to find the significance. As anticipated, the main effect of Trial (image presentation order) was found to be significant. Images presented at the beginning of the trial took significantly longer to interpret than those presented at the end of the trial, irrespective of display type, also known as a learning curve. But the most interesting finding is that the display type significantly affected the learning curve slope (Display×Trial). The statistical results are summarized in Table 3.

Table 3. Analysis of response times

Question	Source	Statistic (**Bold** is significant)	Conclusion (from Tukey *post hoc* [not shown])
Q1 Rank Order	Display	**[F(2, 46)= 30.91, *p*<0.0001]**	**Sprocket** significantly better
	Trial	**[*F*(11,253)=17.44, *p*<0.0001]**	Learning curve existed
	Display×Trial	**[*F*(22,506)=7.99, *p*<0.0001]**	Display type significantly affected learning curve
Q2 Meets Criteria	Display	**[F(2, 46)= 29.86, *p*<0.0001]**	**Sprocket** significantly better
	Trial	**[*F*(11,253)=13.27, *p*<0.0001]**	Learning curve existed
	Display×Trial	**[*F*(22,506)=2.08, *p*<0.0001]**	Display type significantly affected learning curve
Q3 Drop One Dimension	Display	[F(2, 46)= 2.96, *p*<0.0626]	**No significant difference**
	Trial	**[*F*(11,253)=9.37, *p*<0.0001]**	Learning curve existed
	Display×Trial	**[*F*(22,506)=1.98, *p*<0.0001]**	Display type significantly affected learning curve
Q4 Best Dimension	Display	**[F(2, 46)= 8.85, *p*<0.0001]**	**Text table** significantly better
	Trial	**[*F*(11,253)=12.41, *p*<0.0001]**	Learning curve existed
	Display×Trial	**[*F*(22,506)=8.04, *p*<0.0001]**	Display type significantly affected learning curve

Because of significance of the Display×Trial interactions, the time series results are graphed in ***Figure 3***. The graphs show the average time each question took to be answered by the operators. Note that the images were presented in a Latin Squares random order. The x-axis labels are the order the images were presented, i.e., first image presented, second image presented, etc. The experiment was considered long enough at approximately 15-20 minutes per display for the learning curves to be examined. Without fail, the initial reaction to the bar charts across all displays had the longest interpretation times and the sprocket tended to have the lowest mean response time over the course of the experiment. However, interpreting which sprocket had the best value for a given dimension (Q4) had the text and sprocket supplying similar response times.

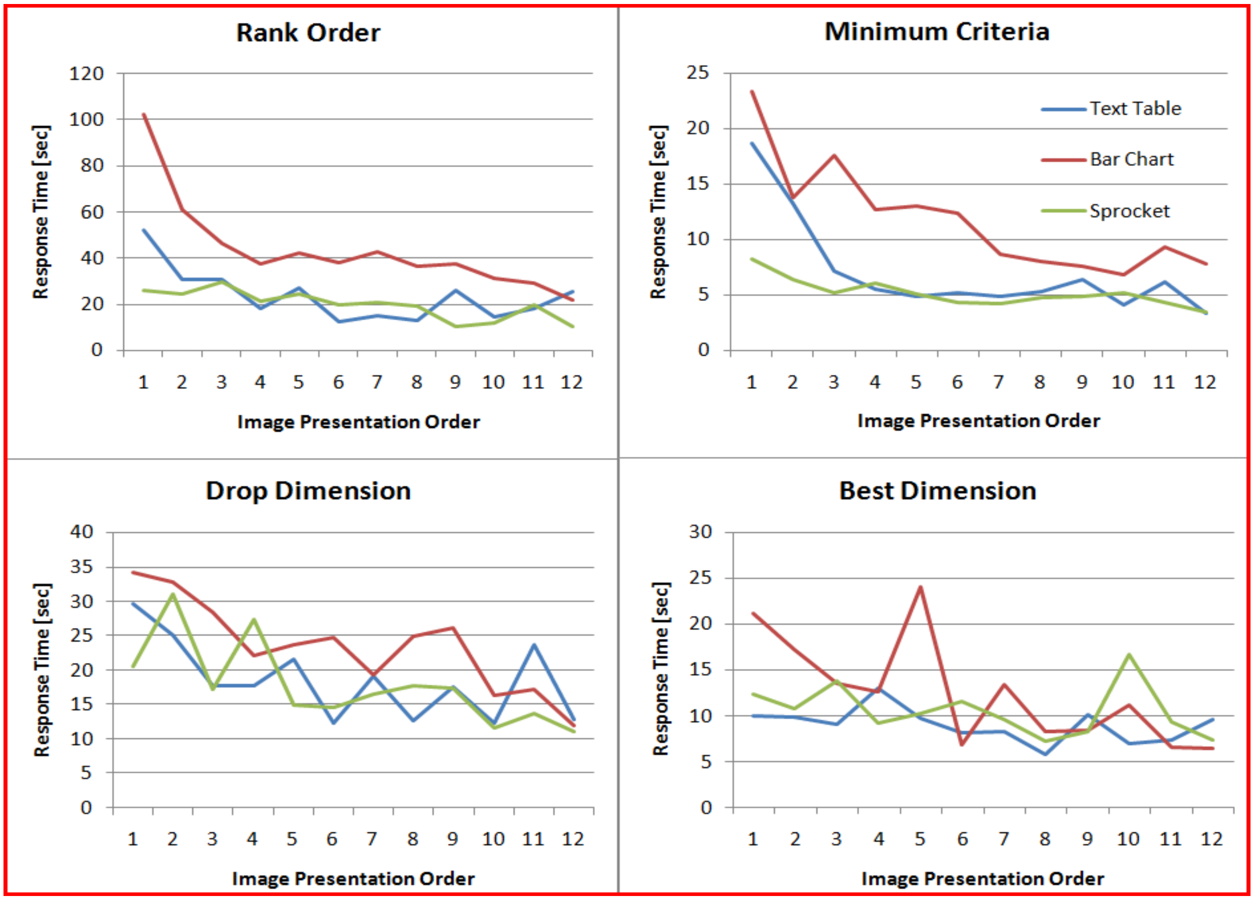


Figure 3. Learning curve for each question

The Sprocket display format offers a spatially compact, concise, single instrument that displays the overall quality of each route and allows ready comparison of their merits. Rescaling the direction and range of dimensions masks their exact value details (which can still be displayed by our software as a "roll-over"), converting it into a less precise value in the angle and radius in the Sprocket. The format exploits the cognitive-perceptual ability to compare size or areas of objects into the cognitive understanding of a mission route's absolute and relative merits. Even more exciting is the Sprocket and the Visual Thinking paradigm from which it was developed, represents only one member of a class of new display formats that exploits the connection between perception and cognition.

References

Arnheim, R. (1969). Visual Thinking, University of California Press.

Arnheim, R. (2004). Visual Thinking, University of California Press.

Endsley, M. R. and E. O. Kiris (1995). "The Out-of-the-Loop Performance Problem and Level of Control in Automation." Human Factors **37**(2): 381-394.

Spence, R. (2007). Information Visualization: Design for Interaction. Essex, England, Pearson Education Limited.

Ware, C. (2004). Information Visualization: Perception for Design. San Francisco, CA, Elsevier.

ASSISTING AIR TRAFFIC CONTROL IN PLANNING AND MONITORING CONTINUOUS DESCENT APPROACH PROCEDURES

A. van der Eijk, M. Mulder, M. M. (René) van Paassen, A. C. in ‘t Veld
Delft University of Technology, Faculty of Aerospace Engineering
Kluijverweg 1, 2629HS Delft, The Netherlands

In advanced noise abatement procedures, the approach of an aircraft is optimized to reduce the noise on the ground. A drawback of many noise abatement procedures is that air traffic controllers are forced to increase spacing, leading to a significant reduction of runway capacity. A display, named the Time-Space Diagram, has been developed to assist controllers in metering, sequencing and merging aircraft flying noise abatement procedures such as the Continuous Descent Approach. Although initial tests were promising, it was recommended that the information could be enhanced by supporting common controller spacing techniques. The improved display was tested in an experiment, in both low and high traffic rate scenarios. Results indicate that the use of the new display results in a significant reduction of controller workload. It also led to a reduction of the number of instructions to the pilots, suggesting a reduced workload on the flight deck as well.

Noise abatement procedures can significantly reduce the noise impact of aircraft during approach (Clarke, 2000; Erkelens, 2002). Unfortunately, the decrease in runway capacity that accompanies Continuous Descent Approach (CDA) procedures restricts the implementation of these noise abatement procedures (Erkelens, 2002). The main reason for this decrease in runway capacity is the inability of air traffic controllers to accurately predict separation between aircraft decelerating at different rates (Reynolds, Reynolds, & Hansman, 2005). As a consequence, to counterbalance the differences in approach time, the separation distances between aircraft performing these procedures are increased substantially (Erkelens, 2002).

Previous research showed that predictions of 4-Dimensional Trajectories (4DTs), shown on an additional display, could support controllers in providing separation during CDA procedures (Tielrooij, In ‘t Veld, Mulder, & Van Paassen, 2008). This additional display, the Time-Space Diagram (TSD), shows predictions of 4DTs in two dimensions: the horizontal axis indicates the aircraft Along Track Distance (ATD) to the runway, while the vertical axis depicts the corresponding Estimated Time of Arrival (ETA) at that distance. In-trail separation between aircraft is then represented by the horizontal distance between two predictions. The initial experimental validation of the TSD demonstrated its potential value in metering, sequencing and merging aircraft in CDAs (Tielrooij et al., 2008). Controllers reported, however, a lack of support for their common spacing strategies. In addition, the TSD would not present conflicts between aircraft on a fixed route and aircraft that were being vectored.

This paper describes the improvements made to the TSD, and the results of a controller-in-the-loop experiment that was conducted to quantify the effects of this display on operator performance, workload and the safety of operation.

The Time-Space Diagram

To make use of 4D trajectory predictions (either sent to the ground or computed on the ground), they have to be presented to the controller in a meaningful way. The Plan View Display (PVD) gives the current positions of aircraft but is not suited to display each aircraft’s future trajectory, because on the plan view these would all overlap. An additional tool could provide the controller with a visual presentation of the future trajectories of aircraft from the current moment down to the runway. The Time-Space Diagram (TSD) provides this visual presentation, see Figure 1.

For each aircraft, the TSD shows the ATD to the runway on the horizontal axis and the ETA on the vertical axis. The connecting line represents a prediction for the future trajectory of that particular aircraft. Each point on the line corresponds to the ATD to the runway versus the predicted time of arrival, at that distance. The slope of the prediction line is an indication for the ground speed of an aircraft. Aircraft flying at a lower groundspeed will have a steeper prediction line than aircraft flying at a higher groundspeed. The space between two prediction lines shows the separation between the two aircraft. The horizontal dimension of this space represents the difference in ATD, the longitudinal separation if these two aircraft are on the same track to the runway (which may not be the case). The vertical dimension of separation space indicates the separation in time.

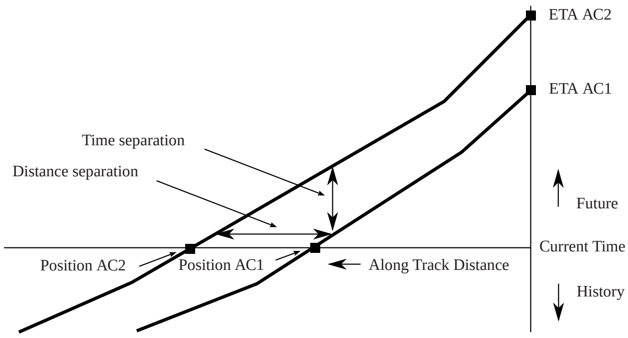


Figure 1: A schematic representation of the TSD (initial design)

Improving the TSD

One of the factors constraining the use of the initial TSD is the presentation of information(Tielrooij et al., 2008). The original and the improved versions are shown in Figure 2. The following changes were hypothesized to make the TSD more clear and self-explanatory.

Layout of the TSD

In the initial layout of the TSD the horizontal axis, representing the ATD, is located at the bottom of the screen, right above the routes, see Figure 2(a). According to the principle of pictorial realism, it is important that the diagram is analogous to the controller's mental model of the system and the physical system itself (Wickens, 1992). Placing the horizontal axis at the top of the screen (underneath the routes) and flipping the vertical axis results in a prediction line that suggests a descending aircraft. The slope of the prediction line is then compatible with the controller's mental model of an aircraft performing a CDA, even though the slope of the prediction line does not represent the vertical speed of the aircraft.

Use of Transparency

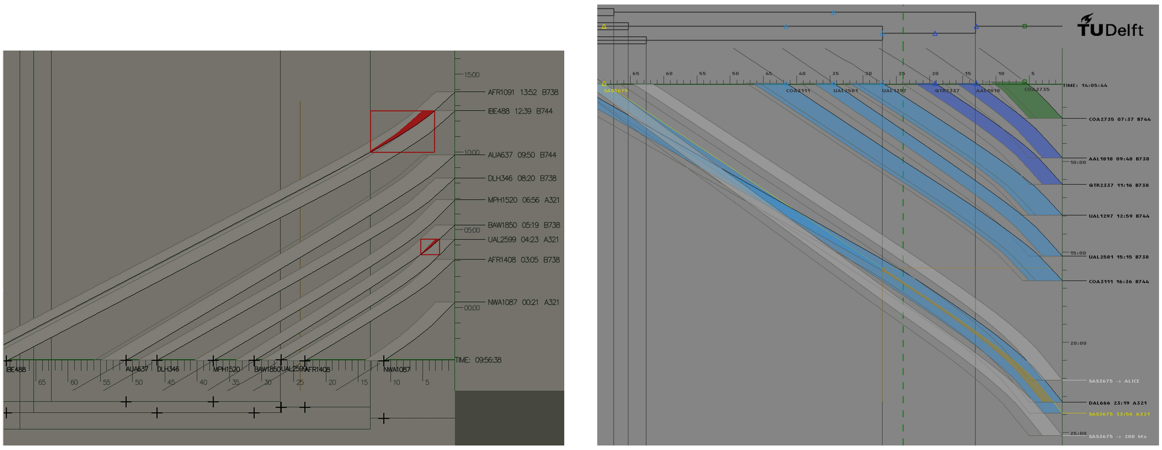
Subjects involved in the initial validation of the TSD reported that the representation of conflicts on the display was often considered problematic (Tielrooij et al., 2008). One of the main problems was overlap: when a conflict occurs, two separation areas overlap and on top of that a red conflict area will be drawn, making the individual predictions hard to distinguish. To solve this issue, transparency is introduced on the TSD. Separation areas and conflict areas are made transparent to a level that in case of a conflict, both the involved separation areas, the prediction line underneath and the conflict area on top, are visible.

Use of Color

On the initial TSD the use of color was limited: separation areas were colored light gray and conflict areas were colored red. All other objects were black and the background was gray. Color, however, could add another dimension to the TSD. In addition, colors could make the TSD more compatible with the PVD, thereby improving the mental model of the controller.

Supporting Spacing Techniques on the TSD

Part of this research was the analysis of the effects common spacing instructions have on the predictions shown on the TSD and how these common spacing strategies would solve a conflict (Tielrooij et al., 2008). The spacing strategies were divided into three categories based on their effects on the 4D trajectory: speed adjustment, changes to the planned route, and temporarily abandoning the planned route.



(a) original TSD (b) improved TSD

Figure 2: Screenshots of the initial and the improved TSD.

Speed Adjustment

When two aircraft tend to become too close to each other, the controller can instruct the trailing aircraft to reduce its speed in order to increase separation. The rate at which spacing increases depends on the difference in speed between the leading and trailing aircraft. On the TSD, the slope of a prediction line indicates the speed of the aircraft, i.e., a speed reduction will lead to an increase in the slope. This results in an increase in ETA, while the ATD remains the same.

Change to the Planned Route

Another solution to increase separation between two aircraft is a change to the route of one of the aircraft. This can be done in two manners: 1) by giving the aircraft a direct instruction to a waypoint further down the route, or, 2) by giving the aircraft the instruction to enter a holding pattern. The waypoint instruction results in a decrease in the ATD, which means that the prediction will shift to the right on the TSD. The holding pattern instruction results in the addition of a specific amount of time to the ETA.

Temporarily Abandoning the Planned Route

Temporarily taking flights off the planned route is an often used strategy to increase separation. The pilot receives a heading instruction (vector) from the controller, which should be followed until enough separation is attained and the pilot is instructed to return to the planned route. However, when an aircraft deviates from its fixed lateral trajectory, the ATD is no longer defined. For the TSD, the ATD is then predicted at the present position and heading, assuming that the aircraft will return and continue the route at that point.

The advantage of this technique is that the controller will continuously receive an indication of the separation on the TSD. When a heading has been instructed by the controller, he/she can watch the TSD and instruct the aircraft to proceed with the approach when sufficient separation is predicted. The increase in ATD depends on two variables: moment of return and deviation angle. A large deviation angle increases the ATD amount of time. An early moment of return will lead to a small increase in ATD.

Experiment

To test the effects of the TSD an experiment has been performed in which controllers perform their task both with and without the additional display. The experiment further aimed at determining the effects of different scenarios on controller performance and workload. An important goal of the experiment was to receive feedback from controllers on the adapted display design.

Method

Subjects and Instructions Eight professional air traffic controllers participated in the experiment. The experience of the controllers ranged from 18 to 38 years. During an extensive briefing, subjects were instructed on the airspace, approach procedure, and the TSD. Subsequently, 90 minutes of training with different display configurations and traffic rates prepared the controllers for the actual experiment. The task of the subjects was to provide ATC services to approaching traffic while *all* aircraft performed a CDA procedure. To do this efficiently and safe, the subjects could use the following instructions: Speed reductions to a minimum of 180 kts in steps of 10 kts, vectors, directs to a specified set of waypoints on the routes, and approach clearance.

Apparatus The apparatus is based on the air traffic simulation in(Tielrooij et al., 2008). For this project not only the TSD was further developed, the ATC simulator was improved as well.

Aircraft, Airspace and Atmosphere Point mass models of three different aircraft types were used: Airbus A321-100, Boeing B737-800 and B747-400, all at their maximum landing weights. To prevent the controllers from applying routine approach operations, learned through years of training and experience with a particular airspace, the airspace had to be unfamiliar. The airspace, airport and routes in the simulator were loosely based on the situation at Sydney's Kingsford Smith Airport, Australia. Approach routes were adapted to provide merging points and space for controller actions. The atmosphere was simulated as an International Standard Atmosphere (ISA) using the ISA relations up to 20,000 m and a logarithmic wind profile.

Independent Variables Two independent variables were defined: *Display configuration* (three levels: baseline PVD, PVD+TSD, and PVD+TSD$^+$) and *traffic rate* (three levels: 15 aircraft per hour, 30 aircraft per hour, and 35 aircraft per hour).

Experiment Design and Procedure After the briefing and training, the measurement phase started. The measurement phase consisted of three blocks; one block for each display configuration. The blocks were randomized using a Latin square matrix to eliminate learning and boredom effects. To further cancel out these effects, the scenarios within a block were randomized as well and breaks were added in between. Before the start of each scenario, controllers could observe the traffic for a while to adjust to the scenario. After each scenario, subjects had to fill in a NASA Task Load indeX (TLX) (Hart & Staveland, 1988) form. At the end of the entire experiment a questionnaire was completed by the subjects.

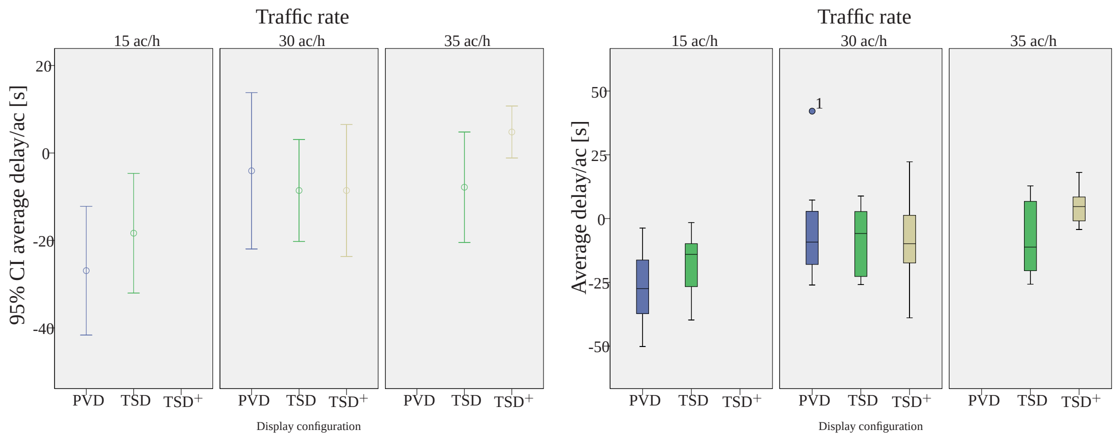
Dependent Measures The effects of the TSD were evaluated in terms of: 1) *performance*, measured by the difference between ETA and ATA and the number of instructions given by the controller to the blip driver, 2) *safety*, measured by the the number of conflicts (loss of separation), and 3) *workload*, measured using NASA TLX.

Results

Performance

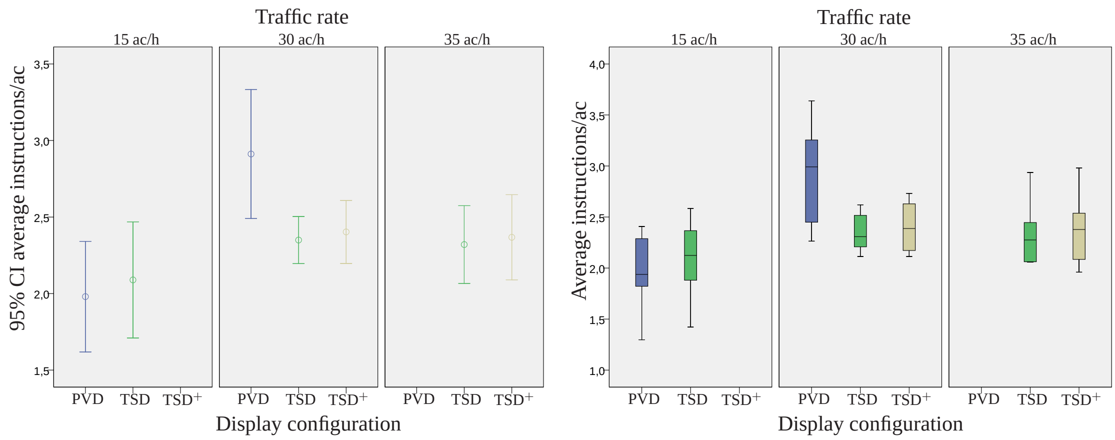
Delay Figure 3 shows the results for the average delay per aircraft. ANOVA results for the complete experiment show that the delay per aircraft was not significantly affected by an increase in traffic rate, $F_{2,12}$ = 2.240, $p > 0.05$. ANOVA results also show that no significant effect was found from the type of display on the delay per aircraft, $F_{2,14}$ = 0.117, $p > 0.05$.

Instructions Figure 4 shows the results for the number of instructions per aircraft. ANOVA results show that with an increase in traffic rate, the number of instructions per aircraft did not significantly change, $F_{2,12}$ = 1.250, $p > 0.05$. Mauchly's test indicated that the assumption of sphericity had been violated for the effects of display configuration, $\chi^2(2)$ = 7.680, $p < 0.03$. Therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = 0.581$). The results show that the number of instructions per aircraft were significantly affected by the display configuration, $F_{1.161,8.130}$ = 5.271, $p < 0.05$. A post-hoc test revealed that this effect can be found between two groups, the cases with additional display (TSD and TSD$^+$) and without additional display (PVD).



(a) Means and 95% confidence intervals (b) Boxplots

Figure 3: Results for the amount of delay



(a) Means and 95% confidence intervals (b) Boxplots

Figure 4: Results for the number of instructions

Safety

During the experiment only three actual conflicts occurred. Two of these, however, were caused by bugs in the simulator software, related to unstable predictions. On the TSD, the aircraft seemed to be in conflict, while on the PVD there was sufficient separation. Therefore, these two conflicts could be neglected. Hence, the results on safety indicate that the display configuration will have no influence on the amount of conflicts.

Workload

Figure 5 shows the results for the z-score of the weighted NASA TLX subjective workload assessment. ANOVA results show that an increase in traffic rate had a highly significant effect on the experienced workload, $F_{2,12}$ = 25.090, $p < 0.01$. A post-hoc test indicated that this effect can be found between the low (15 ac/h) and high traffic rates (30 and 35 ac/h). ANOVA results also show that with the change of display configuration the workload significantly decreased, $F_{2,14}$ = 9.944, $p = 0.02$. A post-hoc test revealed that this effect can be found between two groups, the cases with additional display (TSD and TSD^{+}) and without additional display (PVD).

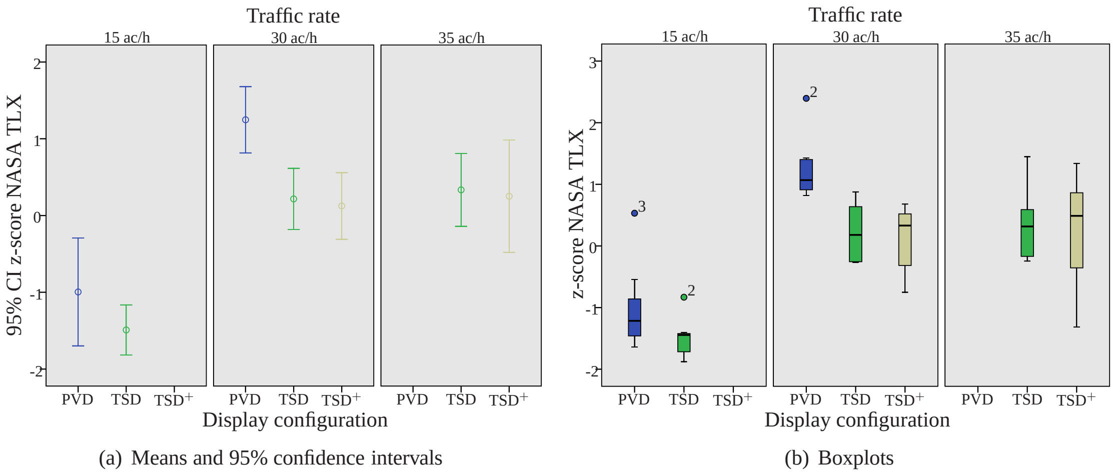


Figure 5: Results of the normalized NASA TLX ratings

Conclusions

By depicting the along track distance versus the estimated time of arrival, the Time-Space Diagram uses 4-dimensional trajectory predictions to present the in-trail separation between aircraft. An experiment was conducted to analyze the effects of the Time-Space Diagram on safety, controller performance and workload. Results show that the Time-Space Diagram significantly reduces controller workload and number of required instructions per aircraft for CDA procedures. No significant effects were found on delay and safety. The availability of additional predictions representing hypothetical instructions were found to have no significant results on the performance, safety and controller workload. Controllers who used these hypothetical predictions, however, stated that planning became easier.

References

Clarke, J.-P. B. (2000). Systems Analysis of Noise Abatement Procedures Enabled by Advanced Flight Guidance Technology. *Journal of Aircraft*, *37*(2).

Erkelens, L. J. J. (2002). Advanced Noise Abatement Procedures for Approach and Departure. In *AIAA Guidance, Navigation, and Control Conference and Exhibit.* Monterey, CA, USA.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload* (pp. 139–183). Amsterdam, The Netherlands: Elsevier Science Publishers (North Holland).

Reynolds, H. J. D., Reynolds, T. G., & Hansman, R. J. (2005). Human Factors Implications of Continuous Descent Approach Procedures for Noise Abatement in Air Traffic Control. In *6th USA/Europe Air Traffic Management Research and Development Seminar.* Baltimore, WA, USA.

Tielrooij, M., In 't Veld, A. C., Mulder, M., & Van Paassen, M. M. (2008). Assisting ATC in Metering, Sequencing and Merging Arrival Streams in Three-Degree Decelerating Approaches. *Proceedings of the 27th European Annual Conference on Human Decision Making and Manual Control, Delft, The Netherlands, June 12-13.*

Wickens, C. D. (1992). *Engineering Psychology and Human Performance.* New York, NY, USA: Harper Collins Publishers Inc. (Second edition)

MANEUVER STEREOTYPES IN AIRBORNE CONFLICT RESOLUTIONS

Esa M. Rantanen
Rochester Institute of Technology
Rochster, NY, USA
Christopher D. Wickens
Alion Science and Technology
Boulder, CO, USA
David Keller
New Mexico State University
Las Cruces, NM, USA

The roles and responsibilities of air traffic controllers and pilots are shifting in the advent of the NextGen air traffic management infrastructure, which also involves high levels of automation. It is important to understand just how large departures from the current ingrained practices the NextGen procedures represent, particularly in extremely safety-critical tasks such as airborne conflict resolutions. Pilots' conflict resolution maneuver preferences have received some attention, but corresponding research on air traffic controllers' practices is almost nonexistent. We analyzed 87 samples of aircraft track data involving conflict alerts and subsequent resolution maneuvers from Atlanta center. Vertical conflict resolution maneuvers were used in the majority of the cases examined. Within the vertical dimension, reductions of current vertical change (climb or descent) were collectively the most frequent resolution maneuver type, but descents were twice as frequent as climbs. Conflict resolution maneuvers furthermore do not seem to be independent from conflict geometries.

The NextGen air traffic control (ATC) and -management (ATM) technologies and procedures will fundamentally change the roles of pilots and air traffic controllers as well as their tasks and task environments. The putative increases in the system capacity and efficiency will be achieved through extensive use of automation, including automated conflict alerting. Although there is already much experience of such systems in both ATC (Conflict Alert, or CA; Nolan, 1998) and in cockpits (Traffic Alert and Collision Avoidance System, or TCAS; Bliss, 2003; Chappell, 1990), procedures for shared conflict avoidance and resolution are still being designed. As researchers and designers consider the implications of these procedures, it is imperative that they remain harmonious with controllers' current techniques of managing traffic and in particular with ingrained separation maneuvers used in response to potential midair conflicts. It is especially important to avoid generating pilots' rules-of-the-road for self-separation, or automation-based conflict avoidance advisories that are at odds with current ATC conflict avoidance techniques. This paper describes an effort to determine the maneuver stereotypes of en route controllers' responses to conflict alerts in operational conditions.

Pilots' Maneuver Preferences

A fair amount of research has been devoted to examining conflict resolution maneuver stereotypes among pilots (e.g., Alexander, Merwin, & Wickens, 2005: Thomas & Rantanen, 2006; Thomas & Wickens, 2008). This is primarily due to the criticality of collision-avoidance maneuvering in response to airborne collision detection systems (e.g., TCAS and Cockpit Display of Traffic Information, or CDTI), often under severe time constraints (seconds) and without coordination with ATC or other aircraft in the vicinity. Note, however, that TCAS always prescribes a maneuver, and it is assumed that if pilots respond at all, they will always comply in a vertical direction specified by the TCAS algorithm. This in contrast to the CDTI, where maneuver choice is up to the pilot as the direction of conflict resolution is not envisioned to be explicitly commanded by the automation. Research on CDTI generally reveals that pilots tend to prefer vertical maneuvers over lateral ones, reflecting perhaps the greater expediency and reduced complexity of such maneuvering (Thomas & Wickens, 2008). However the data are somewhat ambiguous regarding the extent to which the particular geometry of a conflict dictates the direction of a maneuver. Some faint trends were observed by Thomas and Wickens, revealing that head-on conflicts (versus crossing or overtaking) tended to induce relatively more lateral maneuvering.

Air Traffic Controllers' Maneuver Preferences

Investigation of conflict avoidance maneuver preferences among air traffic controllers is substantially more difficult than research on pilots' preferences. There are several good reasons for the paucity of empirical research on this topic, many of which emphasize the differences between pilots' and controllers' tasks and task environments, even at the dawn of the era of distributed control and shared separation responsibility. The main difference between pilots' and controllers' separation responsibilities under mature free flight or NextGen operations is that pilots are primarily concerned about their own aircraft and their attention extends little beyond those other aircraft that pose an immediate or near-immediate threat of loss of separation. Air traffic controllers must concern themselves with the 'big picture' and traffic flows rather than individual aircraft pairs. Their goal is to create conflict-free traffic flows such that they do not need to devote undue attention to individual conflicts. Also, controllers are always responsible for a much larger number of aircraft that any pilot in any situation, and effective management of their own workload is critical to their performance. However, the CA data examined here clearly represent exceptions to this general *modus operandi* and may reveal different conflict resolution patterns that are closer to pilots' demonstrated preferences.

There are very few laboratory studies on controllers' conflict resolution maneuver preferences, and results from such settings must be evaluated against the particular experimental conditions and airspace designs. Rantanen, Yang, and Yin (2006) examined controllers' separation preferences in a simulator study with a simplified airspace and traffic patterns. Vertical separation (i.e., commanding planes to climb or descend, or remain at an intermediate altitude) was much preferred over vectoring (lateral maneuvers). Controllers have also been shown to prefer the vertical dimension for separation in other contexts (Rantanen & Nunes, 2005). Vertical (altitude) or longitudinal (speed) maneuvers typically preserve traffic flows along regular routes and thus reduce the 'problem space' for the controller along those, simplifying their monitoring task. The maneuver preferences observed by Rantanen et al. (2006) undoubtedly reflect the very constrained and relatively small airspace used in the simulation with little room for vectoring (i.e., lateral maneuvering) and few opportunities for routing changes.

Purpose of the Study

In spite of the importance of this topic for aviation safety, and in particular for understanding how controller's tendencies may either reinforce or contradict pilot tendencies, no data appear to exist regarding the actual controller conflict avoidance behavior with live traffic. The intent of this paper is to bring insight to this process, using the operational ATC en route data from controllers responding to conflict alerts at five Air Route Traffic Control Centers (ARTCCs), with emphasis on one large one (see Wickens et al., 2008 for details). Our focus is on two key aspects of the data: (1) what tendencies to controllers show in terms of instructing lateral versus vertical maneuvers, and, within the latter category, climbs vs. descents, and (2) how are these tendencies mediated or influenced by the particular geometry of a conflict. A third aspect of these data pertaining to how controller's responses are mediated by alert reliability (e.g. false alert rate) is reported in detail by Wickens, Rice, Keller, Hughes and Hutchins (2009) in this volume

Method

This research was done on a subset of data from a greater research effort, involving a large set of conflict cases from five ARTCCs: Houston (ZHU), Indianapolis (ZID), Salt Lake City (ZLC), Los Angeles (ZLA), and Atlanta (ZTL). Much of the results reported in this paper is based only on the data from ZTL, which was the only center thus far receiving a geometry x maneuver contingency analysis. However, overall maneuver data (independent of contingency analysis) was available from all five centers, and are reported below. These data were originally provided to researchers at the Federal Aviation Administration (FAA) Technical Center (Allendoerfer, Friedman-Berg, & Pai, 2007) for analysis, and to us by the FAA in cooperation with National Air Traffic Controllers Association (NATCA). Data for each aircraft pair in conflict consisted of predicted point of closest approach, time of alert, and the radar tracks and altitudes of the aircraft, allowing for analyses of the actual conflicts as they were played out (see Wickens et al., 2008, for details). Alas, these data could not be linked to voice transcripts for additional detail (see Allendoerfer et al., 2007).

Each single conflict was defined as an encounter between two aircraft in a pair, which triggered at least one CA (repeated CA onsets for a given encounter might occur as a given pair went in and out of conflict). For each case, approximately six minutes' worth of actual track data of the two aircraft in conflict were recorded; these track data included the x- and y- (latitude and longitude) and z-(altitude) coordinates sampled every 10 seconds. These coordinates were plotted separately for horizontal and vertical trajectories, from where the conflict geometries and maneuvers performed to resolve the conflict could be determined visually. The geometries were classified into three vertical and three horizontal categories; in addition, five classes of maneuvers were defined.

Vertical geometry. The relative vertical behavior of the two aircraft in a pair prior to the alert was categorized as either converging vertically, where one aircraft was climbing and the other descending, or parallel climbs or descents, or both aircraft level.

Lateral geometry. Aircraft trajectories in the horizontal plane were classified into three categories, either converging, diverging, or parallel. In the case of parallel tracks, one aircraft was often overtaking the other. Parallel approaching tracks (e.g., near opposite headings) were classified as 'converging'. Note that diverging lateral tracks could trigger a CA if these involved more rapid convergence on the vertical axis, such that an LOS on the altitude dimension (< 1000 ft) would occur before separation on the lateral dimension (5 miles) is obtained. Note also that vertical and lateral geometries were both applied to every conflict.

Maneuvers. Maneuver type was subdivided into five classes: descend, reduce descent for an aircraft already in descent (e.g., a level off of a descending aircraft), climb, reduce climb, and turn. Either an increase descent or increase climb was simply categorized as a descent or climb, respectively. In the case of joint maneuvers in both the lateral and vertical axes, the CA was assigned to the category of that maneuver which occurred first (earliest). We also reiterate that inferences of an instructed maneuver were made solely from trajectory changes following the CA, since we had no direct access to corresponding voice transcripts (but see Allendoerfer et al., 2007; Friedman-Berg, Allendoerfer, & Pai, 2008). These data were tabulated and analyzed by χ^2-tests for independence.

Results

The maneuvers controllers instructed in response to the impending conflict or as prompted by the CA, as inferred from the aircraft trajectory plots, are depicted in Table 1. A χ^2 goodness-of-fit test on all maneuvers across the five centers (with a null hypothesis of equal proportions of maneuvers) showed significant differences between the five classes of maneuvers, $\chi^2(4, N = 277) = 60.38$, $p < .001$.

Although turns constitute the most frequent single category (36%), these lateral maneuvers occurred much less frequently than those involving vertical trajectory change. This *vertical maneuver domination* was similar to that observed in the previous report and is also consistent with the integrated findings of studies of aircraft (e.g., pilot initiated) conflict avoidance (Thomas & Wickens, 2008).

Descents were commanded twice as frequently as climbs (7% vs. 14%), but modifications to vertical transitions already in progress were equally divided between reductions of climbs and reductions of descents. Collectively, the latter were the most frequent maneuvers. These trends may reflect controllers' concern of the overall fuel efficiency of flights; descending an aircraft is much more fuel efficient than climbing the aircraft beyond its planned altitude, and reductions to climbs and descents already in progress are minimally disruptive to pilots.

Table 1. *Maneuver frequencies across five ARTCCs from where conflict resolution data were obtained.*

	Maneuver										
Center	Climb		Descend		Reduce Climb		Reduce Desc.		Turn		All
ZID	3	(3.37%)	24	(26.97%)	22	(24.72%)	14	(15.73%)	26	(29.21%)	89
ZHU	1	(3.70%)	2	(7.41%)	8	(29.63%)	3	(11.11%)	13	(48.15%)	27
ZLA	3	(5.88%)	4	(7.84%)	10	(19.61%)	4	(7.84%)	30	(58.82%)	51
ZLC	3	(13.64%)	1	(4.55%)	7	(31.82%)	9	(40.91%)	2	(9.09%)	22
ZTL	11	(12.50%)	9	(10.23%)	10	(11.36%)	30	(34.09%)	28	(31.82%)	88
All	21	(7.58%)	40	(14.44%)	57	(20.58%)	60	(21.66%)	99	(35.74%)	277

Our in-depth analysis of contingency between geometry and maneuver was carried out only on 97 CA cases from ZTL. For nine of these there was no maneuver, suggesting that these were false alarms. This roughly 10% non-response rate parallels that reported in the full data set of 497 CA's; the reasons for this are discussed in the Wickens et al. (2009) paper in this volume. Given that the remaining 89 cases involved CA, we expected that most aircraft trajectories would converge. Indeed, a total of 71 aircraft pairs were on either horizontally or vertically converging trajectories and 54 were converging both horizontally and vertically. Conflict resolution maneuvers were more evenly distributed among the maneuver classes. In the majority of cases, controllers either restricted an aircraft's climb (N = 30) or turned the aircraft (N = 28). These data are consistent with the full analysis of the larger 5-center data set, which revealed that vertical maneuvers dominated turns and within the former, reduced climbs were the most prevalent. In particular maneuvers exploiting gravity (reduced climbs and descents) dominated those opposing gravity (climbs and reduced descents) by a ratio of over 2:1.

Contingency Between Geometry and Maneuver Types.

We have discussed the 'main effects' of maneuver type and geometry above (e.g., analyzing the frequency of these categories, independent of the other). Here we focus our discussion on the interaction, or contingencies between the geometry, as perceived by controllers on their display, and the types of maneuvers that were instructed. We examined these contingencies by χ^2 tests for independence. Two contingency tables were created for vertical and lateral geometries and corresponding maneuvers and their combinations. To create these tables we used the three vertical conflict geometry classes and collapsed maneuver classes also into three: turn, [climb or reduce descent], and [descent or reduce climb], for a 3 x 3 table. The rationale for collapsing within the vertical maneuvers was the commonality of the two that worked against gravity, and the two that worked with gravity, as described above.

The results for the vertical geometries approached significance, $\chi^2(4, N = 87) = 8.67$, $p = .069$. The cause of this non-independence is apparent from the data in Table 2; while climbs and reduced descents made up approximately 22% of all maneuvers, these were particularly unlikely to occur in converging vertical geometries (N = 8; 14% of the time). They were also overall disproportionately rarer than other maneuvers, possibly reflecting their fuel inefficiency and disruptive nature for pilots.

Table 2. *Counts of different maneuvers by vertical conflict geometries (expected values in parentheses).*

	Maneuver						
Vertical Geo.	Climb		Desc.		Turn		Total
Converging	8	(13.333)	28	(26.00)	22	(18.667)	58
Level	4	(2.299)	4	(4.483)	2	(3.218)	10
Parallel	8	(4.368)	7	(8.517)	4	(6.115)	19
Total	20		39		28		87

Similarly, three horizontal geometries (converging, diverging, and parallel) were analyzed against the aforementioned three maneuver categories in another 3 x 3 table (Table 3 below). The results were not significant, $\chi^2(4, N = 86) = 3.72$, $p = .44$, but there appears to be a certain degree of dependence between lateral geometry and maneuver tendencies. Turns were much more frequent in converging than in parallel geometries (35% vs. 20%). In both of these analyses some very small expected values (< 5) are noteworthy.

Table 3. *Maneuver counts by horizontal conflict geometries (expected values in parentheses).*

	Maneuver						
Horizontal Geo.	Climb		Desc.		Turn		Total
Converging	15	(14.419)	25	(28.116)	22	(19.465)	62
Diverging	0	(0.930)	3	(1.814)	1	(1.256)	4
Parallel	5	(4.651)	11	(9.069)	4	(6.279)	20
Total	20		39		27		86

We performed one more analysis on combinations of vertical and horizontal geometries (converging—converging, converging—nonconverging, nonconverging—converging, and nonconverging —nonconverging) against the aforementioned three maneuver categories in a 4 x 3 table (see Table 4 below). The test for non-independence was significant, $\chi^2(6, N = 86) = 13.43$, $p = .036$. Turns, representing only 31% of the maneuvers overall, were disproportionately more frequent on geometries with convergence in both axes (40%). Here, we encountered some very small expected values.

Table 4. *A contingency table for combinations of vertical and horizontal conflict geometries and corresponding resolution maneuvers (expected values in parentheses).*

	Maneuver						
Vertical, Horizontal Geometry	Climb		Desc.		Turn		Total
Converging—Converging	7	(10.93)	21	(21.31)	19	(14.76)	47
Converging—Nonconverging	1	(2.33)	7	(4.53)	2	(3.14)	10
Nonconverging—Converging	8	(3.49)	4	(6.80)	3	(4.71)	15
Nonconverging—Nonconverging	4	(3.26)	7	(6.35)	3	(4.40)	14
Total	20		39		27		86

Discussion

In ATC workload management is one of the most critical skills for a successful controller. Consequently, controllers' techniques exhibit certain economy. For example, maintenance of traffic flows is less mentally taxing than keeping track of individual aircraft, and vertical maneuvering is less disruptive to traffic flows than lateral maneuvering. Hence, our results are not entirely surprising: vertical conflict resolution maneuvers (climb, descend, restrict climb or descent) were used in the majority of the cases we have examined. Such maneuvers are often the best solutions to conflicts, especially if the aircraft involved are already in vertical transition. Indeed, reductions of current vertical change (climb or descent) were collectively the most frequent resolution maneuver type. On the other hand, climbs and restricted climbs were the least frequent maneuvers overall in all of our analyses, reflecting the disruptive nature and fuel inefficiency of such maneuvers working against gravity. In the few conflict geometries where they were used in the majority of cases, the difference to other maneuver types was very small. Within the vertical dimension, descents that exploit gravity were twice as frequent as the climbs that oppose it.

We also discovered some indications that conflict resolution maneuvers are not independent from conflict geometries preceding them. Climbs or restricted descents were disproportionately rare in vertically converging geometries, while turns, despite their overall small proportion were frequently employed in resolution of conflicts with converging geometries. We expect these trends to become more salient when the full data set from all five ARTCCs is analyzed, and in much greater detail than was possible here.

It should be kept in mind that 86 cases is not a very large data set when it is divided into 9 or 12 cells in contingency tables. However, the trends apparent in the raw numbers are quite clear and robust. The results reported here are only the first fruits of a continuing research effort, however. We are performing similar analyses on the data from all five centers, and expect to gain a much more detailed insights into controllers' maneuver choices as well as statistically more significant results than here, with only about 20% of the data analyzed. Categorical analysis is common and undeniably valuable way to examine safety data, but its limitations must be acknowledged. Conflict geometries exhibit enormous variability and any classification system necessarily includes very different situations warranting different maneuver choices into the same categories. While this will be less of a problem with the full, 5-center data set, we are also going to treat geometries as a continuous variable allowing more fine-grained measurement of their characteristics.

Finally, we would like to make a case for detailed analysis of operational data, which can reveal patterns and behaviors that could never emerge in simulated laboratory experiments. Routine access to data such as reported here is crucial for the research community to keep up with and contribute to the development of the NextGen systems.

Acknowledgments

This research was supported by grant # OGC #20070271: ADDRESSING THE ALERT PROBLEM IN AIR TRAFFIC CONTROL FACILITIES from the FAA to New Mexico State University. The authors wish to thank Kenneth Allendoerfer, Ken Leiden, Stephen Rice, Krisstal Clayton, and Jill Kamienski for their contributions.

References

Alexander, A. L., Merwin, D., & Wickens, C.D. (2005). Perspective and Co-planar cockpit display of traffic information: implications for maneuver choice, flight safety and mental workload. *International Journal of Aviation Psychology, 15*(1), 1-21.

Allendoerfer, K., Friedman-Berg, F., & Pai, S. (2007). *Human factors analysis of safety alerts in air traffic control* (DOT/FAA/TC-07/22). Atlantic City, NJ: FAA William J. Hughes Technical Center.

Bliss, J. O. (2003). An investigation of alarm related incidents in aviation. *International Journal of Aviation Psychology 13*(3), 249-268.

Chappell, S. L. (1990). Human factors research for TCAS. *Proceedings of the Royal Aeronautical Society Conference on the Traffic Alert and Collision Avoidance System (TCAS)—Its development, testing, Installation and Operational* Use (6-1/6-19).

Friedman-Berg, F., Allendoerfer, K., & Pai, S. (2008). Nuisance alerts in operational ATC environments: Classification and frequencies. *Proceedings of the 52nd Human Factors and Ergonomics Society Annual Meeting* (pp. 104-108). Santa Monica, CA: HFES.

Nolan, M. S. (1998). *Fundamentals of air traffic control (3rd ed.).* Pacific Grove, CA: Brooks/Cole.

Rantanen, E. M., & Nunes, A. (2005). Hierarchical conflict detection in air traffic control. *International Journal of Aviation Psychology, 15*(4), 339-362.

Rantanen, E. M., Yang, J., & Yin, S. (2006): Comparison of pilots' and controllers' conflict resolution maneuver preferences. *Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 16–19). Santa Monica, CA: HFES.

Thomas, L. C., & Rantanen, E. M. (2006). Human factors issues in implementation of advanced aviation technologies: A case of false alerts and cockpit displays of traffic information. *Theoretical Issues of Ergonomics Science, 7*(5), 501-523.

Thomas, L.C., & Wickens, C.D. (2008) Display dimensionality and conflict geometry effects on maneuver preferences for resolving in-flight conflicts. *Human Factors, 50,* 576-588

Wickens, C. D., Rice, S., Keller, D., Hughes, J., Hutchins, S. (2009). Conflict alerts and false alerts in en-route air traffic control: An empirical study of causes and consequences. *Proceedings of the 15th International Symposium of Aviation Psychology*. Dayton, OH, April 27–30, 2009.

Wickens, C. D., Rice, S., Keller, D., Hughes, J., Hutchins, S., & Clayton, K. (2008). *Addressing the alert problems in air traffic control facilities (Part 2: Final report)* (Technical Report 2008-10-2). Las Cruces, NM: New Mexico State University.

CONFLICT ALERTS AND FALSE ALERTS IN EN-ROUTE AIR TRAFFIC CONTROL: AN EMPIRICAL STUDY OF CAUSES AND CONSEQUENCES

Christopher D. Wickens
Alion Science: MA&D Operations, Boulder, Colo.
Stephen Rice
New Mexico State University
David Keller
New Mexico State University
Jamie Hughes
New Mexico State University
Shaun Hutchins
Alion Science: MA&D Operations, Boulder, Colo.

We analyzed the extent to which a high false alert rate of the conflict alerting (CA) system in five ATC facilities was the cause of a "cry-wolf" effect, whereby true alerts of a pending loss of separation were associated with either controller failure to respond or a delayed response. Radar track data surrounding 497 CA's were examined and from these we extracted information as to whether the alert was true or false, whether a trajectory change was (response) or was not (non-response) evident, whether a loss of separation occurred, and the controller response time to the CA. Results revealed an overall 47% false alert rate, but that increases in this rate across facilities was not associated with more non-responses or delayed responses to true alerts, or loss-of-separation. Cry-wolf appeared to be absent. Instead, desirable anticipatory behavior indicated that controllers often responded prior to the conflict alerts.

In June 2006, the National transportation and Safety Board documented a series of accidents - controlled flight into terrain, and mid-air collisions - in which the minimum safe altitude warning (MSAW) and conflict alerts, respectively, announced a pending collision (NTSB, 2006). However, controllers failed to respond or intervene to prevent the accident. Furthermore, anecdotal evidence from a specific accident (midair collision of two Cessna aircraft near San Diego), and from other interviews with controllers (Ahlstrom & Pasnjwani, 2003), suggested the prevalence of controller experience of the "cry wolf" effect (Breznitz, 1983). The "cry wolf" effect is a general syndrome whereby excessive alarms, many of them seemingly unnecessary to the operator (e.g., "false alarms" or "false alerts"), lead to a distrust in the alarm system, and a disregarding of (or late response to) some true alarms.

Linking this well observed phenomenon to the findings of missed alerts in the NTSB study suggests that there may be a causal connection between the two. When examining false alerts in predictive collision alerting systems, certain features of time-dependence (Kuchar, 2000) make these different from other alerts such as cockpit engine warnings (Dixon & Wickens, 2006). In particular, inherent in any dynamic system in a noisy environment subject to cross winds, turbulence, and pilot control inputs, is the problem that prediction becomes less accurate with increasing look-ahead time. Furthermore, an alert may be "false" for two reasons; it may actually predict a loss of separation but extrapolation of the trajectory indicates that an LOS will not occur; or it may correctly predict an LOS, but a subsequent trajectory change (in response to a controller instruction) is implemented so that no LOS is observed. Finally, true "misses" are very rare in CA systems; but these are more often manifest as delayed alerts. Clearly if the alert is delayed so long that there is little time to maneuver the aircraft away from the separation loss, such an event can be seen as equivalent to a miss.

A general conclusion from research which has examined false alerts, when humans can monitor the data in parallel appears to be that, while misses may be catastrophic in a system in which there is no human backup to monitor the raw data in parallel, in systems that allow such parallel human-machine monitoring (Parasuraman, 1987), false alarm-prone systems may often be worse (Dixon & Wickens, 2006; Dixon, Wickens & McCarley, 2007; see Wickens, Levinthal, & Rice, in press, for a summary). This may be particularly true in high workload multi-task circumstances since a false-alarm prone system may not only cause ignorance of true alerts, but, when such alerts are responded to, this response can be quite disruptive of concurrent tasks; either as a result of carrying out the unnecessary alarm-triggered action or of the need to cross check the raw data to establish that the alarm was indeed false. In a further argument for a higher threshold, in many predictive alerting systems such as the conflict alerts studied here, a higher threshold does not necessarily translate to more missed events, but only to a later alerting of true events (a much less catastrophic outcome). Indeed if this alerting look-ahead time still provides adequate time for humans to respond, then the benefit of reducing false alarms would more than offset the shorter time period between the alert and the occurrence of the forecast event (e.g., the pending collision).

The purpose of the current study was to seek evidence for the FA-caused cry-wolf phenomenon from live or "naturalistic" data across five air traffic control facilities in which controllers responded to mid-air conflict alerts (CA's), and across which the CA false alert rate varied. Such live ATC data have never before been examined in this fashion; although it parallels the analysis of weather forecasters (Barnes, et al., 2006), pilots (Bliss, 2004), and health care practitioners (Xiao, et al., 2004), responding to imperfect alerting systems. In this process we must first examine performance of the CA system itself, to assess a FA rate, and then examine the influence of differences in this rate on behavior of the controller, and performance of the controller-CA (human-automation) system as a whole.

In the current research, we addressed the hypothesis that, assuming there to be variability in false alert rate across ATC facilities, those with the higher FA rate, would show greater evidence for "cry wolf" behavior: later responses, and/or more non-responses. In addition, we examine other aspects of controller response to CA's that are either true or false; in particular considering the properties of the alerting system that may lead to a loss of separation, and/or lead to desirable anticipatory behavior.

Methods: CA system analysis

The CA system (FAA, 2003) is designed to predict when two aircraft will close simultaneously, within 5 miles laterally and 1000 feet vertically. Figure 1 presents the schematic for the lateral dimension only. Such closure is known as a **loss of separation (LOS),** shown on the left of figure 1. Hence the CA predicts any LOS that is forecast to occur within a look-ahead time of 75-135 seconds. When the CA system predicts such an LOS, the data tags on the controllers' display start to flash. The algorithm underlying the CA generates a linear extrapolation both on the horizontal (map) plane and the vertical plane, of the current heading and vertical speed of both aircraft (FAA, 2003).

We were provided data from the FAA for 494 conflict alerts, extracted from the busiest 2-hour periods from a sample of 2 or 3 days in each of five en-route ATC centers. Such data (distributed across three different data bases) included for each CA: (1) properties of the pair of trajectories predicted by the CA (e.g., predicted point of closest passage, time of alert), (2) the actual radar tracks & altitude of the aircraft (sampled every 10 sec), and (3) a short analysis of the actual conflict as it was played out (See Wickens, Rice, et al., 2008, for details). The most important element of this third set was a metric (minmax ratio or MMR) describing the inverse severity of the conflict. A value of 0 corresponded to an actual collision and a value of 1 was the threshold for a loss of separation. Higher values indicated passage with greater lateral and vertical separation than the minima. Two key variables provided to us for each center were the "busyness" of the center (the number of encounters per hour (where "encounter" is the point at which the CA algorithm begins to examine track pairs), and the number of CA's during the equivalent period. Table 1 shows these two parameters across the five Centers (row 2 and 3) along with the ratio in row 4 of the total CA's to the total encounters within the center; an estimate of the CA rate. Importantly, Table 1 reveals that what might be defined as the "CA-rate" in the bottom row did not vary substantially across Centers, in spite of the 8-fold variation in "busyness".

Table 1. *Basic data from CA systems.*

ZLC	ZHU	ZLA	ZTL	ZID
1126	1,589	5,529	5679	8,813
22	36	72	435	124
22/4525=.005	124/26440=.004	36/4767=.007	235/38815=006	72/16589=.004

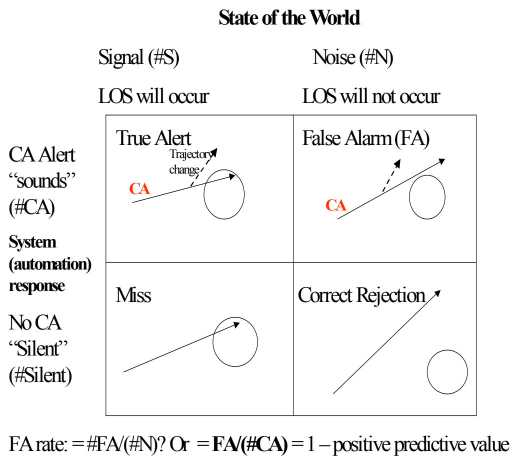


Figure 1: Geometry of CA system shown when an LOS is pending (left column) and not (right), and when the alert is triggered (top) or not (bottom). Within the triggered alerts, controllers may (dashed line) or may not (solid line) initiate a maneuver.

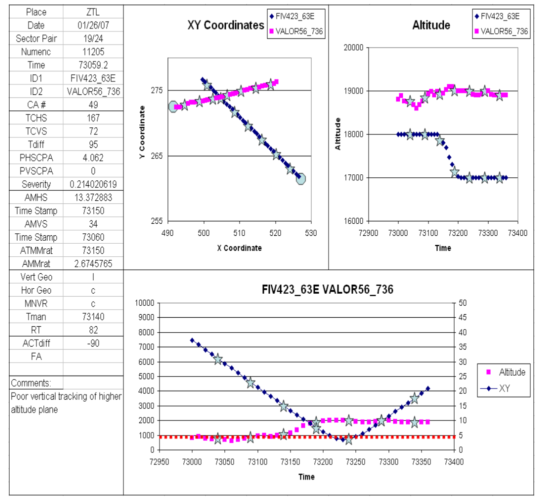


Figure 2: Example of radar tracks used to classify alert as true or false, and to identify controller response. A trajectory change (descent of blue) is clearly visible in the vertical track at the upper right. The lower plot depicts separation on both axes.

Results 1: CA system analysis.

For the CA system, we calculated the false alert rate as the proportion of alerts that were categorized as "false". As noted above, we also distinguished between alerts (both true and false) when controllers did and did not impose a trajectory change, as such change was inferred from the radar tracks. When a change was implemented, a visual analysis was used to extrapolate the pre-change course, to assess if an LOS **would have** simultaneously compromised lateral and vertical minima, had the alteration **not** taken place. This analysis was carried out on ground track data, an example of which is shown in figure 2. The analysis was carried out by two independent observers for two of the centers, and by one of these observers for the remaining three.

We examined the computed FA rate as a function of the CA rate (CA/encounter) across the five centers. Two features became evident from this examination. First, there was considerable variance in FA rate, from a low of 0.28 to a high of 0.58, and on average approximately half of the CA's were "false". This allowed us to examine the cry wolf effect. Second, there appeared to be no relationship between CA rate and FA rate. A separate analysis also revealed that FA rate did not co-vary with overall traffic density. We also analyzed and categorized the geometry of the trajectories of the pairs of aircraft entering into each CA, and of the controller responses (e.g., climb, turn); these analyse can be found in Rantanen & Wickens (2009); and Wickens Rice et al (2008).

Results 2: Controller performance Analysis

Categorical analyses. Before examining the influence of FA rate on manifestations of the cry wolf phenomenon, it is necessary to identify the overall prevalence of those manifestations in our sample of data. Thus, in addition to the dichotomization of true versus false alerts discussed above, - characteristics of the automation - we examined two other important dichotomies which are characteristic of the human (controller): the presence or absence of a response (as inferred from visual analysis of the track data), and the presence or absence of a loss of separation (LOS). As noted in the previous section, it was usually relatively easy to identify whether a distinct change in trajectory was implemented in the time period around a CA (see the descent of the blue aircraft in figure 2), hence allowing inference of the presence and delay of a controller response. However, for a small sample, this classification became quite difficult and so those trials were not included in the data base.

Our analysis revealed that on roughly 10% of the CA's there was no evidence for a controller response, at least as indicated by a trajectory change by either of the two aircraft involved in the CA. These non-responses were statistically more prevalent when the CA was false (18%) than when it was true (1.5% $\chi^2(1, N = 437) = 37.5$, $p < .0001$). Such a result might be anticipated to the extent that the trajectories triggering a false alarm are, by definition, more likely to yield a more distant "closest passage" or miss distance and hence more likely to be considered by the controllers not to require a trajectory change.

Our analysis also revealed that the LOS rate is, like the non-response rate, approximately 10% of the data base. Also, it appears that the two types of outcomes are unevenly distributed across the two types of alerts. Specifically, True alerts are more likely to precede a loss of separation (21%) than are false alerts. (3%; $\chi^2(1, N = 373) = 20.3$, $p < .0001$) Here too, this is a plausible outcome, given that the true alert will occur on a trajectory pair that is more dangerous, and hence slightly more likely to yield the ultimate loss of separation, even *with* a controller intervention.

We then examined the relationship between controller response and LOS, to establish the extent to which non-responses might be associated with a LOS. These observations are collapsed over true vs. false alerts. This analysis indicated that when the controller **did not** respond, this was very unlikely to produce an LOS (5%; and those two events were restricted to a single center), whereas such LOS events were somewhat more prevalent when the controller **did** respond (9%) although the difference in proportion was not significant. ($\chi^2(1, N = 380) = .778$, $p < .378$). We note here that this finding does not necessarily imply that controller responses were counter-productive, since the vast majority of LOS cases occur on true alerts, where there would definitely have been an LOS had the controller not intervened with a trajectory change.

Collectively, the above three analyses provide no evidence for the strongest form of cry wolf effect (non-response leading to a LOS) and indeed the number (2) of such joint events is even less than what the independent product of the two classes of events might predict (10% NR rate X 10% LOS rate = 1% of the CA events = 4). We next sought to determine if there was any causal relation between FA rate, *as it varied across centers*, and either non-responses or LOS events. Figure 3 shows the scatter plot of FA vs. non-response, and reveals a striking and pronounced trend: the greater the false alarm rate in the center, the less controllers tended to respond (r = 0.944; p<.05). However, when the LOS rate was examined as a function of FA rate across Center, there was no trend. This null effect suggests that the increase in non-responses in the more FA-prone Center were not associated with a reduction in safe separation.

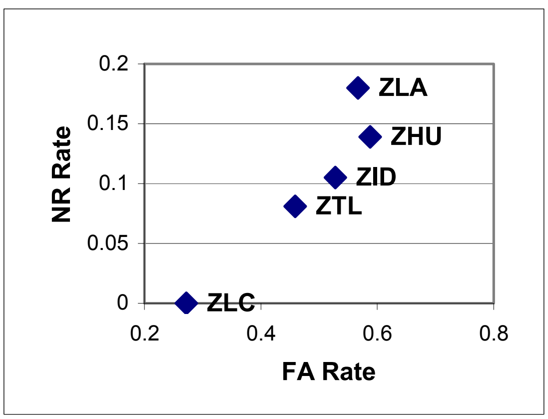


Figure 3: Non response rate as a function of false alert rate.

*Response Time***.** We then examined the second manifestation of the cry wolf phenomenon: the possible delay in controller response time associated with more FA's. Interpreting the delay between the CA and the trajectory change response required consideration of the total transmission lag (TTL).This is the time for the following processes to occur: (1) controller notices a dangerous convergence; (2) controller chooses a trajectory change and communicates this to the pilot; (3) pilot confirms and implements the change with flight controls; (4) the aircraft initiates a sufficient trajectory change to be evident in the radar track This TTL is estimated to be approximately 20-25 seconds (Allendoerfer & Friedman-Berg, 2007). Our analysis revealed that for about 45% of the CA's, controllers must have initiated the trajectory processing (noticing convergence and choosing a maneuver) **before** the CA occurred because the RT was less than 25 seconds. Indeed when we examined the distribution of response times, relative to the CA, we observed a distinct bimodality, with a minimum at around 25 seconds (See Wickens, Rice, et al., 2008). This bimodality, coupled with the estimate of a 25 second TTL, supported the notion that there were two categorically different types of responses: which we labeled **anticipatory** responses and **reactive** responses.

An ANOVA carried out on the ln-transformed RT data indicated that, for anticipatory responses, there was no difference in response time between true and false alerts ($p > .10$); however for reactive responses, true alerts were responded to approximately 73-59=14 seconds more rapidly than false responses ($t(193) = 2.4$, $p < .02$), reflecting the increased urgency of the former. There was no significant difference in RT between LOS and non-LOS encounters, so we can reject the hypotheses that the former resulted because of a delay in response time.

An analysis of three centers' data did reveal a main effect of center, $F(2, 154) = 6.78$, $p < 0.01$, with the highest density center (ZID) showing faster responses (30 s) than either the low (33 s) or mid (36 s) density Centers, an effect observed for both anticipatory and reactive responses. This effect is noteworthy because, whereas increasing density might have been anticipated to slow RT because of greater workload, the faster RT for ZID was observed **despite** its greater traffic density (See Table 1). Finally, within the non-LOS encounters (MMR > 1.0), we correlated RT with the value of MMR to test if later responses were responded with closer (but still above minima) passages. This correlation, examined for the three mid-level Centers was non significant ($p > 0.10$), suggesting that controllers did not compromise safety when their responses were delayed.

Finally, we examined the frequency of anticipatory vs. reactive responses for true vs. false alarms. Analyses of these data reveals that controllers were significantly more likely to anticipate on a true (0.58) than a false (0.37) alert ($\chi^2(1, N = 374) = 5.08$, $p = .024$). This is a plausible finding because the true alert trajectories should signal the impending conflict with greater salience in the raw data of the radar displays.

Discussion

The current data provide little or no evidence that the FA-induced cry wolf effect exists for the en-route CA system, as it is operationally defined by non-response to true alerts, and by later responses to all alerts. More particularly, false alerts do not appear to be responsible for safety-compromise in the ATC centers whose data were sampled. The generality and robustness of this conclusion is supported both by the wide range of center busyness, as well as the large sample size of the data, which provides for powerful statistical conclusions. (That is, the null hypothesis was not accepted simply because of a small N).

Of course ours was not a true experiment with control exerted across all other aspects of the centers. As in any correlational study, confounding variables could have contributed to our results. One such potentially confounding variable is that traffic-induced workload differences between centers could have accounted for effects. Indeed while this is possible, two factors mitigate concern for this confounding interpretation of the result. First, the busiest center (ZID) was only in the middle of the range in terms of both false alerts and non-responses (Figure 3). If we assume busyness is a proxy for workload, then this result would not have been obtained had workload been a responsible factor. Second, the possible confound with workload would have been more problematic had we found that a higher FA rate was associated with more non-responses to true alerts, and/or late responses. In that case we would need to reason as to why workload was not responsible for the effect. But as noted, neither of these associations were observed.

In terms of why FA-induced cry wolf behavior did not appear to be observed here, we note that, while false, most of the alerts in the CA system were not wildly off the mark, and thereby signaled a system whose threshold was set just a little lower than it needed to be, in the conservative interests of preserving safety and avoiding misses or late alerts. Recently Lees and Lee (2007) have found that such alerts can actually be beneficial to performance, in confirming that the system is generally functioning well. In the current case, for the large number of anticipatory responses, one can think of the alerting systems reinforcing the conflict predictions (and trajectory alterations) that the controllers actually made in advance of the alerting system warning.

Acknowledgments.

This research was supported by grant # from the FAA to New Mexico State University. The authors wish to thank Kenneth Allendoerfer, Ken Leiden, Krisstal Clayton, and Jill Kamienski for their contributions.

References

Ahlstrom, V., Longo, M. , & Truitt, T. (2005). *Human factors design guide (DOT/FAA/CT-02-11).*Atlantic City, NJ: Federal Aviation Administration.

Allendoerfer, K. & Friedman-Berg, F. J. (2007). *Human Factors analysis of safety alerts in air traffic control*. Final Report DOT/FAA 07/22. Washington D.C.: Federan Aviation Agency.

Barnes, L. R., Gunfest, E., Hayden, M. H., Schultz, D. M., & Benight, C. (2006). False alarms and close calls: A conceptual model of warning accuracy. *Weather and Forecasting, 22,*1140-1147.

Bliss, J (2003). An investigation of alarm related accidents & incidents in aviation. *International Journal of Aviation Psychology, 13,* 249-268.

Breznitz, S. (1983). *Cry-wolf: The psychology of false alarms*. Hillsdale, NJ: Lawrence Erlbaum.

Dixon, S. R. & Wickens, C.D. (2006). Automation reliability in unmanned aerial vehicle flight control: A reliance-compliance model of automation dependence in high workload. *Human Factors, 48*, 474-486.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse the misses? *Human* FAA (2006).

FAA (2003). Common ARTS computer program functional specification for conflict alerts. NAS-MD-632. Wash DC.: Federal Aviation Administration.

FAA (2007) Human Factors Study of Air Traffic Control Safety Alerts White paper and progress report. Washington DC: Federal Aviation Administration (2007).

Kuchar, 2001, March 27-29). *Managing uncertainty in decision-aiding and alerting system design*. Paper presented at the 6th CNS/ATM Conference, Taipei, Taiwan

Lees, N. & Lee, J. D. (2007). The influence of distraction and driving context on driver response to imperfect collision warning systems. *Ergonomics, 30,*1264-1286.

NTSB, (2006) National Transportation Safety Board (NTSB) Safety Recommendation: A-06-44 through –47. Washington D.C.: NTSB.

Rantenen, E. & Wickens, C.D. (2009) Effects of conflict geometry on controller maneuver selection for conflict alerts. *International Symposium on Aviation Psychology*. Dayton: Wright State University

Sorkin, R.D. (1989). Why are people turning off our alarms? *Human Factors Bulletin*, *32*(4), 3-4.

Wickens, C. D., Rice, S., & Levinthal, B. (2009). Imperfect reliability in unmanned air vehicle supervision and control. In A. W. Evans (Ed.), *Human-robotics interaction in future military systems.*Brookfield, VT: Ashgate.

Wickens, C.D., Rice, S. et al (2008). Addressing the alert problem in ATC Facilities: Final report. Report 2008-10-2. Las Cruces NM.: New Mexico State University.

Xiao, Y, Seagull, F.J., Nieves-Khouw, F. Barczak, N., & Perkins, S. (2004). Organizational-historical analysis of the "failure to respond to alarm" problems. *IEEE Transactions on systems, man & cybernetics – part A: systems & humans. 34,* November; pp 772-776.

PREDICTING THE UNPREDICTABLE:
ESTIMATING HUMAN PERFORMANCE PARAMETERS FOR OFF-NOMINAL EVENTS

Becky L. Hooey[1], Christopher D. Wickens[2], Ellen Salud[1], Angelia Sebok[2], Shaun Hutchins[2] & Brian F. Gore[1]

[1]San Jose State University at NASA Ames Research Center
Moffett Field, California

[2]Alion Science and Technology
Boulder, Colorado

A parameter meta-analysis was conducted to characterize human responses to off-nominal events. The probability of detecting an off-nominal event was influenced by characteristics of the off-nominal event scenario (phase of flight, expectancy, and event location) and the presence of advanced cockpit technologies (head-up displays, highway-in-the-sky displays, datalink, and graphical route displays). The results revealed that the presence of these advanced technologies hindered event detection reflecting cognitive tunneling and pilot complacency effects.

The next generation of the National Airspace System (NextGen; JPDO, 2007) is expected to require new technology to enable operations such as flexible 4-D trajectories, closely spaced parallel approaches, reduced aircraft wake vortex separation standards, equivalent visual operations, precision spacing and merging, and tightly-coordinated taxi operations. Some of the flight deck technologies that are anticipated with the transition to the NextGen include the use of head-up-displays (HUDs), highway-in-the-sky (HITS) displays, datalink, and graphical routing information. To ensure that these new technologies and operations are robust to system perturbations (Burian, 2008), it is important to ensure that they support pilot performance in both nominal and off-nominal conditions. Off-nominal conditions may range from 'less-likely but necessary' operations that are slightly outside the range of normal operations (such as conflict alerts and unpredicted weather events), to very rare events (such as aircraft trajectory blunders and equipment failures). An inappropriate response to an off-nominal event can lead to a cascading effect in the system and disrupt the entire airspace flow. Therefore, a challenge facing the aviation research community is the need to predict pilot performance in the face of off-nominal events.

Due to the unexpected nature of off-nominal events, the opportunities to collect pilot response data in human-in-the-loop (HITL) simulations are often limited to one data point per subject, which both limits the ability to draw valid conclusions and to generalize the findings to other events and scenarios (Wickens, 2001). Human Performance Models (HPMs) are research tools that have been used to evaluate pilot performance under nominal conditions and are often cited as a solution to examine off-nominal scenarios (see Foyle & Hooey, 2008). To date, however, models of off-nominal or unexpected scenarios are limited because insufficient data exist to characterize performance and populate the models. This research effort aimed to extract and extrapolate data from existing HITL studies to inform the development of HPMs of off-nominal scenarios. The scope of this research was limited to off-nominal events with clear, unambiguous onsets and clearly defined responses. It is asserted that human responses to these types of off-nominal events are human performance primitives that transcend task environments and thus are inherently well suited for inclusion as inputs to HPMs.

Method

A comprehensive review of the literature identified 26[1] HITL simulation studies (see References and Gore et al, 2009) that met the following criteria:

- The study was either a simulation or flight test with human pilots as subjects and sufficient detail was provided to discern the method used and the performance data
- Subjects had not received training regarding, or been cued to the possibility of, the off-nominal event
- The off-nominal event was either truly surprising (i.e., one per subject) or very infrequent (e.g., one per condition)
- The off-nominal event had a clear, unambiguous onset (e.g. warning light onset, traffic on runway) and an objective, measurable response (e.g., button press, eye glance, or verbal response)

---

[1] A total of 34 papers were identified and summarized in Gore et al., (2009), however this paper focuses only on the 26 papers that provided miss rate data. Gore et al, 2009 also provides analyses of response latencies.

The review process yielded two types of events: 1) event onsets events, which required pilots to notice the presence of something such as the onset of a warning light or presence of an aircraft on the runway, and 2) error detection events, which required pilots to notice a discrepancy in a cockpit instrument or an invalid clearance from air traffic control (ATC). Both error types were included in these analyses. Events that required diagnosis of multiple cues, as opposed to simple event detection, were not included.

A parameter meta-analysis was conducted to pool pilot response data across multiple diverse HITL studies to increase statistical power and generalizability. The term parameter meta-analysis is used, because unlike a formal meta-analysis that averages *effect-sizes* across studies, it averages quantitative human performance parameters – specifically miss rates of off-nominal *event detection*. Response latencies were also evaluated, however, in most cases there were inadequate data to reach significance in the meta-analysis. These data are not presented here, but are available in Gore et al. 2009. The advantage of this parameter meta-analysis approach is that it produces estimates of response accuracy for each factor (represented as 'costs' or 'benefits' to the probability of detecting the event) rather than simply summarizing average miss rate for each particular off-nominal event. This method has previously been used to evaluate Synthetic Vision System (SVS) displays (Wickens, 2005), and human responses to imperfect diagnostic automation (Wickens & Dixon, 2007).

Analyses were conducted by pooling the event detection miss rate for common conditions across studies and weighting the studies by their sample size. For example, if two studies in one condition had miss rates of 1/5 and 30/50, a single proportion for the studies of 31/55 was extracted. Note that this mean proportion is far closer to the 0.60 value of the second study, than the 0.2 value of the first – but using this weighted approach, the resulting value more closely reflects the proportion of the larger sample size than if both studies had been given equal weighting. Chi-squared tests were used to assess if the relative frequency count of missed vs. non-missed events was statistically equivalent across the level of another variable. Subsequently, where appropriate, further chi-square tests were conducted to determine whether a difference observed might be modulated by a second factor. The modifications may occur when levels of another factor exert very different effects (i.e., a classic two-way interaction), and this modulation can be amplified if the *N* of the different studies contributing to the other factor is very different at its two levels.

Results

An analysis of the probability of a pilot failing to respond to the off-nominal event (that comprises the miss rate data), pooled across all available studies and event types, revealed an overall miss rate of 0.32, a value that is noteworthy for its magnitude above zero. All studies included in our analyses contained a positive indication of the off-nominal event, that is, the events were clearly visible, and hence certainly could be detected if they were expected and attention focused toward their location. This detection rate was further examined as a function of: 1) off-nominal event characteristics and 2) flight deck technology characteristics.

Off-Nominal Event Characteristics: Phase of Flight, Expectancy and Event Location

Three characteristics of the off-nominal events were evaluated: Phase of flight, event expectancy, and event location. These main effects, and interactions among them, are described below. Event characteristics that were also moderated by the absence or presence of flight deck technologies will be described in the following section.

Phase of flight. An analysis of miss rate (that is, the rate that pilots failed to detect an off-nominal event) revealed that across all 26 studies in our analysis, the probability of missing an off-nominal event was highest during departures ($p_{miss} = .50$), followed by cruise ($p_{miss} = .47$), arrival/approach ($p_{miss} = .39$), and taxi ($p_{miss} = .20$; χ^2 (3) = 34.61, $p < .001$). The reader is cautioned in interpreting the departure miss rate, however, as this was comprised of only one study with eight pilots. These miss rates may reflect an expectancy effect as pilots tend to be more vigilant and aware of both the traffic environment and their aircraft status during the arrival and taxi phases than in the cruise and departure phases. They may also reflect a location effect as events during cruise tended to be located on the instrument panel, but during approach the event tended to be out-the-window (OTW).

Expectancy and event location. The effect of expectancy on pilot detection of off-nominal events was assessed by comparing the miss rate from the *first off-nominal event* a pilot experienced to that from all subsequent off-nominal events. As would be expected, the probability of missing the event was higher if it was the first event ($p_{miss} = 0.48$) than for subsequent off-nominal events ($p_{miss} = 0.29$; χ^2 (1) = 24.70 p < 0.001). This produced an **Unexpectancy Cost of 0.19**. Next, the off-nominal events across all available studies were classified as occurring either OTW or head-down in the cockpit. The probability of missing an event was lower when it was OTW (p_{miss} = 0.29) than when it was head down (p_{miss}= 0.39), $\chi^2(1) = 9.88$, $p < 0.01$, yielding a **Cockpit Location Cost of 0.10**. The analysis also yielded an interaction between event expectancy and location. There was a large unexpectancy cost when the off-nominal event was OTW (p_{miss} for first OTW event = 0.50; p_{miss} for subsequent OTW events = 0.23; χ^2 (1) = 39.86, $p < 0.01$; **OTW Unexpectancy Cost of 0.27**) but when the off-nominal event was within the cockpit, there was no difference in miss rate as a function of expectancy (p_{miss} = 0.41 for both). This could reflect that pilots bring their own knowledge of real-world expectancies to the HITL study since in actual operations the frequency, and therefore expectancy, of a head-down event is much greater than for OTW events. In other words, in the simulations, the first cockpit event, was not as truly surprising as the first OTW event.

Flight Deck Technology: HUDs, HITS, Datalink, and Graphical Route Displays

The analyses of pilots' event detection as a function of the presence of various advanced cockpit technologies was largely driven in a bottom-up fashion by the available literature. The technologies reflect a range of technologies that may be incorporated into future advanced cockpits. These include head-up displays (HUDs), highway-in-the-sky (HITS) displays, datalink, and graphical route presentations.

Head-up display (HUD). HUDs are used in current operations for approach and landing, and may be used in NextGen for surface operations and to support low-visibility operations. An analysis using six HITL studies evaluated whether the presence of a Head-up Display (HUD) affected the probability of detecting an off-nominal event (regardless of event location). The probability of missing an event was higher when the pilots were flying with a HUD (p_{miss} = 0.39) than without (p_{miss} = 0.31), $\chi^2(1) = 4.13$, $p<.05$. This produced a **HUD Cost of 0.08**. This HUD effect was modified by the location of the off-nominal event in a manner that reflects the classic Fischer, Haines, and Price (1980) finding that the HUD particularly obscures unexpected OTW events (See also Fadden, Wickens, & Ververs, 1999). When the off-nominal event occurred OTW, the probability of missing the event was greater when pilots were flying with the HUD (p_{miss} with HUD = 0.36), than without (p_{miss} without HUD = 0.27; χ^2 (1) = 4.63, $p < .05$) producing an **OTW HUD Cost of 0.09**. But, if the event occurred head-down in the cockpit, the probability of missing the event was lower (though not significantly) when flying with the HUD (p_{miss} with HUD = .46) than without (p_{miss} without HUD = .51; $\chi^2(1) = .40$, $p = .53$; non-significant **Cockpit Location HUD Benefit = .05**[2]).

Highway-in-the-sky (HITS). A HITS display integrates lateral, vertical, and longitudinal information of the flight path into a perspective path through the air (Wickens & Alexander, 2009). While it may be presented either on a HUD or head-down display, it was presented head-down in all ten studies used in our analysis. The probability of missing an event (all events were OTW) when flying with a HITS display was higher (p_{miss} = 0.45) than when flying without the HITS display (p_{miss} = .22; $\chi^2(1) = 31.03$, $p < .001$). This produced a **HITS Cost of 0.23**, presumably due to the fact that the head-down HITS reduced eyes-out time and induced cognitive tunneling (Fadden, Ververs, & Wickens, 2001; Wickens & Alexander, 2009). The HITS cost remained when we consider only the first, truly surprising OTW event (p_{miss} with HITS = .55; p_{miss} without HITS = .33; χ^2 (1) = 7.01, $p <.01$; **HITS Cost for Truly Surprising OTW Events = .22**).

Datalink. It is expected that NextGen will include datalink communications between pilots and ATC (JPDO, 2007). A great deal of research has evaluated a range of datalink issues such as pilot workload, situation awareness, and heads-down time (e.g., Smith, Polson, Brown, & Moses, 2001). Four studies were identified that compared pilots' ability to detect an off-nominal event (all events were ATC clearance errors) when presented via datalink and/or voice. The probability that a pilot missed a clearance error was more than twice as high when the clearance was presented via datalink alone (p_{miss} = 0.69) than by voice alone (p_{miss} = 0.33) and voice with datalink together (p_{miss} = 0.38; $\chi^2(2)$= 25.73, $p < 0.001$). There was no significant difference in the probability of missing the error between voice and voice with datalink ($\chi^2(1) = 0.12$, $p = 0.72$), so the presence of voice appears to be a buffer, or error-trapping agent, against clearance comprehension errors (see Hooey, Foyle, & Andre, 2001). (The reader is cautioned that the data for voice-only clearance errors are limited to 18 subjects from a single study). A comparison of the voice with datalink and datalink-only conditions yielded a **Datalink-only Cost of 0.31.**

---

[2] Costs and Benefits are provided, even when non-significant, as they are expected to be useful for populating HPMs, the intended purpose of these analyses.

Next, a distinction was made between those clearances that were inappropriate (such as a clearance to turn onto an occupied taxiway creating a nose-to-nose conflict) and those that were impossible (such as a clearance to climb to an altitude below the current altitude). Inappropriate clearances tend to be subtle distinctions that require greater cognitive processing whereas impossible clearances tend to be more salient and obvious. In looking first at inappropriate clearances, the probability of missing a clearance error was much higher when the inappropriate clearance was issued via datalink (p_{miss} = 0.85) than when issued by both datalink and voice (p_{miss} = 0.5; $\chi^2(1)$= 12.27, $p < 0.001$; **Datalink Cost for inappropriate clearances = 0.35**), however, the datalink cost was not significant for impossible clearance errors (p_{miss} with datalink = 0.54; p_{miss} with voice and datalink = 0.44; p>0.1; non-significant **Datalink cost for impossible clearances = 0.1.** Therefore, the pilots caught the more salient impossible errors equally often with or without datalink but were hindered by datalink in detecting the less salient inappropriate errors. This could reflect a criticality difference between the two error types, however there were insufficient data to test this hypothesis.

Graphical routes. Displays that graphically present route information include electronic moving maps for airport surface operations (Hooey, Foyle, & Andre, 2001) or flight procedure rehearsal tools (Arthur, et al., 2004), among others. Four studies were identified that met the meta-analysis criteria and evaluated the effect of graphical displays on pilot detection of off-nominal events. Surprisingly, there was no main effect of the presence of a graphical rendition of the clearance on error detection rates. When the clearance (regardless of delivery method) was accompanied by a graphical presentation within the cockpit, the probability of missing the clearance error was 0.64 as compared to 0.65 when no graphical depiction accompanied the clearance ($\chi^2(1)$= 0.03, p = 0.87; non-significant **Graphical Route Benefit = 0.01**). However, for events in which the clearance was merely inappropriate, but not impossible, it appears as if the graphical presentation did improve event detection (p_{miss} with graphical route = 0.75; p_{miss} without graphical route = 0.86; $\chi^2(1)$=3.6, p = 0.057; **Graphical Route Benefit for Inappropriate Clearance Errors = 0.11**). The graphical route benefit was not observed for impossible clearances, with the trend in the opposite direction (p_{miss} with graphical route = 0.56; p_{miss} without graphical route = 0.49; p > 0.1; non-significant **Graphical Route Cost for Impossible Clearance Errors = 0.07).**

Discussion

This meta-analysis characterized pilots' miss rate for off-nominal events as a function of expectancy, event location, and the presence or absence of various advanced flight deck technologies. It was observed that the miss rate data produced several plausible and significant effects including:

- An overall miss rate of .32
- An unexpectancy cost for first, truly surprising events, especially OTW events
- A cockpit location cost
- A HUD cost, especially for OTW events
- A HITS cost for OTW events
- A datalink cost, especially for inappropriate clearances
- A benefit of graphical routes for inappropriate clearances

While the existence of these and other effects confirms prior work, most critically the current analyses provided robust, stable estimates of their effect size in real-world meaningful units.

An important finding was that the presence of the advanced technologies either hindered off-nominal event detection as was the case for HUDs, HITS, and Datalink, or failed to show a significant benefit for event detection as was expected from the graphical routes. These results may reflect cognitive tunneling effects especially for the HUD and HITS technologies (Fadden, Ververs, & Wickens, 2001; Wickens & Alexander, 2009) and general complacency effects as has been well documented in Parasuraman, Molloy & Singh (1993). This raises a concern for NextGen flight deck design and points to the need for careful consideration of both nominal and off-nominal conditions in the design and evaluation of NextGen technologies and operations. The results of this parameter meta-analysis reveal insights for the development of countermeasures in terms of training, procedures, and on-board alerts and warnings to mitigate the failure to detect off-nominal events. For example, it was seen that when pilots have some forewarning that an event could happen in the simulation studies, the miss rate dropped by 19%. Looking just at OTW events, the miss rate was 27% if pilots were forewarned of the possibility of the event. This suggests that training to remind pilots of the possibility of various events (such as runway incursion 'hot spots' or areas prone to bird strikes), or displays that indicate traffic or weather in the area, even if they are accompanied with high amounts of uncertainty, may reduce the miss rate. The finding that HUD and HITS both reduced event detection could suggest the need to mandate that airlines adopt procedures specifying that when one pilot uses the HUD or HITS, the other pilot must be eyes-out. Finally, the finding that datalink inhibited event detection, especially for inappropriate clearances, is of concern as these clearance errors are the most difficult for both pilots and automation to detect. This result may reinforce procedures that the pilots read the datalink out loud within the cockpit to maximize error detection.

Limitations and Opportunities for Future Research

Each study included in this parameter meta-analysis was conducted with independent research objectives and therefore all differed on important factors relating to the events, flight scenarios, and measurement techniques. One inevitable consequence of any meta-analysis is that the diverse studies may differ from each other on variables other than those used for classification. In some cases this pooling may cause an increase in variance within a category, diluting the strength of an effect. In other cases, it may cause a confound (e.g., studies with a HUD used, on average, pilots with more experience than those without). While it might in some cases have been possible to create an additional category of "experience" (assuming adequate reporting of this variable by the independent researchers) the danger of creating progressively more classification dimensions is that the number of observations within each cell becomes so small that statistical comparisons are challenged. A second limitation is that many of the HITL studies included in the analyses employed a single-pilot, general aviation crew as test subjects. It is possible that two pairs of eyes in the commercial cockpit could reveal a different (presumably lower) miss rate. Finally, it is noted that all data were drawn from HITL simulations and there is always the concern that pilot performance in simulation does not mirror pilot performance in actual operations (see Newman & Anderson, 1994). There is a real need for continued off-nominal event research to further populate the existing off-nominal database to increase the robustness and validity of these findings.

Conclusion

By pooling data across disparate HITL studies, many of which lacked statistical power to draw conclusions and generalize findings when considered individually, we identified several factors that have a robust influence on human performance in off-nominal environments. Three of the variables reported here (Expectancy, Event Location, and HITS) were used to validate a model of visual attention (N-SEEV; Wickens et al., 2009) which then was used to predict pilots responses to off-nominal events in NextGen environments (see Gore et al., 2009). Following HPM efforts will use a larger set of these meta-analysis findings to populate HPMs with valid estimates of pilot performance to estimate response time and accuracy to off-nominal events in the Next Generation Air Space System and to evaluate proposed mitigating solutions.

References[3]

* Alexander, A. L., & Wickens, C. D. (2005). *3D navigation and integrated hazard display in advanced avionics: Performance, situation awareness, and workload* (Technical Report AHFD-05-10/NASA-05-2). Savoy, IL: University of Illinois, Aviation Human Factors Division.

* Alexander, A. L., Wickens, C. D., & Hardy, T. J. (2005). Synthetic vision and the primary flight display. *Human Factors, 47*, 693-707.

* Arthur, J. J., Prinzel, L. J., Williams, S. P., & Kramer, L. J. (2004). *Synthetic vision enhanced surface operations and flight procedures rehearsal tools* (NASA/TP-2004-213008). Hampton, VA: NASA.

* Arthur, J., Prinzel, L. J., 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). Hampton, VA: NASA.

* Bailey, R. E., Kramer, L. J., & Prinzel, L. J. (2006). Crew and display concepts evaluation for synthetic enhanced vision systems. *Proceedings of SPIE*, *vol. 6226.*

Burian, B. K. (2008). Perturbing the system: Emergency and off-nominal situations under NextGen. *International Journal of Applied Aviation Studies, 8*(1), 114-127.

Fadden, S., Ververs, P. M., & Wickens, C. D. (2001). Pathway HUDS: Are they viable? *Human Factors, 43*, 173-193.

* Fischer, E., Haines, R. F., & Price, T. A. (1980). *Cognitive issues in head-up displays* (NASA Technical Paper 1711). Moffett Field, CA: NASA Ames Research Center.

[3] * denotes papers included in the meta-analysis.

Foyle, D. C., & Hooey, B. L. (2008). *Human Performance Modeling in Aviation*. Boca Raton, FL: Taylor & Francis/CRC Press.
Gore, B. F., Hooey, B. L., Salud, E., Wickens. C. D., Sebok, A., Hutchins, S., Koenecke, C., & Bzostek, J. (2009). *Identification of NextGen air traffic control and pilot performance parameters for human performance model development in the transitional airspace*. (Technical Report: NRA NNX08AE87A). San Jose, CA: San Jose State University Research Foundation.
* Helleberg, J. (2005). *Effects of a final approach runway occupancy signal (FAROS) on pilots' flight path tracking, traffic detection, and air traffic control communications*. McLean, VA: The MITRE Corporation.
* Hofer, E. F., Braune, R. J., Boucek, G. P., & Pfaff, T. A. (2001). *Attention switching between near and far domains: An exploratory study of pilots' attention switching with head-up and head-down* (D6-36668). The Boeing Company, October 18, 2001.
* Hooey, B. L., Foyle, D. C., & Andre, A. D. (2000). Integration of cockpit displays for surface operations: The final stage of a human-centered design approach. *SAE Transactions: Journal of Aerospace, 109*, 1053-1065.
* Iani, C., & Wickens, C.D. (2007). Factors affecting task management in aviation. *Human Factors, 49,* 16-24.
Joint Planning and Development Office (2007). Concept of Operations for the Next Generation Air Transport System, v2.0 (June 13, 2007). Retrieved from http://www.jpdo.gov/library/NextGen_v2.0.pdf.
* Johnson, N. R., Wiegmann, D. A., & Wickens, C. D. (2005). *Effects of advanced cockpit displays on general aviation pilots' decisions to continue visual flight rules (VFR) flight into instrument meteorological conditions (IMC)* (AFHD-05-18/NASA-05-6). Savoy, IL: University of Illinois, Aviation Human Factors Division.
* Lorenz, B., & Biella, M. (2006). Evaluation of onboard taxi guidance support on pilot performance in airport surface navigation. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, 111-115.
* 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.
Newman, R. L., & Anderson, M. (1994). *HUD flight testing: Lessons learned*. Presented at the Southeast Section Symposium of the Society of Experimental Test Pilots. Stone Mountain, Georgia.
* Olson, W.A., & Sarter, N.B. (2001). Management-by-consent in human-machine systems: When and why it breaks down. *Human Factors, 43(2),* 255-266.
Parasuraman, R., Molly, R., & Singh, I. L. (1993). Performance consequences of automation induced complacency. International Journal of Aviation Psychology, 3, 1-23.
* Prinzel, L. J., Kramer, L. J., Arthur, J. J., Bailey, R. E., & Comstock, R. J. (2004). Comparison of head-up and head-down "highway in the sky" tunnel and guidance concepts for synthetic vision displays. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, 11-15.
Smith, N., Polson, P., Brown, J. A., & Moses, J. (2001). An assessment of flight crew experiences with FANS-1 Controller-Pilot Data Link Communication in the South Pacific. *Proceedings of the 4th USA/Europe Air Traffic Management Research and Development Seminar, Air-Ground Cooperation Track*. Santa Fe, NM.
* Weintraub, D. J., Haines, R. F., & Randle. R., (1985). Head-up display (HUD) utility, II: Runway to HUD transitions monitoring eye focus and decision times. In *Proceedings of the 29th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica: HFES.
Wickens, C. D. (2001). Attention to safety and the psychology of surprise. Proceedings of the 11th International Symposium on Aviation Psychology. Columbus, OH. Ohio State University.
Wickens, C.D. (2005). Attentional tunneling and task management. *Proceedings of the 13th International Symposium on Aviation Psychology*. Dayton, OH: Wright-Patterson.
Wickens, C.D. & Alexander, A, L. (2009). Attentional tunneling and task management in synthetic vision displays. *International Journal of Aviation Psychology, 19(2), 1-17.*
* Wickens, C. D., Alexander, A. L., Thomas, L. C., Horrey, W. J., Nunes, A., Hardy, T. J., & Zheng, X. S. (2004). *Traffic and flight guidance depiction on a synthetic vision system display: The effects of clutter on performance and visual attention allocation* (Technical Report AHFD-04-10/NASA(HPM)-04-1). Savoy, IL: University of Illinois, Aviation Human Factors Division.
Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8(3)*, 201-212.
* Wickens, C. D., Helleberg, J., & Xu, X. (2002). Pilot maneuver choice and workload in free flight. *Human Factors, 44*(2), 171-188.
Wickens, C. D., Sebok, A., Bzostek, J., Steelmna-Allen, K., McCarley, J. & Sarter, N. (2009). NT-SEEV: A model of attention capture and noticing on the flight deck. In *Proceedings of the 15th International Symposium for Aviation Psychology*. Dayton, OH: Wright State University.

Acknowledgements

This research was supported by a NASA Research Announcement awarded to San Jose State University Research Foundation by NASA's Airspace Super Density Operations Project (NRA # NNX08AE87A; COTR Dr. David Foyle).

NEXTGEN FLIGHT DECK HUMAN FACTORS ISSUES

Ken Funk
Oregon State University
Corvallis, Oregon

Robert Mauro
University of Oregon
Eugene, Oregon

Immanuel Barshi
NASA Ames Research Center
Moffett Field, California

This paper describes a project to compile, from a literature review and preliminary analyses, an initial but reasonably comprehensive list of NextGen flight deck human factors issues. It describes the methodology that was used, presents representative issues from the list that resulted, and makes recommendations to continue work to update the list and use it as the basis for suggested NextGen flight deck standards and design requirements.

The goals of the Next Generation Air Transportation System (NextGen) are to significantly increase the safety, security, and capacity of US air transportation operations. The eight key capabilities deemed necessary to achieve those goals (JPDO, 2007a) will bring major changes to the flight deck, including Internet-like information services, access through them to a common weather picture, integration of weather information into flight deck decision making, negotiated four-dimensional aircraft trajectories, means for equivalent visual operations in low visibility conditions, delegated self-separation, and equipment and procedures for super-density arrival and departure operations.

Plans for NextGen development have been driven largely by technology, and human factors considerations do not appear to be a motivating force behind these and other changes. Thus, the NextGen flight deck could harbor many vulnerabilities to pilot error, jeopardizing the very goals NextGen is meant to accomplish. While past research has applied human factors expert opinion to identify general NextGen human factors issues, as yet, little NextGen-specific human factors analysis has been performed and, to our knowledge, no one has attempted to create a reasonably comprehensive list of human factors issues related specifically to the NextGen flight deck.

Objectives

The objectives of this project were to conduct a preliminary review of literature and perform preliminary analyses to compile and organize an initial but reasonably comprehensive list of NextGen flight deck human factors issues.

Methodology

In the interest of clarity, we defined a NextGen flight deck human factors issue as:

> a statement which, if it should become true in the implementation and operation of NextGen, describes a condition or situation related to flight deck operations in which normal pilot characteristics, capabilities, limitations, and tendencies are very likely to lead to significant problems with NextGen effectiveness, efficiency, or safety.

NextGen is in the early stages of development. The issues that we identified are plausible conditions or situations that could develop as NextGen is implemented and, if they materialize, these issues are likely to lead to serious problems. Because NextGen is still being developed, we cannot be certain that the situations or conditions described in the issues will come to be. But based on what we know about the current air transportation system and plans for implementation of NextGen, they are likely to exist in NextGen unless steps are taken to prevent them. The purpose of identifying issues at this time is to head off those problems by providing input to good, pilot-centered design.

NextGen Human Factors Literature Review

Our first step in identifying NextGen flight deck human factors issues was to search for known issues in the literature which is, as yet, rather limited. We reviewed NextGen issues reports (e.g., Sheridan, Corker, and Nadler, 2006a, 2006b; Murdoch and Press, 2008) and reports on human factors issues with Automatic Dependent Surveillance – Broadcast (ADS-B), a key NextGen enabler, in which GPS-based reports of aircraft's own positions drive traffic displays for both air traffic controllers and pilots (e.g.,Williams et al, 2002; MITRE, 2006). Relevant literature on other aspects of NextGen and Eurocontrol's Single European Sky ATM Research Programme (SESAR) was also examined. We were unable to thoroughly review all of this literature with the time and resources allotted. However, as we reached the end of the project, few new issues emerged as additional literature was examined.

To review this literature, we read the documents for descriptions, either explicitly stated or implied, of conditions or situations that could be related to flight deck operations where normal pilot characteristics, capabilities, limitations, and tendencies would be very likely to lead to significant problems with NextGen effectiveness, efficiency, or safety. We captured these excerpts (in most cases verbatim) in a spreadsheet. To promote consistency in how the issues were posed, we paraphrased distinct issues, as described in our sources, into terse statements having uniform syntax and semantic structure.

For example, Sheridan et al (2006a section 3.1.2.1) suggested that "[m]onitoring and maintaining situation awareness over long and boring periods of nominal operations under automatic control (with a possible need to impose activities for the purpose of maintaining alertness)" was an issue for future NextGen research. From this excerpt (and from others like it in other sources), we identified two NextGen flight deck human factors issues: 1) "Monitoring requirements are excessive" and 2) "Difficult to maintain situational awareness over long, boring periods of nominal operations."

As background for the study, we read the *NextGen Concept of Operations* (JPDO, 2007a) and flight deck-relevant portions of the *NextGen Enterprise Architecture* (JPDO, 2007b) and we reviewed the online *NextGen Enterprise Dataset* (JPDO, 2008). These documents describe NextGen operations, functions, operational improvements, and enablers. Although they are not intended to directly identify flight deck human factors issues, issues arise in the descriptions of NextGen elements. We did not exhaustively analyze these descriptions for issues, but we captured the human factors issues suggested by them in the spreadsheet.

Flight Deck Automation Human Factors Issues Database Application

The past two decades have seen considerable controversy about the safety of flight deck automation and many human factors issues have been raised (e.g., Wiener, 1989). Funk and his colleagues (1999) undertook a comprehensive review of aircraft automation research, aircraft incident reports, incident report studies, and aircraft accident reports, and surveyed pilots and aviation safety experts to develop a comprehensive list of flight deck automation issues. For each issue so identified, they compiled evidence from their sources to support the assertion that it posed a safety problem, and then performed meta-analyses to prioritize the issues for further research. Their findings and supporting data are available on a website (RII, 2007). We reviewed all of their flight deck automation issues to determine which potentially apply to the NextGen flight deck and captured relevant ones in our spreadsheet.

For example, flight deck automation issue 103 is "It may be difficult for pilots to decide what levels of automation are appropriate in specific circumstances, possibly increasing pilot workload." Applying our syntactical and semantical structure for NextGen issues to that, we added the following NextGen flight deck human factors issue to our list: "Automation level decisions are difficult for pilots."

Failure Modes and Effects Analysis

Failure Modes and Effects Analysis (FMEA) is a proven, prospective safety analysis technique that systematically examines a process representation to identify failure modes (ways in which a system can fail), factors contributing to those failures, and their consequences. Pilot error may be considered a kind of failure mode and, from potential errors, flight deck human factors issues may be identified. So we performed a preliminary FMEA to identify more NextGen flight deck human factors issues.

As preparation for the FMEA, we developed a preliminary flight deck functional model, the Oregon NextGen Flight Deck Functional Model version 0.1 (ONFDFM V0.1). The ONFDFM models a general aviation or on-demand air taxi flight from a small airport with an Automated Virtual Tower to a mid-size airport, with parallel runways, in a metroplex. As this was an initial effort conducted with limited resources in a short time frame, we used a simplified functional modeling approach, roughly equivalent to hierarchical task analysis. Elements of the ONFDFM are verb phrases, each one describing the mission or a function or task (low-level function) within the mission. The model is represented as a hierarchical list, like an outline. We modeled the top-level, mission function as *Conduct NextGen flight*. We analyzed the mission function into subfunctions corresponding to mission phases. For each mission phase function, we broke it down into subfunctions subordinate to that, and so on. Rather than detail the entire ONFDFM to a uniform depth, we focused on especially important portions of the model, for example, *Perform departure related activities* and *Perform enroute activities*.

We performed a partial FMEA using the ONFDFM. Due to limited resources and the short time available, we applied the analysis to just two representative parts of the model, *Perform departure related activities* and *Fly enroute free-flight*. The latter is part of *Perform enroute activities* and represents free flight, as opposed to flow corridor, operations.

For each subfunction in these parts of the model, we applied our knowledge of NextGen functionality, our knowledge of present day flight deck operations likely to be similar to NextGen operations, and our knowledge of human operator characteristics, capabilities, and limitations, to identify likely failure modes for the subfunction, (i.e., errors that pilots would likely commit in performing the subfunction). For each error we identified likely effects or consequences of the error. For each specific error identified in the FMEA, we generalized it to one or more broader issue statements. As we identified many similar errors, multiple errors mapped to the same issue, so the FMEA did not produce as many issues as errors. As an example, FMEA applied to *Monitor CDTI for other traffic & ground equipment during taxi out* and other subfunctions led to the identification of the issue, "Use of CDTI to maintain surface separation interferes with visual contact with surface traffic".

Issues Management

As we collected issues using the aforementioned methods, we added them to the spreadsheet. For each issue, we recorded an issue statement (worded using syntactical and semantic structure designed to be reasonably consistent across issues), the source of the issue, a reference (section identifier, page number, item number, or other locater), an excerpt from the text that suggested the issue, additional source and reference information (if the issue was found in more than one source), an optional comment, and information as to whether the issue appeared to be redundant with one or more issues that had already been recorded. To aid in issue classification and organization, we attached one or more descriptive tags or labels to each issue. Because tags were not mutually exclusive and they covered several dimensions of the flight deck domain, they allowed us to categorize and organize the issues in several ways. The tag system will permit more flexible use of the issues list in future research and development.

Because many tags were identified and used in the process, a higher level of organization was required for clarity. So we organized the tags themselves and, by extension, the issues which they designated, into 10 categories. Then for each tag in each category, we formulated a broad issue statement, intended to represent all issues marked with that tag. We organized and set up the issues spreadsheet to present broad issue categories and broad issues, and to filter specific issues by the issue tags and other criteria. We reviewed all the specific issues, edited the specific issue statements for accuracy, clarity, and uniform syntax and semantic structure, then identified and removed redundant specific issues.

Findings and Discussion

Initially, our reviews and analyses yielded 250 specific issues, which, by removing redundant issues, were reduced to 225 specific issues. The specific issues were marked by 81 tags reflecting broad issues, with those broad issues/tags organized into the 10 categories. The following sections present, for each of nine of the categories, a tag (in square brackets) and broad issue representative of that category, a specific issue subsumed by the broad issue, and a list of tags for other broad issues falling in the category. Because these representative issues are also what we believe to be some of the most important ones, we additionally include brief discussions of their significance.

Design, Development, Testing, and Certification Issues

Broad issue:	[development] There is insufficient and inadequate human factors engineering input in the development of NextGen functions and subsystems.
Specific issue:	Inadequate human-in-the-loop fidelity used in development and certification.
Other broad issues:	certification, testing, human-centered design

Although human factors research and design is mentioned in JPDO documentation, it is not clear how NextGen planners intend to address these issues. We are concerned that human-centered design will not be a development priority and that NextGen engineers will rely on their intuition rather than on a comprehensive set of human factors tools and guidelines when designing pilot-system interfaces and tasks.

Issues Related to Pilot-Pilot and Pilot-ANSP Interaction

Broad issue:	[collaboration] NextGen pilot-Air Navigation Services Provider (ANSP) collaboration processes are poorly designed, poorly defined, inefficient, and ineffective.
Specific issue:	Flight plan negotiation processes and mechanisms are poorly designed.
Other broad issues:	voice/data, communication, team

Unless pilot roles, responsibilities, authority, and procedures with respect to collaboration and, especially trajectory negotiation, are clearly defined, designed, and trained, there will be operational confusion, misunderstandings, delays, and errors.

Pilot-Subsystem Interface Issues

Broad issue:	[information] Information on the NextGen flight deck is insufficient or, when available, difficult to access, inadequate, poorly presented to pilots, and often overwhelming.
Specific issue:	Pulled net-centric information is difficult to access.
Other broad issues:	inconsistency, feedback, representation, displays, CDTI, interface, cues, controls

NextGen is an information system. Knowing what information is important to a pilot under a given set of circumstances, how to filter and prioritize it based on context, and how to present that information effectively presents a daunting challenge.

Subsystem-Subsystem Interaction Issues

Broad issue:	[datalink] Pilots lack adequate awareness of automated data exchanges between NextGen ground and air subsystems.
Specific issue:	Pilots lack situational awareness due to automated exchange of flight plan and ... 4DT data.
Other broad issues:	uplink

In the context of a complex flight deck in which multiple tasks are being performed concurrently under a variety of operational stressors, simply giving pilots the option to review and approve automated information exchanges does not guarantee that they will do so, or, if they do, do it quickly and accurately.

Issues Related to Pilot Behavior and Performance

Broad issue:	[attention] Pilots do not properly allocate their attention among information sources and tasks on the NextGen flight deck.
Specific issue:	Both pilots often become involved with NextGen subsystems, which diverts their attention from safety-critical tasks.
Other broad issues:	errors, monitoring, manual skill, overload, decision making, awareness

The number of concurrent tasks on the NextGen flight deck will make it more difficult for pilots to assess the current status of all ongoing tasks and their relative importance and urgency. This will make it more difficult for the flight crew to correctly choose how to allocate their attention and efforts at any given time.

Issues Related to Pilot Roles, Responsibilities, Capabilities, Limitations, and Attitudes

Broad issue:	[authority] Pilot authority on the NextGen flight deck is unclear and/or overly restricted.
Specific issue:	Action responsibility/authority of net-centric information are poorly represented.
Other broad issues:	understanding, reliance, workload, oversight, pilot capabilities, trust, satisfaction, risk, roles, training, stress, responsibility, memory, culture, acceptance

Unless pilot authority is demarcated in general and operationally defined by the design of specific procedures, pilots will be uncertain as to their flight management and control authority in NextGen and therefore less likely to take full advantage of the autonomy, flexibility, and efficiency it promises.

Process and Procedure Issues

Broad issue:	[procedures] Many NextGen processes lack defined procedures or those procedures are poorly designed.
Specific issue:	Temporal and spatial variations in NextGen function require pilots to recognize the need for and use different procedures.
Other broad issues:	processes, intervention, multi-tasking, tasks, flight plan, negotiation

To operate in NextGen, a large number of tasks must be performed using a great deal of equipment. Under these conditions, flight deck procedures cannot be left to the pilots to design *ad hoc*. To avoid inefficiencies and errors, a systematic approach to procedure development should be used.

Flight Deck Subsystem Issues

Broad issue:	[automation] NextGen flight deck automation is overly complex and hard to understand, and its logic and interfaces are poorly designed.
Specific issue:	Automation changes modes without pilot commands to do so, sometimes producing surprising behavior.
Other broad issues:	failure, system control, decision support tools, equipment selection, equipage, standardization, manuals, modes, databases, data entry, complexity, functionality, performance, integration

The level and complexity of automation on the NextGen flight deck will be higher than that of today and even more care must be taken in the development process to assure its usability.

System Issues

Broad issue:	[variations] Temporal and spatial variations in NextGen functionality and subsystems make it difficult for pilots to adapt to different circumstances.
Specific issue:	Temporal and spatial variations in NextGen function require pilots to recognize the need for and use different procedures.
Other broad issues:	trajectory, organization, delay, justice, macroergonomics, system dynamics

NextGen will be a large and complex system, and variations in its functionality over space and time will present challenges to pilots.

Conclusions and Recommendations

Our study was necessarily limited by the short time frame in which it was conducted, by our ability to manage a large number of issues, by the limited amount of definitive information on the NextGen flight deck, and by our own personal knowledge limitations and biases. Nevertheless, we believe that the 81 broad issues, representing the 225 specific issues identified in this project, strongly suggest that the human factors challenges to the effectiveness, efficiency, and, especially, the safety of the NextGen flight deck may be greater than anticipated. With that in mind, we offer the following recommendations for further development of and action on these issues.

1. Create a team of human factors scientists and engineers, flight deck engineers, pilots, and aviation safety experts to collaboratively identify and recommend remediations for NextGen flight deck human factors issues.
2. Create and maintain a web-accessible NextGen flight deck human factors issues database.
3. Create and maintain a NextGen flight deck human factors website to facilitate team collaboration and the dissemination of findings and recommendations.
4. Review other sources for additional issues.
5. Clarify and edit the text of the issues and organize them.
6. Build and maintain a detailed NextGen flight deck functional model (NFDFM), consistent with the emerging NextGen architecture.
7. Use the NFDFM to perform more extensive FMEAs to identify additional issues.
8. Validate and prioritize the issues.
9. Use the NFDFM and FMEA findings to develop detailed NextGen flight deck scenarios for system research, development, and testing.

10. Use the issues list, FMEA results, scenarios, and other findings to develop suggested standards or design requirements for NextGen flight deck equipment and procedures.

Acknowledgments

We gratefully acknowledge the support of the NASA Ames Research Center, through a contract to Decision Research, Inc. We thank Morgan Rogers, for his initial literature review for ADS-B/CDTI human factors issues.

References

Funk, K., B. Lyall, J. Wilson, R. Vint, M. Niemczyk, C. Suroteguh, & G. Owen. (1999). Flight Deck Automation Issues, *Internationa1 Journal of Aviation Psychology, 9*(2), 109-123.

Joint Planning and Development Office [JPDO]. (2007a). *Concept of Operations for the Next Generation Air Transportation System*, Version 2.0. Washington, DC: author.

Joint Planning and Development Office [JPDO]. (2007b). *Enterprise Architecture V2.0 for the Next Generation Air Transportation System*, Version 2.0. Washington, DC: author.

Joint Planning and Development Office [JPDO]. (2008). *NextGen Enterprise Dataset*. http://65.207.77.76:8080/ned2/home.htm, last accessed November 2008.

The MITRE Corporation (2006). *The Impact of Capstone Phase 1 Post-Transition Annual Report – 2005*. Bedford, MA: author.

Murdoch, J. L., & Press, H. (2008). A Discussion of Human Factors Issues Related to the Next Generation Air Transportation System (NextGen). Unpublished manuscript, NASA Langley Research Center, Hampton, VA.

Research Integrations, Inc. [RII] (2007). *Flight Deck Automation Issues* website. http://www.flightdeckautomation.com/fdai.aspx, last accessed September 2008.

Sheridan, T.B., K.M. Corker, & E.D. Nadler (2006a). R*eport of a Workshop on Human-Automation Interaction in NGATS*. Cambridge, MA: US Department of Transportation, John A. Volpe National Transportation Systems Center.

Sheridan, T.B., K.M. Corker, & E.D. Nadler (2006b). *Final Report and Recommendations on Human-Automation Interaction in the Next Generation Air Transportation System*. Cambridge, MA: US Department of Transportation, John A. Volpe National Transportation Systems Center.

Wiener, E.L. (1989). *Human factors of advanced technology ("glass cockpit") transport aircraft* (NASA CR 177528). Moffet Field, CA: NASA Ames Research Center.

Williams K. A. Yost, J. Holland, & R. Tyler (2002). Assessment of Advanced Cockpit Displays for General Aviation Aircraft – The Capstone Program. Oklahoma City, OK: Federal Aviation Administration, Civil Aeromedical Institute.

Risk assessment in aviation

Robert Mauro
Decision Research &University of Oregon
Eugene, Oregon, USA
Immanuel Barshi
NASA Ames Research Center
Moffett Field, California, USA

In aviation, many actions are taken to reduce risk. However, not all risks can be avoided. To effectively manage risk, managers and regulators must evaluate and compare risks associated with different threats. Yet, it is frequently difficult to obtain reasonable assessments of these risks. Traditional approaches often produce unsatisfactory results when the probability of failure is low but the costs of failure are high -- as is often the case in modern civil aviation. Attempts to use a single dimension to evaluate threats often lead to unreliable and contentious assessments. Many risk assessment heuristics and displays can yield misleading and sometimes mathematically incongruous assessments. Furthermore, increases in costs caused by people's reactions to failures are often ignored or grossly underestimated. In this paper, problems with risk assessment in aviation are discussed and a Tool for Risk Identification, Assessment, and Display (TRIAD) designed to address many of these problems is described.

In aviation, safety and efficiency are primary goals. Many of the actions taken by aviation professionals are taken to reduce risk. However, one cannot avoid all risk. Regulators and managers must frequently decide which potential problems to address. To effectively manage risk, one must be able to evaluate the risks associated with different threats and compare them. But it is frequently difficult to obtain precise assessments. To accurately assess the risk associated with a potential failure or other threat, one must consider the possible outcomes that could occur, the likelihood of each outcome, and the consequences that may be associated with each outcome. In this paper, we discuss the assessment of each of these aspects of risk and describe a Tool for Risk Identification, Assessment, and Display (TRIAD) that was designed to assist in their assessment.

Risk is generally defined as a combination of likelihood and consequences -- the more damage that may occur, the greater the risk; the more likely a threat, the greater the risk. In traditional probabilistic risk assessment (PRA), risk is quantified by multiplying an estimate of the amount of potential damage by the estimated probability of the threat (e.g., Bier & Cox, 2007). In many cases, this assessment can be accomplished simply and the obtained result matches our intuitions. For example, a computer manufacturer may be able to estimate the probability that a microchip will fail within a warranty period quite precisely based on laboratory and field data. Calculating the cost of a new chip and the labor required to replace it is also relatively straightforward. Hence, the risk posed to the manufacturer by the potential failure of the microchip can be easily assessed. However, in many cases assessing risk is much more difficult.

Assessing the risk associated with a possible failure or other event becomes more difficult when:

- The event of interest (e.g., a failure) can have many possible outcomes.
- The event under consideration is not repeatable or there is no data from which to directly estimate the probability of the event.
- The event could lead to different types of damage which cannot be easily measured on a common scale.

- The cost of the potential damage or the likelihood of the event is extreme. This is particularly noticeable when dealing with extremely unlikely events that could have catastrophic consequences.

Aviation typically operates under these conditions. For example, an airline may be concerned with a rash of pilot reports of anomalies in the operation of their new flight management systems (FMS). Given the financial and personnel demands of daily operations, management must decide how much time and money to invest in determining the cause(s) of these reports and finding a solution. This requires an assessment of the risk posed by the reported anomalies. In the worst case, an FMS problem could lead to a controlled flight into terrain (CFIT) accident. But there are many other possible outcomes. A CFIT accident is not likely unless the anomaly occurs on approach or shortly after departure. However, encountering an anomaly en route is not without cost. Most of the time, the pilots may notice the anomaly and correct it, but if they don't -- fuel will be wasted, the pilots and airline may be the target of FAA enforcement actions, and there is a (very low) risk of a midair collision. Furthermore, the anomaly could distract the pilots at an inopportune time and cause other problems. All of these possibilities must be considered.

Possible Outcomes

To accurately assess the risk posed by a potential problem, one must first consider the possible outcomes that could result if the problem were to occur. Often, individuals attempt to simplify this task by considering only the worst case. This can be misleading. For example, consider a hypothetical error in an airline's weight and balance calculation program. In the worst case, the aircraft could depart out of balance and encounter an event that causes the aircraft to enter a stall from which recovery is impossible given the weight distribution. However, this is an exceedingly unlikely scenario. A manager might reasonably conclude that this possibility is so remote, that other problems have a higher priority. However, there is a much more likely outcome that should catch the manager's attention. Aircraft that are flown "out of CG" may burn substantially more fuel because of the out-of-balance condition. Although this outcome is not catastrophic, over a large number of flights the cost of the error could be large enough to cause substantial financial damage to the airline. It is also not sufficient to consider only the most likely outcome. In many cases, unlikely outcomes have sufficiently serious consequences and are likely enough to be cause for concern.

Generating lists of possible outcomes requires domain knowledge and creativity. However, in many cases, one can generate outcomes by systematically considering the general classes of factors that are likely to affect the result of a failure or other problem. These factors include:

Phase of flight – The point during an operation at which a problem occurs can have substantial effects on the possible outcomes. For example, the failure of a critical component of a navigation system may have different consequences during takeoff/climb-out, en route, or during descent/landing.

Time – When a problem occurs can have substantial effects on the possible outcomes. For example, the failure of a component may have different consequences during the day, or during the night. Likewise, the same failure could have very different consequences for winter operations than for summer.

Geography – Where a problem occurs can affect the possible outcomes. For example, the failure of a critical component of a navigation system may have different consequences depending on whether the failure occurs over land or during a trans-oceanic flight.

Damage – The physical characteristics of the damage caused by a problem may affect the outcome. The result of a problem may be different if the physical characteristics of the damage (e.g., size, depth, location, and frequency) differ. For example, the damage caused by debris from a turbine engine failure may be different depending on the size and depth of the penetration.

Design Characteristics –The way in which a system is designed will affect the possible outcomes that could result from a problem. For example, the result of the failure of a given system may differ depending on whether the aircraft is equipped with a backup system. Likewise, consequences of a failure could be very different depending on whether the failure is annunciated to the crew or not.

Procedures and training – A problem can have very different outcomes depending on whether or not procedures exist for dealing with it, and on whether or not crews are trained to deal with it. (Note also that procedures and training are often used as interventions to reduce risk.)

Environmental Conditions – Environmental conditions, such as temperature, humidity, wind speed and direction, etc. can affect the result of a problem. For example, the effect of a failure in a cooling system may depend on whether the device is at or below a critical temperature when the system fails. Similarly, a failure in an ice detection system would have very different consequences if the flight is conducted in icing conditions, than if it is conducted in non-icing conditions.

Likelihood

To proceed with a risk assessment, one must estimate how likely it is that each possible outcome will occur. Sometimes, the probability of a given outcome can be estimated quite precisely. For example, one may have engineering data that indicate how often a component fails in practice. But often this is not the case. Many likelihood assessments must be based on expert judgments. In many cases, experts will be reluctant or unable to specify a precise probability for a possible outcome. For example, an engineer may be able to specify the conditions under which a component of a navigation system will fail but no one may know how often those conditions occur in practice. However, even in these instances, it is rarely the case that one knows nothing. It is rarely the case that the probability of an outcome could *plausibly* range from zero to one. Even when one cannot estimate the probability associated with an outcome precisely, one can often offer a "best estimate" and specify a range around that estimate that will confidently bracket the actual probability. This is sufficient to continue with the risk assessment.

Consequences

Because risk is a function of likelihood and consequence, the possible damage that could result from an event must be assessed. In the microchip example used above, it was relatively easy to assess the possible damage because the costs are easy to calculate and only one type of damage, monetary loss, was considered. However, an event could cause many different types of damage that are not easily measured on a single scale. An event could cause property damage, injury or loss of life, or disrupt operations. Furthermore, an event could generate secondary damage through people's reactions to the original event.

People often attempt to simplify the assessment process by trying to use one measure to scale all of the different types of damage. For example, insurance companies and international agreements specify how much the loss of a limb or the death of an airline passenger is worth in dollars. These amounts can then be combined together with estimates of the costs of property damage and lost revenues to arrive at a single monetary value that can be used as *the* measure of the consequences of an accident. However, attempting to create a single scale on which all potential consequences can be arrayed may be counter-productive. For example, people may reasonably disagree with the value attached to life by an insurance company; courts often do. Furthermore, these calculation may lead decision-makers to make trade-offs that they themselves find unacceptable. For example, if an arbitrary monetary value is attached to the value of a life, then the rational decision is to forgo safety investments whenever the costs of those investments exceeds the monetary value of the lives likely to be lost if the investment is not made. Once a monetary value for a life is accepted, the trade-off appears rational although the individuals making the decision may not agree that the value of a life can be reduced to the specified amount.

Disagreements about the validity of an assessment may arise not because of any debate over the possible consequences or their likelihood but only over the value attached to the consequences. To avoid these distractions, it is often better to evaluate the consequences of an event on separate dimensions that are combined only when general agreement on the combination rules can be established. These dimensions may differ by domain. By default, TRIAD provides for the assessment of four types of threats: threats to life and health, threats to property, threats to mission (operational) success, and social amplification.

Social amplification refers to the secondary damage caused by people's reactions to an event (Kasperson et al, 1988). This consequence is often underappreciated. For example, the damage caused by a fatal crash of an airliner includes the value of the aircraft, the damage to life, limb, and property in the aircraft and on the ground, and the loss of revenue caused by the loss of the aircraft and the disruption to the schedule. However, the damage caused by a fatal crash of an airliner also includes the psychological trauma endured by survivors and relatives, increases in fears of flying, and damage to the reputation of the airline and the industry. Some of the costs of this damage are borne by the airline or its insurers either directly in payments to individuals or indirectly in lost ticket sales and decreased stock values. Some of these costs are borne by the industry in decreased travel and calls for increased governmental oversight. Some of the costs are borne by the society as a whole. In many cases, the costs associated with social amplification can substantially outweigh all other consequences.

Combining consequences that are assessed on different dimensions presents another problem. Often, the degree of damage will be evaluated on ordinal scales, but the values are treated as if they were interval or ratio scales. This can cause problems. For example, one is tempted to consider a reduction in a consequence rating from "5" to "3" as being greater than a reduction from "3" to "2" although ordinal scales carry no information about the relative sizes of the intervals between the markers. Hence, an intervention that causes a reduction from "5" to "3" may be seen as much more valuable than one that only reduces the rated hazard from "3" to "2". However, because the intervals between categories are not constant, the improvement reflected by a consequence reduction from "3" to "2" may be greater on some absolute scale than the improvement obtained by reducing the rated consequence from "5" to "3" and this latter reduction may be hardly different from a reduction from "5" to "4" (see Figure 1).

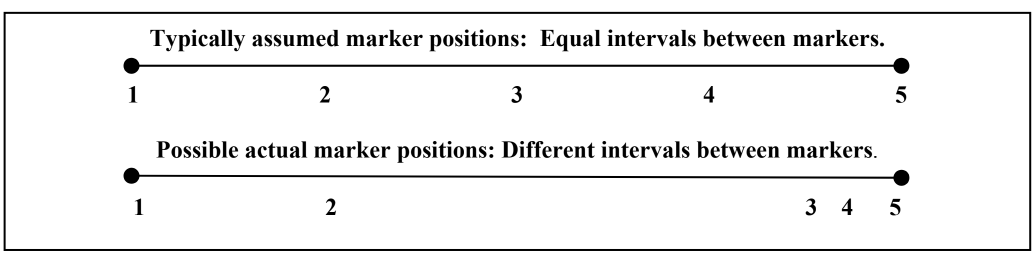


Figure 1. Illustration of possible ordinal values relative to an absolute scale.

This problem is exacerbated when one attempts to combine ordinal scales. Because the same numerals are typically used as markers for relative positions on different scales, users are sorely tempted to treat markers with the same numerical representation as if they were identical and to perform inappropriate arithmetic operations on them. For example, individuals often attempt to multiply the ordinal ratings obtained from two different scales. Consider attempting to combine ratings of threats to life/health and threats to property. If these are made on 5 point scales, the results can be displayed using a matrix like that in Table 1. In general, things get worse from bottom to top, left to right, and along the diagonal from lower left to upper right. However, one cannot easily combine this information into a single summary measure. For example, if the ordinal ratings on each dimension are multiplied, then an outcome rated as Property Damage 5* Life & Health 2 would be considered as risky as an outcome rated Property Damage 2* Life & Health 5 (2*5=5*2=10). But this is not necessarily the case. An incident in which multiple lives are lost but the property damage is $1-$10 million may not be equivalent to one in which there are only minor injuries but the properly damage exceeds $250 million. Neither is it the case that an outcome rated as Property Damage 3* Life & Health 4 (3*4=12) is necessarily worse than one rated as Property Damage 2* Life & Health 5 (2*5=10).

Table 1. *Ordinal Scale Matrix.*

		Property Damage				
		< $ 1 Million	$1 -$10 Million	$10 - $100 Million	$100 - $250 Million	> $250 Million
Life & Health		**1**	**2**	**3**	**4**	**5**
Multiple Deaths	**5**					
Single Death	**4**					
Major Injury	**3**					
Minor Injury	**2**					
Minimal/No Effect	**1**					

Extreme Risks

Assessing outcomes with extreme consequences pose a particularly difficult problem (Kunreuther, 2002). In most cases, the traditional calculation of risk as the product of the probability of an event and the potential consequences appears to approximate our sense of what risk is. For example, a business is likely to treat a high likelihood of a small monetary loss as of roughly equivalent risk to a low likelihood of a somewhat larger loss. However, when the probabilities and/or consequences approach their extremes, the risk estimate produced by the traditional calculation departs from what most people feel it should be. In particular, an event that could cause a catastrophe with very low probability is generally seen as much riskier than an event that is highly likely to cause an outcome with very low cost.

This phenomenon is not entirely psychological. Extreme consequences *are* different. For example, an airline can plan for how to respond to most potential outcomes. But one cannot plan for how to respond if the consequence is the collapse of the company. There is a discontinuity in the risk function at the point at which the consequences become unbearable. One cannot treat the collapse of the company, the destruction of an ecosystem, or the death of a society as simply an outcome with very high costs. This does not mean that one cannot assess extreme risks, only that one should not rely on the mechanical application of any simple risk calculation procedure in all situations.

Risk Displays

The value of a risk assessment depends on its ability to inform decisions. Hence, the manner in which the results are displayed is of considerable importance. Risk assessments are often portrayed by a single point on a two dimensional (probability X consequence) display (see left panel, Figure 2). This display neatly summarizes the assessment but it does not provide many important details. From this display, one cannot determine the precision of the assessment. For example, the “+” in Figure 2 may reflect a very precise value or it may indicate a best guess within a 95% confidence interval that extends from 1 to 5. Only one point is displayed (usually the worst case), although a single event may produce several possible outcomes each of which may occur with different likelihoods and cause different consequences. All of the possible types of consequences are combined on a single scale, but the manner in which they are combined is not clear.

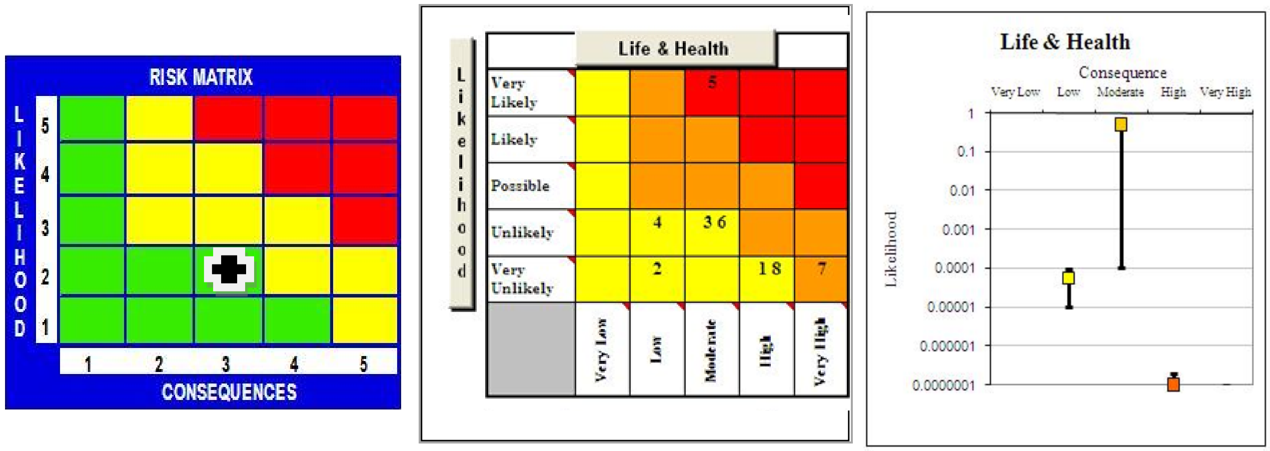


Figure 2. A Common Risk Matrix (left); TRIAD Life & Health likelihood X consequence display (center) and logarithmic risk display (right) showing a possible range of estimates.

In many cases, better decisions may be made if the risks associated with different possible outcomes are displayed, different displays are used for different types of consequences, and confidence intervals around estimates are depicted. TRIAD includes these enhancements (see Figure 2). Different 5 (consequence) X 5 (likelihood) matrices are used to display different consequence dimensions. The evaluators' best estimates of the likelihood and consequence values of each possible outcome (identified by number) are displayed on these graphs (in the center pane of Fig. 2). Auxiliary graphs display the plausible range of likelihood for each outcome (in the right pane of Fig. 2).

Conclusion

In aviation, managers and regulators continually assess risk. However, the heuristics that are commonly used have inherent problems that can render the assessments invalid. Relatively simple steps can be taken to substantially improve the quality of risk assessments even when quantitative data is sparse and traditional probabilistic risk assessment techniques cannot be applied. TRIAD is one tool that can support such comprehensive risk assessment, and can support improved decision making.

Acknowledgements

We thank Nathan Dieckmann, Christina Sheppler, and Ray Vukcevich who assisted with the development of TRIAD.

References

Bier, V. M. & Cox, L. A. (2007). Probabilistic risk analysis for engineered systems. In Edwards, W., Miles, R.F., von Winterfeldt, D. (Eds). *Advances in decision analysis: From foundations to applications*. (pp. 279-301). New York, NY, US: Cambridge University Press.

Kasperson, R. E., Renn, O., Slovic, P., Brown, H. S. (1988). The social amplification of risk: A conceptual framework. *Risk Analysis*. 8, 177-187.

Kunreuther, H. (2002). Risk analysis and risk management in an uncertain world. *Risk Analysis*. 22, 655-664.

FLIGHT DECK DISPLAY TECHNOLOGIES FOR 4DT AND SURFACE EQUIVALENT VISUAL OPERATIONS

Lawrence (Lance) J. Prinzel III[1], Denise R. Jones[1], Kevin J. Shelton[1], Jarvis (Trey) J. Arthur[1], Randall E. Bailey[1], Angela S. Allamandola[1], David C. Foyle[2], Becky L. Hooey[2,3]

NASA Langley Research Center[1]
Hampton, VA

NASA Ames Research Center[2]
Moffett Field, CA

San Jose State University[3]
San Jose, CA

NASA research is focused on flight deck display technologies that may significantly enhance situation awareness, enable new operating concepts, and reduce the potential for incidents/accidents for terminal area and surface operations. The display technologies include surface map, head-up, and head-worn displays; 4DT guidance algorithms; synthetic and enhanced vision technologies; and terminal maneuvering area traffic conflict detection and alerting systems. This work is critical to ensure that the flight deck interface technologies and the role of the human participants can support the full realization of the Next Generation Air Transportation System (NextGen) and its novel operating concepts.

Background

The Next Generation Air Transportation System (NextGen) concept for the year 2025 and beyond envisions the movement of large numbers of people and goods in a safe, efficient, and reliable manner. NextGen will remove many of the constraints in the current air transportation system, support a wider range of operations, and deliver an overall system capacity up to 3 times that of current operating levels. New capabilities are envisioned for NextGen, including four-dimensional trajectory (4DT)-based operations, equivalent visual operations, super density arrival/departure operations, and network-centric operations. National Aeronautics and Space Administration (NASA) research, development, test, and evaluation (RDT&E) of flight deck interface technologies is being conducted to proactively overcome aircraft safety barriers that would otherwise constrain the full realization of NextGen. As part of this work, specific research issues associated with the NextGen Terminal Maneuvering Area (TMA) are being addressed: 1) the impact of emerging NextGen operational concepts, such as equivalent visual operations (EVO) and 4DT operations; 2) the effect of changing communication modalities within a net-centric environment; and, 3) the influences from increased pilot responsibility for self-separation and performance compliance. In the following, an overview of NASA's flight deck interface technology research thrusts for these areas is described.

NASA Collision Avoidance for Airport Traffic

A Collision Avoidance for Airport Traffic (CAAT) research thrust has been formulated to develop technologies, data, and guidelines to enable safe TMA operations. This work expands upon existing research and technologies for tactical and strategic surface operations awareness for the flight crew and also, provides additional, protective Conflict Detection and Resolution (CD&R) functionality for NextGen operations. CAAT integrates airborne and ground-based technologies, which include flight deck displays, conflict detection and alerting algorithms, on-board position determination systems, airport surveillance systems, and controller-pilot data link communications.

Taxi-NASA Head-Up Display

Previous research has shown that the key to preventing surface traffic conflicts is to ensure that pilots know: (a) where they are located, (b) where other traffic is located, and (c) where to go on the airport surface. The CAAT concepts promote these attributes by use of several visual display interfaces including a modified head-up display (HUD) concept based on Taxiway Navigation and Situation Awareness ("T-NASA") research (Foyle, Andre, McCann, Wenzel, Begault, & Battiste, 1996; McCann, Hooey, Parke, Foyle, Andre, & Kanki, B., 1998). The HUD display concepts, sketched in Figure 1 and 2, show current ground speed in digital format, the current taxiway, next cleared taxiway, centerline markers and virtual cones on the taxiway edge. Additional cues are given for turns. These cues consist of turn flags and virtual turn signs (similar to road way turn signs). Hold shorts are displayed with a single line drawn at the hold short location with a virtual stop sign (see Figure 2). A non-conformal taxi director display provides an intuitive display of the relationship between the taxiway centerline and the aircraft's landing gear position. These symbology elements have been shown to significantly enhance situation awareness and navigation precision that would be required for NextGEN equivalent visual operations (EVO). The CAAT system further enhances the HUD visual interface with audible alerts for deviation from the assigned taxi route ("Off Route, Off Route") and unauthorized crossing of a hold line ("Crossing Hold, Crossing Hold").

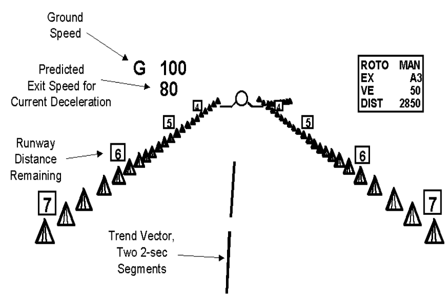


Figure 1. HUD Touchdown Symbology

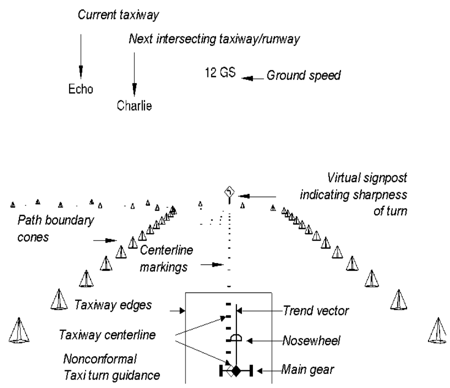


Figure 2. HUD Taxi Symbology

Conflict Detection and Alerting

A goal of CAAT is to provide an additional, protective safety layer of conflict detection and alerting for NextGen operations in the event that the tactical or strategic situation awareness (SA) is not sufficient or human errors or blunders occur. Ownship and traffic data are continually monitored to detect conflicts on the runway, at low altitudes near the airport, and during taxi and ramp operations for multiple classes of aircraft and surface vehicles. Alerts are designed for flight crew awareness and to identify potentially hazardous operational conditions that may require immediate flight crew response (see Figure 3). This work builds from substantial NASA testing for runway conflict detection and alerting (Green 2006, Jones 2002 and 2005, Jones, et. al., 2001, and Jones and Prinzel, 2006), however, low altitude and taxi conflict detection is in the initial development stage.

NASA is also investigating the concept of providing advisories or warnings for potential runway safety hazards. These indications are intended to increase the flight crews' situation awareness about relevant traffic that could affect runway safety. Research is also being initiated regarding the feasibility of providing resolution advisories (RA) for conflicts in the TMA without producing undesired consequences.

NASA Surface Map and Electronic Flight Bag Display Concepts

The increasing unavailability of radio-frequency bandwidth is driving a rapid shift from voice to data-link. By 2030 85% of Air Traffic Services communications are projected to be provided via data-link in the Airport/TMA environments (Eurocontrol, 2005). Net-centric operations hope to capitalize on a data-link environment's strengths. However, previous research has demonstrated numerous flight deck problems, including increased head-down time and pilot workload (e.g., Kerns, 1994; Groce & Boucek, 1987, Prinzo, 1998) which – in a NextGen environment with closer spacing and more pilot responsibility for 4DT separation – could significantly reduce safety margins. Furthermore, there are concerns of loss of "party-line" with data-link (e.g., Midkiff & Hansman, 1992; Pritchett and Hansman, 1995). For these and other reasons, NASA has been investigating the effects of data-link communication and potential visual display technologies that may mitigate, or eliminate, the potential deleterious effects of a voice-by-exception datalink NextGEN TMA environment. The concepts are based on emerging navigation, surveillance, and communicative technologies, such as CPDLC-all, ADS-B (in/out), TIS-B, etc.). The flight deck interface concepts include electronic moving surface map concepts (see Figure 3), head-up, and head-worn displays; and more critically, the information needs and modalities for the flight crew. For instance, the cockpit display of traffic information in a NextGen environment, with the addition of ADS-B intent information, may ameliorate issues of “party-line” information loss or inherent latencies in pilot-ATC communications under Controller Pilot Data-Link Communications (CPDLC), but traffic intent information may be critical to these operations. Unlike flight operations, current surface operations rely heavily on planned holds, following other traffic, and real-time updates to routing and other traffic. Without data-link intent information, these nuances may be lost and NextGen 4DT surface operations performance promises could be unrealized.

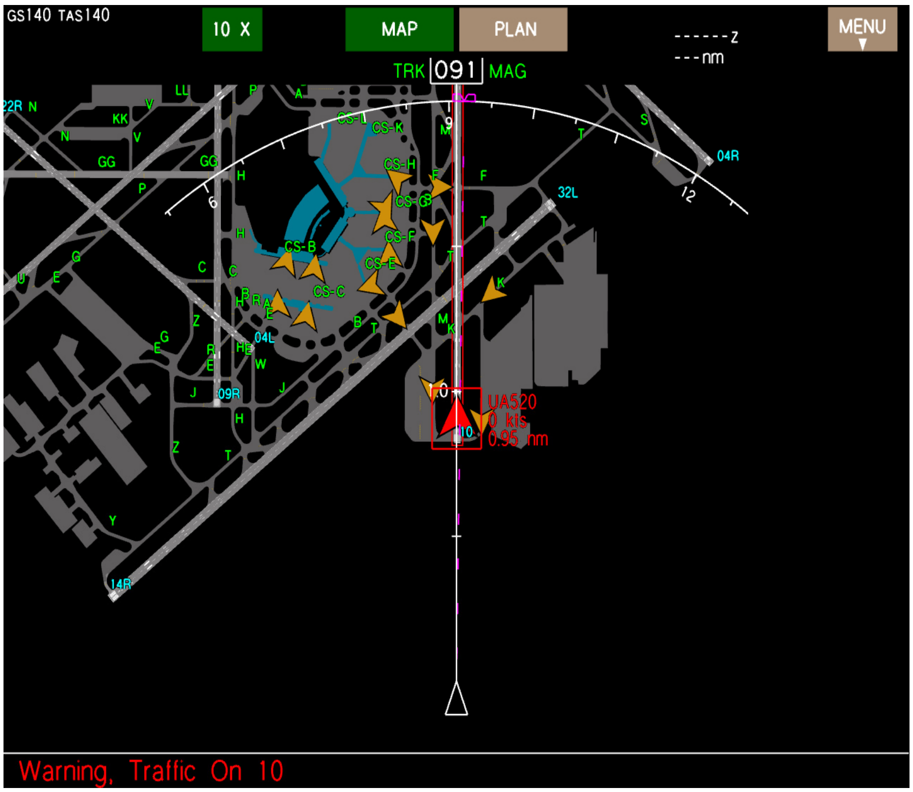


Figure 3. Runway Incursion Traffic Warning Alert (w/ audible alert)

Current research at NASA is focused on advanced surface map display concepts. The NASA surface map display provides traffic and manual query capability of other aircraft intent and graphical depiction of own-ship and target aircraft paths, and automatically prioritizes and selects aircraft, based on threat severity and/or proximity of traffic, and provides prediction and preview capability of other traffic and route conflicts. The surface map is shown in place of the ND when conducting surface operations (only on the pilot-not-taxiing side). The transition to the surface map is automatically done when on approach, the groundspeed is less than 80 knots, and all landing gear is touching the runway. Figure 4 shows the surface moving map with textual and graphical traffic icons displayed, own route graphically depicted in magenta, and the selected traffic’s graphical route and state information (30 sec trend) displayed, graphical (30/60/90 sec) intent prediction. Similar required- and estimated-time-of-arrival information and commanded speeds to meet RTAs are presented on the HUD based on a T-NASA HUD symbology set (Figure 5). These display concepts are supplemented by CPDLC interfaces on the Primary Flight Display (PFD) and Electronic Flight Bag (EFB). The HUD, PFD, and EFB also present 4DT enhanced (FLIR) and synthetic vision display information and advanced tactical and strategic guidance.

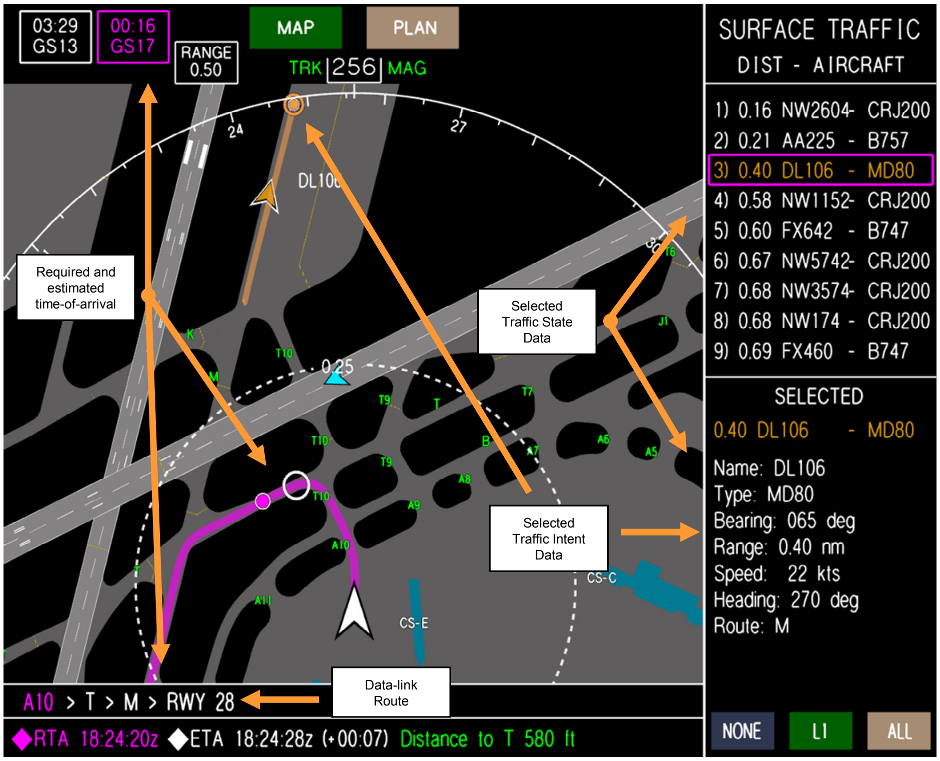


Figure 4. Example NASA Surface Map Display Concept

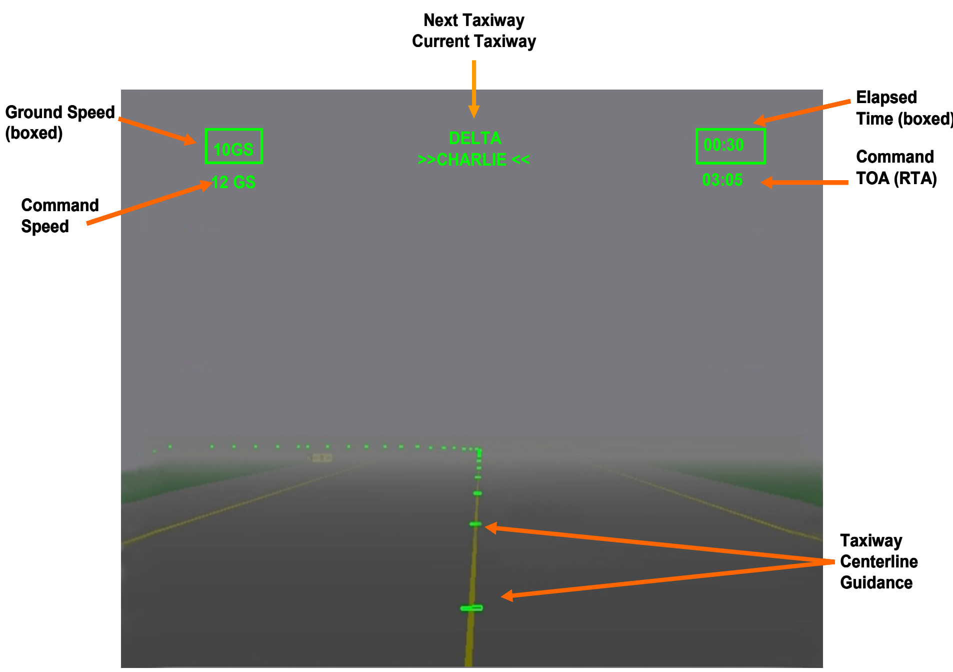


Figure 5. 4-DT Head-Up Display on Surface

NASA Head-Worn Display Concepts

Head-up, conformal information, such as that provided by the T-NASA HUD concepts, provide tactical and strategic awareness for the pilot-flying for safety and performance benefit. A major limitation of the HUD - for ground operations, in particular - is its monochrome form and limited, fixed field-of-regard. A monochromatic display has the inherent problem of being unable to use color for information de-cluttering and information cuing. Coupled with a limited field of regard, the HUD symbology must be carefully designed to optimize the information presentation to the pilot without increasing display clutter. NASA has been investigating emerging Head Worn Displays (HWDs) to resolve these limitations for NextGEN equivalent visual operations. The NASA HWD concept (Figure 6) is a head-tracked, color, unlimited field-of-regard concept that provides a 3-D conformal synthetic vision (SV) view of the airport surface integrated with advanced taxi route clearance, taxi precision guidance, enhanced vision, traffic data, and data-link capability. Simulation research (e.g., Jarvis, Prinzel, et al., 2006) has demonstrated significantly enhanced situation awareness, lowered workload, and taxi efficiency compared to existing head-up and head-down display technologies. The results evince the tremendous potential these displays have for enabling EVO during low-visibility complex terminal and surface operations.

Figure 6. NASA Head-Worn Display Concepts for Surface Equivalent Visual Operations

4DT Guidance Algorithms

NextGen surface traffic management (STM) concepts envision dynamic algorithms to generate speed- or time-based taxi clearances to calculate the most efficient movement of all surface traffic and enable precise surface coordination (see Cheng, Yeh, Diaz, & Foyle, 2004; Rathinam, Montoya, & Jung, 2008). The STM system provides speed or time commands to the pilots at various traffic flow points throughout the taxi route to regulate the required precision of surface traffic movements. The aircraft's taxi speed may be adjusted if the pilot is unable to conform to the speed command, if traffic is unable to comply, creating a reduction in separation, or if the needs of the dynamic airport surface require adjustment.

NextGen STM Concept Development

NextGen taxi operations represent a fundamental paradigm shift to include time-based or speed-based taxi clearances. NASA researchers are helping to define this new paradigm by considering the roles of pilots, ATC, and automation, and by defining procedural and operational requirements. Pilot-in-the-loop studies at NASA have evaluated different concept of operations including issues such as speed vs. time commands and single vs. multiple checkpoints. Advanced display concepts to support to these operations (which may be presented on a head-up display, an electronic moving map, or primary flight display) must ensure that they support pilots' 4DT taxi performance without increasing pilot workload, reducing situation awareness, or promoting excessive head-down time. One recent simulation study revealed significant reductions in time-of-arrival (TOA) error when pilots taxied using error-nulling speed guidance on the primary flight display. Future studies are planned that will evaluate the impact of pilot non-conformance, and STM reliability and system failures.

STM System and Algorithm Development

Since the time-based taxi concept is in its infancy, aviation human factors researchers at NASA are working to impact the design of the STM algorithms so that the resulting STM system does not exceed human performance capabilities. Specifically, pilot-in-the-loop simulation studies are underway at NASA that investigate the effects of: flight deck display bandwidth; number of traffic flow points; and time constraint window size for RTA (see Figure 7, from Foyle, Williams, & Hooey, 2008), as well as the impact of STM re-optimization (due to traffic changes, pilot performance). One recently completed simulation study characterized the distribution of pilots' TOA performance at traffic flow points to inform the development of STM algorithms with regards to the allowable time constraints of the STM system.

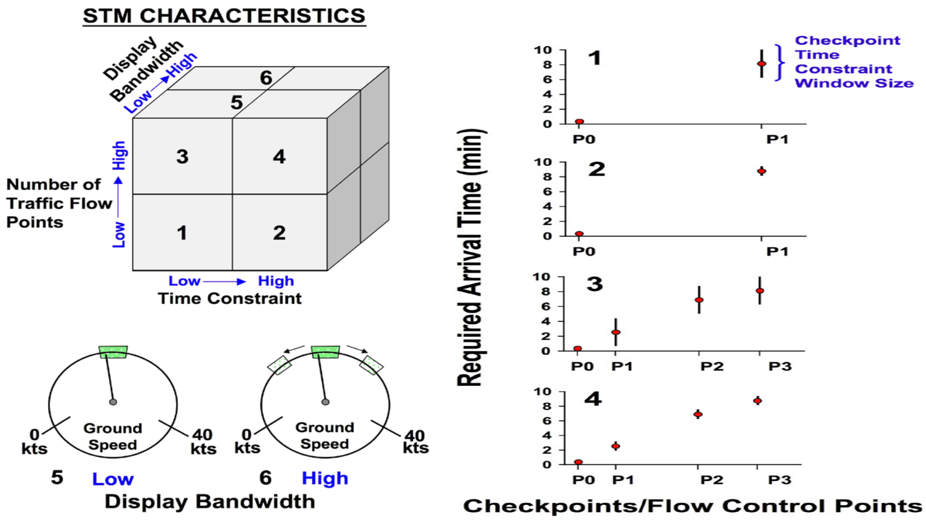


Figure 7. Characteristics of Time-based STM (see Foyle, Williams, & Hooey, 2008)

References

Cheng, V.H. L., Yeh, A., Diaz, G.M., & Foyle, D.C. (2004). Surface-operation benefits of a collaborative automation concept. Proceedings of the AIAA Guidance, Navigation, and Control Conference, Paper AIAA 2004-5409.

Eurocontrol (2005): *Initial Communications Operating Concepts And Requirements For The Future Radio System*. Eurocontrol/FAA Future Communications Study Operational Concepts and Requirements, January 2005.

Foyle, D.C., Williams, J.L., & Hooey, B.L. (2008). Toward the development of information requirements of time-based (4-D) taxi clearances: Pilot-in-the-loop results. NASA Airspace Systems Program Technical Interchange Meeting, Austin, TX March 17-20, 2008.

Arthur, J.J., Prinzel, L.J., Williams, S.P., & Kramer, L.J. (2006). *Synthetic vision enhanced surface operations and flight procedures rehearsal tool.* In J. Verly & J. Guell (Ed.), *Proceedings of Enhanced and Synthetic Vision* (Vol. 6226). Orlando, FL: SPIE.

Rathinam, S., Montoya, J., and Jung, Y. (2008). An optimization model for reducing aircraft taxi times at the Dallas Fort Worth International Airport. *26th International Congress of the Aeronautical Sciences (ICAS)* Anchorage, Alaska, Sep. 14-19, 2008.

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

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

Acknowledgements

The T-NASA and 4-DT guidance research described are conducted at NASA Ames Research Center. For more information, please contact Dr. David C. Foyle (david.c.foyle@nasa.gov) and Becky L. Hooey (Becky.L.Hooey@nasa.gov).

The CAAT, 4-DT surface maps, electronic flight bags, and head-worn display research described are conducted at NASA Langley Research Center. For more information about CAAT, please contact Denise R. Jones (denise.r.jones@nasa.gov); for NASA surface map and electronic flight bag research, please contact Lance Prinzel (l.j.prinzel@larc.nasa.gov) and Kevin Shelton (k.j.shelton@larc.nasa.gov); for NASA head-worn display research, please contact Trey Arthur (j.j.arthur@larc.nasa.gov) and Lance Prinzel (l.j.prinzel@larc.nasa.gov).

Research described in the paper is supported under NASA's Airspace Systems Program, Airportal Project (Safe and Efficient Surface Operations) and the Aviation Safety Program, Integrated Intelligent Flight Deck Project.

The authors would like to acknowledgments the tremendous contributions of many NASA Ames and Langley technicians, programmers, and engineers who support / have supported these research projects. Notable among them are Dr. Robert M. Norman, Lynda Kramer, Steven Williams, Steve Young, Sharon Otero, Glover Barker, Jerry Karwac, and Wei Anderson.

THE COMING PARADIGM-SHIFT IN MAINTENANCE: FROM METALS TO COMPOSITES

Alan Hobbs and Connie Brasil
San Jose State University Research Foundation
Moffett Field, CA

Barbara Kanki
NASA Ames Research Center
Moffett Field, CA

The purpose of this study is to examine the current maintenance practices of airline operators in the detection and repair of damage to composite structures, with the aim of learning lessons that will be applicable to the maintenance of future advanced composite airplanes. A process map was created to capture the events and activities that occur from the moment a damage event occurs, through damage detection, assessment and repair. The study is identifying areas where operational risks may negatively impact the process, where personnel are required to make judgments in the absence of procedural guidance, and areas where future tools or techniques may be of assistance.

The continued airworthiness of aging aircraft is the subject of a research project within the NASA Aviation Safety Program. The aging of aircraft structures is not necessarily an inevitable consequence of the passage of time, but is related to the accumulated effects of flight operations, exposure to environmental conditions, and events during ground handling and maintenance. For example the turnaround of an aircraft at the gate involves the coordinated movements of numerous vehicles and support equipment with the constant potential for contact with the aircraft. Most impacts will be inconsequential, but on occasion, an aircraft structure may sustain damage.

The principle of damage tolerance used in aircraft design ensures that aging-related structural damage can be detected and corrected before it presents a threat to the airworthiness of the aircraft (Goranson, 2007). For example, an awareness of the rate of crack propagation in various metal structures, combined with estimates of the probability of crack detection by an inspector, have allowed inspection schedules to be designed rationally to minimize the risk that an undetected crack will grow to a dangerous length between inspection intervals. From this example, it can be seen that the concept of damage tolerance, that is central to modern aircraft design, relies on two knowledge domains. The first domain is squarely within the field of engineering, specifically knowledge of materials and structures, and the conditions they are expected to encounter during their service life. The second domain is concerned with human performance, specifically the capability of maintenance and engineering personnel to detect, recognize, and rectify degraded conditions before such conditions become dangerous. The application of the damage tolerance concept requires the on-going collection and analysis of in-service data related to these two knowledge domains (Kim, Sheehy, & Lenhardt, 2006).

Aircraft structures have been fabricated from metals for over 70 years, and in that time, aircraft manufacturers and operators have accumulated experience in the design, maintenance, inspection and repair of metallic structures. The failure modes of metallic structures have been studied extensively, and parameters such as the rate of crack propagation can be estimated (Thompson, 2002). The human factors of inspection and damage analysis with metallic structures have also been widely studied (Drury, 1999) and the probability of detection can be estimated for cracks of various lengths under various viewing conditions (Ostrom & Wilhelmsen, 2008).

In line with wider industrial trends, the manufacturers of airline aircraft are now reducing the use of metals in aircraft construction and increasing the use of composite materials. A composite is a material composed of two or more ingredients that are combined at a macroscopic level, and are not soluble with each other (Kaw, 2006). For example a typical matrix composite material is made of woven carbon fibers set in an epoxy resin. Composites have been used in a wide range of products, including boats, consumer goods, military aircraft and advanced general aviation aircraft. Composites are used currently in a variety of structural and non-structural components in commercial airplanes. Examples range from basic fiberglass radomes, honeycomb core engine cowlings, lightweight winglets made of graphite-epoxy materials, and carbon fiber reinforced plastic (CFRP) materials comprising elevators, rudders, ailerons, and spoilers, up to the newest of the glass reinforced aluminum laminate (GLARE) technology utilized in fuselage skins in some aircraft. Composite materials provide advantages including weight savings, increased strength, resistance to corrosion, and aerodynamic efficiency. The next generation of airliners will be characterized by the increased use of composites in primary structures such as the fuselage, empennage, and wings.

Despite the promise and benefits that composite materials hold, they bring a new set of airworthiness issues. Although composites exhibit superior strength in many situations, composite failures can involve mechanisms very different from those of metals. For example, composites failures may involve delamination, fiber breakages, and fluid ingress. Compared to metallic structures, composite materials may also react differently to impacts or abuse, and may experience internal damage while showing little outward sign that damage has occurred. Fatigue cracks in a metallic structure will generally propagate over an extended time period, and the structure may retain much of its strength until an ultimate failure occurs. In contrast, some composite materials experience a sudden loss of strength when damaged.

As composite materials become increasingly important in the airline industry, it is necessary to understand the tasks that must be carried out by operational personnel to ensure the continued airworthiness of aircraft that include composite materials. Some tasks are likely to involve significant perceptual elements, for example, the detection of dents and delamination. Other tasks will largely involve decision-making and communication on matters such as the assessment of damage, and the subsequent repair action. On some occasions, social factors may come into play, for example, the willingness of personnel to report events where they may have caused damage to a composite structure, particularly when no visible damage is apparent. In contrast to trade skills such as welding, the field of composite fabrication and repair is characterized by a lack of standardization and the absence of consistent skill and knowledge requirements for technical personnel. However, as the aviation industry accumulates experience with composite materials, an increasing amount of regulatory standards and general guidance material is being produced by regulatory authorities, the military, and industry groups, notably the SAE Commercial Aircraft Composite Repair Committee (e.g. FAA, 1984; Department of Defense, 2002; Blohm, 2007).

Purpose of the Current Research

The purpose of the current research was to develop a methodology that can be used to examine the information sources, procedures, decisions, tools, expertise, and communication tasks relevant to the maintenance of composite materials on commercial aircraft. This methodology will then be used to help identify task elements that involve human performance-related risks. Such risks could include, but are not limited to; perceptual demands that exceed human capabilities, complex decisions that must be made in the absence of documented guidance, areas where task performance is reliant on expert judgment, situations where social factors such as a culture of blame could interfere with processes, and circumstances where there is a need for tools or technology not currently available. The methodology will be applied to the current state of the practice of managing aircraft composite damage in operations with the aim of identifying current operational risks as well as risks that may carry over to future advanced composite airplanes.

Development of the Methodology

Identifying the Broad Flow of Events

The first step was to identify the broad flow of events in the damage management process. Figure 1 shows five distinct types of events, beginning with events that present hazards to composite materials and moving in time order to ultimate damage mitigation, usually repair. Each stage is likely to involve a distinct population of operational personnel and specific human performance challenges.

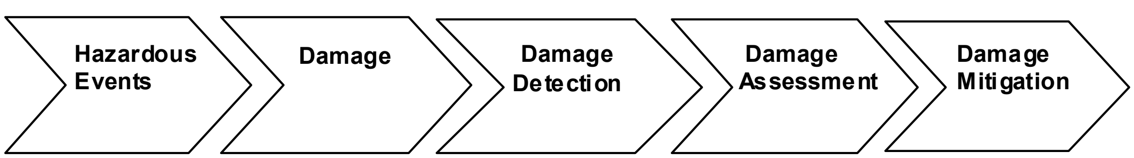


Figure 1. The flow of events in the damage management process.

Hazardous events. Hazardous events are defined as occurrences or conditions that have the potential to damage a composite structure. Hazardous events include bird strikes, in-flight exceedances such as flap over-speeds, maintenance errors such as dropped tools, ramp events, and weather phenomena. It is important to note that although a hazard may be a precursor to damage, a hazard does not necessarily lead to damage. The awareness that a hazard has occurred, however, is an important trigger that may lead to damage detection. The following report illustrates a maintenance-related hazard involving a composite panel on a Boeing 757-200. The incident is one of many that have been submitted to NASA's Aviation Safety Reporting System (ASRS). ASRS is a voluntary, confidential and non-punitive system that enables aviation personnel to report unsafe occurrences and hazardous situations.

> *I was performing an op Job Card on the #2 engine. This op includes an open cowling inspection and then an open up of certain borescope plugs. After the plugs had been installed, the cowlings were closed and some tools were left in the cold stream of the engine unintentionally. I did not realize the tools had been misplaced until after my weekend, which was 4 days later. …. After returning to work after 3 days off I was informed that damage had occurred to the #2 engine thrust reverser, composite panel, as result of the tools.* ASRS Report #463194

Ramp personnel such as baggage handlers and service vehicle drivers may observe hazards on the ramp, such as impacts involving vehicles, loading equipment or jetways. In some cases, the damage resulting from such events may not be clearly visible; as a result, the damage may remain undetected if the hazardous event is not reported.

It has been well established that various organizational factors can discourage the open reporting of incidents. Clearly, punishment of those who report errors or incidents actively discourages personnel from disclosing maintenance incidents. The potentially subtle nature of damage to advanced materials, such as barely visible impact damage or subsurface damage may create dilemmas for personnel who may have unintentionally created the hazard (i.e. dropping a tool) when there are no visible signs of damage and yet reporting the incident may lead to negative consequences for the worker (Boeing, 1994).

At present there are unanswered questions about the human involvement in the detection of, and response to, the events that can damage composite structures. The current research project is considering a range of human factor questions related to these hazardous events, including:

- *Information*: What are the sources of hazardous events? During what stage of operations do they occur, e.g. in-flight, ground handling, maintenance? What signs indicate that a hazardous event has occurred?
- *Procedures*: Are there appropriate and standardized procedures to guide the organization's response to a hazardous event report?
- *Decisions*: What influences whether a person will decide to report a hazardous event, particularly when the event involves a human action?
- *Tools*: Are some events detected via technologies such as on-board quick access recorders?
- *Expertise*: What operational personnel are in a position to detect hazardous events?
- *Communication*: How is information on hazardous events collected, documented and communicated to enable damage detection, assessment and mitigation to occur?

Damage Types. An important distinction can be made between the hazardous event as *cause* and the damage as *consequence*. Delamination, dents, and fiber breakages are examples of the consequential damage that may occur following a hazardous event. Delamination is a failure mode in which layers of the composite matrix separate, with significant loss of mechanical toughness. Common causes of delamination are repeated cyclical stresses and impact events. For example, in 1997, an Airbus A300 experienced an in-flight incident in which the pilot used excessive rudder inputs to steady the plane, imposing high lateral loads on the tail of the aircraft. The aircraft landed safely, and a preliminary visual inspection found no evidence that damage had occurred to the tail fin. In 2001, in response to an accident involving another A300, all A300-600 tail fins that had previously experienced high loads were required to go through ultra sound inspection. Severe delamination damage was found in one of the lugs that attach the vertical stabilizer to the fuselage of the A300 that had been involved in the 1997 incident (NTSB, 2004).

Damage Detection. There are three principal types of inspections during which damage may be detected. The first type is the scheduled inspection performed by maintenance personnel during transit checks, daily checks, and “letter checks” (A, B, C & D). Such checks typically include inspections of known problem areas where damage may occur. Most of these inspections are carried out visually. The pre-flight pilot walk-around can also be classed as a scheduled inspection. The second type of inspection is the “non-directed” inspection, or serendipitous discovery. This is where the technician or inspector was not engaged in actively searching for the damage at the time of its discovery. For example, Goranson (2007) notes that many cracks in aircraft metallic structures are discovered during non-directed inspections. Many cases of non-directed damage discovery have been reported to NASA’s Aviation Safety Reporting System (ASRS), as illustrated by the following example:

> *After removing vertical stabilizer panel (on Airbus A320) to inspect wiring harnesses per our job card, we noticed small cracks propagating from 2 hi-lok fasteners on the front spar next to the transverse load fittings. We informed inspection and they made a write-up. At this time, engineering is still trying to decide what to do. There is no type of NDT [Non Destructive Testing] that will tell us how deep the cracks are…. The reporter said the spar is of composite construction and no non-destructive testing methods or instruments are presently available and no repair processes are in the structural repair manual.* ASRS Report #613739.

The third type of inspection is the conditional inspection. These are initiated in response to a reported event that presents a hazard to the aircraft. Current maintenance procedures include conditional inspections triggered by events such as lightning strikes and heavy landings.

Scheduled and conditional inspections may involve one of three levels of inspection, either general visual, detailed, or special detailed (Kinnison, 2004). General visual inspections are unaided inspections (except for basic support equipment such as ladders and work stands) and are used to detect obvious damage. Detailed inspection involves intense visual inspection, sometimes with the use of lenses or mirrors. Areas may be cleaned in preparation for inspection. Currently 80-90% of inspections of composite structures are visual and that is unlikely to change significantly in the near future (Waite, 2007). Lastly, special detailed inspections are intense examinations of an area involving the use of special non-destructive testing (NDT) techniques. These techniques involve the use of technologies such as ultrasonics, thermography, and x-ray.

Human factor questions related to detection of damage in composites include:

- *Information sources*: How evident are the signs of damage? What signs of damage does the inspector look for?
- *Procedures*: What techniques are currently being applied to the detection of composite damage? What proportion of damage is detected through scheduled inspections/non-directed inspections/conditional inspections?
- *Decisions*: What decisions need to be made during inspections?
- *Tools*: What technologies are used to assist in damage detection and how are they used?
- *Expertise*: What skills, knowledge and training are required to perform inspections?
- *Communication*: How is information on detected damage collected, documented and communicated to enable damage assessment and mitigation to occur?

Damage Assessment. Once damage has been detected, it is necessary to assess its extent, evaluate its implications for airworthiness, and decide on a repair action. Most damage assessment decisions are guided by documentation such as the structural repair manual or maintenance manual. In other cases, engineering staff apply technical knowledge and expert judgment to design a tailored response, particularly when the damaged area is one that rarely sustains damage, or where no standard response is available. In complex cases, the engineering response may require consultation with the original equipment manufacturer.

Human factor questions related to the assessment of damage to composite materials include:

- *Information sources*: What factors are taken into account in decision making?
- *Procedures*: What guidance material is available to assist assessment?
- *Decisions*: What major decisions need to be made about damage assessment? Who is involved in these decisions? To what extent is damage classification a matter of judgment?

- *Tools*: How are NDT technologies used in damage assessment? How is the need for NDT determined?
- *Expertise*: What expertise is required to assess damage?
- *Communication*: How is information on damage assessment collected, documented and communicated to enable damage mitigation to occur?

Damage Mitigation. Damage mitigation is the final step of the process. Mitigation may take the form of a temporary repair such as speed tape, a permanent repair, or the replacement of the damaged component. Composite repairs can require specialized skills and careful attention to conditions such as correct storage of perishable materials, pressure, temperature and curing time. The conduct of a successful composite repair appears to be heavily reliant on accurate human performance, adequate training and appropriate standards and procedures. Deviations from prescribed process can significantly impact the strength of a composite repair (Tomblin, et al., 2007). However the focus of the current study was on the events leading up to the repair activities, rather than the specific activities involved in carrying out the repair.

Development of a Process Map and Interview Protocol

In order to identify the operational risks and human challenges associated with the maintenance of composite structures, a series of site visits are being made to aircraft operators, and interviews are being conducted with personnel who are Subject Matter Experts (SMEs) in the inspection and maintenance of composites. SMEs are drawn from throughout the organization, including those who may observe hazardous events, (including pilots, ramp workers, and maintenance personnel), those who perform scheduled inspections, and engineering personnel involved in damage assessment decision-making.

A process map was developed as a data collection tool to capture the general progression from damage-causing events through damage repair. The structure of the process map ensures that all areas of the operational process are covered during site visits and interviews. The process map includes three potential paths to damage discovery, which are in line with the three types of inspections; scheduled, non-directed and conditional. The process follows a "funnel" pattern, where the early stages can involve a broad range of potential hazard events, as well as many professional and employment groups, from maintenance technicians, to pilots to ramp workers. For example a pilot conducting a walk-around, a professional engineer dealing with a non-standard repair, and the driver of a catering truck who has just bumped an aircraft, will each have a unique contribution to make to the safety of composite materials, but they have their own responsibilities, priorities, and different information needs. As the process continues, and damage is identified and assessed, the process "funnels" down to a narrow range of participants with specialized skill sets and specific knowledge of composite materials.

Following a short introductory discussion, the SME is asked to recall a specific incident involving composite damage that was discovered via one of the three potential detection paths. The incident is then used as a focus for questions as the SME is prompted to identify the people involved at each stage of the process, the tasks they performed, the decisions they made, as well as the information sources, documents, tools, and communication needs at each stage.

Site visits and interviews conducted to this point have enabled the process flow map to be refined to focus on areas of operational risk. It became apparent that personnel tended to have very localized knowledge, in that they could describe their part in the process, but did not necessarily have a good awareness of parts of the process in which they were not directly involved. Therefore the interview protocol was modified to target sections of the process flow according to the roles and expertise of the SME.

Conclusion

The increasing use of composite materials in commercial airline aircraft necessitates an improved understanding of the human involvement in their maintenance. Not only are composite structures significantly different to the metallic structures they are replacing, but the human factors involved in maintaining composite materials may also be significantly different. The lessons learned in the maintenance of existing composite structures on current aircraft are of great potential value as airline manufacturers increase their use of composite materials. The process flow model being developed as part of this study may be the first time that the processes involved in composite materials maintenance has been mapped with a view to identifying the human performance demands of the process and potential operational risks.

We have yet to see how the management of damage in future composite structures will differ from the current processes and practices used for metallic structures and current composites. However, the systematic mapping of current processes, and the gathering of the experiences of operators, will make it possible to identify parts of the existing processes that have the potential to present uncontrolled human performance-related risks, and will thereby predict issues that may arise in the detection, prediction and mitigation of damage in future composite structures. An enhanced understanding of human-related risks may help to inform the development of future technologies, practices, and guidance material, ensuring that the advantages of advanced composite materials are not undermined by uncontrolled process risks.

References

Blohm, C. (2007, November*). Progress Report of the Commercial Aircraft Composite Repair Committee (CACRC).* Paper presented at the Commercial Aircraft Composite Repair Committee meeting, Wichita, KS.

Boeing (1994). *Composite Structure Awareness.* (DVD available through Aircraft Technical Book Company).

Department of Defense (2002). *Composite Materials Handbook*, MIL-HDBK-17. Washington, DC: Author.

Drury, C. (1999) Human Factors in Aviation Maintenance. In D. Garland, J. Wise & V. Hopkin (Eds.). *Handbook of Aviation Human Factors* (pp. 591-606). Hillsdale, NJ: Erlbaum.

Federal Aviation Administration (1984). *Composite Aircraft Structure*. Advisory Circular 20-107A. Washington, DC: Author.

Goranson, U. (2007, September). *Damage Tolerance Facts and Fiction*. Paper presented at the International Conference on Damage Tolerance of Aircraft Structures, Delft Technical University, Delft, The Netherlands.

Kaw, A. K. (2006). *Mechanics of Composite Materials*. Boca Raton, FL: Taylor and Francis.

Kim, Y., Sheehy, S. & Lenhardt, D. (2006). *A Survey of Aircraft Structural-Life Management Programs in the U.S. Navy, the Canadian Forces, and the U.S. Air Force*. Santa Monica, CA: Rand Corporation.

Kinnison, H. A., 2004. *Aviation Maintenance Management*. New York: McGraw Hill.

National Transportation Safety Board (2004). *In flight separation of vertical stabilizer American Airlines Flight 587 Airbus A300-605R, Belle Harbor, New York November 12, 2001*. NTSB/AAR-04/04. Washington, D.C.: Author.

Ostrom, L.T., & Wilhelmsen, C. A. (2008). Developing risk models for aviation maintenance and inspection. *International Journal of Aviation Psychology, 18*, (1), 30-42.

Thompson, R.B. (2002). Nondestructive Evaluation and Life Assessment. In W.T. Becker & R.J. Shipley (Eds.), *ASM Handbook, Volume 11: Failure Analysis and Prevention* (pp. 269-273). Materials Park, OH: ASM International.

Tomblin, J.S., Salah, L., Stevenson, B., Borgman, M., Bohaty, R., Motos, A. & Davies, C. (2007, November). *Bonded Repair of Composite Airframe Laminate and Sandwich Structures*. Paper presented at the Commercial Aircraft Composite Repair Committee meeting, Wichita, KS.

Waite, S. (2007, May). *Perspective on safe maintenance practice - some regulatory concerns*. Paper presented at the FAA/EASA/Industry Composite Damage Tolerance and Maintenance Workshop Amsterdam.

THE EFFECT OF HUMAN FACTORS IN AVIATION MAINTENANCE SAFETY

Andrea M. Georgiou
Middle Tennessee State University
Murfreesboro, Tennessee

Even with the increasing rate of technology innovation, the ultimate responsibility for the safety of a flight lies with humans. According to Boeing, human error accounts for 70% of commercial airplane accidents. This research aims to investigate the human factors that exist in aviation maintenance as well as the extent to which these factors affect safety. Utilizing the National Transportation Safety Board (NTSB) online accident database, the researcher reviewed accidents between 1996 through 2006 caused by maintenance-related errors. The results indicate the top four maintenance errors with the highest number of fatalities were: a).failure to properly complete tasks, b).improper maintenance, c).improper installations, and d). failure to detect or identify problems. In addition, the human factors most prevalent among the attitudes of both Aviation Maintenance Technicians (AMT's) and the Federal Aviation Administration (FAA) officials were demanding deadlines, environmental / personal distractions, and lack of proper use of maintenance manuals or instructions.

Whether a trip is planned for leisure or work, air travel plays a vital role in the day to day lives of individuals worldwide. A vast majority of the population from the working class to the upper class travel through the air transportation industry and are therefore directly affected by aviation safety. The large scope of individuals concerned with safe air travel forces the constant surveillance of accidents and incidents by the Federal Aviation Administration (FAA) government agency, National Transportation Safety Board (NTSB), researchers, and public attention through media. In 2008, Southwest Airlines gained some unwanted attention when the largest fine in FAA history was issued of 10.2 million dollars for allegedly flying at least 117 of its planes in violation of mandatory safety checks (Levine, 2008; Griffin & Bronstein, 2008). Not only does this place increased attention towards Southwest Airlines, but also places scrutiny ten -fold in maintenance departments within all commercial airlines.

Although maintenance-related accidents are far less frequent than accidents caused by pilot error, the end result can be just as fatal. Maintenance personnel, pilots, Air Traffic Controllers (ATC), and Flight Dispatchers are just a portion of the people dedicated to ensure a flight travels safely from departure to arrival. While there are several facets impacting the safety of a flight, it begins on the ground with the Aviation Maintenance Technician (AMT), also known as Airframe and Powerplant (A & P) mechanic.

Overarching Research Questions

1. How many aircraft accidents with at least one fatality have occurred due to maintenance error from 1996 through 2006?

2. What human factors and to what extent do human factors affect a mechanics, AMT, ability to safely conduct maintenance?

3. What are some cost-efficient solutions to decrease the effects of human factors which would result in an increase in aviation safety?

Review of Related Literature

Various Human Factors that Influence Mechanic Performance

As far back as the first powered flight by the Wright Brothers flight in 1903, humans have built and flown aircraft which means human error has always played a role in safety. However, it was not until 1988, when the skin of an Aloha airlines Boeing 737 ripped open in flight, did the FAA conduct the first official safety meeting with respect to aircraft maintenance activities (Lu, 2003). Since then, the boom in human factors research proves that researchers, along with the FAA, understand the influence human factors holds on mechanics performance.

In efforts to place top priority on human factors, in the year 2000 the FAA issued an Advisory Circular (AC) 120-72 titled Maintenance Resource Management training. Within this document, the FAA defines human factors as the scientific study of the interaction between people and machines. The FAA coined the phrase Dirty Dozen, which identifies the twelve most common maintenance-related causes of errors. The Dirty Dozen are as follows: Lack of communication, complacency, lack of knowledge, distraction, fatigue, lack of resources, pressure, lack of awareness, lack of assertiveness, stress, norms, and lack of teamwork.

A less common, yet still insightful, cognitive model of maintenance error was developed by Alan Hobbs with the Bureau of Air Safety Investigation (BASI). Mr. Hobbs research identified the following eight types of errors and the frequency in which they occurred: memory lapse, work-arounds, situational awareness, expertise, action slips, work practice, technical inaccuracy, and perceptual difficulties. The most common error, memory lapse, occurred in 24% of the 127 errors reported by maintenance personnel. Following closely behind at 23%, the second most frequent type of error was the work-around errors. These errors include an individual's knowledge of the correct procedure, but belief it would be all right this time. An example is performing a task in a more convenient manner than that specified in the maintenance manual. Pressures to complete a task within a certain time frame also influence how a mechanic does his/her job. When faced with time pressures, many AMT's decided not to document their actions and failed to perform all the necessary steps in a task (Hobbs, 2000). Unfortunately, there is no way to completely eliminate time pressures because the AMT's that can perform the tasks quickest receive the most business resulting in higher profit.

Due to the fact that human error is inevitable, organizations and companies need to move from blaming an individual worker to implementing a systemic approach to handle maintenance errors (Hackworth, Holcomb, Banks & Schroeder, 2007). In 1996, Boeing developed the Maintenance Error Decision Aid (MEDA) process to "help airlines shift from blaming maintenance personnel for making errors to systematically investigating and understanding contributing causes" (Graeber, nd) . The three principles behind the MEDA principles are: positive employee intent, contribution of multiple factors that contribute to an error, and manageability of errors. With the MEDA process, the traditional way of investigating errors by finding a person to blame is replaced with the new effective method of learning what factors contributed to the error in order to prevent further mishaps.

Methodology

Participants

During the 2008 Mid-South Aviation Maintenance seminar, an FAA official announced the purpose of the research and informed the audience the location in which the researcher was located for voluntary participation in a human factors mechanic survey. AMT's approached the researcher to obtain the survey and was instructed to drop off the survey in an assigned container. Attached to the top of each survey was a university approved consent form along with an explanation for the need of the research. In addition, contact information was provided if the participant had questions regarding the survey or the study. There were 18 surveys collected with the participant average age approximately 46 years old, ranging from 26 through 67 years old. The research process also included open-ended interviews with FAA officials in which the researcher recorded notes in a journal along with tape-record of interviews.

Design Approach and Instruments

The purpose of the study was to learn what type of human factors affect AMT's performance and to what extent have human factors impacted the safety of the aviation industry. Due to the nature of the inquiry process, the researcher determined that a qualitative approach was necessary for the study. Wiersma and Jurs (2005) describe qualitative research as an inductive inquiry process without any preconceived theories or hypotheses for the data collection. The inquiry process included: a) designing and collecting human factors AMT survey, b). conducting interviews with FAA officials, and c). collecting and reviewing the NTSB online aviation accident database. Qualitative data analysis included condensing and organizing the data sets into categories that can be analyzed and placed in emerging categories, themes, and patterns (Gough & Scott, 2000, Wiersma & Jurs, 2005). After all the data was collected, the qualitative data coding began, and the key categories, themes, and patterns are reported in the findings and conclusions of the study.

During data analysis, a triangulation matrix was utilized to ensure focus on the three overarching research questions. The triangulation matrix is listed in Table 1.

Table 1. *Triangulation Matrix*

Overarching Question	Data set	Data Set	Data Set
How many aircraft accidents with at least one fatality have occurred due to maintenance error from 1996 through 2006?	**NTSB online aviation accident database	Researcher field journal	AMT Human factors survey
What human factors and to what extent do human factors affect a mechanics, AMT, ability to safely conduct maintenance?	**AMT Human factors survey	Interviews with FAA officials	Researcher field journal
What are some cost-efficient solutions to decrease the effects of human factors which would result in an increase in aviation safety?	**Interview with FAA officials	Researcher field journal	AMT Human factors survey

*** Indicates the data set largely responsible for answering the overarching question*

Data Analysis

Survey and Interviews

Once all the data was collected, the researcher began the qualitative data analysis. Given that qualitative research analyzes words, not numbers, it is critical to carefully analyze the data and then revisit the data for further analysis for possible categories, trends, and connections between categories (Ratclilff, 2008). In order to stay on course, data was organized into categories relating to the overarching research questions. Quantitative descriptive statistics was incorporated with the analysis of the AMT human factors survey. The procedure for analyzing the data from the interviews with the FAA were also analyzed for common themes as well as any other important responses the researcher felt would address the research questions.

National Transportation Safety Board Aviation Accident Database

To determine the number of maintenance-related aircraft accidents that resulted in at least one fatality for the ten year span of 1996 through 2006, a review of the National Transportation of Safety Board's (NTSB) aviation accident database was necessary. Each maintenance-related accident was copied from the website into a computer document and reviewed for emerging themes and categories. Just because a mechanical failure occurs during flight does not indicate it was the error of an AMT. For example, there were several accidents caused because of engine failure, in-flight separation of parts, and fatigue cracks. These were not accounted for as maintenance-related accidents unless the report specifically cited the fault of maintenance, such as improper or inadequate maintenance.

Findings

Synopsis of Research Findings

With the extensive research of the National Transportation of Safety Board's (NTSB) aircraft accident online database, the most accidents and fatalities occurred within part 91 General Aviation operators. From 1996 through 2006, there were 141 accidents resulting in a tragic 215 fatalities. This should serve as a warning that general aviation needs to improve aircraft maintenance programs. While there were 132 fatalities in part 121 Air Carrier operator, this occurred in only five accidents over a ten year span. Unfortunately, when a part 121 aircraft has an accident the results are generally more severe because of the large number of passengers on board.

Data analysis of the three data sets revealed there were common themes emerging from twelve categories. The top four mechanical errors with the highest number of fatalities in order were: a).failure to properly complete tasks, b).improper maintenance, c).improper installations and d). failure to detect or identify problems that occurred over an extended period of time. The following are the four most common categories of errors along with the attributes assigned for each category:

- I goofed : Accidents that fall under this category listed the probable cause or contributing factors as *failure* to properly complete the maintenance task. This category includes failure to properly torque, lubricate, attach, secure, tighten, adjust, rebalance or balance, and failure to install various parts.
- Failure to maintain: Indicates an accident occurred due to improper maintenance by maintenance person(s) or person acting as a mechanic such as owner/builder. Attributes for this category are as follows: improper maintenance, improper replacement ,misrouting of fuel lines, improper assembly, improper construction ,improper shimmying, misalignments, improper modification, and improper repair.
- Who needs directions?: Accidents that occurred from improper installations indicates the maintenance instructions or directions were not properly followed by the mechanic. A few examples include: improper installation of cylinders, fuel line, oil pump, and magneto contact points.
- Detective needed: The researcher discovered there were several accidents that occurred from failure to detect or identify problems that occurred over an extended period of time. While these issues are not always easy to detect, failure to notice these often subtle issues during inspections can lead to serious repercussions. The accidents occurred from failure to detect or identify fatigue cracks, corrosion, erosion, worn cables, and fretting in propeller blade.

Table 2 lists the categories along with the number of fatalities and type of operations for each category, and Table 3 provides number of accidents for various types of operations along with the associated fatalities.

Table 2. *Number of fatalities and type of operation for each accident category*

Categories	Number of fatalities	Type of operation
I goofed	**123**	Part 91 General Aviation Part 121 Air Carrier Operator Part 135 Air Taxi & Commuter Part 137 Agricultural
Failure to maintain	69	Part 91 General Aviation Part 121 Air Carrier Operator Part 135 Air Taxi & Commuter Part 137 Agricultural
Who needs directions?	53	Part 91 General Aviation Part 135 Air Carrier Operator
Detective needed	43	Part 91 General Aviation Part 121 Air Carrier Operator Part 133 Rotorcraft External Load Part 135 Air Taxi & Commuter Part 137 Agricultural

Table 3. Number of fatalities within the type of operation

Number of fatalities	Type of operation	Number of accidents
215	Part 91 General Aviation	141
132	Part 121 Air Carrier Operator	5
25	Part 135 Air Taxi & Commuter	10
5	Part 137 Agricultural	4
2	Part 133 Rotorcraft External Load	2

The human factors that were most prevalent among the attitudes of both AMT's and the FAA officials were demanding deadlines, environmental and personal distractions, and lack of proper use of maintenance manuals or instructions. According to the AMT human factors survey, the top four distractions are as follows:

- 66% Cold/hot hangar temperatures
- 66% Interruptions while performing a task
- 44% Disorganization (having to track down proper manuals, tools, etc.)
- 38% Lack of resources

The most frequent stresses experienced at work are as follows:

- 61% Demanding Deadlines
- 50% Sick while at work
- 50% Tension among employees and/or employer
- 38% Excessive workloads

Combing the survey results and FAA interviews, the researcher discovered AMT's are not always following the appropriate manuals and rather performing tasks my memory. When asked how frequently do you perform a task from memory if it is a familiar task these were the responses:

- 61% Yes, I perform a task from memory if it is a familiar task.
- 16% No, I do not perform a task from memory even if it is a familiar task.
- 22% On occasion I perform a task from memory if it is a familiar task.

Suggestions for Improving Practice

All of the aforementioned human factors, as well as any other human factor that affects a mechanic's ability to safely perform tasks, must be taken seriously by mechanics, supervisors, FAA and NTSB officials, the United States government, and the general public. Safety should longer be compromised because of the desire to make profit. Maintenance safety training should no longer be voluntary, but rather made mandatory by Federal Aviation Regulations (FAR's). Why do certain errors seem to repeat themselves in the aviation industry? Perhaps errors occur because of pressure from management to complete a task and release the aircraft to the owner, or the AMT has some type or personal distraction that takes his focus off of correctly installing a part. The NTSB accident database does not report what caused the mechanic to make the error, but rather reports the specific error linked to the accident. Simply put- because mechanics are human there will always be human factors affecting their performance. The more awareness and training a person receives the more likely they are to recognize when human factors are affecting performance and take proper action to handle the situation.

References

Federal Aviation Administration. (2002). Advisory circular 120-72-*Maintenance Resource management training.* Washington, DC: Author.

Gough, S., & Scott, W. (2000). Exploring the purposes of qualitative data coding in educational enquiry: Insights from recent research. *Educational Studies, 26*(3).

Graeber, C. (n.d.) *The role of human factors in improving aviation safety.* Retrieved April 16, 2008, from http://www.boeing.com/commercial/ aeromagazine /aero_08/human_textonly.html

Griffin, D., & Bronstein, S (2008, March 7). *Records: Southwest flew 'unsafe' airplanes.* Retrieved April 5, 2008 from http://www.cnn.com/2008/US/03/06/southwest.planes/

Hackworth, C., Holcomb, K., Banks, J., & Schroeder, D. (2007). A survey of maintenance human factors programs across the world. [Electronic version]. *International Journal of Applied Aviation Studies, 7,* 212-229.

Hobbs, A. (2000). Maintenance human factors: Learning from errors to improve Systems. *Proceedings of the Fourth Australian Aviation Psychology Symposium*: *Vol. 2. Aviation Resource Management* (pp.347-355). Vermont: Ashgate.

Levin, A. (n.d.). FAA to check maintenance at all airlines. *USA Today*, Retrieved April 8, 2008, From Academic Search Premiere database.

Lu, C. (2003). An inductive study regarding nonregulatory maintenance resource management (MRM) training (Dissertation, The University of Nebraska at Omaha, 2003 UMI No. 3087733)

National Transportation Safety Board (NTSB). (2000). *Annual review of aircraft accident database. U.S. General Aviation, calendar year 2000*. Retrieved March 8, 2008, from http://www.ntsb.gov/publictn/2004/ARG0401.pdf

National Transportation Safety Board (NTSB) website. (n.d.) Accident database and synopses. Retrieved from http://www.ntsb.gov/ntsb/query.asp

Ratcliff, D. (2008, April/July). Qualitative data analysis and the transforming moment. *Transformation, 25.2 & 3.*

Wiersma, W., & Jurs, S. (2005). *Research methods in education. An introduction (8th ed).* Boston: Pearson Education, Inc

A Network Collaborative Design Construct for the Dissemination of Aviation Safety Research

Brent D. Bowen
Department of Aviation Science, Saint Louis University
St. Louis, MO

Erin E. Block
National Center for Aviation Safety Research, Saint Louis University
St. Louis, MO

Manoj S. Patankar
National Center for Aviation Safety Research, Saint Louis University
St. Louis, MO

Abstract

The constructs for collaborative network building include common tenets for the establishment of communication channels not only within the network but for constituencies external to the network. These constituencies are beneficiaries of the resulting knowledge which emerges and is disseminated. The Safety Across High-Consequence Industries (SAHI) conference was formed in 2003 for the purpose of bringing together safety leaders from multiple fields within the high consequence industries of healthcare, nuclear power, aviation, and others. Through SAHI four multinational conferences have been convened and resulted in bodies of safety knowledge available through widely distributed proceedings. In the process of generating and establishing the SAHI conference, an informal collaborative network of industry and academic leaders was formed with the goal of enhancing industry safety. Originating as an informal grouping of concerned parties, this collaborative network has evolved through several iterations and is currently being built around a more structured, technologically-based networking solution that formalizes the relationships and advantages that have been built through the previous generations of network collaboration. The next steps in this maturing evolution include the founding of the International Journal of Safety Across High-Consequence Industries and the formal establishment of the SAHI Collaborative Network. This paper serves the purpose of chronicling the development of SAHI and establishing the foundation for the launch of the International Journal of Safety Across High-Consequence Industries as a component of the National Center for Aviation Safety Research.

Development of the SAHI Collaborative

Several years ago, researchers from the health care and aviation industries wondered whether there were any safety best practices that could be transferred from aviation to health care. As the discussion evolved, the idea of a multidisciplinary conference focused on bi-directional transfer of best practices was born. In March 2004, Saint Louis University took the leadership role in hosting the first conference on Safety Across High-Consequence Industries (www.parks.slu.edu/sahi) with a goal to connect aviation, health care, industrial safety and other critical-incident industries. The founding organizers of SAHI: Jeff Brown, Tom Bigda-Peyton, Lou Halamek, Jim Bouey, and Manoj Patankar envisioned a forum for effective scientist-practitioner integration with regard to safety research and the role this type of social collaborative network could play in promoting safety.

The Safety Across High-Consequence Industries (SAHI) conference brings together professionals from the medical, public health, and aviation industries to discuss solutions to current challenges and directions for future research in safety. Convened every 18 months, the conference promotes cross-industry discussions of safety and a balanced, scientist-practitioner approach to addressing safety issues. Of significant importance, the SAHI conference is designed to provide a unique forum for researchers and practitioners to share their research results, present experiential case-studies and above all, forge new friendships that foster collaborative problem-solving. (SAHI Conference Program 2008) SAHI conferences, held at Saint Louis University include: 1st SAHI: March 9 & 10, 2004; 2nd SAHI: September 20-22, 2005; 3rd SAHI: March 13-15, 2007; 4th SAHI: September 10-12, 2008.

In a review of the most recent SAHI conference, Block and Bigda-Peyton (in press) identified the most critical safety concerns facing high-consequence industries. Conference attendees noted several key issues affecting safety: the need for improved understanding of open-systems interactions; the role of organizational culture in safety attitudes and behaviors; the importance of various approaches to organizational change management and their impact on long-term change sustainability; and the role of interpersonal communication in affecting the success of safety efforts. In consideration of these critical issues, researchers with the National Center for Aviation Safety Research and the Department of Aviation Science at Saint Louis University began formal development of the SAHI Collaborative Network.

Establishing a Formal Network Structure for the SAHI Collaborative

Collaborative networks are formed to bring together synergistic relationships for the purpose of provisioning an optimal foundation from which to pursue common goals. (Metz, 2007). Many collaborative networks begin through an informal or ad hoc gathering of parties based on common interests. Such was the case for SAHI until the introduction of structure to optimize and expand upon the initial successes. The formalized concept of a collaborative network for SAHI originated from work by Bowen and Lu (2004) on the development of the policy research construct. Within that exploration, Bowen and Lu conceived a process representation that allows for a methodological representation of a working construct for the purpose of building a model for applications in an environment such as SAHI. Through the application of the policy research construct and the research by Metz, the idea of a formalized SAHI collaborative network was generated. To operationalize this concept, Block conceived elements that would form an organizational structure to provide a sustainable and viable entity. (Bowen & Block, 2008)

Today the Safety Across High-Consequence Industries Collaborative Network has been established as an international collaboration with more than 100 active participants. Members and potential members are primarily to be found among safety leaders in various organizations, safety researchers at other institutions, and members of government offices concerned with safety. Joining the founding group, an organized core of multidisciplinary professionals brought further vision to the SAHI concept. These include Psychology (Sabin and Block), Business (Van Slyke and Miller), Aviation Education (Bowen and Kelly) among others. The resulting evolved Network provides members opportunity to interact and share ideas through participation in a knowledge-exchange forum for researchers and practitioners. This (primarily) electronic network further allows members to seek assistance with safety issues, share best practices, and engage leaders and researchers across industries in improving safety.

The Network centers on the 12 key contributors to the SAHI conferences, who serve as an active steering committee providing guidance and oversight for the conference and the activities of the Network. General membership in the Network is voluntary and is generally extended at the request of the potential member. At certain times the active steering committee may encourage particular researchers or safety leaders to join the Network if they have not done so. Members and potential members are primarily to be found among safety leaders in various organizations, safety researchers at other institutions, and members of government offices concerned with safety. While the primary focus of the National Center is on aviation safety research, the practical focus of the Network encourages interest and effort across high-consequence industries such as health care, nuclear power, environmental, as well as aviation.

Transition from Informal to Formal Network Accelerated by a National Research Center

The SAHI Collaborative is a key element of the recently established National Center for Aviation Safety Research (NCASR or National Center) at Saint Louis University. Improving safety in high-consequence industries continues to be a significant priority for both industry leaders and safety researchers (Block & Bigda-Peyton, in press). The Safety Across High-Consequence Industries Network Collaborative is thus established to be an extension of the National Center. The National Center has been crafted to be a dynamic organization which can transform and adapt to changing national priorities in aviation safety research focused in the areas of:

1. Business Case for Safety Management Systems
2. Safety Culture
3. Multi-risk Assessment
4. NextGen Safety Assessment
5. Incident Investigation
6. Maintenance Aviation Safety Action Programs

The goal of the National Center for Aviation Safety Research at Saint Louis University is to serve as the central resource for practitioners, researchers, and consultants to develop sustainable safety initiatives across air transportation, as well as other high-consequence industries. The National Center will sponsor experimental as well as applied/action research in aviation, health care and other high-consequence industries; publish a globally disseminated research journal; host the Safety Across High-Consequence Industries Conference; and develop specific training programs for multiple industries. (National Center, 2008).

Elements and Activities of the Maturing SAHI Collaborative

The goal of the Collaborative Network is to promote effective scientist-practitioner integration with regard to safety research. To accomplish this goal, efforts of the Collaborative Network are structured around 3 functions: 1) as a link between SAHI key contributors and industry/research leaders in safety; 2) as an entry mechanism for incorporating new industries and organizations into safety discussions and participation; and 3) as an outreach to industry, government, and the scientific community that is focused on aviation safety practices, but is firmly based in aviation safety research and draws from research of the NCASR. Specific industry partners such as airlines, air traffic control facilities, aviation maintenance organizations, health care facilities, nuclear power plants and others serve as the field sites for research. Lessons learned from one industry may be tested for transferability into another industry to maximize the benefits of multidisciplinary research and development efforts. (Parks, 2008)

In connection with the three functions of the Collaborative Network, a significant number of Network activities will focus on promotion of, and attendance at, the international Safety Across High-Consequence Industries conferences. These conferences are an ideal time for Network members to interact and share ideas with key contributors, to invite industry leaders to attend in hope of their future participation in the Network, and to outreach in promotion of the scientist-practitioner approach to safety program initiatives. In between conference meetings, members will publicize the work of SAHI and National Center in their organizations and encourage other organizational leaders concerned with safety to participate in the Collaborative Network through active, topical working groups. (Bowen & Block, 2008)

In addition to these activities, Network members are expected to participate in a knowledge-exchange forum for researchers and practitioners. This (primarily) electronic forum allows members to seek assistance with safety issues, share best practices, and engage leaders and researchers across industries in improving safety. Network members will have an opportunity to engage colleagues/researchers in an ongoing dialogue on relevant safety issues. This approach moves safety-critical discussions beyond the approximately annual meetings of the SAHI conference into the realm of a constant, iterative process of safety improvement. By participating in an electronic collaborative entity, Network members will be able to truly share information in a real-time format that encourages peer-to-peer learning; this truly embodies the key message of the fourth SAHI conference, in which industry safety leaders agreed that they "don't compete on safety."(Patankar, 2008)

Participants in the SAHI Collaborative will include a cadre of academic fellow appointees for the purpose of participation in graduate education. These fellows will form the nexus of multidisciplinary clusters that facilitate graduate seminars within Parks College. It is envisioned that fellow clusters will emerge in each area of SAHI focus. These areas include aviation, healthcare, power generation and transmission, and other high-consequence fields with a common core element of safety systems management. (Bowen, Lehrer, Patankar, & Block, 2008).

Research Dissemination Through Creation of a Multi-national Journal

In addition to the SAHI Collaborative Network, the National Center has launched, with world-wide and well-established expertise, the International Journal of Safety Across High-Consequence Industries (IJSAHI). The goal of the Journal is to cross boundaries so that overall systemic safety can result through the integration of research and industry practice. The foundation relationships and targeted outcomes are represented in an open conceptual design construct with intent to foster diverse membership growth and dissemination of aviation safety research world-wide. Initial foci include but are not exclusively limited to, the following fields: Aviation, Engineering, Health Care, Manufacturing, Nuclear Power, Security, Technology, and Transportation. Topical areas covered include:

- Systems Safety: Research and Practice, Scientific Process, Strategies, Initiatives & Outcomes
- Advanced Technology Systems: Design, Technology Integration & Improvements, Forecasting, Information Systems, Data-mining
- Culture: Ethics, Business, Management, Regulation, Safety Systems and Society, Policy Development & Implementation
- Human Factors: Engineering, Logistics, Collaboration, Simulation, Risk Management & Mitigation
- Education: Training, Communication, Learning Styles, Psychology, Case Study, Reporting Systems, Information Transfer & Collaboration
- Economics: Fiscal Implications, International Relations

Through the journal, a global network of aviation safety research dissemination has been created, to be linked electronically in an environment that fosters ongoing collaboration in addition to the multinational conference meetings. The inaugural issue of *The International Journal of Safety Across High-Consequence Industries* was launched in Spring 2009 at the International Symposium on Aviation Psychology. (Bowen & Fink, 2008)

Conclusion

The SAHI Collaborative Network and the IJSAHI will contribute to meeting the ongoing research and educational goals of the National Center for Aviation Safety Research. An action research model is employed to extract data, conduct modeling, and develop concepts for deployment and dissemination under the National Center's direction. The research results will continuously feed back to the programs of the NCASR and subsequently improve systemic safety. Development of both the Collaborative Network and the IJSAHI are innovative methods for creating a cross-industry focus on safety that moves beyond basic processes to incorporate system-wide issues. Participation in both the Collaborative Network and the IJSAHI by academic and industry community members is welcomed and encouraged. (Block & Bigda-Peyton, in press). Through Network participation critical issues will be addressed and result in effective scientist-practitioner integration with regard to safety research. The Safety Across High-Consequence Industries Collaborative Network will continue to bring together professionals from the medical, public health, power and aviation industries to discuss solutions to current challenges and guide directions for future research in safety.

Acknowledgements

The presentation of this paper was sponsored in part by the National Center for Aviation Safety Research at Parks College of Engineering, Aviation and Technology at Saint Louis University. Visionary leadership of the founding members of the SAHI concept (Jeff Brown, Tom Bigda-Peyton, Lou Halamek, Jim Bouey, Manoj Patankar) are recognized for innovation in safety research and practice. The authors thank Mary Fink for implementation of the journal concept and Amanda Pope for creative activity regarding SAHI, the NCASR, and the IJSAHI. The next SAHI Conference is scheduled for March 10-12, 2010 at Saint Louis University. For more information on joining the SAHI Collaborative Network or submitting work for review and possible inclusion in the IJSAHI, please contact Dr. Erin Block at eblock@slu.edu.

References

Block, E. E. & Bigda-Peyton, T. (in press). Critical issues facing high-consequence industries: Outcomes of the fourth annual Safety Across High-Consequence Industries conference. *International Journal of Safety Across High-Consequence Industries.*

Bowen, B. D. & Block, E. E. (2008, November). White paper on development of the SAHI Collaborative Network Construct: A Formalized Approach. *Prepared for Parks College of Engineering, Aviation and Technology. 3 pps.*

Bowen, B. D. & Fink, M. M. (2008, October). International Journal of Safety Across High-Consequence Industries call for papers. *Prepared for Parks College of Engineering, Aviation and Technology. 2 pps.*

Bowen, B. D., Lehrer, H., Patankar, M. S., & Block, E. E. (2008, October). White paper on graduate program development at Parks College of Engineering, Aviation and Technology. *Saint Louis University. 12pps.*

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.

Metz, N. (2007). Determining contributors to network effectiveness in networks with public-private partners: An examination of AGATE and SATS. *Dissertation Abstracts International, 68* (11), (UMI No. 3284044).

National Center for Aviation Safety Research established at Parks. (2008, December). *Parks Today, 81(2),* 5.

Parks College of Engineering, Aviation and Technology. (2008). *Fourth Safety Across High-Consequence Industries Conference.* [Program]. St. Louis, MO.

Patankar, M. S. (2008, September 19). SLU's SAHI Conference calls for 'No competition on safety'. *St. Louis Business Journal.* Retrieved March 13, 2009, from http://stlouis.bizjournals.com

SAHI. (n.d.). Retrieved March 13, 2009 from Saint Louis University, Parks College of Engineering, Aviation and Technology Web site: http://www.slu.edu/x25875.xml

TRANSFER OF SKILLS FROM MICROSOFT FLIGHT SIMULATOR X TO AN AIRCRAFT

Mark N. Callender, Wayne A. Dornan, Wendy S. Beckman, Paul A. Craig, Steve Gossett
Middle Tennessee State University
Murfreesboro, Tennessee

In the spring of 2008, with funding from the Aircraft Owner's and Pilot's Association, Middle Tennessee State University performed a study to evaluate the transferability of skills from Microsoft Flight Simulator X (MSFSX) to an aircraft for novice flight students. Nine students practiced tasks in six MSFSX Flight Lesson modules until the modules were successfully completed. The number of iterations required by students to accomplish each module satisfactorily was recorded. These students, along with nine others which comprised the control group, received flight training in a DA-40 for the same six maneuvers. They were subsequently evaluated on the number of attempts required to perform each maneuver successfully. The Transfer Effectiveness Ratio was utilized to calculate the transfer of training from MSFSX to the aircraft for each maneuver. The data suggest that the MSFSX packaged Flight Lessons modules have the capability to improve novice student performance in an aircraft.

With fuel, insurance, and maintenance costs increasing, the cost of flight training is continuing to rise as well. These increases, added to an already expensive endeavor, make affording flight training a more difficult task for flight students. To counter these effects, simulation has become widely used to support flight training curricula as a lower cost alternative. Several types of simulation devices are available and approved for training by the Federal Aviation Administration (FAA). Flight simulators and flight training devices (FTDs) are devices that provide high levels of realism with full sized cockpits and visual systems; the difference being that flight simulators provide force cueing (CFR, 2007, Part 61) while FTDs do not. These devices very closely replicate the aircraft they are meant to model; however, the cost of these devices also more closely model the prices of the aircraft they represent. This relegates their acquisition and use to larger flight training operations such as the military, airlines, and university aviation programs. The FAA has approved the use of lower cost options in the form of personal computer-based aviation training devices (PCATDs). PCATDs may be used for up to ten hours of instrument instruction in both Parts 61 and 141 (FAA, 1997). They are much simpler than simulators and FTDs, consisting of a computer, a monitor, flight and engine controls, and a means by which to control other devices such as flaps and radios. The cost of these devices make them much more accessible to smaller flight training operations; however, several thousand dollars (Koonce & Bramble, 1998) is still out of reach for most individuals. Non-FAA approved PCATDs are commercially available in the form of flight simulation games utilizing off-the-shelf gaming joysticks, yokes, and rudder pedals. With home computers becoming more commonplace, adding these store bought simulation systems can be done for less than $100. For anyone able to afford flight training, this cost is minimal. Advances in computer and simulation technology have brought these "games" from relatively humble beginnings into very realistic representations of flight, rivaling FAA approved systems. Although these inexpensive systems are not approved and cannot be logged, they may still benefit flight students. Currently the average time required for an individual to complete the Private Pilot Certificate is in excess of 75 hours (FAA, 2006), although the minimum Part 61 time required is only 40 hours (CFR, 2007, Part 61). If a training device were able to prepare flight students to more efficiently utilize their time in an aircraft, their aircraft training time could be significantly reduced; thereby, reducing the cost of flight training.

In the spring of 2008, Middle Tennessee State University (MTSU) began a several month long research project, funded by the Aircraft Owner's and Pilot's Association's (AOPA's) Air Safety Foundation (ASF), in an effort to assess Microsoft Flight Simulator X 's (MSFSX's) effectiveness as a training aid for ab initio pilots. The study followed eighteen subjects from zero flight experience to successful completion of selected Private Pilot tasks (FAA, 2002) in MTSU's Diamond DA-40 aircraft. The effectiveness of MSFSX was determined based upon the established metric, the Transfer Effectiveness Ratio (TER) (Roscoe & Williges, 1980). This research program differed from other studies in that MSFSX is a non-FAA approved, inexpensive, commercially available system which can be used independently of a flight instructor; and, the tasks evaluated were directed at ab initio pilots.

Transfer Effectiveness

Training aids are of benefit only if the experience they provide positively transfers to the aircraft. Positive transfer means that time spent using the training aid reduces the amount of time spent training in the aircraft. Neutral transfer indicates that use of the training aid had no effect on training time in an aircraft, while negative transfer implies that more time was spent in the aircraft than otherwise necessary, possibly due to poor habits imparted by the training aid. One method of determining the relative value of training aids is the TER (Roscoe & Williges, 1980). This metric compares two groups and their number of attempts at a particular task in the actual environment until acceptable performance has been reached. One group has the opportunity to practice the task by using a training aid. The other does not. The number of attempts taken to achieve proficiency by using the training aid then normalizes the difference between the numbers of attempts each group made in the real environment. Symbolically, the TER is given below:

$$TER = \frac{Y_o - Y_x}{X}$$

Y_o represents the control group's average number of attempts at a task in the actual environment until proficiency, given no prior experience. Y_x represents the experimental group's average number of attempts at the same task in the actual environment until proficiency, given prior experience utilizing a training aid. X represents the experimental group's number of attempts in the simulated environment until proficiency is reached. The TER directly indicates the number of attempts saved in the real environment relative to the number of simulated attempts. With information about the average time for each attempt and cost per hour of the aircraft and simulator, the TER also indicates time and cost savings achieved by simulation (Callender, 2008). Many research programs investigating transfer effectiveness of flight simulators and FTDs look only at TERs; however, these devices are very expensive, requiring substantial per hour fees. When this factor is analyzed, higher TER values become necessary in order to justify the use of simulation even with positive transfer for certain tasks. This is where lower cost simulation products become advantageous. They require much lower positive TER values to begin providing cost savings to flight students.

Method

Participants

This MTSU study solicited volunteers from the local area (Murfreesboro, TN). Eligibility for the study required that participants have no prior flight training, little to no experience using MSFSX, and comfort using a computer. Preference was given to individuals who answered affirmatively to having a strong desire to learn to fly. From the group of volunteers meeting these requirements, eighteen were randomly selected to participate. Nine participants were placed in the control group and trained in MTSU DA-40 aircraft by MTSU certified flight instructors (CFIs). Nine other participants were placed in the experimental group to receive training using the MSFSX package followed by training in the DA-40. The participants were not enrolled in a collegiate flight training program.

Apparatus

The purpose of this study was to evaluate the effectiveness of MSFSX; therefore, this software constituted the main component of the experimental simulation system. MSFSX is unique in that it has built in interactive lessons utilizing a virtual flight instructor. This system, unlike most other flight simulators, FTDs, and PCATDs, provides instruction, with feedback, without outside assistance; therefore, no CFI was necessary for operation of this system. The hardware consisted of a Dell Optiplex 745 personal computer which met the minimum requirements of the software, with a 19" flat panel display, and a Saitek PS33 Aviator joystick with integrated throttle levers. The system was placed on a table top with a chair for the participant and a chair for an observer. The aircraft used by both groups were MTSU Diamond DA-40s equipped with round dial primary instrumentation. MTSU CFIs provided the necessary instruction for the aircraft training flights.

Training Curricula

The training curriculum used for the experimental group was based upon available lesson modules within MSFSX. The selected lesson modules corresponded to six predetermined Private Pilot tasks (FAA, 2002). Each lesson module consisted of a text based description/explanation of the lesson with the expectations for successful completion clearly stated. Each lesson began with audio instruction from the virtual instructor usually followed by a visual demonstration of the task. The participant was then asked to perform the task within the prescribed tolerances. Exceeding the tolerances resulted in a visual alert in the form of a message at the top of the screen and a verbal alert from the virtual instructor. Lesson modules were completed in a specified order, with completion of one lesson being prerequisite to completion of the next. Participants in the experimental group first completed all of the relevant MSFSX lessons before transitioning to the aircraft, while control group participants immediately began training in a DA-40. The same six tasks were trained in the aircraft in the same order as that prescribed for MSFSX. Instruction in the aircraft was given by two MTSU CFIs following a script in order to standardize instruction to all participants. The CFIs verbally introduced/explained a task, demonstrated the task, and asked the participant to perform the task to certain standards. The standards used mirrored those within the MSFSX lessons.

Data Collection

The tolerances within MSFSX were the basis for evaluation both within the simulation and in the aircraft. During the MSFSX training, an observer recorded, on a data collection form, the number of attempts it took a participant to complete a task without exceeding any parameter indicated by the program. In the aircraft, the CFI first identified a tolerance exceedance and then recorded the number of attempts it took a participant to complete a specified task without tolerance exceedance on a similar data collection form. Both the MSFSX observers and the CFIs were given training within MSFSX or a DA-40 FTD, as appropriate, in recognizing and recording tolerance exceedances prior to working with participants.

Design

This experiment utilized a control group and an experimental group. The control group received training in the DA-40 aircraft only. The experimental group received training in both the aircraft and MSFSX. The independent variable was whether or not a participant received prior preparation in MSFSX. The dependent variables were the number of attempts until successful completion of the six tasks trained. With only six dependent variables, t tests were performed, following F tests for variance, in order to assess whether significantly fewer attempts were required by the experimental group to achieve proficiency at the prescribed tasks.

Results

Since there were only two groups being compared and a relatively small number of tasks evaluated, simple *F* and *t* tests were used to evaluate the difference between the mean numbers of attempts for each group. Table 1 lists the mean number of attempts for each group by flight task. Table 2 lists the TER and *p* values for each task. Only one task showed a statistically significant difference in the number of attempts taken by each group in the aircraft; however, five out of six tasks resulted in positive TER values. The lack of significant differences between the majority of the means may be due to the small sample sizes of the groups coupled with the large variances within some of the tasks. In the case of Power-Off Stalls, the negative TER value may be indicative of negative learning effects due to the stall lesson within MSFSX, or it may also be due to the small sample sizes.

Table 1. *Average attempts to complete six piloting tasks in an aircraft*

	Experimental		Control	
Task	M	SD	M	*SD*
Straight-and-Level Flight	2	1.5	2.11	1.54
Constant Airspeed Climb	1.22	0.44	2.33	1.32
Constant Airspeed Descent	1.56	0.53	1.67	0.87
Slow Flight	1.56	0.73	2.11	1.62
Power-Off Stall	1.89	2.03	1	0
Steep Turn	2.78	2.82	3.56	2.6

Table 2. *Transfer Effectiveness Ratios (TER) for six piloting tasks*

Task	TER	*p*
Straight-and-Level Flight	0.04	0.88
Constant Airspeed Climb	0.36	0.03
Constant Airspeed Descent	0.03	0.75
Slow Flight	0.08	0.36
Power-Off Stall	-0.25	0.22
Steep Turn	0.23	0.55

Discussion

Positive TERs indicate that beneficial transfer of training occurred. The magnitude of the TER represents the extent to which this transfer occurred. That FTDs and flight simulators may provide significant positive transfer has been shown in recent studies (Macchiarella, Brady, & Lyon, 2008); however, positive TERs do not necessarily translate to financial benefit to the student pilot. Given the high acquisition and operational costs of flight simulators and FTDs, flight training institutions must charge substantial per hour fees for their use. This leads to a minimum positive value of TER at which a cost benefit will be seen by a flight student. If a task to be trained has a TER lower than this minimum value, although the transfer remains positive, training this task in the simulator will not necessarily benefit the student financially. MSFSX, with acquisition cost for the software and joystick under $100 and no operational costs thereafter, significantly reduces the minimum TER required to provide positive financial benefit to student pilots. The acquisition cost for the software places it within the reach of many flight schools and flight students unable to afford more expensive systems. Student pilots, utilizing MSFSX at home, can train more conveniently and frequently than otherwise possible. It has been shown that when the time spent training particular tasks in simulation increases, the transfer effectiveness decreases (Roscoe & Williges, 1980). This means that as the time spent using simulation increases, the amount of benefit gained in the aircraft does not increase proportionally. The TER is therefore reduced. For higher priced systems, this decrease in transfer effectiveness limits the amount of time that it is cost effective to spend in simulation. However, for MSFSX, even though the transfer effectiveness would also likely decrease as more time is spent, the lack of operational cost would allow extended use to provide ever increasing transfer without additional cost as a concern. This increased transfer could lead to pilots becoming more knowledgeable and proficient before attempting a task in an aircraft. This increase in skill level may be able to reduce the average time required to achieve the Private Pilot Certificate, which would also reduce the cost of obtaining the certificate.

Conclusion

The results of this study suggest that positive transfer is achieved when using MSFSX prior to training in an aircraft. An expansion of this study with larger sample sizes and more pilot tasks should be used to verify these findings. This study was performed in a highly controlled environment; however, MSFSX was designed to be used by individuals independently. The study summarized above constituted Phase I of a two phase AOPA-funded project. Phase II will follow participating flight students from non-collegiate flight training programs from zero time through receipt of their Private Pilot Certificate. Study participants will receive MSFSX, a joystick, and rudder pedals to use in their homes throughout their flight training. The average number of hours these students take to receive the Private Pilot Certificate will be compared to the average flight hours of students at the same training facilities who do not enroll in the study. Phase II is currently underway.

References

Callender, M. N. (2008). Transfer of cost effectiveness as guides to simulator/flight training device use. *Collegiate Aviation Review*. 26(1), 28-32.

Code of Federal Regulations. 14 CFR Chapter I Part 61 Certification: Pilots, Flight Instructors, and Ground Instructors. Retrieved February 23, 2009, from http://ecfr.gpoaccess.gov.

Federal Aviation Administration. (1997). *Qualification and approval of personal computer-based aviation training devices*. Advisory Circular (AC) 61-126, Author.

Federal Aviation Administration. (2002). *Private pilot practical test standards for airplane*. (FAA-S-8081-14AS). Author.

Federal Aviation Administration. (2006). *Student pilot guide*. (FAA-H-8083-27A). Author.

Koonce, J. M., & Bramble, W. J. (1998). Personal computer-based flight training devices. *The International Journal of Aviation Psychology*. 8(3), 277-292.

Macchiarella, N. D., Brady, T., & Lyon, B. S. (2008). An application of high fidelity ftds for ab initio pilot training: The way ahead. *Collegiate Aviation Review*. 26(1), 67-75.

Roscoe, S. N., & Williges, B. H. (1980). Measurement of transfer of training. In S. N. Roscoe (Ed.), *Aviation psychology* (pp. 182-193). Ames: Iowa State University Press.

COUNTERMEASURES TO MITIGATE EFFECTS OF FATIGUE AMONG FLIGHT ATTENDANTS: TO IMPROVE TRANSPORTATION SAFETY AND PRODUCTIVITY

Lori J. Brown
Western Michigan University, College of Aviation
Kalamazoo, Michigan, United States
John Niehaus
Western Michigan University, College of Aviation
Kalamazoo, Michigan, United States

As airlines restructure and cut corners to make ends meet, flight attendants are experiencing a new industry trend that must be put to rest. At many carriers, flight attendants are forced to work to the point of exhaustion because of poorly scheduled duty time, lengthened duty days, or flagrant company violations of schedules. Research efforts on human factors: including the effects of fatigue, sleepiness, sleep disorders and circadian rhythms—on transportation safety has become a top priority. Research has identified key findings concerning fatigue in the flight attendant occupation, where sleep deprivation and disruption of circadian rhythms are known to occur (Testimony of Patricia A. Friend, 2007). With models, new technology, and convenient logical interface tools, we can anticipate worker fatigue and improve safety. Decreasing fatigue and its associated errors, we would enable operational improvements to further meet business requirements of today's airlines, especially in these lean times.

As the deep concessions demanded of flight attendants during the recent and ongoing financial turmoil of the airline industry have taken hold; it has become clear that airline management hopes to keep crews working longer duty days, with greatly reduced time off between said duties. As stated by the AFA, "Some air carriers are routinely taking advantage of a "reduced rest" provision in the Federal Aviation Administration's Flight Attendant Duty Time and Rest Regulations which allows the minimum rest of nine hours to be reduced to eight." Flight Attendants have reported that in some cases they have forgotten to perform critical safety functions, including the arming of doors and even fallen asleep on the jump-seats.

The NTSB itself has recognized the danger posed by fatigue in the transportation industry, and has recommended setting work hour limits for transportation operators based on fatigue research, in the areas of pilot fatigue, air traffic control, and some research on maintenance fatigue. There is no doubt that pilot, air traffic control and maintenance fatigue is of serious concern; however, the industry also needs to realize the flight attendant fatigue is also a serious concern, particularly in the era of heightened security awareness (Testimony of Patricia A. Friend, 2007).

Research has shown that such work environments provided by the aviation industry, can result in an inability to get to sleep (which may lead to further disruption of the circadian rhythm) and to the accumulation of sleep debt. The results of these potentially cascading effects show themselves as a decrease in performance. Sleep loss has been shown in several studies to create waking neurobehavioral deficits; which include vigilance degradations, increased lapses of attention, cognitive slowing, short term memory failures, slowed physical and mental reaction time, rapid and involuntary sleep onsets, decreased cognitive performance, increased subjective sleepiness, and polysomno graphic evidence of increased sleep pressure (Nesthes, & Schroder, 2007).

A web-based survey conducted post 9/11, assessed the fatigue of flight attendants working for a major U.S. airline (Sherry & Philbrick, 2004). This web-based survey revealed pervasive fatigue on a number of dimensions using multiple measures. The authors concluded that the studied cohort was "clearly one of the most fatigued populations we have studied." The data from this study detailed that the average amount of sleep reported was 6.4 hours, an amount known to cause fatigue problems, particularly if continued over a number of days.

According to the Association of Flight Attendants CWA (AFA-CWA), the Federal Aviation Administration (FAA) finally delivered the flight attendant fatigue study to Congress, who requested it at AFA-CWA's urging in 2007. Originally due back to the Transportation and Infrastructure Committee in June 2005, the FAA had been ignoring the requests of AFA-CWA and Congress to release the results for over a year.

Patricia Friend, AFA-CWA International President said "Fatigue has been overlooked for too long which is what makes this study even more vital." The results confirm that flight attendants are frequently "experiencing issues consistent with fatigue and tiredness" and that "fatigue appears to be a salient issue warranting further evaluation." According to recommendations cited in the report, "based on the incident reports, flight attendant comments, and the outcomes from the sampling of actual duty and rest time, it appears that the opportunities for adequate rest for flight attendants need to be further evaluated."

Modeling to Minimize the Effects of Fatigue on Cognitive Performance
Different bio-mathematical models of fatigue are available for use by flight attendants. The following is a list of a few of the most accepted models and tools, including a very short description of each: (Neri D., & Nunnely S. 2004)

1. The Two-process Model (Achermann, 2004) is based on the assumption that there is a linear interaction between a sleep/wake dependent homeostatic and circadian process that generates the timing of sleep and waking.
2. The System for Aircrew Fatigue Evaluation (SAFE) (Belyavin and Spencer, 2004) is a program used to assess the fatigue implications of aircrew schedules and uses the QinetiQ alertness model.
3. The Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE) Model (Hursh, Redmond, Johnson, Thorne, Belenky, Balkin, Strom, Miller, and Eddy, 2004) is based on the assumption that there are three components: a sleep reservoir, circadian rhythm, and sleep inertia that combine additively.
4. The Fatigue Avoidance Scheduling Tool (FAST) is a fatigue assessment tool based on the above mentioned, SAFTE. This model predicts the effectiveness of humans based on the amount of sleep and allows users to determine the best schedule to avoid fatigue. This allows airlines additional risk management, and can be used as a safety and accident tool, training tool, and to predict performance for various work schedules (Hursh, S.R., Redmond, D.P., Johnson, M.L., Thorne, D.R., Belenky, G., Balkin, T.J., et al., 2004).

Countermeasures
We can use results garnered from previous fatigue studies to suggest potential countermeasures to sleep and circadian issues that flight and cabin crews encounter. Each individual crew member will benefit from these countermeasures differently, and will need to later decide which garners the best results for them. This is why education about fatigue and countermeasures is a crucial element of training. In order to maximize the success for each individual crew member, researches suggest, trying different combinations for different periods of time to discover what is the most effective (Fatigue Countermeasures Group, 2005).

One of the most crucial countermeasures is the early recognition of fatigue in yourself or other crew members. Individuals must recognize fatigue in order to address it. Since it is difficult for people to estimate their own alertness and fatigue levels, more objective criteria may help in assessment. Some of the signs that may be caused by fatigue are: forgetfulness, poor decision making, slower reaction time, decreased vigilance, communication difficulties, fixation, lethargic, and moodiness. If any of these signs are apparent, the individual can employ an alertness strategy. Alertness strategies can be categorized as:

<u>Preventive strategies:</u> Those used before flying or between flights to reduce the effects of fatigue, sleep loss, and circadian disruption. These are strategies that are employed prior to checking in for a trip, or during layover time. These techniques can help ensure restorative sleep and minimize circadian rhythm interruptions. At home: get the best possible sleep before flying, try to get at least 8 hours of sleep and use strategic naps. These techniques can help to decrease the likelihood of the crewmember starting the trip with a sleep deficit.

<u>Operational strategies:</u> Used during flights to maintain alertness and performance. The only things that can reverse physiological sleepiness, is a sleep period or nap. Strategic caffeine consumption while on duty to acutely increase your alertness can be effective, though is not recommended within several hours before going to sleep. Stay hydrated and be sensible about nutrition. Move, stretch, exercise (walk about the cabin), this is an advantage that flight attendants have over a pilot, the feasibility to get exercise. Caffeine, activity, artificial indoor lighting, or other stimulation, can mask sleepiness, and help you maintain a level of alertness until you can get sleep. These strategies do not necessarily affect the underlying physiological mechanisms of fatigue, but focus on managing fatigue during operations. Primarily, these short-term strategies help to stave off, or mask underlying physiological sleepiness

(Fatigue Countermeasures Group, 2005). It is important to note that, when an individual uses two or more of the countermeasures together, it can produce a “synergistic” approach, maximizing alertness and performance; thereby, increasing safety and productivity.

Herbal Countermeasures
Valerian root is the strongest of the herbal relaxants. It is used by some people who cite its calming effects to treat insomnia, stress, nerve disorders, headaches, gastrointestinal and respiratory problems, and smooth muscle cramps. Valerian root should not be used in high doses or for prolonged periods of time, as it can cause symptoms such as headache, nausea, and restlessness.

Kava kava is used by some as a muscle relaxant and to treat depression, nervous anxiety, insomnia, restlessness, as well as a host of other conditions. Side effects with frequent usage in high doses include weakness, leg paralysis, lack of motivation, and inflammation of body and eyes which leads to scaly rashes that can turn into ulcers.

Melatonin is a naturally occurring hormone produced by the pineal gland in the brain. Since its secretion increases at nighttime, and is correlated with the sleep/wake cycle, melatonin is being studied as a treatment for insomnia. Many companies claim that melatonin fights stress, aging, jet-lag, high blood pressure, and immune system deficiencies. However, not much is known about long-term side effects; so any use of melatonin should be under a medical doctor’s supervision. Melatonin is sold as a dietary supplement and is not approved by the FDA, so it is not regulated for purity (Fatigue Countermeasures Group, 2005).

High Lux Lights
Originally aimed at the treatment of seasonal affective disorder (SAD) and winter blues, NatureBright® Photodynamic Therapy (PDT) products have recently been applied to people with mood and cognitive problems, shift work fatigue, jet-lag, disturbances of the sleep-wake cycle, and premenstrual syndrome (PMS). These and similar products have been involved in many studies relating to shift (night) workers whom amount to an estimated 270 million workers (Leger, Philip, Jarriault, Metlaine, Choudat, 2008), many of which are flight attendants, pilots, and air traffic controllers.

The February ninth 2006 edition of the Harvard Gazette mentioned the medical school’s own research on light therapy, stating that “the eyes are part of a light reception system that can keep you alert when sleep starts to fog your brain” (Cromie, 2009, p. 2). The Harvard study also suggested that “light may be a powerful countermeasure for the negative effects of fatigue for people who work or study at night” (Cromie, 2009, p. 2). This is essential for pilots and flight attendants of whom, the safety of hundreds of passengers depend every night.

Researchers have been able to demonstrate that bright light pulses of 10,000 lux, at about 30 minutes a day, were able to help adjust employees to new circadian rhythms (Leger, et. al., 2008). The light entering the retina is said to affect neurons in the suprachiasmatic nuclei (SCN) of the hypothalamus, which is the compound that affects circadian light/dark cycles in humans. These neurons secrete a chemical called vasopressin, which studies have shown is a neuropeptide involved in synchronizing these cycles. When a person is undergoing bright light therapy, the properly tuned, high lux wavelength of light enters the retina and is thought to re-energize inactive neurons in the SCN. These neurons once again begin secreting vasopressin, allowing the subject to redevelop a normal sleep cycle (Forbes, Morgan, Bangma, Peacock, Adamson, 2009).

Western Michigan University and Nature Bright have collaborated on a pilot study designed to research the feasibility of high lux lights to mitigate fatigue for pilots, flight attendants, and air traffic controllers. The ethoses of the study are focused on “Long Haul” schedules for flight attendants and pilots. Some of the light products tested are:

Sun Touch Plus® Light and Ion Therapy system: The desk top device emits powerful 10000 Lux Sky Effect light and high density negative ions. Pilots, flight attendants, and air traffic control could consider using light therapy in their rooms before flight/shift to increase concentration, and reduce fatigue.

Dia® portable light therapy unit: The portable device emits 5000 Lux Sky Effect light (LED source) and is rechargeable. If pilots and flight attendants need to use light therapy when they are away or during flight, this portable unit can be used to help reset circadian rhythms.

Sky Effect® lighting: The overhead fluorescent lamp emits bright blue-enriched white light. Replacing office lighting for the Sky Effect lighting can create a therapeutic environment for increasing alertness and calmness. The crew and airports can use the Sky Effect light for calming travel anxiety, and increase staff concentration throughout.

At completion of the pilot study, a research project will be designed to look further at the mitigating effects of neurons secreting vasopressin, allowing the subject to redevelop a normal sleep, reduce fatigue, elevate mood, and increase business productivity.

New Technologies and Innovations

Boeing and Airbus now have the technology to integrate crew alertness systems into modern aircraft. These systems are designed to alert the crew if no crew activity is detected within a specified time limit; some even measure blink rate. After the silent "crew response" advisory is triggered, an aural warning is triggered. This continuous aural warning is sufficient to wake a pilot. Most of these systems work in conjunction with the Flight Management Computers. Although this integration may prove to be effective for pilots, where does this leave the flight attendants?

Crew-centered technologies such as integrated sensor/software systems are able to monitor the state of awareness of crew members and automatically assess fitness for duty in real time. These systems embody two innovations: utilizing true 3D sensor data and employing sophisticated machine-learning techniques to assess state of awareness, fatigue, and vigilance. This crew monitoring system utilizes a new high-speed, high-resolution sensor based on Structured Light Illumination (SLI) capable of capturing 3D scenes within the cockpit and cabin in real time; along with an advanced pattern recognition system that can monitor facial expressions and body language. These crew monitoring systems will constantly gauge the state of awareness of each individual crew member. The monitoring can also be customized to individual crew members; potentially providing very sensitive and accurate detection of crew mental and emotional states. This system may be combined with remote sensors for monitoring vital functions to enhance effectiveness.

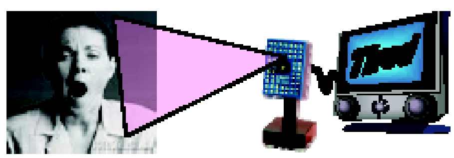


Figure 1: Monitoring crew members for fatigue

The Facial Action Coding System (FACS), a scheme for determining human emotional states, was developed over a span of years. There has also been work that defines a continuum of these measurable states on two axes: state of awareness (mapped between arousal and sleep) and state of well-being (mapped between pleasure and displeasure).The FACS approach has had successes in various arenas, particularly psychology (Michigan Aerospace Corporation, 2007).

1. Ekman, P., Friesen, W., Hager, J. (2002). Facial Action Coding System (FACS): Manual & Investigator's Guide, *A Human Face.*
2. Ekman, P., Huang, T., Sejnowski, T., Hager, J. (1992, August 01). Final Report to NSF of the Planning Workshop on Facial Expression Understanding.

Figure 2: Six of the seven basic emotions (Pantic, M., Context-sensitive Facial Expression Analysis, Man-Machine Interaction Group Electrical Engineering, Math and CS Delft Univ. of Tech).

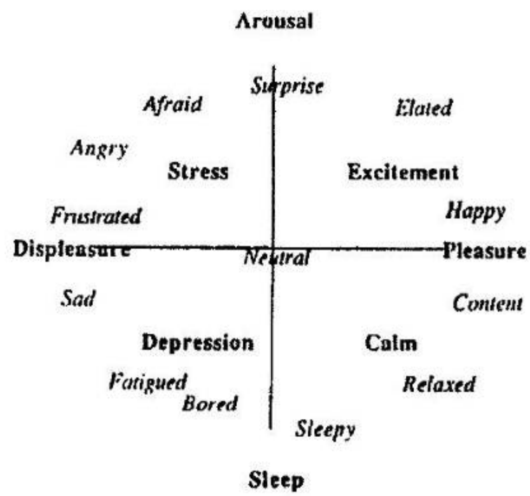


Figure 3: Emotions mapped by states of awareness

3. Gocke, R., (2005, April) Thermal Imaging and Facial Expression for Affective Sensing, GIW '05, Sydney, 27-28. Pantic, M., Patras, I., Rothkrantz, L. (2002) Facial Action Recognition in Face Profile Image Sequences, IEEE.

Crew Alertness Monitoring System (CAMS) – a sophisticated monitoring system for real-time assessment of the functional state of all crew members. This system will integrate the best available approaches and develop additional technologies as needed to attain performance levels deemed adequate for assessment. The platform will be flexible for easy addition of inputs to enhance performance. Figure 4 illustrates the process for integrating all bio-inputs in order to assess the functional state of an individual.

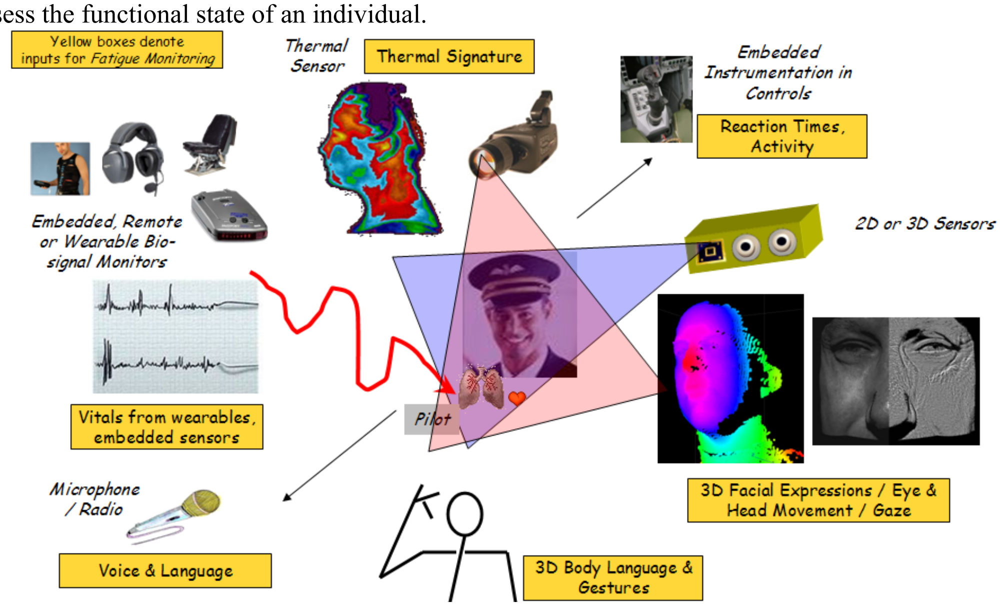


Figure 4: Process for integrating all bio-inputs in order to assess the functional state of an individual.

In summary
Recommendations for further study include: (House Rpt.108-671)
1) A scientifically based, survey of flight attendants to assess the frequency with which fatigue is experienced, the situations in which it appears, and the consequences.
2) A focused study of aviation incident reports in order to determine what role fatigue played in already reported safety incidents.
3) The need for research on the effects of fatigue. This research would explore the impact that rest schedules, circadian factors and sleep loss have on flight attendants' ability to perform their duties.

4) The determination and validation of fatigue models for assessing how fatigued a flight attendant will become.
5) Development of training material to reduce the level of fatigue that may be experienced by crews and to avoid factors that may increase fatigue levels.

With a convenient logical interface tool, we can anticipate worker fatigue, optimize schedules, reduce risk of error, and improve safety. We can also isolate fatigue related events through the use of mitigations and countermeasures, staffing analysis and workforce planning (Nesthes, D.J., Schroder, 2007).

It is abundantly clear that flight attendant fatigue is real; it is a problem, and one that is growing. Some may argue that an error caused by flight attendant fatigue is not as serious one caused by a pilot. However, an error caused due to flight attendant fatigue can lead to a tragic loss of life in the event of an in-flight emergency or during an evacuation. The recent ditching of US Air Flight 1549 is a perfect example of the link between safety and the flight attendant. To effectively address fatigue, we must combine regulations with operational practices, countermeasures, and education.

References

Cromie, W.J. (2006, February 09). When the Blues Keep You Awake. *Harvard Gazette*.

Ekman, P., Friesen, W., Hager, J. (2002). Facial Action Coding System (FACS): Manual & Investigator's Guide, *A Human Face*

Ekman, P., Huang, T., Sejnowski, T., Hager, J. (1992, August 01). Final Report to NSF of the Planning Workshop on Facial Expression Understanding.

Fatigue Countermeasures Group (2005), Human Factors Research and Technology Division, NASA Ames Research Center. Flight Attendant Fatigue, from http://human-factors.arc.nasa.gov/zteam/.

Forbes D, Morgan DG, Bangma J, Peacock S, Adamson J. Light Therapy for Managing Sleep, Behavior, and Mood Disturbances in Dementia. *Cochrane Database of Systematic Reviews* 2004, Issue 2. Art. No.:CD003946. DOI:10.1002/14651858.CD003946.pub2.

Gocke, R., (2005, April) Thermal Imaging and Facial Expression for Affective Sensing, GIW '05, Sydney, 27-28.

Hursh, S.R., Redmond, D.P., Johnson, M.L., Thorne, D.R., Belenky, G., Balkin, T.J., (2004). Fatigue models for applied research in war fighting. Aviation, Space, and Environmental Medicine.

Leger, Damien, Philip, Pierre, Jarriault, Philippe, Metlaine, Arnaud, Choudat, Dominique (2008, May 2005). Effects of a combination of napping and bright light pulses on shift workers' sleepiness at the wheel: a pilot study. *European Sleep Research Society*, DOI: 10.111/j.1365-2869.2008.00676.x

Michigan Aerospace Corporation, (2007), Crew Alertness Monitoring System

Nesthes, D.J., Schroder (2007, July), Flight Attendant fatigue Study, Dot/FAA/AM-07/21, FAA Office of Aerospace

Pantic, M., Context-sensitive Facial Expression Analysis, Man-Machine Interaction Group Electrical Engineering, Math and CS Delft Univ. of Tech

Pantic, M., Patras, I., Rothkrantz, L. (2002) Facial Action Recognition in Face Profile Image Sequences, IEEE.

Sherry, P., & Philbrick, K., (2004, May). Report on American airlines professional flight attendants fatigue survey. Paper presented at the meeting of the *Association of Professional Flight Attendants*, Dallas, TX.

Testimony of Patricia A. Friend, International President of the Association of Flight Attendants. CWA-AFL-CIO, Subcommittee on Transportation and Infrastructure, (2007, November 01).

The Departments of Transportation, and Treasury, (2007) *Appropriations Bill*, House Rpt.108-671.

SIMULATOR SICKNESS IN THE FLIGHT SCHOOL XXI TH-67 FLIGHT MOTION SIMULATORS

Catherine Webb, M.S.[1]
Julie Bass, SSG[1]
David Johnson, Ph.D.[2]
Amanda Kelley, Ph.D.[1]
Christopher Martin, M.D.[3]
Robert Wildzunas Ph. D.[1]

[1]US Army Aeromedical Research Laboratory, Fort Rucker, AL
[2]Army Research Institute, Rotary Wing Aviation Research Unit, Fort Rucker, AL
[3]US Army Aeromedical Center, Fort Rucker, AL

Simulator sickness (SS) is a common problem during flight training and can affect both instructor pilots (IPs) and student pilots (SPs). This study was conducted in response to complaints about a new rotary wing flight simulator. To investigate, Simulator Sickness Questionnaire (SSQ) data were collected from 129 SPs and 73 IPs. Analysis of these data helped direct recommendations based on the scientific literature for reducing SS. One year later, a post-test collected SSQ data from 50 SPs and 25 IPs. To test the effectiveness of the recommendations, a 2 (experience) x 2 (time) between-subjects Multivariate Analysis of Variance was used. There was a main effect of time and experience for the nausea, oculomotor and total scores of the SSQ. While it may never be possible to completely ameliorate SS in the new simulators, the recommendations that were implemented did reduce SS symptoms.

The phenomenon of simulator sickness (SS), a form of motion sickness caused by physical and/or visual motion in a simulator, has been well documented. Compared to motion sickness, the symptoms of SS tend to include more visual disturbances than gastrointestinal manifestations. The most accepted theory of SS is the sensory conflict theory proposed by Reason and Brand (1975), which suggests that sickness results when the vestibular, visual, and proprioreceptive senses perceive motion information that conflicts with expectations based on past experience of actual flight (Crowley & Gower, 1988).

The purpose of this study was to assess reports of SS in a new rotary wing flight simulator. Recommendations were provided to reduce, or preferably, eliminate the SS problems, and a post-study was conducted to evaluate the effectiveness of the recommendations. It was hypothesized that IPs would report more SS (in terms of prevalence and severity) than SPs. An additional hypothesis was that adherence to the recommended guidelines would reduce SS.

Methods

Equipment

The TH-67 Flight Motion Simulator is a full motion flight simulator manufactured by FlightSafety International (Broken Arrow, Oklahoma). Each simulator has a three channel panoramic visual system and a six-degree-of-freedom motion system. These simulators are used in Phase 1 of the US Army's Flight School XXI for instrument and military skills training. The same group of simulators was used over the course of the entire study.

Questionnaire

The Simulator Sickness Questionnaire (SSQ) is a well validated pen-and-paper questionnaire designed to detect the prevalence and severity of 16 possible symptoms generally associated with SS (Kennedy, Lane, Berbaum, & Lilienthal, 1993). Participants rate the severity of symptoms on a scale ranging from 0 (none) to 3 (severe). In addition to a total severity score, the SSQ yields a nausea, oculomotor, and disorientation subscale score.

Participants

Two hundred and two helicopter pilots from Fort Rucker, Alabama (73 IPs and 129 SPs) participated in the pre-study. Data from three participants (1 IP and 2 SPs) were excluded from the analysis due to insufficient data. Demographic data is presented in Table 1.

Seventy-five helicopter pilots from Fort Rucker, Alabama (25 IPs and 50 SPs) participated in the post-study. Demographic data is presented in Table 1. Of the 25 IPs in the post-study, 17 also participated in the pre-study; however they were not matched to evaluate individual change because the data was de-identified.

Table 1. *Participant Characteristics.*

	Pre-Study		Post-Study	
	IPs	SPs	IPs	SPs
n	73	129	25	50
Mean Age (*SD*)	51.1 ± 8.3	24.8 ± 3.2	51.9 ± 8.1	26.7 ±3.8
Mean Flight Hours (*SD*)	6541.7 ± 4515.8	48.6 ± 194.9	7770.4 ± 5195.8	27.6 ± 92.5

Procedure

Pre-Study

For the pre-study, data was collected over three, 5-day class cycles. On the first day of data collection, each IP was assigned two SPs. For each simulator session, one student flew the simulator, while the other student observed from the rear area of the simulator cabin. After 2 hours (hr), the students changed roles. The IP remained in the front seat during both sessions. On each day of the class cycle, the students and their IP completed the SSQ immediately after the simulator flight period. The participants did not use the same individual simulator for all 5 days of data collection; simulator assignments were based on availability. A total of 950 SSQs were completed in the pre-study.

Recommendations

After analyzing the data from the pre-study, a number of recommendations to reduce SS were provided to the directors of the flight training program. The recommendations that were implemented and incorporated into the training program were: simulator flights were reduced from 4 to 3 hr (1.5 hr per student); pilots were instructed to close their eyes before freeze/reset; and unusual or unnatural maneuvers were limited. The course was reduced from 5 days to 3 days since most of the hover training and ground work were removed from the program of instruction entirely. There was an effort to avoid improperly calibrated simulators (e.g., misalignment, out of focus, luminance mismatch, distortions) until repaired. And finally, emphasis was placed on stressing the importance of proper rest/health discipline, and giving instructors enough time to adapt and maintain adaptation.

Post-Study

Based on results of the pre-study, recommendations were made and implemented during the post-study. Procedures were similar to those in the pre-study however the class cycle was shortened from 5 days to 3 days; thus data was collected over a 3-day class cycle. Additionally, the time each student flew the simulator was reduced from 2 hr each to 1.5 hr. The SSQ was completed at the end of the simulator session on each of the 3 days of the class cycle. Data from 225 SSQs were collected in the post-study.

Results

Pre-Study

In the pre-study, participants completed the SSQ across one 5 day class cycle (i.e., five administrations). The most commonly reported symptoms overall included eyestrain, general discomfort, headache, and difficulty focusing. Regardless of severity, 72% of IPs and 91% of SPs reported at least one symptom over the course of the five sessions. As for the profile of the SSQ subscales, disorientation symptoms predominated, followed by oculomotor symptoms. Over the course of the 5 days, mean total SSQ scores ranged from 10 to 45.

Post-Study

One year following the pre-study, after the recommendations were implemented, the post-study was conducted. Instructor pilots and SPs completed the SSQ across one 3 day class cycle (i.e., three administrations). In the post-study, the most commonly reported symptoms included eyestrain, general discomfort, nausea and burping. With regard to frequency data, 64% of IPs and 90% of SPs reported at least one symptom, regardless of severity, over the course of the 3 days. The profile of the SSQ subscales was the same as that of the pre-study, with disorientation symptoms predominating, followed by oculomotor symptoms. Over the course of the 3 days, mean total SSQ scores ranged from 10 to 17.

Effectiveness of Recommendations

To determine the effectiveness of the recommendations in reducing SS, a 2 x 2 between-subjects Multivariate Analysis of Variance (MANOVA) was conducted. The two independent variables were experience (IP or SP) and recommendations for SS reduction (pre-study or post-study) and the four dependent variables were the differences in nausea scores, oculomotor scores, disorientation scores, and total scores of the SSQ. Differences in each SSQ subscale score and total score from the first administration to the last administration were calculated for each participant. Of particular interest was the comparison of the difference scores from the pre-study to those of the post-study. The MANOVA showed a significant main effect of experience, $F(4, 270) = 3.055$, $p = .017$, and a significant main effect of the recommendations, $F(4, 270) = 2.628$, $p = .035$. There were no significant interactions. Levene's test of equality of error variance showed that this assumption was violated. To account for this violation, the data were subsequently analyzed using independent *t*-tests (equal variances not assumed) and a Bonferroni correction was applied to reduce the risk of a Type 1 error ($p = 0.05/6 = 0.0083$).

Independent samples *t*-tests revealed a significant main effect of experience on nausea, oculomotor, and total SSQ difference scores (Table 2), such that IPs had significantly larger (more negative) difference scores, and thus experienced more SS than SPs. There was also a main effect of the recommendations on nausea, oculomotor and total SSQ difference scores (Table 2), indicating that difference scores were more negative in the pre-study. This signifies that those SS symptoms were more severe over the class cycle in the pre-study than in the post-study.

Table 2. *Mean Difference Scores (± SE).*

	Experience				Time			
	IP	SP	*t*	*p*	Pre-Study	Post-Study	*t*	*p*
Nausea	-13.14 ± 3.18	-2.19 ± 1.46	-3.13	.002	-8.79 ± 1.80	1.27 ± 2.48	-3.28	.001
Oculomotor	-12.05 ± 2.80	-2.88 ± 1.40	-2.93	.004	-8.32 ± 1.64	-0.20 ± 2.31	-2.87	.005
Disorientation	-15.09 ± 4.12	-3.81 ± 1.95	-2.48	.014	-10.08 ± 2.29	-1.67 ± 3.66	-1.95	.054
Total	-15.19 ± 3.55	-3.26 ±1.58	-3.07	.003	-10.26 ± 1.97	0.00 ± 2.87	-2.95	.004

Note: Negative difference scores indicate SSQ scores increased from the first administration to the last. Positive difference scores indicate SSQ scores decreased from the first administration to the last.

Discussion

According to Stanney, Kennedy, and Drexler (1997), simulators producing mean total SSQ scores greater than 15 are a concern, and scores greater than 20 indicate a "problem simulator." Consequently, the simulators used in flight training program could be classified as problem simulators for the pre-study, but not so for the post-study. In addition, the profile of the three subscales indicated that disorientation symptoms predominated in both the pre- and post-study, which is atypical of SS, in which oculomotor symptoms are most frequently observed. High disorientation scores are correlated to postural instability following simulator sessions (Kennedy, Berbaum, & Lilienthal, 1997), which raises concerns regarding ground safety (e.g., exiting the simulator, driving home from the simulator session, and even flying aircraft).

Rotary wing aircraft are known to cause higher rates of simulator sickness compared to fixed wing aircraft (Johnson, 2005). Reviews of rotary wing flight simulators found the occurrence of SS ranged from 13 to 70% (Wright, 1995). The occurrence of SS for both the pre- and post-studies (64 to 91%) are high compared to other frequency rates published in the literature for military flight simulators. There are several possible explanations or factors that may have contributed to the high frequency rate. For example, the logistics of the flight training program require an SP to be in the back of the FMS while another SP is in control. Degree of control is an important factor influencing SS, as sickness decreases as the amount of control increases (Johnson). Also, this study included several IPs with many thousands of hours of flight experience, another factor well known to increase susceptibly to SS (Johnson). Lastly, data was not collected regarding the prior histories of motion/simulator sickness in the participants.

Consistent with previous SS literature, in both the pre- and post-studies, IPs reported significantly higher SSQ scores than the SPs for all four SSQ subscale scores. While this finding was expected on the first day of simulator flight, the IPs showed an increase in SS symptoms over the 5-day course in the pre-study and the 3-day course in the post-study which was unexpected. Despite the role flight experience plays in SS, IPs would be expected to adapt to a simulator over time. There are a number of factors which may have contributed to this unexpected finding such as lack of control over previous day activities (simulator versus actual flight) and variability in instructor schedules. This is, of course, speculation and additional research will need to further identify the root cause of the absence of adaptation in the IPs.

According to Johnson (2005), the best current solution to SS is adaptation (i.e., developing a tolerance to the stimuli that produce sickness). This study revealed evidence of adaptation in the nausea SSQ score in the post-study. Perhaps, for the post-study, the 3-day class cycle was not long enough to adapt significantly to the other symptoms of SS. However, it is important to note that the implemented recommendations were in fact improving adaptation for both IPs and SPs as evidenced by the significant changes in difference scores.

Limitations

Although every effort was made to ensure the recommendations provided were implemented, factors such as costs and practicality limited the implementation of some recommendations. Additionally, some behaviors continued that were not recommended, such as positioning the SP in the back seat when not flying. In addition, data was unavailable as to which TH-67 FMS each individual participant used each day. This lack of consistency in simulator use introduces a potential confound to the study thus limiting the precision of conclusions. Future studies should track simulator use/assignment to determine if SS is more prevalent and/or severe in a particular FMS. Finally, in the pre-study, data was collected over three class cycles whereas data was only collected over one class cycle in the post study. Thus, the violation of the homogeneity of variance assumption was potentially due to the unequal sample sizes of the pre- and post-studies. Future studies should aim to ensure equal sample sizes when comparing group differences.

Conclusion

Flight simulators are a safe and cost effective alternative to actual flight and are an invaluable tool for training SPs. However, as the Army relies on simulator technology, it cannot afford to ignore the lessons of the past. These studies provide evidence that adherence to well documented simulator practices within the task, simulator, and individual domains can reduce the prevalence and/or severity of SS in emerging flight simulation systems. Although the optimal solution to the SS problem lies in addressing and evaluating SS during a simulator's design and development stages, these recommendations can be used as interim solutions to reduce SS.

Acknowledgments

The authors wish to thank the staff and instructors at Warrior Hall for their cooperation during this study and SFC Daniel R. Fuller for his role in data collection. Also, we wish to thank Dr. Loraine St. Onge for her extensive and constructive feedback on early drafts of this paper. The view, opinions, and/or findings contained in this paper are those of the authors and should not be construed as an official Department of the Army position, policy, or decision. Note: SSG Bass is currently stationed at Fort Hood, TX. LTC Wildzunas is currently at the US Army School of Aviation Medicine. LTC Martin is currently at Walter Reed Army Medical Center.

References

Crowley J. S., & Gower, D. W. (1988). Simulator sickness: I'm ok, you're ok, it's the simulator that's different. *United States Army Aviation Digest, 1-88-11,* 9-11.

Johnson, D. M. (2005). *Introduction to and review of simulator sickness research.* Fort Rucker, AL: United States Army Research Institute- Rotary Wing Aviation Research Unit; ARI Report No. 1832.

Kennedy, R. S., Berbaum, K. S., & Lilienthal, M. G. (1997). Disorientation and postural ataxia following fight simulation. *Aviation, Space and Environmental Medicine, 68,* 13-17.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire (SSQ): an enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3,* 203-20.

Reason, J.T., & Brand, J.J. (Eds.). (1975). *Motion sickness.* New York: Academic Press.

Stanney, K. M., Kennedy, R. S., & Drexler, J. M. (1997). Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (pp.1138-1142). Santa Monica, CA: Human Factors and Ergonomics Society.

Wright, R. H. (1995). Helicopter simulator sickness: *A state-of-the-art review of its incidence, causes, and treatment.* Alexandria, VA: United States Army Research Institute for the Behavioral and Social Sciences; ARI Research Report 1680.

FRACTAL TIME SERIES ANALYSIS OF HUMAN HEARTBEAT INTERVALS IN A CHANGE BLINDNESS TASK

Sheldon M. Russell
Gregory J. Funke
Consortium Research Fellows Program
Alexandria, VA

Camilla C. Knott
Aptima, Inc
Dayton, OH

Benjamin A. Knott
Air Force Research Laboratory
Wright-Patterson AFB, OH

Fractal analysis is a set of techniques used in the study of complex systems to understand patterns of variability that are often observed in non-linear and natural systems. These techniques have been successfully applied to the study of human physiology, such as gait, heart rate variability, and respiration (West, 2004). Understanding variability in physiological measures may provide insight into human performance, individual workload, and stress. As an exploration of the possible contributions of fractal analysis to human performance, heart rate data from an existing research study was re-analyzed using methods common in fractal analysis of time series data. Results indicate that variability was measurably different based on experimental manipulations, illustrating the potential utility of fractal analytic methods in understanding human performance. Researchers are encouraged to explore these methods for their own research.

Although a relatively new addition to the sciences, the study of non-linear dynamical systems (formerly known as chaos theory) has yielded interesting findings in a variety of scientific disciplines (an easily accessible overview of this body of work can be found in Gleick, 1998). As the study of complex systems progressed, Mandelbrot (e.g., 1983) introduced fractal geometry, a new type of mathematics used to describe the unique nature of non-linear systems. One of the more prolific findings has been that many natural systems have data that is fractal in structure (Newman, 2005). Fractal data sets are those that obey a power law distribution and have a persistent pattern that is independent of the scale of observation (Mandelbrot, 1983).

The application of these new methods to non-linear systems has yielded a variety of interesting findings in the short time they have been studied (Newman, 2005). While the bulk of scientific contributions in the area have been in the earth and physical sciences, some relationships have emerged in the area of human physiology – examples include the variability observed in heart rate and walking gait (West & Griffin, 2004).

Many existing data sets may already lend themselves to reanalysis using fractal methods, allowing further exploration without significant design and data collection costs. The requirements of frequency analysis of time series data are relatively easy to meet, in that the time series must be from one measurement source (i.e., one subject) and data collection resolution must be high enough to produce a minimum of 300 data points for inclusion in the analysis (a comprehensive overview of time series analysis can be found in Malamud & Turcotte, 1999).

Interestingly, study of the fractal structure of the human heartbeat has yielded important diagnostic information related to heart health. Analysis of inter-beat interval (IBI) data using frequency analysis reveals that a frequency pattern known as "pink noise" (or "1/f noise") is typically observed in a healthy heartbeat (when fitted to a power law this ratio is expressed as an exponent of 1). Research indicates that differences in the frequency response patterns of an IBI time series can signal the development of a serious heart malfunction, such as cardiopulmonary arrest or myocardial infarction, in the near future (an overview of frequency analysis of IBI data can be found in McSharry & Malamud, 2005).

Measures of heart rate variability have previously been used as indices of cognitive workload and stress (see, e.g., O'Donnell & Eggemeier, 1986, for a review). Researchers utilizing these measures typically compare mean interbeat variability across conditions for evidence of differences attributable to experimental manipulations. Overall, decrements in heart rate variability have been observed in association with increases in mental workload (Kalsbeek, 1971; Kalsbeek & Ettema, 1963).

Since fractal analysis of IBI has been demonstrated to be diagnostic in hospital patients, it is possible that the fractal dynamics of a heartbeat time series may also provide a sensitive index of workload and stress, and yield new insights into human performance. Of particular interest is the possibility that fractal indicators of workload may be more sensitive than averaged measures when comparing across individuals due to the computational process required for analysis. This process essentially normalizes the data, thereby reducing individual differences often associated with physiological measures, which should result in increased sensitivity (Malamud & Turcotte, 1999).

Several goals were established for the present exploratory study. The first was to test if a significant fractal structure could be observed in IBI data from an existing data set, and to determine if differences in the fractal properties of that data could be attributed to experimental manipulations. The second goal was to compare the results of the fractal analysis to a standard measure of workload as an index of convergent validity. Demonstrating similar statistical results using each of these measures will help to establish the utility of fractal analysis for understanding workload.

Methods

Participants

Three men and seven women between the ages of 18 and 30 (M = 22.7 years, SE = 1.33) served as paid participants in this study. Participants completed the experiment in groups of two, yielding a total of five groups.

Experimental Design

Data for the current study were drawn from a change detection experiment by Knott et al. (2007). In their study, participants were required to monitor an array of geometric shapes for small shifts in spatial location. Participants completed the experiment individually and in a dyadic "team" condition.

Though the original featured a more complex experimental design, data included in the current study was drawn from a subset of variables manipulated by Knott et al. (2007). Specifically, data in the current study was derived exclusively from the team condition and was organized around two of the original factors: set size (6 or 48 icons) and communication condition (permitted, restricted). Both of these were within-groups variables. The set size factor determined the number of shapes participants were required to monitor. The communication factor specified if participants were allowed to communicate with their teammate during task performance.

Stimuli & Apparatus

A flicker task display was created using custom JAVA scripts. Each display consisted of an even number of randomly distributed red, blue, and green squares at 1° visual angle. A trial consisted of four images presented in the order and duration shown in Figure 1: display A (750 ms), mask (250 ms), either display A, or A' with one square changed in position (750 ms), mask (250 ms). This sequence repeated until the participant responded by key press to indicate whether or not a change had occurred. If a change was detected, then participants also indicated by mouse click which square changed in position. Pentium 4 desktop computers were used to present two identical displays to participant dyads on 18" NEC Multisync LCD 1800 flat screen monitors.

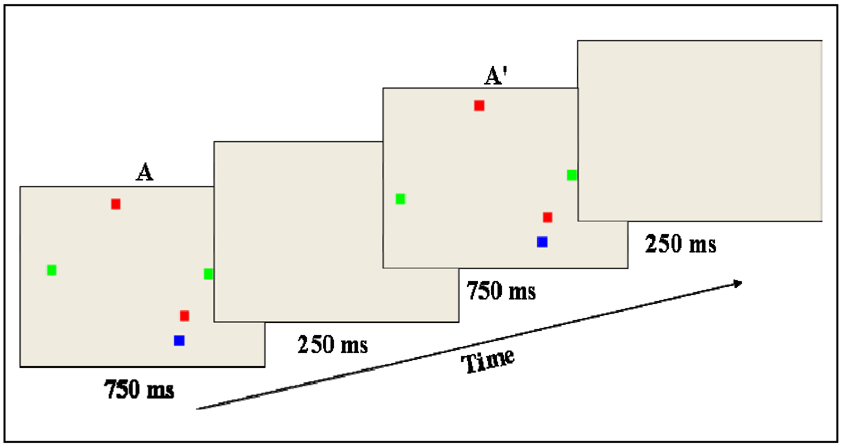


Figure 1. The figure denotes the flicker task with a change in the set size 6 condition. In this trial, the leftmost square moved closer to the screen edge on the third frame. This figure is from Knott, Nelson, McCroskey & Miller (in press) and is reprinted with permission from the authors.

Heart rate information was collected using UFI EZ-IBI software and equipment which recorded the interval between R-wave peaks at a rate of 1000Hz. This system required three surface electrodes to be attached to the participant to detect the electrocardiogram (ECG) signal. Electrodes were attached to the sternum, the left side of the body below the 5th rib (approximately 6 inches from the armpit), and to the right wrist (as a ground). The skin below the electrodes was prepared by applying NuPrep ECG skin abrasion gel to the site, gently wiping the skin with gauze, and then cleaning the site with alcohol prior to application of the electrodes. Heart rate data was recorded to Microsoft Access, and was exported into Microsoft Excel for subsequent analysis. A time series for one participant is displayed in Figure 2.

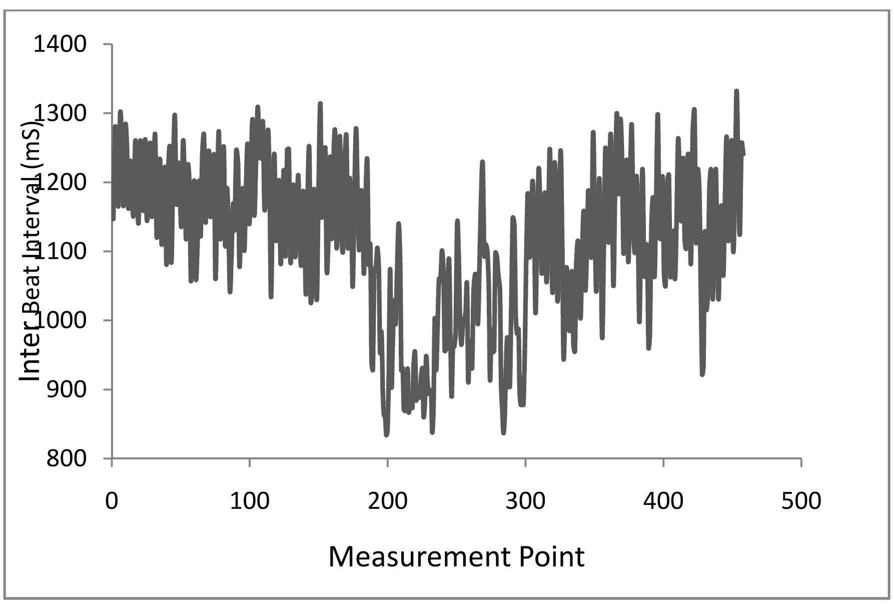


Figure 2. Inter-beat interval (IBI) time series for one participant in the communication-restricted condition, with a set size of 48 icons. IBI is measured in milliseconds. Values on the abscissa correspond to the serial count of measurements recorded during this block of trials.

Procedure

Teams completed 36 trials in each experimental condition, for a total of 144 trials in each experimental session. Both set size and communication condition were blocked factors; set size was blocked within communication condition. All teams completed trials in the communication-restricted condition followed by trials in the communication-permitted condition to ensure that participants could not carry over team-level strategies between conditions.

Following the conclusion of each trial block, participants completed the NASA Task Load Index (TLX; Hart & Staveland, 1988), a standard measure of cognitive workload which is widely used in human performance research (Wickens & Hollands, 2000). The TLX was employed in this experiment to provide a global index of mental workload associated with task performance, and to establish convergent validity with the fractal analytic technique employed in the current study.

Results

Fractal Analysis

Data from each block of trials was imported into Autosignal, version 1.6, and analyzed using the power spectral density method, described as follows. First, a Fourier transform for unevenly spaced data (i.e., the Lomb Periodogram method; McSherry & Malamud, 2005) was performed. Output from this analysis was then imported into Microsoft Excel, and trimmed such that the largest data point included in each data set was equal to or less than half the length of that data set.

The trimmed data was then plotted on log x/log y axes and two best fitting power lines were plotted with their equations displayed. The exponent value of the line for the frequency values from -.3 to -.4 Hz was recorded for subsequent analysis. This frequency range was selected because a visible "break" in the slope of the data which occurred at -.4Hz, indicating that frequency values less than -.4Hz were white noise and did not have a significant fractal structure (this pattern of frequency structure has also been observed by Voss, Schulz, Schroeder, Baumert, Caminal, 2009). The output from a power spectral density analysis for one participant is depicted in Figure 2.

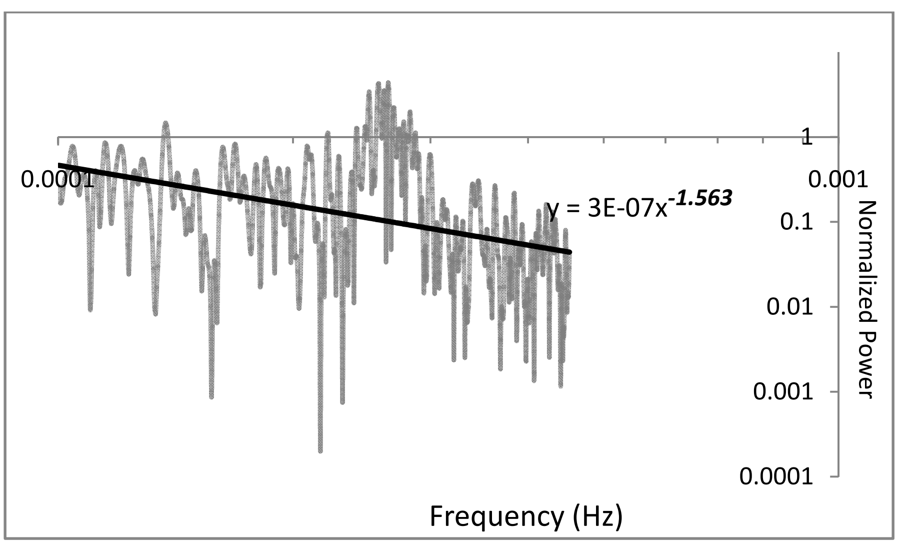


Figure 3. Power spectral density plot of the time series data of Figure 2, fitted to log-x log-y axes. A power function has been fitted to the data and the exponent is displayed for the equation.

Statistical analysis

The observed fractal exponents ranged from -.935 (the most variable) to -3.289 (the most ordered). Exponents for each participant were compiled and the resulting data were analyzed by means of a 2 (set size) × 2 (communication condition) repeated measures analysis of variance (ANOVA). The goal of this analysis was to determine if differences in fractal exponents could be attributed to the experimental factors. A statistically significant difference between conditions in this analysis would indicate that experimental manipulations were associated with a shift in heart rate variability (i.e., from less to more ordered).

Results of the analysis indicated a statistically significant main effect for communication condition, $F(1, 9) = 10.47$, $p < .05$. No other sources of variance in the analysis were significant ($p > .05$). Participants' IBI was significantly more structured in the communication-restricted condition (mean exponent = -2.26, *SE* = .19) compared to the communication-permitted condition (mean exponent = -1.79, *SE* = .14).

Mental workload

To test the effects of the experimental conditions on participants' evaluation of task workload, participants' mean global TLX workload ratings in each condition were calculated. These ratings were then tested for statistically significant differences between conditions by means of a 2 (set size) × 2 (communication condition) repeated measures analysis of variance (ANOVA). The results of this analysis indicated statistically significant main effects for communication condition, $F(1, 9) = 7.86$, $p < .05$, and for set size, $F(1, 9) = 12.95$, $p < .05$, but their interaction was non-significant ($p > .05$). Participants rated their workload as higher in the communication restricted condition ($M = 37.87$, $SE = 4.88$) compared to the communication permitted condition ($M = 28.75$, $SE = 5.08$). In addition, participants rated their workload as higher in the set size 48 condition (M = 37.87, SE = 4.85) compared to the set size 6 condition ($M = 28.75$, $SE = 4.90$).

Discussion

The results suggest that in an experimental setting, differences in fractal properties can be observed in heart rate variability that correspond to experimental manipulations when using frequency analysis techniques. Furthermore, fractal exponents can be observed in existing data sets that fit the necessary requirements for the analytic techniques. In the current analysis there was a statistically significant difference present between subjects who were in communication during the block of trials and those that had restricted communication. It is noteworthy that this result does not match the analysis of heart rate data conducted by Knott et al (in press). The results of the previous analysis suggested a statistically significant increase in heart rate between set size (heart rate was faster when the set size was 48 icons when compared to 6).

In interpreting the significance of the current findings, one must be even more cautious than normal. Although existing data sets can meet the necessary criterion for time series analyses, problems may arise. In this case, a necessary implementation with the order of conditions used by Knott et al (2007) created a confound for the current analysis. Specifically, communication restricted conditions were always collected before communication allowed conditions. This was important to prevent carryover effects of strategy, but it is difficult to distinguish if the observed fractal changes are due to the effect of communication condition, are a measure of fatigue, or some other effect of order. Another perplexing finding is that subjective workload was reported as higher when communication was restricted. Although order confounds this finding as well, it suggests that the increase in the fractal exponent is not related to an increase in subjective workload, and in fact the reverse may be true, although future research should address this relationship.

Regardless of the origin of the changes, the present analysis suggests that power spectral density analyses of fractal structure may be sensitive to experimental manipulations in a laboratory setting. Given their abundance in natural systems, it is likely that increased study of fractals will bring new findings for human performance research, particularly in physiological measures of workload and fatigue. Further research in this area should focus on designing experiments with fractal analysis techniques in mind, as well as further exploring the corresponding changes in fractal structure of heart beat intervals based on experimental manipulations of workload and fatigue.

Acknowledgements

The authors would like to thank the authors of Knott et al (in press) for providing generous access to their original data. The authors would also like to acknowledge the valuable input of Chris Barton and the other members of the Complexity Research Group at Wright State University.

References

Gleick, J. (1999). *Chaos: Making a new science.* New York, NY: Penguin.

Hart, S.G., & Staveland, L.E. (1988). Development of a multi-dimensional workload scale: Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North-Holland.

Kalsbeek, J.W.H. (1971). Sinus arrhythmia and the dual task method in measuring mental load. In W.T. Singleton, J.G. Fox, & D. Whitfield (Eds.), *Measurement of man at work* (pp. 101-113). London: Taylor and Francis.

Kalsbeek, J.W.H., & Ettema, J.H. (1963). Scored irregularity of the heart pattern and the measurement of perceptual or mental load. *Ergonomics, 6*, 306-307.

Knott, C.C., Nelson, W.T., McCroskey, M.K., & Miller, B.T. (2007). Do you see what I see? Effects of communication on scanning strategies in change detection by individuals and teams of observers. *Proceedings of the Human Factors and Ergonomics Society, 51*, 1186-1190.

Knott, C.C., Nelson, W.T., McCroskey, M.K., & Miller, B.T (in press). *Analysis of Eye Movements in Change Detection with Teams Using a Simulated Tactical Situation Display* (Tech Report No: Pending). Wright-Patterson AFB, OH: Air Force Research Laboratory.

Malamud, B.D. & Turcotte D.L. (1999). Self-affine time series: I. Generation and analyses. *Advances in Geophysics, 40,* 1-90.

Mandelbrot, B.B. (1983). *The Fractal Geometry of Nature*. New York, NY: W.H. Freeman and Company.

McSharry, P.E. & Malamud, B.D. (2005). Quantifying self-similarity in cardiac inter-beat interval time series. *Computers in Cardiology, 32,* 459-462.

Newman, M.E.J. (2005). Power laws: Pareto distributions and Zipf's law. *Comtemporary Physics, 46*, 323-351.

O'Donnell, R., & Eggemeier, F.T. (1986). Workload assessment methodology. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance*, (pp. 42-1 to 42-49). New York, NY: Wiley.

Voss, A., Schulz, S., Schroeder, R., Baumert, M., & Caminal, P. (2009). Methods derived from nonlinear dynamics for analyzing heart rate variability. *Philosophical Transactions of the Royal Society, 367,* 277-296.

West, B.J., & Griffin L.A. (2004). *Biodynamics: Why the wirewalker doesn't fall.* Hoboken, NJ: Wiley.

Wickens, C., & Hollands, J. (2000). *Engineering Psychology and Human Performance*. Upper Saddle River, NJ: Prentice-Hall Inc.

SITUATIONAL AWARENESS ASSESSMENT IN FLIGHT SIMULATOR EXPERIMENT

Henk van Dijk, Koen van de Merwe and Rolf Zon
National Aerospace Laboratory NLR
Amsterdam, the Netherlands

Within the HILAS (Human Integration into the Lifecycle of Aviation Systems) project a flight simulator experiment was performed. The aim of the experiment was to study and select relevant Human Factors tools for pilot Situational Awareness assessment. One specific scenario was designed in which a malfunction of the aircraft was simulated: an Indicated Air Speed discrepancy. The malfunction was introduced during flight and slowly progressed over time while researchers monitored if and how pilots detected the discrepancy. Pilot behaviour was studied during the scenario; i.e. pilots' Situational Awareness was assessed via eye trackers and rating scales.

HILAS[1] stands for Human Integration into the Lifecycle of Aviation Systems. The objective of the Flight Deck Technologies (FDT) strand within the HILAS project is to create a Human Factors (HF) related set of tools that can be used to design and evaluate flight deck technologies. Such an HF set of tools can be applied as an instrument for HF certification. The importance of HF certification and its added value is explained by Jorna (2007) and McDonald (2007). The flight simulator experiment discussed in the current paper forms a part of the FDT strand.

Flight Simulator Experiment

The flight simulator experiment performed in the HILAS FDT strand comprises two phases. The results of the phase 1 experiment were previously discussed by Zon & Roerdink (2007). The current paper discusses part of the phase 2 experiment. The specific aim of the experiment discussed in the current paper was to study and select relevant HF tools for pilot Situational Awareness (SA) assessment.

Participating pilots

Six crews of each two airline pilots (i.e. a captain and a first officer) participated in the experiment. There are two pilot tasks in the simulation: pilot flying (PF) and pilot not flying (PNF). The tasks to be carried out by PF and PNF match normal operations and are varied between captain and first officer.

Flight simulator

GRACE (Generic Research Aircraft Cockpit Environment) is a generic flight simulator, representing a modern large two-engine fly-by-wire airliner (see GRACE cockpit in Figure 1). GRACE has a number of standard configurations. For the current experiment the Airbus A320 configuration was selected. A high fidelity simulator such as GRACE allows researchers to perform realistic experiments in a fully controlled environment.

[1] The Human Integration into the Lifecycle of Aviation Systems (HILAS) project is part of the 6th framework programme for aeronautics and space research, sponsored by the European Commission. The overall aim of the HILAS project is to develop a model of good practice for the integration of human factors across the life-cycle of aviation systems. The project contains four strands of work (see for further information http://www.hilas.info/mambo/).

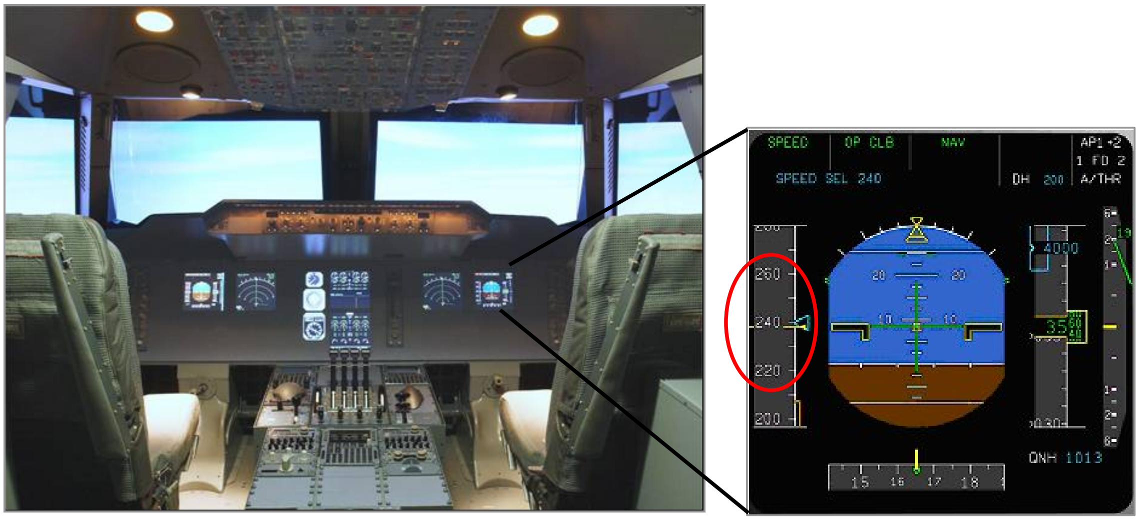


Figure 1. GRACE cockpit on the left side and PFD on the right side. The speed tape on the PFD is circled in red.

Experimental scenarios

The flight consisted of a trip from London Heathrow to Amsterdam Schiphol (starting in cruise, ending in landing). A specific scenario was designed in which SA was hampered. This was done by simulating a malfunction of the aircraft during the flight: an Indicated Air Speed (IAS) discrepancy was introduced. The discrepancy was indicated by the two available Primary Flight Displays (PFDs); i.e. one display showed the correct air speed while the other showed a lower, false air speed (see Figure 1 for a PFD with the speed tape circled in red). Once the discrepancy was initiated by the simulator after about 10 minutes in flight, it slowly progressed over time while researchers monitored if and how the pilots detected the discrepancy, and if and how the pilots figured out what the correct air speed was. As far as the crew concerns, they were flying a normal flight until the malfunction was detected. The flight duration of this scenario was 25 minutes.

It was expected –as the discrepancy progressed over time– that most of the crews detected the malfunction earlier than after the 2 minutes on which an engine display warming was given. As it turned out, none of the six crews discovered the specific discrepancy on both PFDs before the engine display indicated the air speed discrepancy. The crews were informed of the malfunction with an auditory warning after 2 minutes. Consequently, the analysis focussed only on the time that the crew needed to figure out the correct air speed (i.e. after the engine display warning), and not on the time they needed to detect the malfunction. The period after the warning until the moment of detection of the malfunction is referred to as post-period. In the analysis, this post-period is compared to a reference period that has the same length as the post-period and takes place immediately before the onset of the IAS discrepancy (see Figure 2 for an illustration of this time-line).

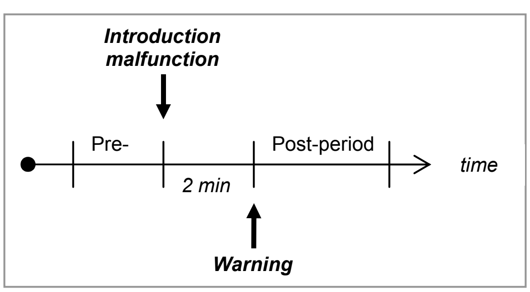


Figure 2. Illustration of the time-line for the analysis.

HF tools

During this SA scenario, pilot behaviour was studied using the following HF tools:

Eye tracking

The eye tracker that was used in the current experiment is the Applied Science Laboratories (ASL[2]) 6000 with Ascension Technologies optical head tracker. Two trackers were placed in the cockpit (one for each pilot). The pilots wore a headband on which the optronics were mounted.

Dwell time on the PFDs (crewmember's own PFD and cross check) and entropy were used as eye tracking measures for the PF and PNF. Dwell time provides information regarding the amount of time spent viewing the PFD and can be interpreted as a measure of attention. Entropy provides information regarding the search strategies used by the crewmembers and can be interpreted as a measure of randomness of the viewing pattern.

ISA

The simple rating technique called Instantaneous Self Assessment (ISA) was used in the current scenario to measure the pilots' overview of the situation on that particular moment in the scenario; i.e. the ratings were self assessed during the flight. The pilot was asked (every other 2 minutes) to respond to the rating scale presented on the touch screen display (an electronic flight bag) placed in front of him/her by assessing his/her current situation overview (5 being very high and 1 being very low; see ISA rating scale in Figure 3).

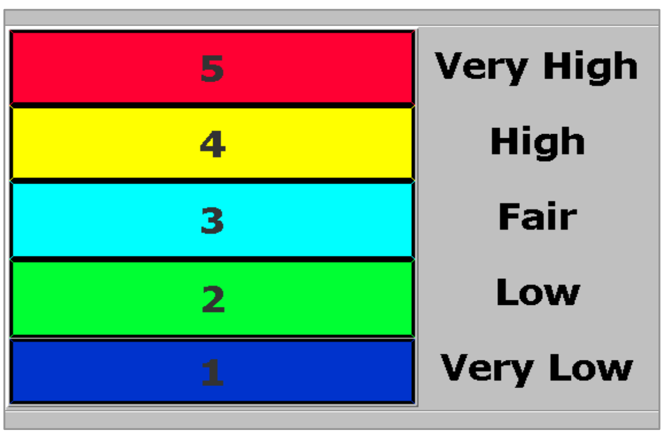


Figure 3. ISA rating scale displayed on an electronic flight bag.

Results

An α of 5% was used for significance testing and Cohen's *d* was used as a measure of effect size.

Eye tracking

The difference in viewing behaviour between the period in which the crew was looking for the error (post-period) and the reference period before the error was introduced (pre-period) was analysed. A paired-samples t-test showed significantly longer time spent dwelling on the crewmember's own PFD in the post-period compared to the pre-period ($t(9) = -2.326$, $p < .05$, $d = -.74$). Similar results were found for cross check behaviour ($t(9) = -4.005$, $p < .01$, $d = -1.27$) and all PFDs ($t(9) = -2.916$, $p < .05$, $d = -.92$).

[2] See for further information http://www.a-s-l.com/.

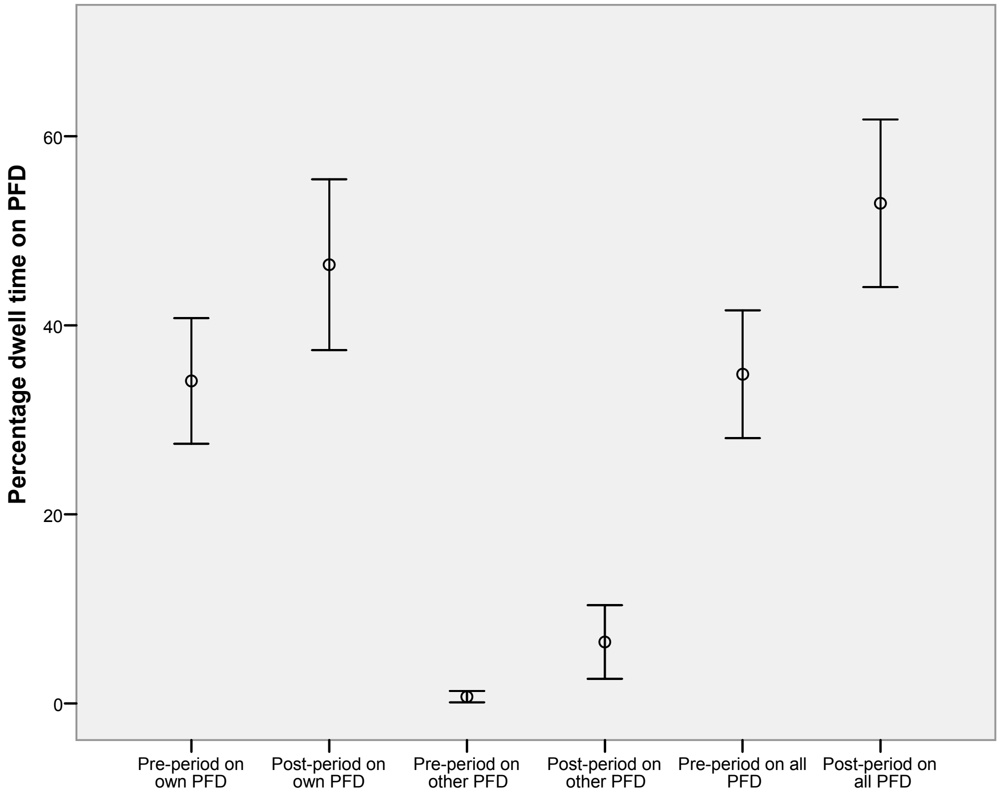


Figure 4. Percentage dwell time on the crewmember's own, other and all PFDs in the pre-period vs. the post-period. All differences between the pre- and the post-period were statistically significant.

The relationship between the time it took for the crew to discover the IAS discrepancy on the PFDs and the amount of time spent looking at the different PFDs was also analysed. It was assumed that the malfunction could only be discovered by cross checking both PFDs and comparing the information presented on them. A significant negative correlation was found between the discovery period and the amount of time spent cross checking the other PFD ($r(12) = -.613, p < .05$). This means that the more time the crewmember spent on cross checking the other PFD, the less time it took to discover the IAS discrepancy. There was no significant correlation found for the time spent on one's own PFD.

The scanning patterns of the crews were investigated using a t-test. The results showed a significant increase in entropy during the post-period ($t(10) = -2.347$, $p < .05$, $d = -.71$). This means that the patterns followed by the eye movements during the post-period were less systematic than during the pre-period indicated searching behaviour.

ISA

The average of multiple ISA answers per crew were investigated using a t-test ($n = 28$). ISA ratings were significantly lower in the post-period than in the pre-period ($t(27) = 2.780$, $p < .05$. $d = .52$). This means that the crew reported that they experienced a reduced SA in the post-period.

Key Findings

The current paper discussed the phase 2 high fidelity simulator experiment within the HILAS FDT strand. The experiment studied and selected relevant HF tools for pilot SA assessment. A specific SA scenario was designed in which a simulated malfunction was introduced during the flight; i.e. after a certain amount of time in flight, an IAS discrepancy was set on and slowly progressed over time.

Eye measures comprise a number of variables that can be measured and recorded with an eye tracker. The measures can be used as indicators of fatigue, mental or visual workload and also as indicators of attention. The eye tracking measures in the current experiment revealed the eye scanning patterns; i.e. where did pilots focus on during the period in which they tried to discover the correct air speed and how quickly did they make the discovery. As such, and supplemented by the ISA ratings, this informed the researchers about the perceived SA of the pilots during the course of the flight.

Eye tracking

Eye tracking can be considered as an indirect measurement of pilot attention and focus on tasks. The basic assumption is that looking at a certain location means that attention is focussed on this particular location. The analyses of the eye tracker data primarily focussed on the two PFD locations since these locations were the designated displays on which the IAS discrepancy was revealed. Of course, this is a simplification of the reality as there are other methods in finding out the correct air speed (e.g. checking with air traffic controller, looking at ground speed on navigation display).

The dwell time results indicate an increase of dwell time after the engine display warning on the pilot's own PFD and the other's PFD (cross check). This was as expected since the warning of an air speed discrepancy on the engine display encourages in looking at the different PFDs (as air speed is indicated here). Interestingly, the dwell time on the other's PFD correlates negatively with the time it takes to discover the correct air speed (the duration of the post-period decreases when the time used for the cross check increases). This result corresponds with the notion that the cross check is evident in discovering the correct air speed.

The entropy results reveal a higher entropy value in the post-period compared to the pre-period. This was as hypothesized. Random screening after the engine display warning increases because the pilots are searching for the solution.

ISA

Self rating techniques such as ISA are frequently used to elicit subjective estimates of SA from participants. These rating scales are typically administered post-flight. The primary advantage of such techniques is their low cost, ease of implementation and non-intrusive nature. However, self rating techniques administered post-flight suffer from a number of disadvantages that are associated with reporting SA data "after the fact". These include that participants are prone to "forgetting" periods of the flight when they possessed a poor level of SA, and more readily remember the periods when they possessed a superior level of SA (Endsley, 1995). Therefore, in the HILAS phase 2 experiment, the ISA technique was implemented in such a manner that ratings could be assessed during the course of the flight; i.e. pilots could rate their SA on a particular given moment in the scenario. This helps to pinpoint the ups and downs in the SA of the specific pilot. As it turns out, assessing ISA in flight using an electronic flight bag is an easy-to-implement and easy-to-use technique in an experiment. The crew did not report any problems (e.g. distraction of the flying task or intrusiveness of the tool) and rated their SA immediately after the scale popped up on the electronic flight bag.

The ISA results (for PF and PNF) show a decrease in SA after the engine display warning. This was as expected since this particular SA scenario is set up to hamper SA.

Summarizing, crew behaviour –specifically the eye scanning patterns– has been proven to change due to the compromised SA. This is clearly indicated by the results of the eye tracking measures dwell time and entropy. The ISA results add to this the perceived SA of the crew during the course of the flight. Together, eye tracking and ISA provide a more complete picture than both measures independently.

Acknowledgements

The authors would like to thank the HILAS FDT strand members for their contribution in the experiment. Further, we would like to thank the European Commission for sponsoring this research.

References

Jorna, P. G. A. M. (2007). Human performance enhancements: from certification to HCI innovation. In: D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics, HCII 2007, LNAI 4562* (pp. 697-704). Berlin Heidelberg: Springer-Verlag.

McDonald, N. (2007). Human integration in the lifecycle of aviation systems. In: D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics, HCII 2007, LNAI 4562* (pp. 760-769). Berlin Heidelberg: Springer-Verlag.

Zon, G. D. R., & Roerdink, M. I. (2007). HCI testing in flight simulator: set up and crew briefing procedures. Design and test cycles for the future. In: D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics, HCII 2007, LNAI 4562* (pp. 867-876). Berlin Heidelberg: Springer-Verlag.

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

PROVIDING EVIDENCE OF A MULTIPLE-PROCESS MODEL OF TRUST IN AUTOMATION

Stephen Rice
Gayle Hunt
New Mexico State University
Las Cruces, New Mexico

This study focuses on the effects of human responses to computer automation aids. Previous research has shown that different types of automation errors (false alarms and misses) affect human trust in different ways. False alarms tend to negatively affect operator compliance, whereas misses tend to negatively affect operator reliance. Participants were asked to determine whether an enemy target was present or absent in a series of images, a task similar to what a UAV operator might be asked to perform. A diagnostic aid provided recommendations before participants viewed each image. Reliability and type of automation error were manipulated in order to provide data to determine which of four theoretical models is most accurate. Analyses provided conclusive evidence that a multiple-process theory of operator trust is the only model which accurately explains behavior outcomes in this type of situation. A discussion of theoretical and practical implications of this finding is included.

The use of automation has accelerated so rapidly that it has outpaced the formation of a comprehensive theoretical understanding of human interaction with automated systems. Specifically, it is alarming to take notice of the lack of theory explaining how errors in automation affect human trust in this technology, which has an impact on the level of human dependence on automated systems. Although it is possible to find theoretical literature on this topic (e.g. Meyer, 2001; 2004; Parasuraman & Riley, 1997), it is nonetheless limited and must be studied more extensively in order to achieve a full understanding of the human cognitive process during human-automation interaction. Both trust (a cognitive state) and dependence (a behavior) are key factors in these studies, although they are not always correlated.

Many researchers have studied the cognitive and perceptual benefits of the use of automation on multi-tasking. In particular, it is presumed that when automation can be held responsible for completing a task, this should free up cognitive resources for the operator to focus on a different task at hand. The number of simultaneous tasks performed should equal the number of automated tasks plus the task being performed by the operator (e.g. Dixon, Wickens, & Chang, 2005).

Recently, there has been some attention on research focused on transferring cognitive resources away from an automated system and toward another task. An example of this research might be visual search and supervisory control. Many times, these two tasks are performed together (e.g. Dixon & Wickens, 2006). A real world example of this would be when a UAV operator must monitor all controls while simultaneously searching images for enemy targets (e.g. Maltz & Shinar, 2003).

While presently there are multiple people employed to operate a single UAV, there is an eventual goal to assign only one operator to each UAV, making it essential to understand the cognitive process of performing multiple tasks as well as working with, trusting, and depending on automation to lighten the load. It is also crucial to study automated systems that are not perfectly reliable and discover their implications on operator trust and performance of human-automation teams.

It is thought that automation takes place over four stages that are somewhat related to stages of human information processing (Parasuraman, Sheridan & Wickens, 2000). The first stage is information synthesis, which occurs by directing focus to particularly important environmental factors. The second stage is diagnosis, when automation provides an assumption about the information that has been taken in. Some examples may include warning alarms, which serve to focus the operator's attention on important events by utilizing auditory and/or visual warnings. The third stage is selection of response, and the fourth stage is execution of that selection. This paper will focus primarily on automation diagnosis, the second stage.

When stage 2 diagnosis is being utilized, it is possible that the operator may not have access to the information that is being processed. Instead the operator may only have access to the information provided by the automation. In the situation where an operator has raw data to confirm or disconfirm automation warnings, the performance outcome may be very different than when an operator has no raw data and must choose whether or not to follow the automation blindly (Sorkin & Woods, 1985). Further, human dependence on automation systems may be much more fickle when there is no raw data to allow the operator to confirm the diagnosis, which may cause a grave failure. This study will focus only on diagnostic automation with raw data readily available to the operator.

Although reliable automation warnings can greatly increase positive performance (e.g. Dixon, Wickens & Chang, 2005; Dixon & Wickens, 2006; Dixon, Wickens & McCarley, 2007), diagnostic automation is seldom perfect. Most often, the information being processed by the automation is imperfect, or the automation must make an assumption about an event that has yet to occur. Because diagnostic automation is often flawed, it is critical to understand how this factor affects the operator's trust and dependence on the system.

Signal detection theory is a useful tool in the understanding of automation errors. According to this theory, there are four possible outcomes of a diagnostic automation: hits (correct warning), misses (incorrect non-warning), false alarms (incorrect warning), or correct rejections (correct non-warning). Among these four possibilities, two of them (misses and false alarms) are automation errors. It may seem like a simple assumption that these errors are equally harmful and produce consequences of equal severity, but research has indicated differently. Data supports the idea that false alarms may be more harmful than misses (e.g. Bliss, 2003), and that the two error types actually cause very different consequences, especially in relation to operator trust (e.g. Dixon & Wickens, 2006; Maltz & Shinar, 2003; Meyer, 2001; 2004; Wickens & Dixon, 2007).

One method of discriminating between false alarms and misses has been offered by Meyer (2001; 2004). According to Meyer, false alarms have a negative effect on compliance, whereas misses have a negative effect on reliance. Compliance refers to how the operator responds to a warning, and reliance is how the operator responds to no warning. It can be inferred that this is an indication of each type of error affecting single but separate cognitive processes, as seen in Figure 1b. While this may seem reasonable, there has been much data indicating that false alarms affect *both* operator reliance and compliance (Dixon & Wickens, 2006; Dixon, Wickens & McCarley, 2007; Wickens, Dixon, Goh, & Hammer, 2005).

Some of the strongest evidence to date suggests that false alarms have just as much of a negative effect on reliance as do misses (Dixon, Wickens, & McCarley, 2007). This data is conceptualized in Figure 1c.

Finally, further evidence supports yet another model. Rice and McCarley (2008) found that not only do false alarms have an effect on both compliance and reliance, but the same is the case for misses. It was found that misses have a negative effect on both areas as well. This evidence may be conceptualized in two ways, as demonstrated in Figures 1a and 1d.

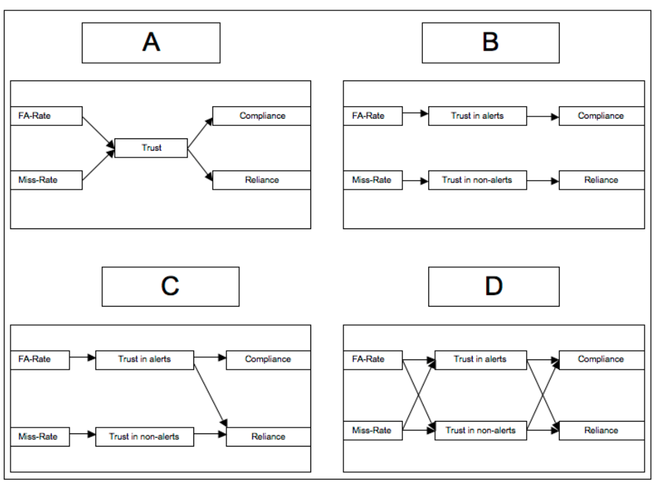


Figure 1. A) Single-process model; B) A selective two-process model; C) Mandler's two-process model; D) non-selective two-process model. Adapted from Dunn & Kirsner (1988).

With so many different models attempting to explain how errors in automation affect the behaviors of operators, there must be a way to distinguish between them and determine which best fits the data. One solution to this problem is to perform a state trace analysis on the data (Bamber, 1979; Dunn & Kirsner, 1988).

A state trace analysis plots two dependent variables against each other. The goal of this comparison is to observe if a monotonic relationship exists within the data. If a monotonic relationship is revealed, any change to one dependent variable will have a similar effect on the other. Should this be the case, a single process model, such as that in Figure 1b, would be supported. However, if the data demonstrates a non-monotonic relationship, (or a reversed association; Dunn & Kirsner, 1988), this would support a multiple process model, which indicates that at least two cognitive processes are in play (Bamber, 1979).

In order to simulate a task common to UAV operators, participants were asked to perform a target detection task by searching a set of aerial images of Baghdad for the presence or absence of an enemy tank. Assistance in this task was provided to participants in the form of a diagnostic aid which ranged in reliability from 95 to 55 percent in 5% increments. The automation was set to produce either false alarms or misses. This yielded 18 conditions total; 9 for false alarms and 9 for misses. Participants were fully informed of the type of errors the automation would make, as well as the reliability of the automation, and they were given clear feedback following each response.

There were four hypotheses related to this study: 1) better overall performance would occur in conditions with higher automation reliability as compared with lower automation reliability; 2) false alarm conditions would have strong selective effects on compliance and weaker non-selective effects on reliance, which would affect both how often the participant agrees with the automation and how quickly this decision is made; 3) miss conditions would have strong selective effects on reliance and weaker non-selective effects on compliance, again affecting both the rates of agreement and the response time; and 4) the use of a state trace analysis would uncover a non-monotonic relationship between compliance and reliance measures, supporting a multiple process theory to explain the outcome.

Method

Participants included 380 undergraduate students from New Mexico State University (230 females, 150 males), with a mean age of 20.3, who were compensated with partial course credit for their participation. All participants were screened for normal or corrected-to-normal eye sight as well as for color vision.

Images were presented on a 1024 x 768 resolution 20 inch flat-screen Dell monitor and computer with a refresh rate of 60 Hz. The monitor was roughly 40 degrees by 40 degrees in visual angle. One hundred images were presented to each participant, half of which consisted of target-absent stimuli and half of target-present stimuli. The target-absent stimuli were made up of 50 unchanged aerial images of Baghdad. The target-present stimuli were made up of the same 50 target-absent images, but with a tank digitally inserted into the image. The visual angle of the tank was roughly 2 degrees by 2 degrees, displayed with the turret facing one of 8 randomly assigned directions, N, NE, E, SE, S, SW, W, or NW. The presentation of images was randomized.

Participants were seated in a chair with their heads positioned by a chin rest 20 inches from the computer monitor. Each participant signed a consent form and proceeded to read directions presented on the screen. Directions included a sample picture of the target tank as well as full information regarding the type of error and reliability of the automation. Participants were asked to proceed quickly and with as much accuracy as possible. When participants felt confident with their instructions, they began the experiment.

As per the instructions, participants were aided by an aid which presented a recommendation before each image appeared on the screen. Twenty participants were randomly assigned to each of the 18 conditions.

Each trial was preceded by a small fixation, displayed for 1000 ms. Next was a screen providing the automation recommendation for the upcoming image, displayed for 1500 ms. The screen was then replaced with a randomly selected image from the previously mentioned set of 100 images. The image remained on the screen until the participant made a decision indicated by pressing the F or J key about the absence or presence of a tank (respectively). Following this, participants were presented with a feedback screen, displaying their accuracy and response time for that decision, as well as their cumulative accuracy for all the images they were previously presented with.

Results

General analyses (*d'*, C, and RT) are offered first, subsequently followed by additional analyses regarding compliance and reliance concerns. Finally, a state trace analysis was performed so as to test the theoretical models represented in Figure 1.

Sensitivity refers to a participant's ability to differentiate target-present from target-absent images, quantified by using the signal detection measure of sensitivity, *d'*. A two-way ANOVA on the imperfectly reliable conditions with Automation Error Type and Reliability as factors indicated that performance increased as the reliability of the automation increased, $F(8, 342) = 14.65$, $p < .001$, but showed no reliable effect of Automation Error Type, $F(1, 342) = 1.23$, $p > .05$, nor a reliable interaction of Automation Error Type and Reliability, $F < 1.0$, $p > .05$. These findings denote that although increased reliability rates did improve overall accuracy, automation error type did not affect accuracy.

Participants' response bias was calculated using the signal detection measure C. A two-way ANOVA on the imperfectly reliable conditions with Automation Error Type and Reliability indicated that participants in the False

Alarm conditions (*M* = -0.06), had a more liberal response bias than they did in the Miss conditions (*M* = 0.23), $F(1, 342) = 58.12$, $p < .001$; that is, participants were more apt to indicate that the target was present, regardless of their true performance sensitivity. There was no significant main effect of Reliability on participants' response bias, although it was marginally significant, $F(8, 342) = 1.87$, $p = .063$. An interaction between Automation Error Type and Reliability, $F(8, 342) = 2.77$, $p < .01$, indicated that the False Alarm condition was more likely to generate a liberal response bias among participants as the automation reliability increased.

Agreement rates and RTs were measured with the assumption that when participants trusted the automation, they would respond quickly in agreement. Only RTs from correct trials were integrated into the data analysis.

Compliance rate refers to the frequency in which participants agreed with the automation when it reported that a target was present. A two-way ANOVA performed on the imperfectly reliable conditions, with Automation Error Type and Reliability as factors, revealed a main effect of Automation Error Type, $F(1, 342) = 38.81$, $p < .001$, and a main effect of Reliability, $F(8, 342) = 4.87$, $p < .001$, with a significant interaction, $F(2, 66) = 2.15$, $p < .05$. These results indicate that participants in the False Alarm conditions were less likely than those in the Miss conditions to agree with the automation when it reported that a target was present, particularly when the automation was less reliable.

Reliance rate refers to the frequency in which participants agreed with the automation when it reported that the target was not present. A two-way ANOVA performed on the imperfectly reliable conditions, with Automation Error Type and Reliability as factors, revealed a main effect of Automation Error Type, $F(1, 342) = 25.64$, $p < .001$, but no main effect of Reliability, $F(8, 342) = 1.87$, $p > .05$, and no interaction, $F(2, 66) = 1.14$, $p > .05$. The significant main effect of Error Type indicates that participants in the Miss conditions were less likely than those in the False Alarm conditions to agree with the automation when it reported that a target was not present.

Compliance response time refers to the speed in which participants agreed with the automation when it reported that a target was present. A two-way ANOVA performed on the imperfectly reliable conditions, with Automation Error Type and Reliability as factors, revealed a main effect of Automation Error Type, $F(1, 342) = 48.15$, $p < .001$, no main effect of Reliability, $F(8, 342) = 1.37$, $p > .05$, and no significant interaction between Automation Error Type and Reliability, $F(2, 66) = 1.48$, $p > .05$. These results indicate that participants in the False Alarm conditions were slower to agree with the automation when it reported that a target was present, as compared to those in the Miss conditions.

Reliance response time refers to the speed in which participants agreed with the automation when it reported that a target was not present. A two-way ANOVA performed on the imperfectly reliable conditions, with Automation Error Type and Reliability as factors, revealed a main effect of Automation Error Type, $F(1, 342) = 33.70$, $p < .001$, with no main effect of Reliability, $F(8, 342) = 1.57$, $p > .05$, and no significant interaction between Automation Error Type and Reliability, $F(2, 66) < 1.0$, $p > .05$. These results indicate that participants in the Miss conditions were slower to agree with the automation when it determined that a target was not present, as compared to those in the False Alarm conditions.

The data above expose a pattern of false alarm rates affecting participant compliance more so than reliance, whereas miss rates affected participant reliance more so than compliance. However, in regards to the theoretical models discussed in the Introduction, this behavioral data cannot conclusively determine which model is correct. In order to test these issues, state trace analyses were performed on agreement rates and RTs, as seen in Figure 2a and 2b.

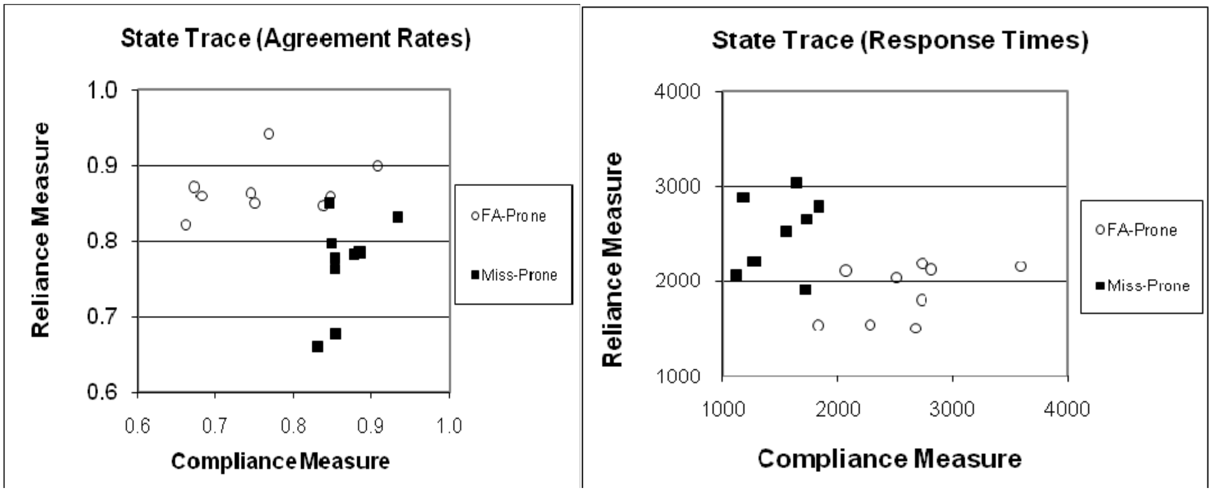


Figure 2. State Trace Analyses on a) Agreement Rates (%); and b) RTs (sec).

This analysis uncovers a non-monotonic relationship between the dependent variables. Spearman rank order correlations on the data revealed an $r = -0.28$ for Compliance agreement rates against Reliance agreement rates, and $r = -0.35$ for Compliance RT against Reliance RT. This information conclusively supports the notion that false alarm-prone and miss-prone automation affect at least two different cognitive processes, which translates to a different effect on operator behavior.

Discussion

All four theoretical models described in the introduction assume that there are only two types of automation errors (miss and false alarm) and two types of responses to these errors (compliance and reliance).

Recall the argument by Meyer (2001; 2004) presented in the introduction (Figure 1b). The data in this current study conclusively disconfirm the extreme version of this theory (Note: we agree that Meyer never actually advocated an extreme version). Analyses clearly show that false alarm rates affect both compliance and reliance. Furthermore, the theoretical model in Figure 1c has also been disconfirmed, as miss rates also affected both compliance and reliance.

With the models in Figures 1b and 1c no longer viable, we are left with the task of determining the correct model between Figures 1a and 1d. The only way to distinguish between these two models is with the use of a state trace analysis. As explained in the introduction, a monotonic relationship provides support for a single-process model, while a non-monotonic relationship proves a multiple-process model.

A very clear non-monotonic relationship emerged from the data. In short, an increase in value for one did not result in an equal increase for the other. Clearly, there are at least two separate cognitive processes involved in responding to automation false alarms and automation misses. This finding now disconfirms the theoretical model displayed in Figure 1a and clearly confirms the model in Figure 1d.

As predicted, higher automation reliability yielded better human-automation performance overall, when compared to less reliable automation. This effect is consistent with previous research (Dixon & Wickens, 2006; Dixon, Wickens & McCarley, 2007). Thus, highly reliable automation should be used whenever possible, as it may be argued that reduced reliability rates could possibly do more harm than good (Wickens & Dixon, 2007).

Designers must take care when establishing the bias (more false alarms vs. more misses) of automation systems. They must not falsely assume that either type of error will produce the same effect. While the current data indicate that overall human-automation performance is not differentially affected by the automation bias, it is clear that performance during target-present or target-absent trials *is* differentially affected. The discrepancy between degradation of compliance (associated more strongly with false alarms) and degradation of reliance (associated more strongly with misses) must be carefully considered when programming the bias of an automated system.

Designers must consider which type of human error is more devastating—missing a target or falsely reporting a target. In a situation like airport security screening, it is much more dangerous to miss a target object than it is to falsely detect a target object. On the other hand, regarding an event such as a fire alarm, it is much more dangerous to have constant false alarms, as people may become subject to the "cry-wolf" effect (Breznitz, 1983). In situations like these, designers must adjust the automation bias according to the least harmful potential outcome.

Acknowledgments

The authors wish to thank Jackie Chavez for her help in collecting data. The authors also wish to thank Jason McCarley for extremely helpful comments and suggestions during the process of this study. This research was funded by an Air Force grant (Index #111915). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author.

References

Bamber, D. (1979). State trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology, 19*, 137–181.

Bliss, J. (2003). An investigation of alarm related accidents and incidents in aviation. *International Journal of Aviation Psychology, 13*(3), 249-268.

Breznitz, S. (1983). *Cry-wolf: The psychology of false alarms*. Hillsdale, NJ: Lawrence Erlbaum.

Dixon, S. & Wickens, C. (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., & Chang, D. (2005). Mission control of multiple unmanned aerial vehicles: A workload analysis. *Human Factors, 47*(3), 479-487.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse the misses? *Human Factors. 49*(4), 564-572.

Dunn, J.C. & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review, 95*(1), 91-101.

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*, 563-572.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors, 46*(2), 196-204.

Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, and abuse. *Human Factors, 39*(2), 230-253.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics, 30*(3), 286-297.

Rice, S. & McCarley, J. (2008). The Effects of Automation Bias and Saliency on Operator Trust. *International Congress of Psychology*.

Sorkin, R. D., & Woods, D. D. (1985). Systems with human monitors, a signal detection analysis. *Human-Computer Interaction, 1*, 49-75.

Wickens, C.D. & Dixon, S. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomic Science, 8*, 201-212.

IMPLEMENTATION ISSUES FOR UAV CAMERA VIEW TRANSITION DISPLAY AID

Nicholas Wright
711th Human Performance Wing, Supervisory Control Interfaces Branch
Dayton Ohio
Heath Ruff, Brian Mullins
Ball Aerospace & Technologies Corporation
Dayton Ohio
Antonio Ayala, Gloria Calhoun, Mark Draper
711th Human Performance Wing, Supervisory Control Interfaces Branch
Dayton Ohio

An operator supervising multiple unmanned aerial vehicles (UAVs) will be required to switch attention between UAVs, each potentially involving different scenario environments and task requirements. A transition aid is now under evaluation that employs synthetic vision technology to enhance an operator's situation awareness when switching between UAVs. Instead of discretely switching from the camera view of one UAV to the camera view of another, the algorithms driving the transition automatically provide a display format that uses a "fly out, fly in" metaphor over several seconds to transition between the two camera views. This paper will describe the many parameters that need to be specified in the implementation of each segment of a transition format (e.g., what heading, path, rate, duration, and ending point). Results from studies conducted to evaluate specific parameters will be summarized as well as an evaluation of a candidate transition in a multi-UAV control station environment.

Supervisory control of multiple UAVs will be a particularly time-critical, cognitively demanding task (Scott, Mercier, Cummings, and Wang, 2006). Even with highly autonomous UAVs, operators will need to respond to changes in mission requirements and intermittently collaborate and communicate with others in the distributed control network. Moreover, the operator will need to switch attention between UAVs, each potentially involving very different scenario environments (terrain, threat environment, mission objectives, weather, etc.) and task requirements. Not only is there a potential for negative effects associated with task interruptions (Speier, Vessey, and Valacich, 2003; Monk, Trafton, and Boehm-Davis, 2008) and the mental effort required in context acquisition after the switch (St. John, Smallman, and Manes, 2007; Zijlstra, Roe, Leonora, and Krediet, 1999), there is also the potential for negative transfer of context to occur, such that the specific information and tasking involved in the previous mission might delay or degrade the operator's ability to effectively perform tasks in a new mission. For instance, if the operator has a mental model of friendly forces being south of the target in the first mission, will the operator inappropriately apply this model to the new mission? A transition tool that enhances a UAV operator's situation awareness when switching between two camera views would be beneficial.

Previous research has demonstrated an improvement in task switching when a transition between two- and three-dimensional views of the same scene is provided (Holland, Pavlovic, Enomoto, and Jiang, 2004) and improvements in spatial judgments with transitions between different perspective-rendered views of the same scene (Keillor, Trinh, Hollands, and Perlin, 2007). The use of smooth transitions between two- and three-dimensional views has also been explored by Nielsen, Goodrich, and Ricks (2007) and, for air traffic control displays, by Azuma, Daily, and Krozel (1996). To date, efforts have primarily focused on transitioning between ground-based camera views of the same object/scene. Using augmented reality technology, the user is provided computer-generated views not served by the physical cameras to help retain context and spatial relationships with respect to the scene when transitioning between the viewpoints. The results from these efforts inform the design of a transition display for multi-UAV applications that involve more than one UAV viewing the *same object/scene* from different viewpoints. The present paper describes research to develop a display format which helps a UAV operator transition between camera views for applications requiring two or more airborne vehicles monitoring *different objects/scenes*. The goal of this application is to help the operator dissociate from the context/spatial relationships associated with the first UAV/camera view and rapidly acquire needed situation awareness of the new UAV camera view, reducing the potential for negative transfer of context to occur.

UAV Camera View Transition Display Aid Approach

The Air Force Research Laboratory (AFRL) transition display aid provides a display format that dynamically changes between previous and current camera views in a semi-continuous manner, rather than discretely switching the camera views. This dynamic transition takes several seconds and uses a "fly-out, fly-in" metaphor utilizing synthetic vision technology (Figure 1). There are numerous design issues to consider in implementing a transition format. For the fly-out and fly-in segments, what altitude/heading should the virtual camera (VC) start and end at, what path should the VC take, at what rate and acceleration should the VC move, and for how long? Regarding the traverse segment, if the operator is transitioning between two camera views of the same target, then this segment would be important to help retain context and spatial relationships. However, for the targeted application where the camera views are changing from one geographical area to another, showing the scene between the two environments may be of less interest. Manipulation of the various parameters for each segment can change how the transition is perceived by the operator and thus its utility. Another question is the degree to which the operator should have control over the transition parameters in each segment. Research is underway to evaluate these design issues. This short paper will summarize research to date, focusing on the design issues pertaining to the fly-in segment.

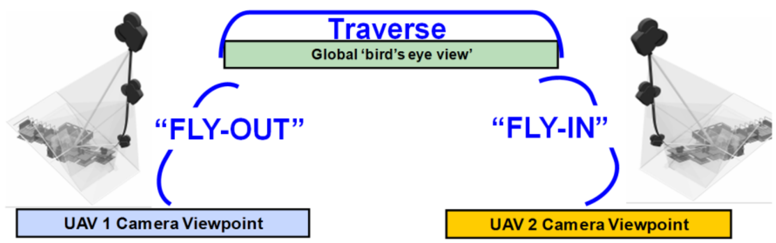


Figure 1. Illustration of each segment of a transition aid format. A three-dimensional perspective of synthetic ground imagery from varying altitudes is provided, switching from a view (determined by the camera's orientation/viewpoint on the current UAV), to a global view not tied to any one UAV, and then back to a view determined by the camera's orientation on the newly selected UAV. During this transition, points of interest are highlighted with overlaid, geo-registered computer-generated symbology.

Duration of the Fly-in Segment

The duration of the fly-in segment must be long enough to provide operators with the necessary visual cues to rapidly acquire situation awareness. However, the time spent viewing the transition format will delay initiating tasks with the real world imagery. A pilot study was conducted to examine several fly-in parameters, including fly-in duration (Lefebvre, Wright, Ayala, Draper, Calhoun, Ruff, and Mullins 2008). AFRL Open Scene Environment (OSE) visualization software was used to present participants with a synthetic camera view that moved along a preset path in an urban environment. Six participants viewed 12 pairs of fly-in segments with three duration time periods: 2, 4, and 6 seconds. The order of trial blocks with each fly-in duration was counterbalanced. Participants were asked to rate their preference for each time duration after completion of all trials. A Friedman Two-Way Analysis of Variance was performed on paired comparisons data from relative judgments of the fly-in durations, and the results revealed a strong trend towards significance ($\chi^2(2) = 4.750$, $p = 0.093$). Results from a post-hoc Wilcoxon Signed Ranks Test indicated that preference ratings for 4 seconds were more favorable than ratings for 2 seconds ($Z = 2.201$, $p = 0.028$; Figure 2).

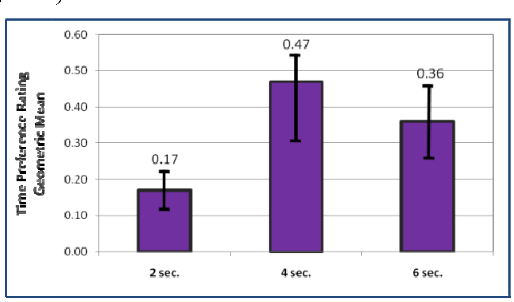


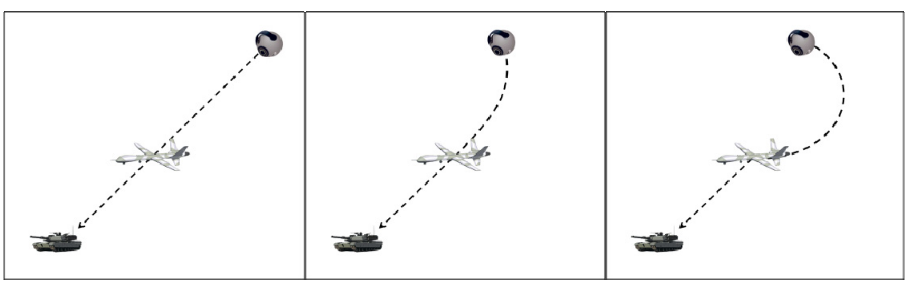
Figure 2. Post-session preference rating for each fly-in duration time.

Path of the Fly-in Segment

The transition's fly-in path segment must provide adequate cues for a viewer to rapidly develop a cognitive map of spatial points of interest, without having the information flow distracting or causing other negative effects. In initial developments (Lefebvre, et al., 2008), the VC was focused on the area surrounding the new UAV. After reviewing pilot study data and consulting with UAV operators, it was determined that the focus should be on improving awareness of the environment surrounding the sensor viewpoint, not the environment surrounding the UAV (which is already represented in the Tactical Situation Display) because the UAVs often do not fly directly above their target. Thus, the VC's center point at the beginning of the fly-in path was changed to the sensor's center point. By incorporating this with a new stare point lock-on tool, the camera center point remains on the image center regardless of the fly-in path.

Several start-points were examined for the VC at the beginning of the fly-in. A path that starts the VC behind and above the actual sensor was chosen as it allows the VC to fly towards the target throughout the fly-in in a natural manner. Using this approach, three new fly-in concepts (linear, shallow curve, and deep curve) were developed and evaluated (Figure 3; Lefebvre, et al., 2008). All three fly-ins were divided into two phases: decent from start point to the UAV and zoom-in to sensor's viewpoint. The VC in the linear fly-in started at a point 12,000 ft away from the sensor along a vector from the sensor's center point through the UAV. The UAVs were at an altitude of 4500 ft AGL (standoff distances from the ground targets were adjusted to maintain camera pitch of ~45 degrees). This meant that the VC started at an altitude of 13,500 ft AGL. By using a FOV of 72 degrees, it was possible to get the sensor's entire relevant area of influence in the VC's view at the start point. The first portion of the fly-in (decent) lasted 5.15 seconds and used an exponential function of 0.6 so the fly-in started slowly and sped up in the middle. The second portion of the fly-in (zoom-in) lasted 0.85 seconds and used an exponential function of -0.9 to slow down the zoom-in as it approached the end (i.e., the new sensor viewpoint).

The two curved fly-ins (see Figure 3) started 12,000 ft away from the UAV as well and had 72 degree FOVs. These two fly-ins always started at an angle of 77.5 degrees above horizontal with respect to the UAV (as opposed to 45 degrees). The curved fly-ins flew towards the target along paths created using cubic Bezier curves to provide multiple perspectives of the area. One fly-in followed a shallow curve that approached the vector from the target through the UAV as it flew (first stage took 5.2 seconds with an exponential function of 0.2; second stage took 0.8 seconds and exponential function of -2.0). The other fly-in followed a deeper curve that went well beyond the vector and approached horizontal flight as it flew in towards the camera (first stage: 4.8 seconds with an exponential function of 0.1; second stage: 1.2 seconds and used an exponential function of -3.0).



Linear Shallow Curve Deep Curve

Figure 3. Illustrations of three fly-in concept camera paths evaluated. All had a total duration of 6 seconds.

Method

Data were collected from 12 participants as they viewed manipulated synthetic camera views that moved along a pre-set path in an urban environment (created with AFRL's OSE visualization software). In Part 1, participants were instructed to create a mental map of the area (and location highlighted by synthetic flags) as they viewed a series of fly-ins (twice with each of the three fly-in concepts). During each fly-in, 8-10 red and blue distractor flags and 5 target flags (each with a unique color) were presented. After each fly-in, a target color was requested and the participant drew a line on a form from the center, representing the new camera view, to the location of the requested flag color. The absolute angle formed by the intersection of the recalled vector of the target and the actual vector from the center of the area was used to measure accuracy from 0 to 180 degrees. Subjective data on the visual appeal and situation awareness provided by the fly-ins were also collected. In Part 2, additional subjective assessment of the fly-in concepts was collected. Participants viewed six paired fly-ins, with each pair comprised of two different fly-in concepts. After each pair, participants compared the two fly-in concepts in terms of situation awareness, visual appeal, and preference.

Results

Statistical analyses of the subjective data collected in both Parts 1 and 2 failed to find significant differences in situation awareness, visual appeal, and preference ratings for the three fly-in concepts. In contrast, the analysis of the objective performance measure in Part 1 was informative. Results indicated that the mean accuracy (difference in the angle between the marked location of the requested flag and the real location of the flag) across fly-in concepts just missed being statistically significantly different at the 0.05 level ($F(2,22) = 3.485$, $p = 0.055$). Post hoc hypothesis test results showed that participants more accurately indicated the location of the requested flag with the linear fly-in concept compared to the shallow and deep curves' paths ($F(1,11) = 12.634$, $p = 0.005$; Figure 4). It appears that participants were able to create better cognitive renditions of the areas with the linear fly-in due to the fixed perspective it utilized. With the linear fly-in, as the camera flew in, the flags' orientations relative to the fly-in path vector remained fixed, whereas with the curved fly-ins, their orientations changed constantly as the perspective changed.

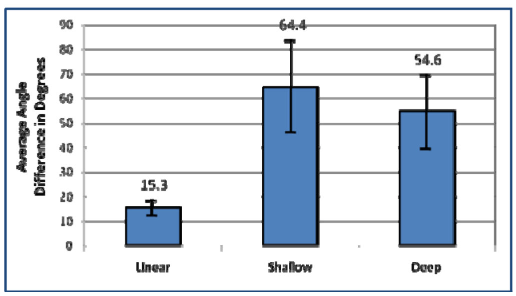


Figure 4. Accuracy in indicating target flag location with each fly-in path concept evaluated.

Fly-in Segment Evaluated in Multi-UAV Simulation

The fly-in segment was evaluated as part of a transition implemented in a multi-UAV simulation environment (see Draper, Calhoun, Ruff, Mullins, Lefebvre, Ayala, and Wright, 2008). Figure 5 illustrates its implementation and shows the duration and rate of change (exponential factor) for each segment. The operator's initial view of a house is from the camera mounted on UAV 1. The camera view switches from a (simulated) live video feed to a purely synthetic environment from a VC which then changes altitude and zoom to give the impression of a smooth continuous fly-out that starts slowly, speeds up in the middle, and ends slowly. The view then switched immediately from the environment surrounding the first target to the environment surrounding the target of the second UAV. The parameters for the fly-in phase provided a short delay at the top of the fly-in and then acceleration towards the target, finally slowing down as the target is approached.

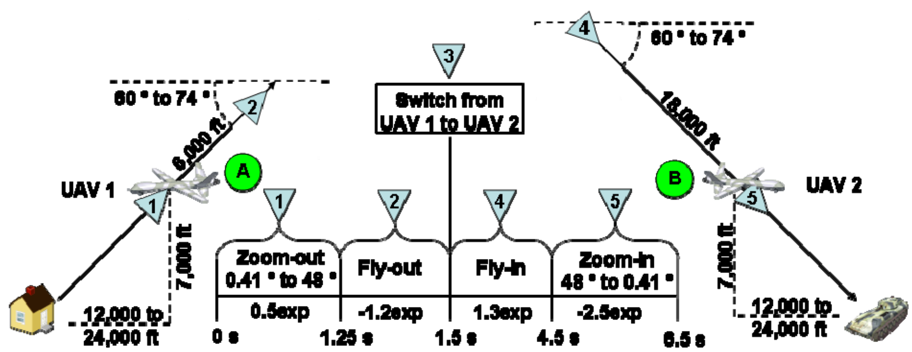


Figure 5. Illustration of parameters used to implement UAV camera view transition display aid.

Besides recording the participants' impressions of the fly-in segment, another objective was to determine if the transition helped the participants' overall situation awareness and improved their performance on a target search/designation task after switching to a new UAV/camera view. This evaluation also manipulated the mission scenario to determine if the utility of the transition depended on whether the previous mission was a static, surveillance-type mission, or a dynamic, close air support mission. Finally, this study was designed to determine if the presence of the transition aid had any negative effects on participants' completion of secondary mission-related tasks. This is important because if the transition aid degrades performance on any task, then its candidacy for multi-UAV control applications is questionable, even if improvements in situation awareness after switching to a different UAV/camera view are realized.

Method

The experiment utilized the Vigilant Spirit multi-UAV operator control station testbed (Rowe, Liggett, and Davis, 2009) which included a full-size camera view from the currently selected UAV (simulated with MetaVR's Virtual Reality Scene Generator), thumbnail camera views from all the UAVs, Tactical Situation Displays (TSDs) showing the location of four UAVs (Global TSD) as well as a close in view of the selected UAV (Local TSD), and windows used for secondary mission-related tasks. Thirteen participants performed eight 12-minute trials, four with the transition format present when switching UAVs/camera views and four without the transition. Each trial consisted of multiple dynamic (close air support) and static (surveillance) missions. Participants received a verbal prompt and chat message when mission transitions should occur and this information identified the next UAV and mission type.

- Dynamic Missions (2 minutes): participants were tasked with locating and designating two enemy tanks, as well as remembering information about the tanks and the relative location of other symbology. At the end of each dynamic mission, participants were asked a question to detect negative transfer of context, as accuracy would depend on knowledge of the current camera view, as opposed to the previous camera view. Response time and accuracy for the tank designation and question response were recorded, as well as the efficiency with which the camera was moved.

- Static Missions (2-5 minutes): participants were tasked with monitoring the video feed for the selected UAV and typing "truck" in the UAV chat window when a truck appeared in the video. The percent of trucks detected was recorded, as well as response time. During static missions, the camera view was automatically zoomed in all the way and the joystick was inactive.

During the static missions of each scenario, participants were required to complete several secondary tasks, representative of the type and range of activities anticipated for multi-UAV supervisory control. These included: a) click on respective UAV thumbnail when prompted to switch UAV sensor monitoring, b) retrieve information (e.g., altitude from UAV summary window), c) monitor Global TSD for an unreported aircraft symbol, d) monitor health and status matrix, and e) monitor audio stream, for tasks associated with the assigned call sign. Completion time and accuracy measures were recorded for each task.

Results

Participants' questionnaire ratings indicated they had more situation awareness in trials with the transition aid format, compared to trials without the transition aid ($F(1,12) = 5.493$, $p = 0.037$). However, participants failed to answer the administered question that measured context-specific situation awareness more accurately with the transition aid. Response time to the question also did not differ significantly as a function of whether the transition aid was present or not ($F(1,12) = 1.522$, $p = 0.241$). While the transition was not found to hinder performance on secondary tasks, it also did not measurably improve performance on the key task – the average time to locate/designate targets was only slightly faster (2.3 seconds) when the transition was utilized ($F(1,12) = 2.054$, $p = 0.177$). The transition, however, did improve the target designation task in terms of camera movement efficiency. Participants' initial camera movement was more accurate (by approximately 12 degrees) when the transition aid was presented compared to when it was not presented ($F(1,12) = 5.969$, $p = 0.031$).

The results from this full mission simulation evaluation indicated several potential enhancements which may increase the utility of the transition display aid for switching between UAV camera views (Draper, et al., 2008). Briefly, the transition format needs refinement, ranging from the speed of various transition segments to whether or not the operator has direct control over transition parameters. The findings also indicate that research is needed to determine which station display(s) should present each information element required for multi-UAV control. Additionally, this experiment showed that there are numerous factors that may influence the utility of a transition aid, including the nature of the missions involved and the users' strategy. Follow-on research is underway to address potential enhancements and other issues identified as a result of these evaluations.

References

Azuma, R., Daily, M., & Krozel, J. (1996). Advanced human-computer interfaces for air traffic management and simulation. In *AIAA Flight Simulation Technologies Conference Proceedings*, pp. 656-666.

Draper, M., Calhoun, G., Ruff, H., Mullins, B., Lefebvre, A., Ayala, A., & Wright, N. (2008). Transition display aid for changing camera views in UAV operations. In *Proceedings of the Humans Operating Unmanned Systems (HUMOUS)*. (http://conferences.telecom-bretagne.eu/humous08/papers/)

Holland, J. G., Pavlovic, N. J., Enomoto, Y., & Jiang, H. (2004). Advantage for visual momentum not based on preview. In *Proceedings of the Human Factors and Ergonomics 48th Annual Meeting* (pp. 1803-1807). Santa Monica, CA: Human Factors and Ergonomics Society.

Keillor, J., Trinh, K., Hollands, J.G., & Perlin, M. (2007). Effects of transitioning between perspective-rendered views. In *Proceedings of the Human Factors and Ergonomics 51st Annual Meeting* (pp. 1322-1326). Santa Monica, CA: Human Factors and Ergonomics Society.

Lefebvre, A.T., Wright, N.F., Ayala, A., Draper, M.H., Calhoun, G.L., Ruff, H.R., & Mullins, B.P. (2008). *Get-in-the-Zone (GITZ) Display Format for Changing Camera Views in Multi-UAV Operations*. Technical Report submitted for publication. Wright-Patterson Air Force Base, OH: Air Force Research Laboratory.

Monk, C.A., Trafton, J.G., & Boehm-Davis, D.A. (2008). *Journal of Experimental Psychology: Applied*, *14*(4), 299-313.

Nielsen, C.W., Goodrich, M.A., & Ricks, R.W. (2007). Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics*, 23(5), 927-941.

Rowe, A.F., Liggett, K., & Davis, J. (2009). Vigilant Spirit Control Station: A Research Testbed for Multi-UAS Supervisory Control Interfaces. In *Proceedings of the 15th International Symposium on Aviation Psychology.*

Scott, S.D., Mercier, S., Cummings, M.L., & Wang, E. (2006). Assisting interruption recovery in supervisory control of multiple UAVs. In *Proceedings of the Human Factors and Ergonomics 50th Annual Meeting* (pp. 699-703). Santa Monica, CA: Human Factors and Ergonomics Society.

Speier, C., Vessey, I., & Valacich, J.S. (2003). The effects of interruptions, task complexity, and information presentation on computer-supported decision-making performance. *Decision Sciences*, *34* (4), 771-797.

St. John, M., Smallman, H.S., & Manes, D.I. (2007). Interruption recovery tools for team collaboration. In *Proceedings of the Human Factors and Ergonomics 51st Annual Meeting* (pp. 229-233). Santa Monica, CA: Human Factors and Ergonomics Society.

Zijlstra, F.R., Roe, R.A, Leonora, A.B., & Krediet, I. (1999). Temporal factors in mental work: effects of interrupted activities. *Journal of Occupational and Organizational Psychology*, *72*, 163-185.

VIGILANT SPIRIT CONTROL STATION: A RESEARCH TESTBED FOR MULTI-UAS SUPERVISORY CONTROL INTERFACES

Allen J. Rowe
Kristen K. Liggett, Ph.D.
Jason E. Davis
Air Force Research Laboratory
Wright-Patterson AFB, OH 45433

Since its inception, the unmanned aerial vehicle (UAV) has adapted to military life and has subsequently become an integral part of modern day warfare. Although unmanned, this technology remains dependent on human interaction for optimal function. Bridging the gap between rapidly advancing technology and the human, the Vigilant Spirit Control Station (VSCS) serves as a multi-faceted facilitator in areas ranging from research to combat missions. The result, consequentially, is an increase in the efficiency of the program by enabling a single operator to supervise multiple vehicles. Streamlining technology is tantamount to the program's success. Developed with this in mind, VSCS effectively integrates sophisticated advancements for the purpose of strengthening the collaborative relationship between the operator and the UAV, and ultimately serves to propel this multi-purpose asset into the next decade.

Although there have been UAVs in existence since before manned flight, it was during the Vietnam era that the use of UAVs as surveillance vehicles significantly emerged (Krock, 2002). Today, UAVs have become a multi-purpose asset used by all branches of the military. In addition, UAVs are being used by state and local governments for such tasks as border patrol, search and rescue, forest fire monitoring, disaster response, and air traffic control. Commercially, UAVs are being considered for power line inspection, monitoring traffic, and filming in Hollywood (Frederick, 2006). On the military side, UAVs have become an integral part of modern-day warfare. Typical missions include intelligence, surveillance and reconnaissance (ISR), target acquisition, suppression of enemy air defenses, and combat missions. To support this wide variety of missions, UAVs carry many different payloads, from various sensors (electro-optical, short-wave infrared, etc.) to a range of armament.

Regardless of the UAV mission, the human interaction with these vehicles is of utmost importance. True, the vehicles are unmanned, but the operations of the vehicles always include a human component, and thus the need for a ground control station (GCS). It is through the interfaces in the GCS that operators perform tasks to ensure successful operations. These tasks include controlling the vehicle, to monitoring the information that the vehicle is gathering and transmitting back to the GCS. Therefore, an important link between the vehicles and the operators are the interfaces provided to execute the mission. The ratio of one operator controlling or supervising one vehicle may seem challenging enough, however, due to the high demand of qualified UAV operators (Hoffman & Kamps, 2005), current trends are moving toward a single operator supervising multiple vehicles. This adds to the importance of robust interfaces that leverage common components across various vehicles, payloads, and missions. As the services work toward interoperability, the development of a common GCS is one of the first steps (Osborn, 2009). Therefore, designing interfaces with a flexible software architecture, a standard way of communicating, a consistent look and feel for performing the majority of tasks, and a subset of tailored interfaces to support "specialty tasks" (i.e., automated aerial refueling), would facilitate this goal. The objective of this paper is to describe VSCS – a UAV GCS interface testbed. First, an overview of the VSCS philosophy will be provided, followed by examples of its implementation in a number of different programs designed to support various missions.

Vigilant Spirit Control Station Overview

VSCS originated several years ago with a primary goal of developing graphical user interface (GUI) concepts to effectively supervise up to four lethal UAVs. This thrust in the late 1990's received attention from the Defense Advanced Research Project Agencies (DARPA) Unmanned Combat Air Vehicle (UCAV) program. A Cooperative Research and Development Agreement (CRDA) was quickly established between the Air Force Research Laboratory's (AFRL) Human Effectiveness Directorate and the UCAV program's prime contractor, Boeing. This relationship helped to pave the way for a series of developments over the next several years that would help VSCS gain momentum in the arena of supervisory control of multiple UAVs by a single operator. During the development of such a system to accommodate the diverse missions and vehicle payloads across multiple vehicle platforms, it became apparent that an advanced intuitive user interface needed to be developed that provided a single common solution. VSCS was developed as a robust research testbed allowing researchers to explore a variety of supervisory control interface concepts to aid in addressing these issues. As illustrated in Figure 1, VSCS was designed around an open architecture allowing researchers access to the development tools needed to concentrate on the variety of scenarios concerning effective control and supervision of multiple UAVs. VSCS comprises a multitude of tools to aid both the researcher and UAV operator, such as a suite of advanced innovative operator interfaces; a simulation environment to aid in stimulating a synthetic environment for the modeling of various vehicle payloads, sensors, and human factors testing tools; dynamic mission planning (DMP) interfaces for interacting with vehicle supervision and control; a robust and flexible software architecture that allows for multiple configurations to accommodate diverse missions across a multitude of vehicle platforms; and finally the interoperability and communication across these vehicle platforms and the associated GCSs.

Figure 1. VSCS Components

Flexible Software Architecture

VSCS has been designed to be used in various types of environments and configurations and for control of multiple vehicle platforms. Developed within a research organization, the software is required to support human-centered experimentation. These tests introduce software requirements for running participants through preplanned trials, collecting usage data, and providing mechanisms to display diverse user interface designs on the fly. More mature research can include conducting live flight tests, for which the GCS must have an ability to communicate with various commercial UAV platforms and also be implemented with concern for potential safety of flight issues. Finally, a robust modeling and simulation framework is needed to either drive laboratory-based research or to test systems prior to flight test. To meet all of these sometimes conflicting requirements, VSCS has been designed to be extremely flexible.

VSCS uses several interrelated mechanisms to achieve its required level of flexibility. The first is a set of Extensible Markup Language (XML) based configuration files that, when properly organized, define what VSCS refers to as a *mission*. A mission contains many items that can be configured: the UAVs under VSCS control and the payload and capabilities of those vehicles; pre-flight defined items such as points and areas of interest, real-world entities to be tracked, and imagery; symbology to be used across GUI elements; and many other settings and scenario-specific items. Closely related to a mission is the concept of a display layout, which is an XML-based specification of the types of GUI elements on the VSCS display and their sizing and positioning. Additionally, VSCS provides numerous extension points that allow for the integration of new GUI components and also various types of algorithms and non-graphical functionality. All of these can be loaded by the GCS without modifying any core source code, through the use of appropriate mission and display layout files.

The data file-driven nature of VSCS is one way that the software can easily support working with different types of UAVs in a variety of scenarios. Depending on the mission and display layout chosen by the operator at startup, any number of UAV exercises can be executed, and prosecuted efficiently by equipping the operator with a specially adapted interface toolset. Another way that these files are used is to provide an efficient means of conducting human-in-the-loop studies. For instance, in preparation for an experiment, a set of missions could be created that allow for altering aspects of the battlespace between trials, adjusting components of the GCS display, or both. Through the use of a test operator console, the person conducting the study can start and stop trials, effectively loading new missions automatically across both VSCS and simulation components, in a sequence that achieves the study's goals.

Interoperability

Another feature of VSCS that opens it up for a wide array of uses is the way that it communicates with other systems. The primary interface that will be addressed in this discussion is the one between the GCS and the UAV that it is controlling. VSCS has adopted the data link interface defined in NATO Standardization Agreement

(STANAG) 4586 for UAV command and control (NATO Standardization Agency, 2008). This standard states that its aim “is to promote interoperability of present and future UAV systems […]”. The STANAG 4586 aims to define a common set of functions that, when implemented on a particular unmanned aerial system (UAS), allow any similarly designed UAV GCS to control that asset to a certain degree. A complete systems architecture is also specified that allows for unobtrusive implementation of the standard in a manner that allows each UAV system to retain any proprietary or custom communications protocol while still being STANAG-compliant. This is accomplished through what is referred to as a Vehicle Specific Module (VSM).

From VSCS’s perspective, all outgoing vehicle command and control and incoming vehicle telemetry and status is conducted through the use of applicable STANAG messages. Assuming the vehicle being controlled does not natively understand these STANAG messages they must first pass through a VSM. This VSM translates the data contained in the STANAG messages into equivalent UAV-specific messages that are then sent to the vehicle for uplink commands (or vice-versa for downlink telemetry and status). While the STANAG provides the functions necessary for basic interoperability, there can still exist occasion to provide platform-specific extensions to the standard for advanced functionality and to alleviate potential safety of flight concerns. For the most part, however, VSCS has been able to leverage STANAG 4586 to achieve a high level of interoperability between several types of vehicle platforms, both virtual and physical.

The VSCS operator interface incorporates a flexible modular design that can be configured to accommodate various mission and payload requirements. The following sections will cover details regarding the core capability interface tools available within VSCS to aid the operator in these functions. As noted in previous discussions, VSCS software architecture provides developers a robust environment for the development of mission and payload specific operator interface tools for specific vehicle platforms that lie outside of VSCS core capability.

Mission Management

Supervisory control of multiple systems requires intuitive and robust operator interfaces to effectively perform all mission management functions. To address this need, VSCS includes a suite of tools to aid the operator during these missions. These are depicted in the vehicle Alert and Summary tool, a tactical situational display (TSD) to provide advanced mapping capability, the command and control interfaces, and dynamic mission planning (DMP) interfaces. Figure 2 depicts a typical mission management display setup. A brief description of each of these will be provided. For further detailed information, please refer to the VSCS Operator Manual (Williams, Feitshans, and Rowe, 2002)

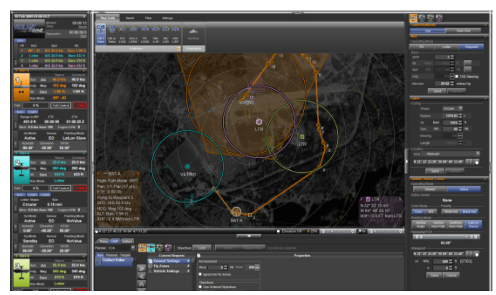


Figure 2. Mission Management Display

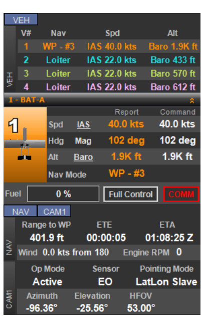


Figure 3. Alert & Summary Tool

The vehicle Alert and Summary tool provides a quick look assessment of pertinent UAV information tailored to the current mission phase (Figure 3). Each UAV is depicted in a dedicated pane providing unique features to aid the operator in quickly distinguishing the various UAVs under the operator’s control. There are four key elements used to provide cues to the operator when performing basic mission management functions. These are color, glyphs, IDs, and callsigns. Color is used throughout the system to uniquely identify each UAV and its associated data, such as flight plans, loiter locations, and sensor information. Glyphs typically indicate vehicle platform and provide basic navigation information such as vehicle heading, altitude and airspeed. Each UAV is assigned a unique ID, such as 1, 2, 3, and so forth, and compliments the glyph to provide another situation awareness (SA) measure for quickly locating a designated UAV. Finally, a unique vehicle callsign is issued based on the current mission and is used in much the same manner as typical manned aircraft missions. A quick crosscheck mechanism is also provided to show basic navigation parameters for all UAVs the operator is currently controlling in this mission phase, such as, navigation mode, airspeed, and altitude. This information is shown at the top of the Summary tool and uses many of the key indicators described to designate each UAV. Payload information, such as sensor cameras, weapons, or radar systems is also displayed for each UAV. Various alerting mechanisms are provided in this panel to indicate loss of communications, loss of global positioning system (GPS) information, low fuel/battery life, or other mission specific alerts.

The TSD provides a common mission operating context consisting of standard aviation charts and geo-referenced imagery (Figure 4). The mapping functionality of the TSD is analogous to using standard commercial tools such as Google Maps, MapQuest, or military standards such as FalconView. The TSD makes available mission information such as UAV position, air and ground tracks, airspace management aids, mission plans, sensor viewing locations (footprints) if available, and target locations. A wealth of additional information can be displayed on the TSD if source data is available for the given UAV platform given their geo-registered coordinates are known. UAV information is consistently shown with the same standard features as those found on the Summary tool as previously discussed. As mission complexity increases, overlapping symbology and clutter become an inherent problem the operator must manage. To aid the operator in dealing with this increased information management, several de-cluttering options are available to reduce this visual overload and reduce symbology clutter. Color is used extensively to aid in reducing information overload as well as techniques to filter several elements based on mission phase. As an example, UAV routes can be filtered based on phase of flight and waypoints visited to significantly reduce this clutter, as well as labels and extraneous information.

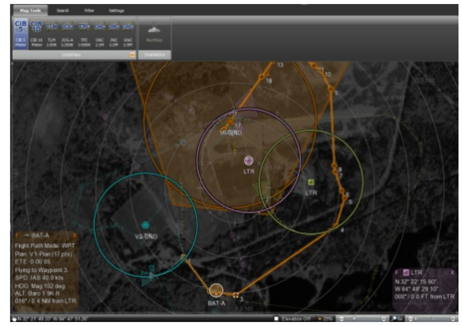


Figure 4. TSD

UAV command and control is accomplished through the use of a variety of standard mission and vehicle platform specific tools (Figure 5). Depending on the vehicle platforms for a particular mission configuration, the appropriate command and control tools are made available in a standard interface tightly coupled to the TSD. These tools provide vehicle navigation control, loiter management, sensor payload management, and DMP. Several techniques are used to assist the operator in accomplishing the goals of a particular mission. Standard keyboard and mouse input is always provided as a redundant input mechanism, but more intuitive techniques can be utilized for frequently used and common functions. The use of voice input, Hands on Throttle and Stick (HOTAS), gaming controllers, and touchscreens are being explored to provide an enhanced user experience. Graphical alternatives are used directly on the TSD when appropriate to aid the operator in quickly issuing a command to a particular UAV (Figure 6). Simply selecting a UAV on the TSD and gaining access to a context sensitive suite of commands and manipulating the constraints and parameters of that particular command directly on a geo-registered map simplifies the complex task of managing multiple UAV in a complex mission scenario (Williams, Hughes, Feitshans, Rowe, and Williamson, 2005).

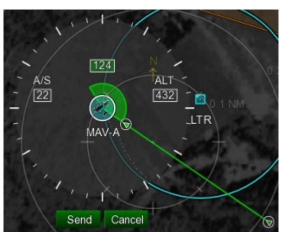


Figure 6. TSD Widget

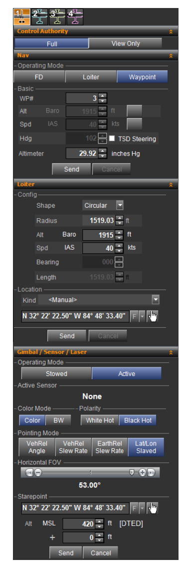


Figure 5. Command & Control Tool

Payload Management

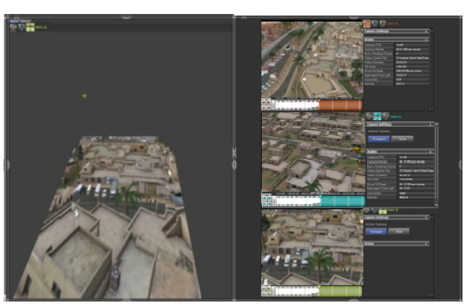


Figure 7. Sensor Payload Management Display

In addition to providing effective mission management functionality, the introduction of various vehicle payloads such as electro-optical (EO) and infra-red (IR) sensors, as well as a diverse set of weapon systems, poses a unique challenge for the multi-UAV operator to maintain SA and successfully execute the desired mission. VSCS includes a suite of payload management tools in conjunction with cooperative vehicle mission planning to accomplish the dual role of air vehicle operator as well as a mission payload operator. These include the use of digital video recording (DVR) capabilities and mosaic functionality (Figure 7). Video mosaic techniques, as shown in the left side of Figure 7, have shown potential in aiding the operator during high workload situations and provide enhanced SA as well as providing a more stabilized view of the target during tumultuous conditions (Feitshans, Rowe, Davis, Holland, and Berger, 2008). Figure 8 provides a close-up view of the DVR tool. Depending on the vehicle platform, image stabilization could introduce undesirable artifacts that could be eliminated or reduced by allowing the operator to freeze frame or "rewind" the video to inspect it more closely when searching for targets of interest. Providing an intuitive user interaction such as clicking directly in the video to designate geo-registered targets or pausing the video through the use of commonly known interfaces found on commercial digital video disk (DVD) devices can reduce operator workload while maintaining SA on several other UAV's during a particular mission.

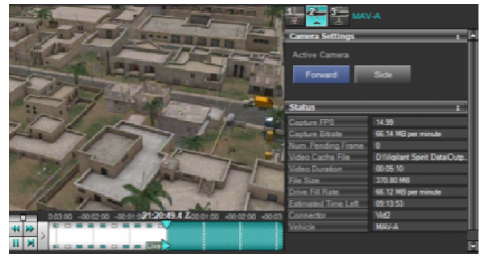


Figure 8. DVR Tool

Weapon and stores payload management is also a challenge while trying to maintain supervisory control over multiple UAVs. Limited research has been accomplished in this area and further information can be found in the following publications (Williams et al., 2002; Williams, Venero, and Linhart, In Press). The use of DMPs has shown great promise in reducing operator workload while maintaining cooperative control of multiple UAVs. VSCS has been designed to interface with these planners through the interoperability mechanisms discussed earlier. To date, VSCS has interfaced with various cooperative control algorithms developed within the Air Vehicles Directorate, as well as Commercial off the Shelf (COTS) products such as Operations Research Concepts Applied Planning and Utility System (Williams et al., In Press).

VSCS Example Program Applications

VSCS has had many opportunities to be involved in a wide variety of unique projects. From advanced research and demonstrations to full-mission usability assessments, these programs have led to a multitude of research results and lessons-learned to further enhance the VSCS testbed. These programs include UCAV, the Joint Unmanned Combat Air System (JUCAS), Long Range Strike (LRS) (Williams, et al., In Press), Automated Aerial Refueling (AAR) of multiple UAVs (Williams, Burns, Feitshans, Rowe, and Davis, 2008), and the multi-aircraft management aspects of the Predator Program. Internal AFRL sponsored research programs such as Cooperative Operations in Urban Terrain (COUNTER) and Multi-UAV Supervisory Control Interfaces Technology (MUSCIT) have increased the VSCS sensor and mission management capabilities to include unique toolsets for effectively monitoring multiple small and micro UAVs to provide enhanced urban telepresence (Feitshans et al., 2008; Patzek, In Press).

To address more fundamental basic and applied research questions, a variety of AFRL-sponsored part-task evaluations have recently been performed. One effort is evaluating a transition aid designed to rapidly build operator SA when switching between UAV missions and their associated sensor views (Draper, Calhoun, Ruff, Mullins, Lefebvre, Ayala, & Wright, 2008). The Multi-Aircraft Video – Human/Automation Target Recognition (MAV-HATR) studies focused on effectively monitoring multiple video feeds for target identification and tracking (Carretta, Patzek, Warfield, Spriggs, Rowe, Gonzalez-Garcia, & Liggett, In Press). Another line of research is investigating the use of enhanced symbology to portray navigation and status information for multiple UAVs. This involves the development of glyphs. Finding the optimal size, information portrayal, and reduced clutter to intuitively show context sensitive mission information is the focus of a series of studies to support multi-UAV supervisory control. Figure 9 illustrates a notional glyph that could be displayed on a TSD or Summary tool to aid the operator during the course of the mission. Enhanced tools are constantly being developed and evaluated to refine their effectivness during all phases of a diverse set of missions. One such tool includes providing terrain shading as well as a 3D perspective viewing capability directly on the TSD to enhance awareness for potential hazardous terrain (Figure 10). Another tool is a temporal display, which provides cues to upcoming events as well as visual deconfliction of potential hazards or retasking of UAVs during inherently

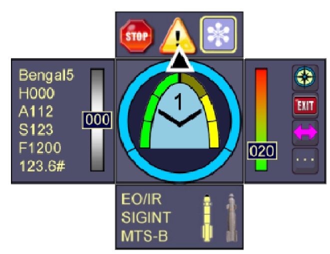


Figure 9. Notional Glyph

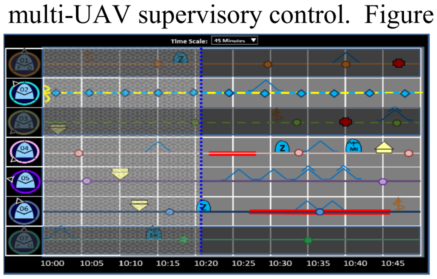


Figure 11. Notional Temporal Display

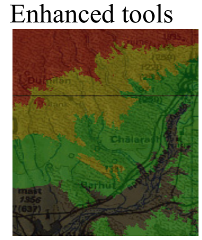


Figure 10. Terrain Shading

beneign or idle periods of time during the mission (Figure 11). The list of useful tools to aid the operator in multi-UAV control are constantly being developed and refined by the VSCS team. Combining the right mix of tools during complex and potentially stressful environments are the focus of VSCS.

Summary

The goal of VSCS is to provide the UAV community with a research testbed to continue to push the envelop of advanced multi-UAV supervisory control. This is accomplished by providing a robust software architecture and interoperability capability. It has enabled VSCS to be used throughout several research and flight test projects. The success of VSCS is evident in the wide spread utilization of this research testbed throughout several government sponsored organizations to promote multi-UAV supervisory control across diverse missions to provide one common solution.

Acknowledgements

The authors would like to acknowledge Mr. Gregory Feitshans, VSCS Program Manager. VSCS is currently under the management and direction of the 711 Human Performance Wing, Human Effectiveness Directorate, Supervisory Control Interfaces Branch, 711 HPW/RHCI.

References

Carretta, T. R., Patzek, M. J., Warfield, L., Spriggs, S. E., Rowe, A. J., Gonzalez-Garcia, A., & Liggett, K. K. (In Press). Target Acquisition Involving Multiple Unmanned Air Vehicles: Interfaces for Small Unmanned Systems (ISUS) Program Final Report (AFRL-RH-WP-TR-2009-XXXX). WPAFB, OH.

Draper, M., Calhoun, G., Ruff, H., Mullins, B., Lefebvre, A., Ayala, A., & Wright, N. (2008). Transition display aid for changing camera views in UAV operations. In *Proceedings of the Humans Operating Unmanned Systems (HUMOUS)*. (http://conferences.telecombretagne.eu/humous08/papers/)

Feitshans, G. L., Rowe, A. J., Davis, J. E, Holland, M., & Berger, L. (2008). Vigilant Spirit Control Station (VSCS) "The Face of COUNTER" . AIAA Guidance, Navigation, and Control Conference.

Frederick, M. (2006). UAVs Take to the Skies. *Space News*. 23 Jan. 2006. 26 Jan. 2009 <http://www.space.com/spacenews/>.

Hoffman, J. C., & Kamps, C. T. At the Crossroads: Future "Manning" for Unmanned Aerial Vehicles. *Air & Space Power Journal AFRP 10-1 XIX (2005)*. Air & Space Power Journal Spring 2005. 01 Mar. 2005. 28 Jan. 2009 <http://www.airpower.maxwell.af.mil/airchronicles/apj/apj05/spr05/hoffman.html>.

Krock, L. Timeline of UAVs. *Nova.* Nov. 2002. 27 Jan. 2009 <http://www.pbs.org/wgbh/nova/>.

NATO Standardization Agency. (2008). Standard Interfaces of UAV Control System (UCS) for NATO UAV Interoperability, STANAG 4586 Edition 2 Amendment 1. Department of Defense.

Osborn, K. DoD To Set UAV Standards by Summer. *Defense News* 19 Jan. 2009. 28 Jan. 2009 <http://www.defensenews.com/story.php?i=3907656>.

Patzek, M. J. (In Press). Multi-UAV Supervisory Control Interface Technology (MUSCIT). In *Proceedings of the 15th International Symposium on Aviation Psychology*. Dayton, OH.

Williams, R. Burns, R. S., Feitshans, G. L., Rowe, A. J., & Davis, J. E. (2008). Vigilant Spirit Control Station Operator Manual for Automated Aerial Refueling (AFRL-RH-WP-TR-2008-XXXX). WPAFB, OH.

Williams, R. Feitshans, G., & Rowe, A. (2002). User's Manual for the AFRL Unmanned Combat Air Vehicle Operator-Vehicle Interface (Version 3.0). WPAFB, OH.

Williams, R., Hughes, T., Feitshans, G, Rowe, A., & Williamson, D. (2005). Operator-Vehicle Interface Concepts for Control of Multiple Unmanned Aerial Vehicles (AFRL-HE-WP-TR-2005-0159). WPAFB, OH.

Williams, R., Venero, P., & Linhart, C. (In Press). Long Range Strike Aircraft Dynamic Mission Planning Project Final Report. WPAFB, OH.

ROBOT OPERATOR SPECIFICATIONS DERIVED FROM THE OCCUPATIONAL INFORMATION NETWORK

Michael D. Coovert

University of South Florida and MDC & Associates, Inc.

Tampa, FL 33620

Linda R. Elliott

Army Research Lab/HRED

Ft. Benning, GA 31905

The use of robots in aviation is widespread, for use as targets, decoys, remote sensing, reconnaissance, and increasingly for combat missions. Robots come in all forms and capabilities, from handheld micro air vehicles to hypersonic versions capable of high altitude long distance missions. At the same time, ground-based robots have proven effective for both military and civilian applications such as manufacturing and remote sensing / manipulations. Certainly, Talon and Packbot robots have proven their worth in battle conditions. Just as there is great variance in the type of robot being developed and utilized, so too is there tremendous variation operator requirements. Thus, it is essential the individual possess the requisite knowledge, skills, abilities and other characteristics to do so. Our work specifies the requirements for a prototypical robot operator through an examination of the U.S. Department of Labor/Employment and Training Administration's Occupational Information Network (O*NET). The applicable aviation occupations include UAV operations and also many other robots used in the aviation domain. Results yield the following for operator characteristics. *Knowledge*: The high frequency types are mechanical, production and processing, computers and electronics. *Skills*: high language component with high active learning, active listening and reading comprehension, critical thinking and mathematics. *Abilities*: Seven specific types cognitive ability are deemed important: problem sensitivity, information ordering, oral comprehension, deductive reasoning, oral expression, inductive reasoning, and written comprehension. Our paper documents the full operator profile that can be used for a variety of purposes including selection, training, and human factors design and specifications.

There is a long history for classification systems to be employed in describing worker characteristics and requirements for occupations in the United States. These specifications are important as they lead to standards for equipment design, human factors requirements, and training needs, among others. The U.S. Department of Labor successfully employed the *Dictionary of Occupational Titles* for many years, but in light of changes to occupations and the need to access and update this information in a timely fashion drove the development of a new system, the Occupational Information Network (O*NET).

O*NET represents a comprehensive assessment of various jobs and occupations found in the United States. The data about each occupation were gathered by taking both a job-oriented and a worker-oriented perspective for the requirements to perform work. This provides complete coverage of the knowledge, skills, abilities, and other worker characteristics required for successful job performance. The O*NET framework is presented in Figure 1.

Information in O*NET is very rich and meant to fulfill the needs of a variety of purposes, such as an older worker interested in changing occupations or locations, a young individual preparing to entering the workforce and wanting to know the skill requirements for various occupations, or a high school vocational counselor providing guidance to students.

We examine the O*NET database to glean a picture of the cognitive and task demands for a prototypical robot operator so that information might be used to address a variety of purposes ranging from selection and training to human factors. The information gleaned from O*NET can serve as one source of input into a taxonomy of cognitive and task demands for the warfighter robot operator. The other source of input is the scientific literature in general and the human-robot literature in particular.

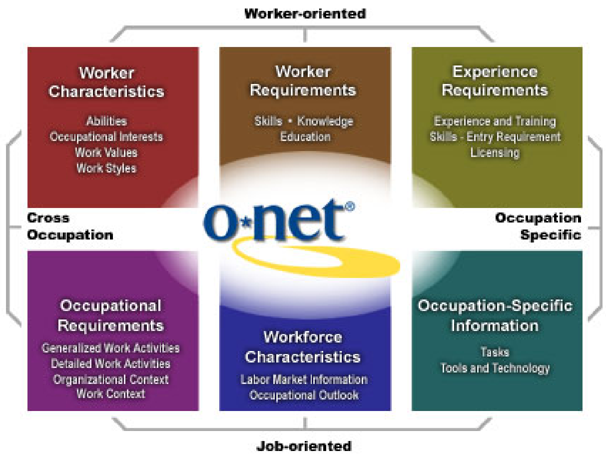


Figure 1. Content model of O*NET (figure reproduced from the O*NET Resource Center, used with permission. http://www.onetcenter.org/content.html#cm2).

The paper is organized as follows. First, a description of the O*NET coding scheme is given to provide an understanding of the hierarchy of jobs and occupations within the database (for a detailed description of any aspect of the construction and validation of O*NET, see the O*NET final report (Peterson et al., 1997). Second, the search strategy and occupations utilized is listed. Data from eleven different occupations was employed to generate the profile for the prototypical robot operator occupation. Third, a description of job knowledge, skills, and abilities as defined in O*NET is provided along with a summary for the robot operator occupation. Fourth, work activities defined by O*NET are provided and those that lead into our robot operator are listed. Although perhaps not as directly pertinent as knowledge, skill, and abilities for human factors specifications, information on work activities is still useful and informative. Finally, the prototypical robot operator occupation worker characteristics and task activities are listed by highlighting those that are most important across existing robot occupations.

*O*NET Occupational Code*

Each occupation requires a specific mix of knowledge, skills and abilities to accomplish the required tasks and activities. O*NET uses a set of descriptors and for each occupation and at the highest level there are six of these. They are expanded to 277 for which unique measurable information is gathered. The information about any one occupation in O*NET is linked in a relational database via an occupational code. The following quote from the O*NET website describes how the occupational classification scheme is applied.

"Each item in the hierarchy is designated by a six-digit code. The hyphen between the second and third digit is used only for presentation clarity. The first two digits of the SOC code represent the major group; the third digit represents the minor group; the fourth and fifth digits represent the broad occupation; and the detailed occupation is represented by the sixth digit. Major group codes end with 0000 (e.g., 33-0000, Protective Service Occupations), minor groups end with 000 (e.g., 33-2000, Fire Fighting Workers), and broad occupations end with 0 (e.g., 33-2020, Fire Inspectors). All residuals ("Other," "Miscellaneous," or "All Other"), whether at the detailed or broad occupation or minor group level, contain a 9 at the level of the residual. Detailed residual occupations end in 9 (e.g., 33-9199, Protective Service Workers, All Other); broad occupations which are minor group residuals end in 90 (e.g., 33-9190, Miscellaneous Protective Service Workers); and minor groups which are major group residuals end in 9000 (e.g., 33-9000, Other Protective Service Workers):

33-0000 Protective Service Occupations
  33-9000 Other Protective Service Workers
    33-9190 Miscellaneous Protective Service Workers
      33-9199 Protective Service Workers, All Other"

O*NET provides a readily accessible and searchable means for gathering information about thousands of occupation titles across hundreds of occupation titles in the US.

Occupational Relevance Score

The O*NET database contains hundreds of occupations and many thousands of jobs, with new ones being added daily. One of the primary advantages of the database is the ability to compare existing jobs to new job, or jobs in one occupation to those in another. To accomplish this comparison, O*NET employs a relevance score, which is defined as: **"**Relevance Score - The search strategy used in the keyword search employs a combination of occupational information, such as associated alternate titles, description, and tasks. A raw score is calculated based on the number of matches across the different data elements and their respective weights. This maximum score becomes the normalization factor. The scores are translated to a 0 to 100 relevance ranking by the following formula: relevance ranking = (score / maximum score) * 100. Thus, the occupation with the highest relevance ranking will be 100. Those occupational titles receiving less that the maximum score will receive a lower ranking. The lowest possible ranking is 0." (Occupational Information Network, 2009).

*KNOWLEDGE, SKILLS, ABILITIES and WORK ACTIVITIES as DEFINED in O*NET*

The knowledge, skills, and abilities definitions come from the O*NET website. See Occupational Information Network (2009) http://www.onetcenter.org, or Peterson et al. (1997) for complete information.

Knowledge

Success in any job depends on having a background set of knowledge. Thirty-three specific knowledge areas are defined in O*NET. Occupations are rated on the extent to which each specific type of knowledge is important to that occupation. knowledge area as defined in O*NET an organized sets of principles and facts applying in general domains. Examples from the 33 include: biology, computers and electronics, design, geography, mechanical, personnel and human resources, physics, and transportation.

Skills

As with knowledge, it is important to identify the extent to which particular skills are necessary in an occupation. O*NET defines six major categories of skills: Basic, complex problem solving, resource management, social, systems, and technical. Each category is elaborated from into one to eleven specific type of skill. The six categories are defined as follows. 1. *Basic Skills*: Developed capacities that facilitate learning or the more rapid acquisition of knowledge. 2. *Complex problem solving skills*: Developed capacities used to solve novel, ill-defined problems in complex, real-world settings. 3. *Resource management skills*: Developed capacities used to allocate resources efficiently. 4. *Social skills*: Developed capacities used to work with people to achieve goals. 5. *Systems skills*: Developed capacities used to understand, monitor, and improve socio-technical systems. 6. *Technical skills*: Developed capacities used to design, set-up, operate, and correct malfunctions involving application of machines or technological systems.

Abilities

O*Net defines four major categories of abilities: *cognitive*, *physical*, *psychomotor*, and *sensory*. Each category has from nine to twenty-one specific types. The list of *cognitive abilities* is very comprehensive – with twenty-one unique types identified and linked to occupations. They cover the gamut of cognitive functioning, ranging from the perceptual level (e.g., perceptual speed) through memorization, and higher-order processing such as that found in deductive and inductive reasoning, oral and written comprehension, and mathematical reasoning. *Physical abilities* are essential for jobs with a high physical component. These deal with gross physical characteristics (e.g., trunk strength) and are anticipated to be of little importance to the typical robot operator position. Physical abilities are defined as: Abilities that influence strength, endurance, flexibility, balance, and coordination. *Psychomotor abilit*ies are the third class of abilities utilized in O*NET to classify jobs. These deal with more micro-level abilities (e.g., control precision, finger dexterity) than the physical abilities class. As robot operators must interact with the robot through an operator control unit of some type (computer interface), it is anticipated these are more likely aligned with the prototypical robot operator occupation. Psychomotor activities are defined as those abilities that influence the capacity to manipulate and control objects. The final class of abilities in O*NET are *sensory abilities*. Although sensory abilities are important as they feed into higher level perceptions and cognitions, they are anticipated to play a smaller role in direct overall importance to the job than other skill classes. Sensory abilities are defined as abilities that influence visual, auditory, and speech perception.

Work Activities

The final set of O*NET descriptors used here is *work activities*. There are four major categories of work activities defined in O*NET: *information input*, *interacting with others*, *mental processes*, and *work output*. Each, as defined in O*Net is now listed. 1. *Information input*: Where and how are the information and data gained that are needed to perform the job. 2. *Interacting with others*: What interactions with other persons or supervisory activities occur while performing this job? 3. *Mental processes*: What processing, planning, problem-solving, decision-making, and innovating activities are performed with job-relevant information? 4. *Work output*: What physical activities are performed, what equipment and vehicles are operated/controlled, and what complex/technical activities are accomplished as job outputs?

Summary

O*NET provides a comprehensive approach for evaluating an occupation from the perspective of worker and job requirements. Our purpose is to employ the rich data of O*NET to determining the cognitive and task demands on a warfighter utilizing one of many different robots with various operator control units. This information, combined with that from the scientific literatures will help with the establishment of taxonomy of such demands. Once established, various human factors purposes can be served, such as: equipment design, workload reduction/optimization, interface development (scalability issues), and task specification (e.g., for collaborative technologies; individual vs. team issues).

Method

Three searches in O*NET were conducted, two focusing on robot occupations and, given the interface often used to interact with a computer, one search utilized the term computer. The robot occupational searches employed 'robot operator' 'robot' as search terms. Robot operator is clearly targeted at our interests, but a search was also conducted on robot as it is more general and would turn up specifics that might need further investigation. The two searches are summarized in the table below. The top ten occupations are the same regardless of which term is utilized—although the rank-order changes. Warfighters often control robots via a computer interface. Due to this, a search was conducted on a 'computer operator' occupation. In sum, for the knowledge, skills, abilities, and work activities for the prototypical robot operator job were identified based on a search employing three occupational terms (robot operator, robot, computer operator) which yielded eleven occupations. These are listed in Table 1.

Table 1. *Occupations, relevance scores, and rank orders for search terms*.

	Search ='robot operator'		Search='robot'	
Occupation code and title	Relevance score	Rank order	Relevance score	Rank order
51-4122.00 Welding, Soldering, and Brazing Machine Setters, Operators, and Tenders	100	1	94	2
51-4121.06 Welders, Cutters, and Welder Fitters	97	2	68	6
17-3024.00 Electro-Mechanical Technicians	89	3	100	1
51-4011.00 Computer-Controlled Machine Tool Operators,	68	4	67	10
51-9031.00 Cutters and Trimmers, Hand	62	5	68	6
17-3023.03 Electrical Engineering Technicians	59	6	71	3
17-3027.00 Mechanical Engineering Technicians	59	7	71	3
17-3023.01 Electronics Engineering Technicians	58	8	71	3
49-2094.00 Electrical and Electronics Repairers,	57	9	68	6
27-1021.00 Commercial and Industrial Designers	55	10	68	6
43-9011.00 Computer Operators	15	15	<5	<5

Results

Specifying Requirements for a Prototypical Robot Operator Job

Now that the relevant occupations that can be used to pull data from have been identified, we proceed through the data examining the frequency each knowledge, skill, or ability type is mentioned for the jobs. That is listed in the following tables along with the percentage of occupations in our set requiring it. For evaluation purposes, a frequency of 5 which equates to a percentage of 45 to be significant enough to be considered an important characteristic for a robot operator. For example, mechanical knowledge is listed as a requirement for nine of the 11 jobs in our sample. This equates into it being a requirement for 82% of the sample jobs.

Prototypical Robot Operator: Knowledge, Skills, Abilities, and Work Activities

The ground work has been laid for developing the requirements for a prototypical warfighter who must operate a robot. These requirements are derived from the perspective of knowledge, skills, abilities, and work activities contained in the nations Occupational Information Network. O*NET includes information on worker and job requirements, and the extent to which jobs are related based on the profile of these requirements. Worker knowledge, skills, abilities, and work activities provided the data for the warfighter robot operator profile.

Knowledge refers to an organized set of principles and facts that apply in general domains. *Mechanical knowledge* is the most important and operators need to be able to understand the design, use, repair, and maintenance aspects of the robots. *Production and processing* is identified as important, but as it deals more with manufacturing is not as central to a warfighters understanding. *Knowledge of computers* and electronics is essential as well. Operators need to have an understanding of the hardware and software of the robots and perhaps some application programming. Similar to this is *knowledge of engineering and technology* where the warfighter operating a robot is able to apply principles, techniques and procedures for equipment design. Knowledge of the *English language* is essential, but all warfighters will have this and it is not unique to the robot operator. The final knowledge requirement is *mathematics* as it helps in many aspects of robot operation.

There are six major skill categories, but only two (basic and technical) contain specific skills important to the warfighter robot operator. Basic skills refer to the developed capacities that facilitate knowledge or its acquisition. The basic skill requirements are five: *active learning, active listening, reading comprehension, critical thinking,* and *mathematics*. *Active learning* deals with the ability to understand the implications of both current and future problem-solving and decision-making. This is clearly an important capability for the warfighter involved with operating a robot. *Active listening* is a skill where the individual gives full attention to what others are saying, taking time to understand the points being made, and not interrupting at inappropriate times. The third basic skill for the warfigher robot operator is the capacity to understand written sentences and paragraphs in work related documents – *reading comprehension*. A fourth basic skill requirement is *critical thinking* which involves the ability to use logic and reasoning to identify strengths and weaknesses of alternate solutions, conclusions, or approaches to problems. This seems almost intuitive as a required skill for a warfighter operating a robot. The final skill is *mathematics* and is useful for solving and considering alternative approaches to problems.

Technical skills are the second skill class important for the warfighter robot operator to posess. Technical skills are the developed capacity to design, set-up, operate and fix technological systems. Four are specifically required: troubleshooting, equipment maintenance, equipment selection, and operation monitoring. *Troubleshooting* deals with determining the cause of operating errors and deciding what to do about it. Clearly this is an important skill for a robot operator. *Equipment selection* and *equipment maintenance* are two skills focusing on managing tools and equipment for a job and for selecting the correct apparatus and maintaining it during use. Finally, *operation monitoring* is perhaps the most essential skill as it deals with the capacity to watch gauges, dials, and other indicators to ensure the robot is operating properly. This is important for both line of sight and non line of sight operation.

Abilities are the final worker-oriented attribute to assess for the warfighter robot operator. Of the four major categories of abilities in O*NET, three are important. Cognitive abilities are the most essential, at least in terms of number as seven separate ones are considered important. Cognitive abilities influence the acquisition and application of knowledge to problem solving. Beginning with *problem sensitivity*, this is the warfighter robot operators ability to tell when something is either wrong or likely to go wrong. It does not focus on solving the problem, but on the ability to recognize that there is a problem. Second is *information ordering* which involves the ability to arrange things in a certain order or pattern according to specific rules. *Deductive reasoning* is another required ability and it is the capacity to apply general rules to specific problems to produce answers that make sense. Similarly, *inductive reasoning* is also essential. Separate from deduction, inductive reasoning is the ability to combine pieces of information to arrive at general rules or conclusions. Other cognitive abilities focus on the oral and written. *Oral comprehension* states it is important for the operator to be able to listen to and understand information and ideas presented in spoken words and sentences. Similarly, the *oral expression* ability states the need to be able to communicate information and ideas in speaking so others will understand. Finally, the warfighter robot operator needs the ability to read and understand information and ideas presented in writing. This final ability is *written comprehension*.

Psychomotor abilities refer to the capacity of an individual to manipulate and control objects. For the warfigher operating a robot three psychomotor abilities are important: Arm-hand steadiness, control precision, and finger dexterity. *Arm-hand steadiness* is the ability to keep one's hand and arm steady while moving the arm or holding the arm and hand in one position. *Control precision* is the ability to work quickly and repeatedly adjust the controls of a machine or vehicle to one or more exact positions. The third psychomotor ability is *finger dexterity* which is the ability to make precisely controlled and coordinated movements of the fingers of the operator's hands to grasp, manipulate, or assemble objects. In looking at these three psychomotor abilities it is easy to see how they are important for the typical human-robot interface consisting of a joystick, mouse, or similar interaction controller.

Finally are sensory abilities which are those that influence visual, auditory, and speech perception. For the warfighter who is the operator of the prototypical robot, only near vision is considered essential. *Near vision* is the ability to see details at close range. While it is easy to understand how this ability is important to all operators, it is also acknowledged that other sensory abilities will be important depending on the particular robot or task. For example, if utilizing a 2D or 3D auditory interface, *auditory attention* (ability to focus in the presence of distracters) would be important when engaging in a targeting task. Another example is *hearing sensitivity* for the same type of interface.

Work activities are the final domain to be considered for our warfighter robot operator. There are four primary categories of work activities and each is considered for relevance to the warfighter. Information input, interacting with others, mental processes, and work output are all important for the robot operator. Functional levels of each are provided in our full technical report.

Summary

Our goal is to identify the cognitive and task demands for a warfighter who must operate a robot. There is, however, tremendous variance in the type of robots deployed and the interactional devices associated with robots. The strategy taken here is to leverage information and data available across thousands of workers, hundreds of jobs, and many occupations where workers interact with robots. These have been evaluated in O*NET for the knowledge, skills, abilities, and work activities that are essential for successful performance. By aggregating across jobs and then selecting those knowledge, skills, abilities and work activities that are most prominent it is possible to create a profile for a prototypical robot operator. The profiles can be used for a variety of purposes including input into a cognitive and task demand structure that feeds a variety of human factors needs and goals.

Acknowledgements

This work was supported by government contract number DAAD19-01-C-0065, task order 83. The views, opinions, and/or findings contained in this report are those of the authors and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

References

Occupational Information Network, O*NET Online, DOI=http://www.onetcenter.org.

Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., Fleishman, E.A., & Levin, K.Y., (1997). *O*NET Final technical report, Volume I [and] Volume II [and] Volume III*. American Institutes for Research, Washington, DC.

AVIATION AUTOMATION DESIGN AND IMPLEMENTATION - THE NEED FOR HUMAN FACTORS CONSIDERATIONS

Karl Fennell
Airline Pilots Association – International
Herndon, Virginia
Shawn Pruchnicki – Ohio State University
Columbus, Ohio
David McKenney, Airline Pilots Association - International
Helena Reidemar, Airline Pilots Association - International
Kevin Comstock, Airline Pilots Association – International

This document outlines the Air Line Pilots Association's (ALPA) aviation automation concerns and expresses the recommendations of both experienced pilots and experts alike. Refer to our statement of position document for additional details at www.ALPA.org under Safety, HFT (Human Factors and Training). Safe and effective aviation automation* is only possible when human factors principles are utilized properly. We strongly encourage engineers, regulators, and operators to apply human factors considerations at every stage of aviation automation hardware, software, and procedure design. Occasionally procedures or products are implemented without these considerations. This inattention can make usage problematic and has produced unintended consequences resulting in accidents and incidents. Incorporation of human factors considerations early in product/procedure design will help to avoid repetition of past mistakes and will ensure that automation maintains and increases the level of aviation safety in the future.

Automation is and will continue to play an important role in the evolution of global air transportation. Due to the complexity that arises from decreased traffic separation and increased use of automation, the dangers of coupling - tight integration and interdependence - increase as well. When an automated coupled system fails, the failure can escalate and cause catastrophic breakdown of the entire system. Appropriate human involvement can provide flexibility to counter problems with overly integrated automation.

In order to design against these failures it is essential to follow clear human factors guidelines that allow pilots and air traffic controllers/managers to interact with each other. It is important to design the automation in a way that complements the strengths of the human and automation and protects against their limitations. To paraphrase the director of the NASA

* We define automation functionally for commercial flight in terms of how it is used for flight purposes. The three purposes include the following: Control Automation, Information Automation, and Management Automation.

Aviation Program, Amy Pritchett (May 2008, Human Factors in NextGen, Arlington, TX), automation cannot handle the complexities of the Next Generation Air Transportation System without humans at the center as integral components and still maintain our current level of safety.

Automation is evolving into new roles to enable new aviation systems to function with increased utilization and control. This places new demands on a system already under pressure. The new automation components are themselves a potential source of error and risk. This is especially true if automation design and implementation has not adhered to established human factors principles.

Aircraft will still need to be flown by pilots. Piloting tasks will change and evolve, including mastering new types of automation and responsibilities. The ultimate responsibility for flying a safe aircraft will remain with pilots.

Automation Philosophy

A well-trained and well-qualified pilot has been, is, and will be the critical center point of aircraft safety systems and an integral safety component of the entire commercial aviation system (ALPA Unmanned Aircraft System (UAS) Policy, May 2007). This system includes not only the crew and aircraft hardware/software but also the operator, regulator and all policies and procedures employed.

The pilot in command has the final authority and responsibility to assure the safe outcome of the flight. It is imperative that the pilot is able to completely control the aircraft during all phases of flight. A design is unacceptable if the aircraft or the transportation system within which it is intended to operate would prevent the pilot from exercising complete control at all times.

The pilot must continue to be the decision maker at the center of the aircraft operation. This provides needed flexibility in a tightly coupled automated system. This also ensures vested human involvement and responsibility. A pilot is certified by regulating authorities with strict mandates for command and operation of the aircraft. These standards help maintain the high levels of safety required in commercial aviation.

Design and implementation of automated systems must focus on augmenting the benefits and strengths of humans while protecting against natural limitations. In general, automation should solve problems, not create additional problems in new or existing systems.

Design

The most effective automation design places appropriate emphasis on human capabilities and limitations. This type of design focuses on several foundational human factors issues. These include appropriate feedback, meaningful alerts and warnings, proper level of automation, and optimum level of pilot involvement per task. Human factors design requirements and end product goals must be established prior to design conceptualization with input from pilots and other users. This allows necessary modifications prior to actual development.

Automation systems should also be designed for the environment in which they will be operating. This includes the ability to handle short-notice changes necessary to accommodate factors such as varying weather conditions, changes in routing due to the presence of other aircraft, and degraded automation performance or failure. This must be done without placing unmanageable demands on the pilots and air traffic system. Automation systems should also be designed so they are compatible with the original equipment manufacturer (OEM) Flight Deck design philosophy in which they will be used, including those systems added to the flight deck after initial certification.

Pilots must be able to control every level of automation. For this to work correctly, the automation should provide the pilots with clear indication of both its present status and expected future state. The automation must provide adequate time for the pilots to intervene in the operation if necessary. Alerts and warnings must balance too many false alarms with too few critical warnings. Too many false alarms can result in lack of trust; too few actual alarms can result in missing critical failures and false security.

Standardized procedures and Crew Resource Management (CRM) should be considered in conjunction with automation design. The intended procedures must be communicated to the users. Consideration must also be given to the sequence, synchronization of procedures and time criticality of any task. A procedure may be benign when performed in normal sequence, but hazardous if performed slightly out of sequence.

Evaluation and Certification

Every new automation component or tool will require an operational evaluation and should be conducted with the participation of the end user, i.e. line pilots. The operational evaluation should include the accomplishment of a thorough risk analysis that leads to a risk mitigation plan. This must be accomplished before any automation system is introduced into the aviation domain.

Scenarios should be built to evaluate the automation function in the operational context in which it will be utilized. Evaluations should be objective with reproducible metrics. These evaluations must be accomplished prior to certification and accepted for use.

Clear evidence must show that pilots and controllers are able to use the automated system or procedure at acceptable error rates – prior to implementation. This evidence should include empirical tests with sufficient statistical power and external validity to guarantee reliable results. The evidence must demonstrate that typical operators are able to use the equipment to perform both normal and emergency operations. Testing should also show that any actual operational errors or the precursors to those errors are both low risk and only occur at low rates that do not pose risk for actual operations.

Training

The objective of training should be to provide pilots with a complete and accurate model of the automation system. This enables pilots to correctly identify and predict the system's actions and to control them during normal and abnormal situations. Training should not be used as an attempted substitute for poor Human Factors design.

Airline specific automation philosophy should be standardized across fleets to the maximum extent possible as long as it does not conflict with the OEM flight deck design philosophy. This reduces transition errors, increases consistency across fleets, improves transitioning pilot performance, and allows for standardized assessment of potential safety issues.

Any flight automation maneuver or procedure introduced in initial, recurrent, or special training that requires motor skills or complex sequenced actions must be trained in full motion simulator with enough repetition to promote retention and provide the opportunity to demonstrate proficiency.

Specific benefits can be achieved with the consideration and application of human factors to automation in aviation. Current operations will become safer by trapping and eliminating system design flaws. Future operations will be able to meet demands such as increased capacity and efficiency while increasing the existing level of safety.

Acknowledgements

We extend our thanks to many who helped with this project including Nancy Law (ALPA), Brian Townsend (ALPA), Linda Orlady (ALPA), Rip Torn (ALPA), Kent Lewis (ALPA), Rob Hill (ALPA), Charles Billings (Professor Emeritus - Ohio State University), Robert Mauro (University of Oregon), Chris Wickens (Professor Emeritus - University of Illinois), Peter Polson (Professor Emeritus- University of Colorado), and Asaf Degani (NASA Ames).

References

Abbott, K, Slotte, S, Stimson, D (Co-chairs of the Federal Aviation Administration Human Factors Team) (1996). The Interfaces Between Flightcrews and Modern Flight Deck Systems, Federal Aviation Administration

Billings, Aviation Automation: *The Search for a Human Centered Approach*, 1997, Lawrence Erlbaum Associates, Mahwah, New Jersey)

Blackmon, M. H. (2004). Cognitive walkthrough. In W. S. Bainbridge (Ed.), *Berkshire Encyclopedia of Human–Computer Interaction* (Vol. 1, pp. 104–197). Great Barrington, MA: Berkshire Publishing.

Burian, B, Barshi, I & Dismukes, K. (2005) Center we have a problem: Emergency and abnormal situations in aviation, Presentation given at the 13th International Symposium on Aviation Psychology, Oklahoma City, Oklahoma)

Degani, A. (2004). *Taming HAL: Designing interfaces beyond 2001*. New York: Palgrave-MacMillan. Degani, A., Heymann, M.

Dekker, S., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology, and Work, 6*, 79-86.

Dekker, S. (2006). *Field guide to understanding human error.* Hampshire, England: Ashgate Publishing Limited.

Dixon, S., Wickens, C.D., & McCarley, J.M. (2007) On the independence of reliance and compliance: are false alarms worse than misses. *Human Factors, 49*, 564-572.

Fennell, K, Sherry, L, Roberts, R. Feary, M. (2006) Difficult Access: The Impact of Recall Steps on Flight Management System Errors, *International Journal of Aviation Psychology*, 16(2).

Funk, K., Lyall, B., & Riley, V. (1996). *Perceived human factors problems of flight deck automation,* Final Tech. Rep. prepared for the FAA. Washington, DC: Federal Aviation Administration)

Funk, K., Lyall, B (2000). A Comparative Analysis of Flightdecks With Varying Levels of Automation, Federal Aviation Administration Grant 93-G-039, Final Report.

Hellenic Republic Ministry of Transport and Communication, Air Accident Investigation & Aviation Safety Board (AAIASB) Aircraft Accident Report Helios Airways Flight HCY 522, Boeing 737-31S, at Grammatiko, Hellas on 14 August 2005, 11/2006

Heymann, M., & Degani, A. (2007). Formal analysis and automatic generation of user interfaces: Approach, methodology, and an algorithm. *Human Factors.* 49.

Heymann, M, Degani, A, Barshi, I, (2007). Generating Procedures and Recovery Sequences: A Formal Approach, Technion, Israel Institute of Technology, Haifa, Israel , NASA Ames Research Center, Moffett Field, California, USA

Hollnagel, E., Pruchnicki, S., Woltjer, R., & Etcher, S. (2008). A functional resonance accident analysis of Comair flight 5191. Proceedings of the 8th International Symposium of the Australian Aviation Psychology Association (AAvPA), Sydney, Australia.

Mauro, R. & Barshi, I. (2003). Training Smart: Using principles of cognitive science in aeronautical education and training. *Proceedings of the American Institute of Aeronautics & Astronautics.* Reno NV.

Parasuraman, R. Molloy, R. & Singh, I.L. (1993). Performance consequences of automation-induced “complacency”. *The International Journal of Aviation Psychology, 3*, 1-23.

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

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model of types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, 30*, 286-297.

Perrow, C. (1984). *Normal Accidents: Living with High-Risk Technologies*, New York: Basic Books.

Polson, P., & Smith, N. (1999). The cockpit cognitive walkthrough. Proceedings of the Tenth Symposium on Aviation Psychology. (pp. 427-432) Columbus OH: Ohio State University.]

Polson, P. S. Irving, J. Irving (1994) Applications of Formal methods of Human Computer Interaction to Training and Use of the Control And Display Unit. Tech Report 94-08, University of Colorado.

Pruchnicki, S., LaRoche, G., Corrie, S (2005). The Use of Color in Aircraft Flight Deck Displays, ALPA Statement of Position, Air Line Pilots Association

Sarter NB, Mumaw RJ, Wickens (2007) Pilots' monitoring strategies and performance on automated flight decks: an empirical study combining behavioral and eye-tracking data, CD, 1: *Hum Factors*, 49(3):347-57. Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI.

Shappell, S., Detwiler, C., Holcomb, K., Hackworth, C., Boquet, A. & Wiegmann, D. (2007). Human Error and Commercial Aviation Accidents: An Analysis Using the Human Factors Analysis and Classification System. Human Factors, 49, (227-242).

Sheridan, T. B. (2002). *Humans and automation: Systems design and research issues.* New York: Wiley.

Sherry, L., Feary, M., Polson, P., Mumaw, R., & Palmer, E. (2001). A Cognitive Engineering Analysis of the Vertical Navigation (VNAV) Function. NASA Technical Memorandum 210915.

Sherry L., Feary, M., Polson, P., & Fennell, K. (2003). Drinking from the fire hose: Why the flight management system can be hard to train and difficult to use. NASA Technical Memorandum, Moffet Field, CA

Sherry L., Fennell, K., Feary, M., & Polson, P. (2005). Analysis of Human-computer Interaction in Response to FMS Messages. Presented at the Thirteenth Symposium on Aviation Psychology. Oklahoma City, OK.

Sheridan, T, Nadler, E (2006). A Review of Human-Automation Interaction Failures and Lessons Learned, U.S. Department of Transportation Research and Innovative Technology Administration, Volpe National Transportation Systems Center, NASA Airspace Systems Program, Final Report, October 2006,

Wiegmann, DA , University of Illinois at Urbana-Champaign, Final Report, U.S. Department of Transportation, Federal Aviation Administration; Airplanes Statistical Summary of Commercial Jet Airplane Accidents Worldwide Operations1959 -2006, Boeing Commercial Air Group).

Wiener, E. L.Curry, R. E. (1980). Flight deck automation: Promises and problems. *Ergonomics, 23*, 995-1011.

Wiener, E. (1988). Cockpit automation. In E.L. Wiener and D.C. Nagel (Eds.), Human factors in aviation (pp. 433-461). San Diego: Academic.

Woods, D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R. Parasuraman & M. Mouloua (Eds). *Automation and human performance: theory and application. Human factors in transportation* (pp. 3-17) Mahway NJ: Lawrence Erlbaum Associates.

MULTI-UAV SUPERVISORY CONTROL INTERFACE TECHNOLOGY

Michael Patzek, Ph.D.
Doug Zimmer
Gregory Feitshans
Mark Draper, Ph.D.
Supervisory Control Interfaces Branch
711th Human Performance Wing
Air Force Research Laboratory

Tom Hughes
Infoscitex Corporation
Dayton, OH

John Flach
General Dynamics
Dayton, OH

The Air Force Research Laboratory's Supervisory Control Interfaces Branch (711HPW/RHCI) is conducting an advanced technology development program, entitled Multi-UAV Supervisory Control Interface Technology (MUSCIT). This program is focused on human systems integration; developing and integrating controls, displays, and decision support aids that enable a single operator control station to control multiple unmanned aerial vehicles (UAVs) in the performance of dynamic intelligence, surveillance, and reconnaissance (ISR) tasks as well as close air support (CAS) missions. This 5-year program, which began in 2007, employs a spiral development approach that consists of repeated analysis, design, development, virtual simulations, and flight tests to evaluate, refine, and mature advanced control station designs. The program will demonstrate effective human supervisory control and multi-UAV mission execution across a variety of mission situations and complexity and will identify key human factors challenges that must be overcome for fully enabled multi-UAV control by a single control station. This paper provides an overview of the MUSCIT program and details program goals, technology challenges, developmental approach, and expected products.

Unmanned aerial vehicles (UAVs) have proven effective in performing numerous military functions, serving as both strategic and tactical assets. Recent experience in both Iraq and Afghanistan highlight both the tremendous operational utility as well as the significant operational and technical challenges associated with fielding unmanned aerial systems. These challenges however do not come as a surprise to the research community charged with advancing the state-of-the-art in warfighter capabilities. In November 1996 the Air Force Scientific Advisory Board (SAB), under the direction of the Air Force Chief of Staff, published a report (Worch, Borky, Gabriel, Heiser, & Swalm, 1996) which concluded that "UAVs have significant potential to enhance the ability of the Air Force to project combat power in the air war". However, this report identified several human factors issues and challenges including addressing human-machine function allocation, establishing human performance data and criteria, and maintaining adequate crew situation awareness given unavailable sensory-perceptual cues, overconfidence, automation complacency, and/or boredom. The fact that the human operator will be removed from direct interaction with the air vehicle does not eliminate the human element from the system. In fact, such a concept arguably increases the complexity of the human-machine coordination issues.

In 2003 the SAB reiterated many of the same themes in a second study addressing the technology challenges associated with the development and deployment of UAVs to support current and future combat and ISR mission requirements (Johnson & O'Neil, 2003). The SAB noted that human-systems integration is not being adequately addressed in current system acquisitions or research programs and identified mission management "as the most significant technical challenge for future UAV systems". The SAB suggested numerous benefits that can be realized with mission management technology, not the least of which was a reduction in the operator to vehicle ratio required to effectively employ these systems.

A number of research and development efforts are focused on addressing these issues as well as other emerging needs that include 1) enhancing rapid response capability for performing ISR and close air support missions, 2) improving persistence capability for simultaneous coverage of multiple regions or areas of interest and 3) increasing the span of control with a single control station. Air Combat Command's Predator Multi-Aircraft Control (MAC) effort represents a recent attempt at single control station, multi-UAV operations using a modified Predator ground station under a fairly rigid mission concept of employment (Eggers & Draper, 2007). Within this concept a single pilot would control up to four vehicles while each sensor operator (up to four) managed and monitored a single video sensor feed. Though this concept works well for relatively stable missions (e.g., monitoring a fixed location) the demands on control quickly increases when one of the missions escalates toward a more dynamic task (e.g., tracking a moving target). In such cases, a second pilot is often summoned to control the remaining static missions while the first pilot manages the vehicle involved in the dynamic mission. If a second mission were to turn dynamic, crew workload saturation becomes a possibility.

The MAC concept illustrates that in a multi-vehicle control context, further progress is needed to increase mission flexibility and effectiveness on a per vehicle basis. To increase mission effectiveness, crew performance and capability enhancements are needed reduce the attention and workload demands on operators. Technology development and advanced designs are required to facilitate more timely and effective operator situation assessment, keeping operators "in-the-loop" and able to effectively direct the mission and provide highly accurate situation assessments and command decisions.

To address the above needs, AFRL's Supervisory Control Interfaces Branch (711HPW/RHCI) is conducting a 5-year advanced technology development program entitled Multi-UAV Supervisory Control Interface Technology (MUSCIT). The goal of the MUSCIT program is to investigate and develop technologies that will enable the flexible, highly effective control of multiple UAV assets from a single control station for the conduct of tactical ISR and CAS missions. A key aspect of MUSCIT is that it is focused not only on individual technologies, but the *integration* of those technologies into a coherent crewstation design. MUSCIT integrates new control/display technologies, new decision support aids, and novel multi-UAV architecture to maximize flexible, fault tolerant control of multiple tactical ISR UAVs for expanded missions. Candidate interface concepts, focused heavily on mission and sensor management, will then be prioritized in terms of demonstrated value under realistic mission simulations and flight tests. Expected payoffs include:

1. Reduced operator-to-vehicle ratio performing UAV ISR and CAS missions
2. Increased mission effectiveness (e.g., faster response time to time-critical events), flexibility
3. Increased operator effectiveness with manageable workload
    a. Better mission and system situation awareness for multi-UAV operations
    b. Decreased error in searching for and identifying targets and in switching between UAV control
4. Technology integration prototypes and guidelines
    a. Potential upgrades to existing systems
    b. Designs for new systems
    c. Candidate common control station components & procedures across UAV platforms
5. Reduced logistics footprint and system lifecycle costs

These expected payoffs provide some insight into the technical challenges that MUSCIT faces in developing an effective UAV supervisory control interface. For example, one challenge involves determining and supporting the appropriate levels and types of human-automation interaction for mission and sensor management across a variety of mission situations. In working this area, the MUSCIT team needs to be cognizant of human performance tendencies and issues associated with automation such as complacency, bias, vigilance decrement, mode confusion, loss of "knowledge of intent", cognitive overload, and attention / cognitive "tunneling". Another technical challenge is ensuring the operator interface is capable of providing necessary, timely information for maintaining situation awareness and effective decision-making across different situations/contexts. In other words, the interface should make it easy for the operator to acquire, assess, decide and implement actions. This may include a support system that locates, selects, and/or filters information based on the context to help streamline the information gathering and assessment process. Initial assessments have shown that there can be significant visual demands in performing target acquisition tasks. Therefore the MUSCIT team is investigating concepts to offload, or assist, the visual channel for both acquiring information and commanding actions in order to reduce the visual scan requirements and enhance overall sensory throughput.

MUSCIT Technical Approach

MUSCIT is investigating baseline and advanced technologies, and developing integrated crewstation designs, across different levels of multi-UAV mission complexity and crew composition, as illustrated in Figure 1. Two different advanced crewstation designs will be developed: a single operator design and a two-person crewstation. Spirals 1 and 2 will exercise advanced operator crewstations employing single and multiple UAVs within the context of relatively static and stable mission scenarios (e.g., point and area surveillance). Such missions typically require less time-critical tasking and dynamic mission re-planning. Spirals 3 and 4 will investigate challenges associated with increases in the degree of mission complexity reflected in more dynamic scenarios such as time sensitive targeting, dynamic target tracking, and CAS. Such missions are expected to place greater demands on both mission and sensor management tasks. Evaluating the advanced designs in this manner will aid in the capabilities assessment and refinement of promising supervisory control interface and control technology for a wide range of mission complexity.

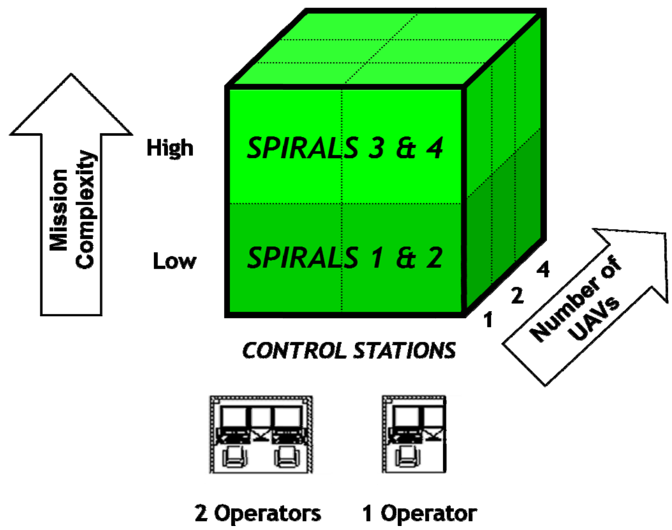


Figure 1. MUSCIT Approach

MUSCIT employs a spiral approach to develop the supervisory control operator interface and assess crew and mission-level performance. Each spiral begins with an analysis of existing and future missions and associated functions and tasks. With this knowledge, specific operator challenges are identified and candidate interface solutions researched. Part of this effort involves an analysis of where automation may aid the operator (within the constraints of current automation technology maturity). Vetted interface concepts are then integrated into unified control station designs, building on a foundation of AFRL's Vigilant Spirit multi-UAV control station testbed (see below). During this process, many informal usability assessments are conducted and refinements made, leading to a design that undergoes high-fidelity human-in-the-loop simulation testing and eventually limited flight tests. Subject matter experts from the participating commands will be sought to participate in the spiral simulation and flight tests. Flight tests will be conducted with surrogate UAVs to exercise the mission functionality and tasks. A cross-service control station working group serves to identify and critique mission scenarios, design attributes, and assessment methods.

MUSCIT Control Station

Serving as an initial design prototype for the MUSCIT program is the 711HPW/RHCI Vigilant Spirit multi-UAV control station (VSCS) framework. Over the past several years, VSCS has evolved in response to several programs that require an operator control interface to multiple UAV systems (Feitshans, Rowe, Davis, Holland & Berger, 2008). Through this evolution VSCS has emerged as a robust operator interface in support of simultaneous control of multiple UAVs as well as an effective test-bed for the investigation of issues associated with multi-UAV control. The MUSCIT program will evaluate and further develop the VSCS baseline over the course of the program.

VSCS has been developed upon an open architecture and facilitates the integration of new features and capabilities. Such features and capabilities are identified and/or developed based on analysis of the specific task demands associated with the operational mission defined for each developmental spiral. The current MUSCIT baseline configuration (see Figure 2) has four main components; vehicle status, tactical situation display (TSD), vehicle and payload management, and sensor exploitation. The vehicle status area allows the operator to maintain situation awareness of the UAVs the operator is controlling. The TSD allows the operator to maintain battlespace awareness. The vehicle payload and sensor management area allow the operator to control the aircraft and the payloads they are carrying. Finally, the sensor exploitation area allows the operator to view, manipulate and interpret the sensor feeds coming from the UAVs.

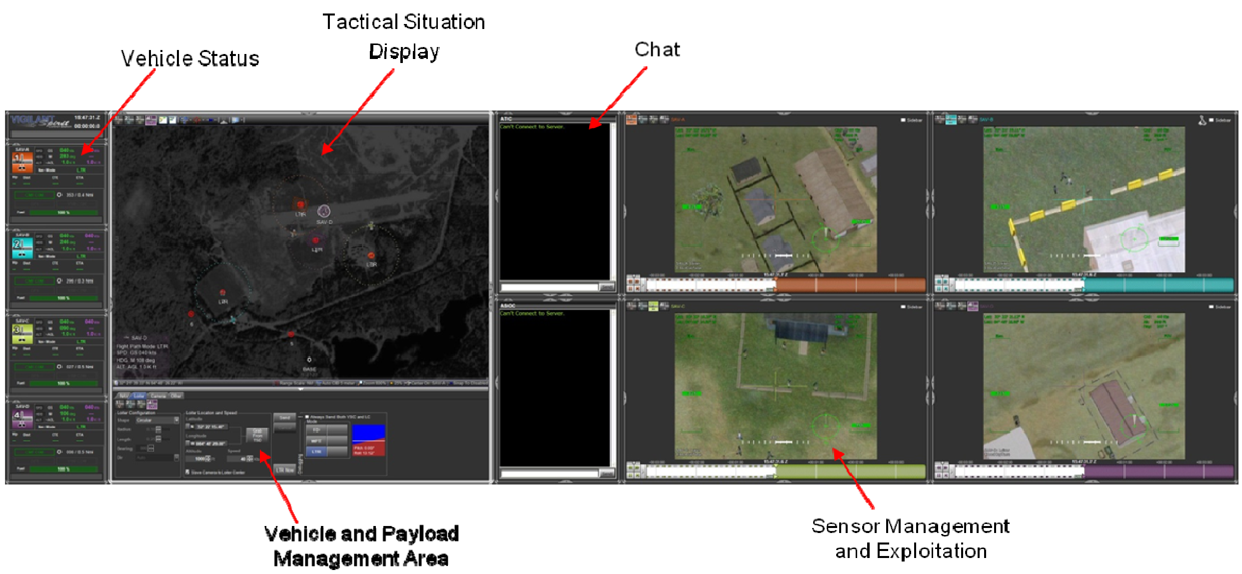


Figure 2. Sample layout of the control station.

Automation concepts, decision aids, control mechanization and associated operator interface concepts and information visualization techniques are integrated into the VSCS baseline based on insights gained during mission analyses, input from the user community, refinement of interface concepts derived for other applications, and lessons learned from simulation assessments. These concepts are then assessed during part task and full mission simulation and flight test assessments, providing both quantitative and qualitative measures of performance and effectiveness of the operational utility of the integrated control station.

MUSCIT Simulation Environment

In addition to the development and integration of control station technologies, the development of a realistic and robust simulated operational environment for the assessment of multi-UAV control poses several unique challenges to the MUSCIT program, not the least of which is providing an appropriate representation of the battlespace being sensed by onboard sensors. Within a Tactical ISR mission, UAVs are often assigned to observe, monitor and/or track ground entities (e.g., dismounts and/or vehicles) operating within a particular area of interest. In representing these ground entities, the simulation must capture not only the visual characteristics of these entities but details of their behavior as well. Simulations require imagery of sufficient detail to detect entities as well support the ability to identify distinguishing features or behaviors that characterized these entities as either a targets or non-targets. In short, to gain insight into the demands and challenges associated with multi-UAV concepts of operations it is important that one shapes the conditions of observation of simulations in such a way that preserves the essence, difficulty and complexity of the task to be accomplished (Woods and Hollnagel, 2006).

In creating these entity behaviors, the MUSCIT team must make tradeoffs between the desire to carefully control the environment to achieve the necessary repeatability across experimental trials and the level of interdependence of behaviors as entities react and adapt to evolving situations. In initial spirals where missions focus primarily on observation of static points of interest, the level of interdependence of behavior would be expected to be minimal. As such, detailed scripting of entities and entity behavior seems appropriate. In future spirals as mission focus more on direct contact of forces, such interdependence becomes more complex and behaviors more unpredictable. In such scenarios it may become necessary to implement agent models of individual entities that dynamically react to evolving situations in a manner that is appropriate to the anticipated motivations and characteristics of these entities. In some cases it may become necessary to enable third-party control of select entities to enhance both the realism of the simulation environment but also ensure the scenario is executed as necessary to achieve assessment objectives. This unpredictability will stress controlled experimentation efforts.

In creating its simulation environment, the MUSCIT program has developed a simulation architecture that includes the FLexible Analysis Modeling and Exercise System (FLAMES®) as a means of representing ground entities. FLAMES® is a family of computer software products that provides a framework for computer programs that simulate the physical and cognitive behavior of complex entities that act and interact in time and space. FLAMES® communicates to other components of the simulation architecture through a Distributed Interactive Simulation (DIS) interface. Entities within FLAMES® can either be scripted to run in a deterministic manner, adaptively controlled as computer generated agents, or be dynamically controlled by other third-party human controllers. For the visual representation of the battlespace, MUSCIT simulations employ the Virtual Reality Scene Generator™ (VRSG™) developed by MetaVR, Incorporated. VRSG™ is a real time computer image generator designed to visualize geographically expansive and detailed worlds on personal computers. The images generated are displayed in the sensor exploitation area on the control station.

To generate scenarios for virtual simulations, the scenarios are first created using FLAMES®. Individual entities are developed and their movements are scripted using FLAMES®. These scenarios are then saved and run for the trials. As FLAMES® runs during the trial the entity state information is passed via DIS (Distributed Interactive Simulation) packets to VRSG™ to be displayed as 3D models in the virtual scene. The result is a high fidelity 3D virtual world that contained entities whose movements are repeatable across sessions.

Flight Test Environment

To support upcoming flight test exercises, the MUSCIT program will utilize MLB Company Bat 3 UAVs equipped with Cloud Cap Technology, Inc. Piccolo II autopilots and TASE stabilized camera gimbals. The equipped Bat 3 platform (see Figures 3 & 4) has a 6 foot wingspan, contains a retractable sensor, and has nominal 5 hour flight duration. The Bat 3s will be used in flight tests to investigate issues with multi-UAV control and the operator interface unique to the flight test environment, verify results found during simulation tests, and help to inform the development of our future simulation environment to more accurately reflect the demands and constraints associated with UAV control in the field. As with the simulation environment, a significant challenge for flight test is creating an effective representation of the battlespace that captures the task demands associated with the mission being investigated. Creating and replicating significant and interesting surveillance events remains both a coordination and logistics challenge for the MUSCIT program.

Figure 3. MUSCIT's Bat 3 UAV.

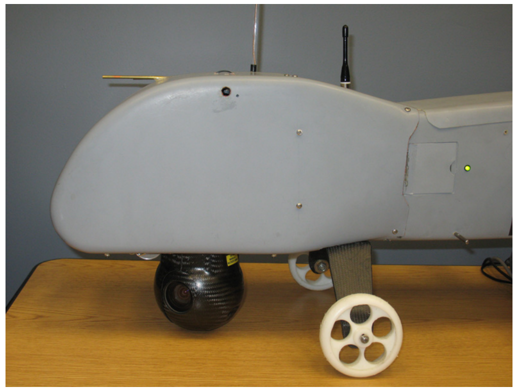


Figure 4. TASE gimbal deployed on BAT 3.

Summary

The MUSCIT program, building upon state-of-the-art UAV operator interface research, is pursuing a spiral approach to identify and validate integrated advanced control station technology for conducting multi-UAV ISR and CAS missions. Through repeated analysis, design, simulation and flight testing, the program develops and evaluates advanced operator interface concepts for single and multi-UAV supervisory control using mission and sensor management measures of performance as well as mission effectiveness measures across a variety of mission conditions. The potential payoffs from this effort include increased UAV span of control, increased mission effectiveness, improved cooperative UAV operations, and increased UAV control station commonality. In addition to the control station design prototypes that are produced for each spiral, the program will provide documentation on the details of the technologies and integrated designs along with the associated design rationale and prioritized human factors challenges that can be leveraged for existing and future UAV systems.

References

Eggers, J.W. and Draper, M.H (2007). Multi-UAV Control for Tactical Reconnaissance and Close Air Support Missions: Operator Perspectives and Design Challenges, NATO Report RTO-HMF-135.

Feitshans, G.L., Rowe, A.J., Davis, J.E., Holland, M. and Berger, L. (2008). Vigilant Spirit Control Station (VSCS): The Face of COUNTER. *AIAA Guidance, Navigation and Control Conference and Exhibit* (AIAA-2008-6309).

Johnson, R.O. and O'Neil, M.R. (2003). *Report on Unmanned Aerial Vehicles in Perspective: Effects, Capabilities and Technologies. Volume 1 Summary*, Washington DC: Air Force Scientific Advisory Board.

Woods, D.D. and Hollnagel, E. (2006). *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering.* Taylor & Francis, Boca Raton, FL.

Worch, P.R., Borky, J., Gabriel. R.F., Heiser, W. & Swaim, T.S. (1996). *Report on UAV Technologies and Combat Operations. Volume 1 Summary*, Washington DC: Air Force Scientific Advisory Board.

COGINITVE TASK ANALYSIS OF DISTRIBUTED NETWORK-CENTRIC INFORMATION FOR THE PROMOTION OF SHARED SITUATIONAL AWARENESS WITHIN COLLABORATIVE UAS OPERATIONS

Jeffrey A. Getzlaff
Kutta Technologies, Inc.
2075 W. Pinnacle Peak Rd.
Phoenix, AZ 85028

Allen J. Rowe
Thomas R. Carretta
Air Force Research Laboratory
Wright-Patterson AFB, OH 45433

A hybrid model of cognitive task analysis coupled with activity theory and team cognition was evaluated to determine human-computer interface (HCI) design factors that promote Shared Situational Awareness (SSA) within a collaborative unmanned aerial system (UAS). A computer testbed simulation was created for use with participants in a time-sensitive Intelligence, Surveillance, and Reconnaissance (ISR) and weapons engagement mission testing scenario. A cognitive analysis was performed which consisted of a Hierarchical Task Analysis (HTA), Applied Cognitive Task Analysis (ACTA), time-sensitive activity analysis, and coordinated team cognition. Results from testing indicated that the promotion of situational awareness (SA) was enabled by network-centric updates among users in a collaborative UAS. The major cognitive task determined was maintaining SA of the big picture while performing the mission task at hand. Recommendations include the automation of a region of interest for network-centric updates, active filters for decluttering, and the synchronization of entities portrayed on HCIs.

The utilization of the Global Information Grid (GIG) and the introduction of functional concepts such as Horizontal Fusion (HF), Enterprise Services (ES), and the Distributed Common Ground Control Station (DCGS) 10.2 will enable Network-Centric Warfare (NCW) in the 21st century. Additionally, the implementation of complex adaptive systems will assist in the fusion of ISR data from multiple collection platforms and enable multi-INTelligence (INT) data fusion products. The effect of publishing and consuming data from a Network-Centric Environment (NCE) by a UAS assists in the identification and tracking of targets or points of interest.

Figure 1. Collaboration of multiple UAS within a C2ISR COI from network-centric updates utilizing HCIs for shared situational awareness

The collaboration and synchronization of multiple heterogeneously located UAS Command and Control (C2) will enable optimum time on station and sensors on target for identification and persistent surveillance (DoD, 2007). One of the key components to these functional concepts is a NCE with HCIs for increased SA and collaborative decision making. To enable this effect networked team members must maintain a shared understanding of the battlefield as dynamic events occur without overloading their workload or cognitive process.

Figure 1 illustrates a Concept of Operations (CONOPS) of multiple collaborative UASs identifying and tracking a target for persistent ISR within a C2ISR Community of Interest (COI). Network-centric information updates within the COI from ISR data and multi-INT fused data products promote the creation of a Common Operational Picture (COP) among the HCIs of the networked users in the system. To realize the benefits of a NCE, a user processes individual and shared situational awareness (SSA) in their cognitive domain for knowledge building and situational understanding of the battlefield. However, the sheer magnitude and type of data that can be presented to a user at one time could potentially overwhelm the user's cognitive process adding to the "Fog of War." This paper presents a cognitive demands analysis methodology for the promotion of UAS SSA and testing results for HCI design considerations.

HCI Analysis Methodology

A review of various testing methods of cognitive demands, user inputs, time-sensitive performance, and system functionality was performed to determine an optimum yield of a hybrid HCI analysis methodology. The following are overviews of the determined high opportunity researched methodologies.

The HTA methodology is beneficial in determining the goals and inputs a user takes on a system. This system-centric approach lends itself well to Universal Modeling Language (UML) Use Case creation for requirement generation and for interface design analysis. However, the limitation of the narrow focus of the task and no high level view of the cognitive aspects on the user usually requires this methodology to be coupled with other analysis methods (Crystal & Ellington, 2004).

The next analysis methodology investigated was the ACTA. This analysis method is a streamlined version of the more robust Cognitive Task Analysis (CTA) and consists of three interview methods of test participants and/or subject matter experts (SMEs). The interviews are composed of a task diagram, knowledge audit, and simulation overview. The task diagram interview identifies the demanding cognitive elements of the task in relation to an overview performance of the task. The knowledge audit elicits probes of a user's experiences, prediction of events, situational awareness, and perception of the environment. The simulation interview enables visibility into the cognitive process of a user through a challenging simulation scenario. The ACTA methodology captures the cognitive elements of the participants and task skills required for judgment and decision making (Militello & Hutton, 1998). A cognitive demands table highlights the difficult cognitive elements from the three interview methods in relationship to system goals and functionality. Analysis of the table focuses on determining relationships which input into HCI design criteria recommendations. Overall, this methodology provides inputs to cognitive demands of a task. However, this methodology lacks the capability to represent the mental model of the participant in relationship to individual and shared situational awareness and team coordination and cognition.

The third analysis method investigated was activity theory. This methodology views the activity rather than the performance of individual tasks and can be conceptualized as a work process method. The activities performed are related to other activities to yield an effect. This methodology seemed promising in uncovering new behaviors and activities in relation to the time-sensitive Joint Targeting Cycle (JTC) and dynamic targeting model. Limited in scope and new in implementation, this methodology requires coupling with known existing task methodologies.

The last analysis method researched was team coordination and cognition in relationship to shared SA among team members. The Endsley model of situational awareness (Endsley, 2000) and the Office of Naval Research's (ONR) structural model of team collaboration were analyzed as a potential cognitive process models for team collaboration. Figure 2 illustrates the resultant hybrid cognitive process model for team collaboration. The individual and system level task factors are not represented in order to focus on the components of SA, collaboration, perception, communication, decision, actions, and the cognitive process.

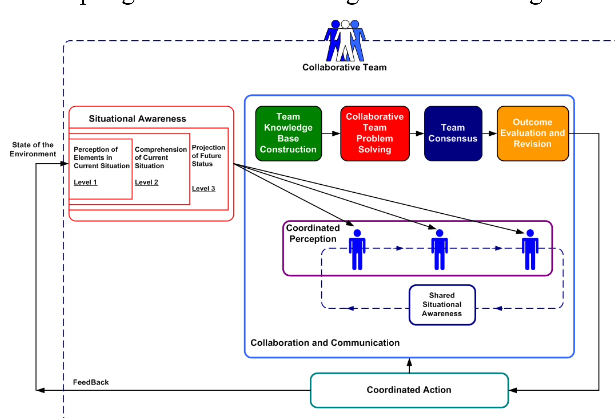


Figure 2. Hybrid model of team collaboration

During collaborative team problem solving, the team utilizes SSA, collaborative knowledge, and shared understanding to propose different Course of Actions (COAs). Individual team members use their own mental models and knowledge to assist in building collective team cognition. Within the team consensus state, team members negotiate to determine the best COA and utilize team shared understanding and collaborative knowledge from SSA. In the last structural stage, the perception of the team mission goal is evaluated in relation to the chosen COA. Measurement of the cognitive process of team members is enabled through the introduction of a roadblock transformation (Cooke, DeJoode, Pedersen, Gorman, Connor, & Kiekel, 2004) to normal operations to observe coordinated perception and action of team members.

Bonaceto and Burns' (2003) roadmap for cognitive engineering in system engineering was utilized in the creation of a hybrid analysis method from the above researched methodologies. The ranking of UAS C2 challenges of "smaller" organizations, "better" coordination, and "faster" execution to high opportunities for cognitive measurement methods was employed to create the resultant hybrid HCI analysis methodology model.

Figure 3 illustrates the resultant HCI cognitive task analysis methodology for determining design factors, levels of automation, and portrayal of information from network-centric updates. Within the Venn diagram is HTA for representation of the goal-oriented system view of tasks a user takes on the system. Additionally, ACTA aids in determining the cognitive elements of a user employing the system (e.g., decision making and judgments). A task diagram interview, knowledge audit, simulation interview, and cognitive demands table are performed for the ACTA. Activity theory takes into account the workflow process and relates to the time-sensitive targeting model (Office of the Joint Chiefs of Staff, 2007): Find, Fix, Track, Target, Engage, and Assess.

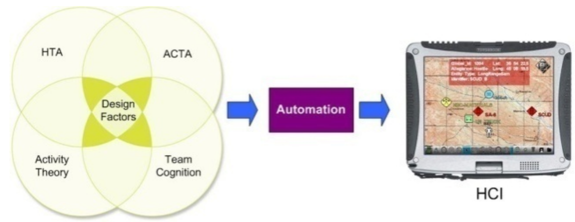


Figure 3. HCI cognitive task methodology for network-centric HCIs with automation

The researched hybrid model of team collaboration is utilized to determine the team cognition and coordination and the amount of SSA achieved. A roadblock transformation within a simulation scenario is presented to observe and measure the coordinated team efforts and shared SA. All these analysis methods are related to the information, cognitive, and physical domain to create design criteria for an HCI with network-centric updates. Also determined from the analysis is the level of software automation required to account for workload and projection of future status.

HCI Simulation Testbed

Because access to actual USAF UAS operations is limited to research, a simulation testbed was created to initially test the HCI analysis methodology and to serve as the simulation for the ACTA. The created testbed is a modification of a Phase II Small Business Innovative Research (SBIR) Distributed UAV Access System that was integrated with the Vigilant Spirit Control Station (VSCS) lab at the 711 HPW/RHCI at Wright-Patterson Air Force Base. It should be noted that the created cognitive task model, specifically ACTA, can be utilized to contrast expert and novice participant groups by conduction testing using the same simulation. Therefore, a second sample group of Predator Operations Center (POC) personnel is planned to be performed and contrasted to the initial sample group results presented in this paper.

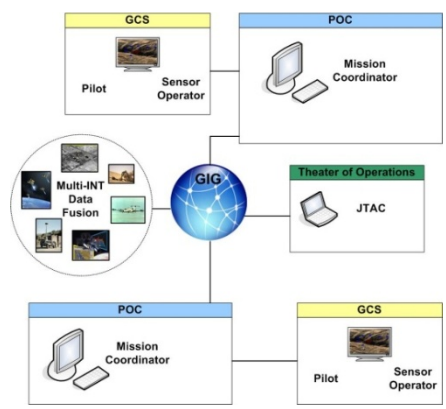


Figure 4. Computer simulation testbed

Illustrated in Figure 4 is the created simulation testbed with emulated components of a POC. Connected to the simulation are a Joint Terminal Attack Controller (JTAC) and simulated network-centric data updates from multi-INT data fusion. Contained within a POC are a Mission Coordinator (MC), Senior MC (SMC) and Mission Commander (MCC). For testing purposes the MC role was selected to analyze due to the tasks of mission planning, coordination of imagery collection, threat detection, and communication with the personnel within the Ground Control Station (GCS). Within the GCS the pilot and sensor operator share a Tactical Situational Display (TSD) for updates of the battlefield and promotion of a COP.

Testing Scenario

The sample participant group consisted of commercial airline pilots, a small UAV pilot, RC pilot, FAA DER, and engineers with a background in UAV CONOPS. The participants were asked to play the roles of a MC and a sensor operator or pilot in the GCS. Previous research of a POC task overview (Drury & Darling, 2007) has shown that the high level task of targeting has the most cognitive demands on a user. The research performed concentrated on tasks in relationship to team collaboration for ISR and target engagement. The knowledge audit consisted of participants utilizing a 2D/3D HCI displaying threats and friendlies and interview probes in relation to the promotion of a COP. The participants were allowed to utilize the HCI for a set time then asked to recall from memory the battlefield environment and relate it to a collaborative ISR or weapons engagement UAS mission.

A human-in-the-loop simulation was performed for a time sensitive scenario. This simulation was utilized to probe the participant's cognition and decisions relating to the hybrid model of cognitive tasks. From a previous Situational Awareness Global Assessment Technique (SAGAT) with a computer testbed simulation it was determined that freezing the simulation to probe for questions was a hindrance to the overall simulation tempo. Therefore, for the cognitive task simulation participants were actively engaged and challenged for questions probing their knowledge with a textual dialog for input of their answers.

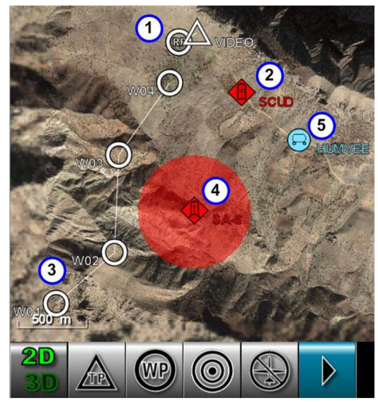


Figure 5. Simulation scenario HCI

The simulation scenario consisted of five main events with interview questions probing the participant's cognition after the occurrence of the incident in the simulation (see Figure 5). The first event consisted of an imagery request of video along a mountain road (1). The second event was the discovery of a SCUD launcher threat from the video and posting of the entity data to the GIG. This update was displayed in the HCIs of the participants with an audible cue (2). The first UAS maintained persistent surveillance and tracking of the target while the second UAS created a mission route for target engagement (3). During ingress to the target, a threat of a SA-6 from multi-INT data fusion was posted into the system and displayed on all participant's HCIs within the Collaborative Unit (CU) (4). Finally, after a modified mission route was created avoiding the SA-6 and the UAV was enroute to the SCUD target a friendly force was posted into the system and displayed in close proximity to the target of interest (5). In addition to the interview questions, team coordination and collaboration was observed during the simulation events.

Testing Results

The task overview interview resulted in five steps in relationship to the performance of a UAS CU: entering the group, status and location, communication within and out of the CU, joint operational roles, and notification to exit the group. The most cognitively demanding steps were the joint operations of surveillance and weapons engagement while maintaining situational awareness of the big picture and location of other UASs.

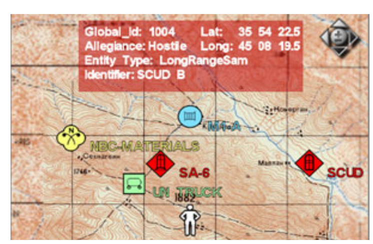


Figure 6. Knowledge audit TSD

The Endsley SA model was coupled with the ACTA components of the big picture, job smarts, and self monitoring in the analysis of the participants performance of the knowledge audit. Figure 6 represents the 2D view of the Tactical Situation Display (TSD) HCI utilized for testing. From analysis of the results, participants utilized roads from topographic features for recall and spatial relationship of entities to form a mental picture. Also, satellite imagery and a 3D digital elevation model assisted in the perception of entities in current environment. The comprehension of the current situation in relation to an ISR or weapons engagement mission highlighted the need for entity positional updates and indicators of last direction traveled. Some participants perceived the UN truck was in danger while others thought the friendly M1 tank was moving to strike the SA-6. Additionally, to assist in the comprehension of the current situation route traces of UASs within the CU were utilized. In order to maintain a COP a task that was identified as important was the comparison of video to the TSD. To optimize the performance of tasks between members of a CU the entities display on the TSD should be synched, thus enabling a COP among the HCI of the users. Participants projected that ISR or a weapons engagement in the area should take into account the SA-6 in the close proximity to the SCUD launcher with friendlies and neutrals in the area. Active filters also were employed on the TSD to filter friendlies and threats on the battlefield to assist in decluttering the display. Testing results indicated that the automation of entities displayed to a dynamic region of influence based on the UAV's position would assist in promoting a COP.

For each event in the simulation interview, participants were asked questions to query their judgment, decision making, and SA. These questions consisted of assessment of the current situation, items which led to actions, error a person could potentially make, future projection of the battlefield, and next actions to perform. Results from the probing of the participants during the simulation are listed in Table 1.

Table 1**.** *Simulation Summary*

Event	Actions	Assessment	Critical Cues	Potential Errors
(1) Imagery Request	Approve the imagery request based on security and priority.	No threats in area of imagery request route.	Situational awareness display of entities and terrain.	Does not know availability of UAS in CU for tasking.
(2) SA of Scud Launcher	Post NC update of location of target and communicate status.	Analysis of video stream for status of entity.	Current operational state of target (i.e., moving, preparing to launch, or abandoned).	Missing identification of target in video.
(3) SCUD update and route request	Creation of mission route for target engagement.	Location of threat, communication with JTAC.	Location of SCUD in HCI. Location of other UAS flight patterns.	Does not know terrain feature in area or location of other UASs in CU.
(4) SA-6 Threat	Change route continue communication with JTAC, CU, and higher command.	Comparison of threat location to mission route.	Location of threat zone to UAV mission route.	Creating a route that violates the threat zone of the SA-6.
(5) Friendly Update	Report friendly to JTAC and personnel in CU. Establish communication with friendly force.	Vicinity of friendly to SCUD.	Direction of friendly travel on HCI. Current distance from target.	Friendly in weapon engagement area.

Table 2 illustrates a cognitive demands table for HCIs among a collaborative group of UASs based on the testing results. This table relates the cognitive elements to difficulties, HCI cues and strategies, and common errors.

Table 2. *Cognitive Demands*

Cognitive Element	Why difficult	Cues and Strategies	Common Errors
Maintaining a COP	Dynamically changing battlefield with multiple threats and friendlies.	NC updates and communication between UAS CU.	Unaware of battlefield entities from the performance of the task at hand.
Projection of future status of battlefield	Require knowledge of narrow focus picture in relation to larger view.	Updates of the status of entities.	Not having the current state of the entity.
UAS coordination and collaboration	Multiple skill levels and members in CU. Some personnel are only told on a need to know basis.	Tone of dialog in communication of team members.	Incorrect data due to relaying of information.

Discussion and Conclusions

One of the most cognitively challenge tasks determined was maintaining situational awareness of the big picture and determining how it related to the task at hand (e.g., planning a mission route, tracking a target). Analysis of the testing results in relation to the time-sensitive targeting model identified the activity of communication as a key component in reducing the cycle time of target detection to engagement. Specifically, the automation of communication between the pilot, MC, and JTAC for weapons engagements based on rules of engagement (ROE). Maintaining a COP between CU team members was enabled through network-centric updates to their respective HCIs. During the dynamic events of the SA-6 and friendly force update within the test simulation the coordination among team members was observed. Key elements determined were the ability to communicate among the team members, share information, and collaboratively come to a team consensus of the COA to take. One of the enablers of team collaboration and decision making was the positional display of an entity on the HCI with a unique identifier (e.g., Global Unique ID) among team members.

Recommendations

Based on the analysis of the testing results and participant feedback, a supervisory HCI for use with a UAS CU within a COI is recommended. This HCI should employ an automated smart pull of data from the GIG from the UAVs region of interest. Additionally, it is recommended that the HCI contain automated communication links to members within the CU and JTAC, automated and manual declutter filters, and the ability to send data and display received data from a user's HCI. These NC HCIs could be utilized by a MC or functional components created and incorporated with legacy HCIs (e.g., FalconView). The realized effect of the utilization of collaborative UAS operations with cognitively developed HCIs is a robustly-networked Air Force performing information sharing and decision making at an increased tempo for accomplishment of mission goals.

Acknowledgements

This technical paper was performed under a Phase II Small Business Innovative Research grant "Network-Centric Communications: Distributed Real-Time UAV Access," Contract FA8650-06-C-6620.

References

Bonaceto, C. A., & Burns, K. J. (2003). *A roadmap for cognitive engineering in systems engineering,* The MITRE Corporation.

Cooke, N. J., DeJoode, J. A., Pedersen, H. K., Gorman, J. C., Connor, O. O., & Kiekel, P. A. (2004). *Coordinated Awareness of Situation by Teams: Measuring Team Situational Awareness of a Communication Glitch*. Technical Report, Cognitive Engineering Research Institute.

Crystal, A., & Ellington B. (August, 2004). Task analysis and human-computer interaction: approaches, techniques, and levels of analysis, *Proceedings of the Tenth Americas Conference on Information Systems*, NY.

DoD (2007). 'Persistent ISR Planning and Direction.' *Joint Integrating Concept Version 1.0*, DoD, 29 March 2007.

Drury, J. L., & Darling E. (2007). *A "thin-slicing" Approach to Understanding Cognitive Challenges in Complex Command and Control*, The MITRE Corporation.

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* (pp. 3-32). Mahwah, NJ: Lawrence Erlbaum Associates.

Militello, L. G., & Hutton, R. J. B. (1998). Applied Cognitive Task Analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 41(11),* 1618 -1641.

Office of the Joint Chiefs of Staff (2007). 'Joint Targeting - Pub 3-60,' *Joint Chiefs of Staff*, April 2007, pp. II-1 – II9.

MANUAL AND COOPERATIVE CONTROL MISSION MANAGEMENT METHODS FOR WIDE AREA SEARCH MUNITIONS

Thomas R. Carretta
Lamar Warfield
Michael J. Patzek
Air Force Research Laboratory
Wright-Patterson AFB, OH USA

Wide Area Search Munitions (WASM) combine the attributes of unmanned aerial vehicles with those of traditional munitions. The WASM concept envisions artificially intelligent munitions that communicate and coordinate with one another and with human operators to effectively perform their tasks. This study examined target acquisition for unaided operators with that of an automated cooperative controller for a complex task involving the prosecution of ground-based targets. Participants completed nine trials for each control mode (manual and cooperative) by number of WASMs (4, 8, and 16) combination. Target hit rate was not affected by control mode or number of WASMs. However, target acquisition efficiency degraded under manual control and as the number of WASMs increased. Workload was greater for the manual mode and increased as number of targets increased. Self-ratings of the ability to perform a simultaneous attach were lower for the manual mode and decreased as the number of WASMs increased.

Future unmanned aerial systems are expected to be more autonomous than those that are currently operational. In these systems, a single operator may be expected to monitor and exert executive control over several unmanned systems (Barbato, 2000; Clough, 2002; Prieditis, Dalal, Arcilla, Groel, Van Der Bock, & Kong, 2004). The US Air Force is considering advanced system concepts that could deploy multiple semi-autonomous unmanned weapons systems into the battle zone. One such system, the Wide Area Search Munitions (WASM), is a hybrid that combines the attributes of an unmanned aerial vehicle (UAV) with those of traditional munitions. The WASM concept envisions artificially intelligent munitions that communicate and coordinate with one another and with human operators to perform their tasks more effectively. WASMs can be deployed individually or in groups from larger aircraft and are capable of searching for, identifying, and attacking targets. Cooperative control concepts have been proposed to enhance coordination among these systems leading to optimal resource allocation (Goraydin, 2003; Scerri, Liao, Lai, Sycara, Xu, & Lewis, 2004; Schumacher, Chandler, & Rasmussen, 2002; Schumacher, Chandler, Rasmussen, & Walker, 2003). Research into strategies for controlling them presents a challenging problem that is being addressed by simulating WASMs as accurately as possible and evaluating them in human-in-the-loop (HITL) simulations and concept of employment scenarios. The Low Cost Autonomous Attack System (LOCAAS) was the first generation of such search munitions and served as the basis for the WASM testbed used to conduct HITL simulations.

Researchers have applied teamwork theory to build large teams that can accomplish complex goals using completely distributed intelligence. Algorithms have been developed to evaluate the ability to simultaneously deploy 200 WASMs to search and destroy ground-based targets in a coordinated support role with manned aircraft (Scerri et al., 2004).

The objective of this study was to examine target acquisition performance for unaided human operators with that of an automated cooperative controller in accomplishing a complex task involving the prosecution of ground based targets with WASMs. This purpose of the study was to provide empirical data on an operator's ability to simultaneously manage multiple WASMs while performing a target search, identification, and weapon assignment task. This information will provide valuable insights into concepts of employment and technology requirements for future munitions and semi-autonomous systems (e.g., how much automation is acceptable, information requirements, need for decision aiding software, manpower and personnel qualification requirements).

Method

Participants

Twelve full-time civilian and military employees stationed at Wright-Patterson AFB OH participated in this study. The sample consisted of 12 men who ranged in age from 20 to 45 years with a mean of 30.3 years. All participants reported being in good to excellent health and having vision correctable to 20/20, normal color vision, and normal peripheral vision. Most participants indicated that they had prior simulator (67%) and video game (92%) experience. Participation was voluntary and no compensation was offered in exchange for participation in this study.

Measures

Task performance and questionnaire data were collected.

Task performance measures. Several objective measures of target acquisition performance were collected. The *Number of High Priority Targets Attacked* and *Number of False Alarms* are self-explanatory. *Mean Time on Target* is the average of the actual time on target for the WASMs. *Mean Time on Target Error* is the average error between the actual time on target and requested time on target; that is, how close the attacks were to the requested time. This score could be computed only for the cooperative control condition. *SD of Time on Target* is the standard deviation of the actual time on target compared with mean time on target (i.e., how close the attacks were to each other). *Time to Plan* is the time from when the first target was selected to attack authorization or cancellation. *Time to Complete* is the time from authorization to when the last target was attacked.

Questionnaires. The questionnaires were a demographic data/background questionnaire, confidence ratings, the National Aeronautics and Space Administration Task Load Index (NASA-TLX; Hart & Staveland, 1988), and a post-test questionnaire that elicited a self-assessment of the ability to perform a near simultaneous attack under the manual and cooperative control conditions and comments regarding the operator interface.

Figure 1. WASM experimental station.

Equipment

Figure 1 shows a test participant interacting with the experimental station. Participants were seated directly in front of a 13.3 inch CF-73 Panasonic laptop computer that presented the simulated WASMs attacking targets on a Falcon View map. Still images of potential targets were displayed on a poster next to the laptop to aid the participants during target acquisition. Participants used a mouse with a scroll wheel to designate targets and make weapon assignments. A second laptop computer was placed nearby where participants entered questionnaire responses.

Procedures

The study began with a pre-briefing, informed consent, and the biographical questionnaire. The pre-briefing provided information regarding the purpose of the study, equipment, controls, and displays to be used, procedures, and the mission scenario. Following the pre-briefing, training was conducted to achieve familiarity with test equipment, procedures, and tasks. Participants completed three practice trials for each control mode (manual vs. cooperative control) by number of WASMs (4, 8, or 16) combination using a representative target set. Prior to starting the test trials, participants were fitted with electrodes to measure electrical brain, eye, and heart activity[1]. There were nine test trials for each control mode by number of WASMs combination. Immediately following each test trial, participants rated the level of confidence in their target acquisition decisions and subjective workload. After conclusion of the final test session, participants completed the post-test questionnaire regarding their experience.

Analyses

Analyses compared the objective and subjective data on the target acquisition task for manual versus cooperative control over three levels of mission complexity (4, 8, 16 WASMs). Related samples t-tests and repeated measures analyses of variance were performed since participants were exposed to all control mode by number of WASMs combinations.

Objective measures of performance included number of hits, number of false alarms, and target acquisition efficiency scores. Subjective measures were overall workload, confidence in target acquisition decisions, and self-assessment of the ability to accomplish near simultaneous attack. It was assumed that task difficulty would increase going from cooperative control mode to manual control mode and as the number of WASMs increased from 4 to 8 to 16. As a result, all analyses were performed using a .05 Type I error rate and a directional hypothesis.

Results

Target Acquisition Performance

Number of hits and false alarms. It was expected that performance under the cooperative control mode would equal or exceed that under the manual mode. Contrary to expectations, the number of high priority targets attacked was not affected by control mode. Although we intended to examine number of false alarms, we were unable to because the rate was extremely low with only 2 false alarms across all participants.

Time on target, time to plan, and time to complete measures. Means and standard deviations for the time on target, time to plan, and time to complete measures are presented in Table 1. It should be noted that mean time on target error (i.e., average error between the actual time on target and requested time on target) cannot be computed for the manual control mode because a requested time on target cannot be specified in manual mode.

No statistically significant effects were observed for *Mean Time on Target* for control mode, number of WASMs, or their interaction. *Mean Time on Target Error* (i.e., how close the attacks were to the requested time) generally increased as the number of WASMs/targets increased ($F(2, 10) = 6.96$, $p < .05$). The low value for the 8 WASM condition may have occurred due to the closer placement of targets in this condition relative to the 4 WASM/targets condition.

SD Time on Target Error (i.e., how close the attacks were to each other) was affected significantly by level of control ($F(1, 11) = 40.69$, $p < .01$), number of WASMs/targets ($F(2, 10) = 49.63$, $p < .05$), and their interaction ($F(2, 10) = 11.30$, $p < .01$). An examination of the means in Table 1 showed that time between attacks was greater for the manual versus cooperative control mode and generally increased as the number of WASMs/targets increased.

---

[1] The physiological data had not been processed and analyzed in time for inclusion in this paper.

Significant effects were observed for both *Time to Plan* and *Time to Complete* for control mode and number of WASMs/targets. *Time to Plan* was greater for manual control ($F (1, 11) = 20.70$, $p < .01$) and increased as the number of WASMs/targets increased ($F (2, 10) = 19.76$, $p < .01$). *Time to Complete* was *less* for manual control ($F (1, 11) = 490.81$, $p < .01$) and increased as the number of WASMs/targets increased ($F (2, 10) = 6.89$, $p < .01$). At first, it appears counterintuitive that *Time to Complete* was lower for the manual versus the cooperative control mode. However, it should be noted that in the manual control mode, target authorization and attack occur separately for each WASM/target combination and once authorization has occurred, the WASM takes a direct flight path to the target. In the cooperative control mode the attack does not occur until all target/WASM combinations have been authorized and it is necessary for some WASMs to employ longer flight paths to enable simultaneous attack.

Table 1. *Means and Standard Deviations: Number of High Priority Hits, Time on Target, Time to Plan, and Time to Complete Scores.*

		Cooperative Control		Manual Control	
Score	N WASMs	Mean	SD	Mean	SD
N High Priority Hits	4	3.33	0.00	3.27	0.12
	8	6.66	0.00	6.55	0.38
	16	12.30	0.09	12.52	0.33
Mean Time on Target	4	494.00	83.88	573.84	327.90
	8	488.57	55.83	446.71	67.35
	16	540.15	75.55	552.56	288.37
Mean Time on Target Error	4	2.04	1.22	-------	-------
	8	1.30	0.53	-------	-------
	16	8.58	4.44	-------	-------
SD Time on Target Error	4	2.24	2.11	10.17	4.21
	8	1.45	1.44	17.58	7.16
	16	9.09	6.16	27.43	11.89
Time to Plan	4	22.47	4.00	39.40	15.66
	8	36.01	7.63	61.26	26.83
	16	70.16	13.71	105.24	51.05
Time to Complete	4	117.22	11.89	63.06	10.45
	8	124.63	7.49	65.64	5.43
	16	148.09	26.76	74.96	10.90

N = 12

Confidence Ratings in Target Acquisition Decisions

Examination of the mean confidence ratings indicated an overall high level of confidence, with a mean score across all level of control by number of WASM/targets conditions of 4.75 out of a possible 5. Although confidence ratings varied, they were in the “fairly confident” to “very confident” range for all level of control by number of WASMs/targets combinations, even for the manual control mode with 16 WASMs/targets, which had a mean of 4.33 out of a possible 5. Although there was a trend toward greater confidence for decisions made using the cooperative control mode, this trend was not statistically significant. It should be noted that the observed power for this test was low, suggesting that if a larger sample were tested the effect might reach statistical significance. Mean confidence level was related significantly to the number of WASMs/targets ($F(2, 10) = 9.52$, $p < .01$). An examination of the means showed a general trend toward lower confidence as the number of WASMs increased, especially for the manual control mode.

Subjective Workload

Subjective workload was measured using the NASA TLX. As previously discussed, the NASA TLX has 6 subscales (Mental, Physical, Temporal, Performance, Effort, and Frustration) that are combined to create an overall workload index. Examination of the means revealed a consistent trend toward increased workload going from the cooperative control mode to the manual control mode and from 4 to 8 to 16 WASMs. This trend was statistically significant for the Total workload score and for all of the NASA TLX scales except Physical workload. For Total workload, significant effects were obtained for control mode ($F(1, 11) = 32.06$, $p < .01$), number of WASMs/targets ($F(2, 10) = 13.16$, $p < .01$), and their interaction ($F(2, 10) = 8.09$, $p < .01$). Mean Total workload for the cooperative control mode was relatively low with values of 13.91, 15.37, and 21.20 respectively for 4, 8, and 16 WASMs/targets. Mean Total workload for the manual control mode was 28.81, 38.97, and 51.15 for 4, 8, and 16 WASMs/targets.

Post-Test Questionnaire

Following completion of the test trials, participants completed a post-study questionnaire regarding their experience. They rated ease with which they were able to use the operator interface to identify targets and their ability to classify the priority level of targets. Both ratings were on a five point scale: 1 – poor, 2 – fair, 3 – good, 4 – very good, and 5 – excellent. Although ratings for ease of use and ability to classify the target priority level varied, the mean ratings for both approached "very good." Ratings for ease of use ranged from 3 to 5 with a mean of 3.92; those for ability to classify the target priority level ranged from 2 to 5 with a mean of 3.83.

Participants then rated their ability to perform a simultaneous attack using the cooperative control and manual control modes for the 4 and 16 WASMs/targets conditions. Ratings were on a five point scale: 1 – poor, 2 – fair, 3 – good, 4 – very good, and 5 – excellent. There were significant effects for control mode ($F(1, 11) = 66.00$, $p < .01$), number of WASMs/targets ($F(1, 11) = 61.90$, $p < .01$), and their interaction ($F(1, 11) = 28.94$, $p < .01$). Inspection of the means showed a strong trend toward lower ratings of ability to perform a simultaneous attack for the manual control mode and for the 16 WASMs/targets condition. The means for the cooperative control mode were 4.83 and 3.83 for the 4 and 16 WASMs/targets. The means for the manual control mode were 4.17 and 1.50 for the 4 and 16 WASMs/targets.

Participants had the opportunity to provide open-ended comments regarding the WASM interface and procedures. Seven of the 12 participants made one or more comments. These focused on ways to improve the manual control mode and the interface design. Suggestions regarding the manual control mode included adding the ability to insert waypoints and timing points to improve simultaneous attack. Suggestions regarding the interface design focused on providing multiple data input options in addition to the mouse and using a larger screen or multiple screens.

Discussion

Participants were able to acquire and attack nearly all of the targets even under the most demanding condition, that is, manual control of 16 WASMs. As expected, unaided operators were not able to achieve simultaneous attack of the targets as efficiently as the cooperative controller. Time between attacks was greater for the manual versus cooperative control mode and generally increased as the number of WASMs/targets increased. The decrement in performance efficiency between the manual and cooperative control modes is important under the circumstance when it is crucial to limit the amount of time an adversary has to respond to a first attack. Even in the least demanding condition involving 4 WASMs/targets, participants' ability to manually perform a near simultaneous attack was degraded compared to the cooperative control mode. These results also are reflected in participants' self-assessments of workload and their ability to perform a near simultaneous attack.

Additional studies are needed to examine factors that may affect performance differences between the manual and cooperative control modes. For example, the extent to which targets are clustered (or dispersed) in the search area may affect the relative efficiency of the manual and cooperative control modes. Also, it would be informative to examine additional numbers of WASMs/targets (1, 2, 3, … n) to better determine performance differences between the manual and cooperative control modes.

References

Barbato, G. (2000). Uninhabited combat air vehicle controls and displays for suppression of enemy air defenses. *CSERIAC Gateway. 11 (1),* 1-4.

Clough, B. T. (2002). *UAV swarming? So what are those swarms, what are the implications, and how do we handle them?,* AFRL-VA-WP-TP-2002-308. Wright-Patterson AFB, OH: Air Force Research Laboratory, Air Vehicles Directorate, Control Sciences Division.

Goraydin, O. (2003). *Analysis of cooperative behavior for autonomous wide area search munitions.* Unpublished Master's thesis. Wright-Patterson AFB, IH: Air Force Institute of Technology.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA TLX (task load index): Results of empirical and theoretical research in human mental workload. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). New York: Elsevier.

Prieditis, A., Dalal, M., Arcilla, A., Groel, B., Van Der Bock, M, & Kong, R. (2004). *Smart Swarms: UAVs that think.* Lockheed Divisions, Inc., Davis, CA.

Scerri, P., Liao, E., Lai, J., Sycara, K., Xu, Y., & Lewis, M. (2004). Coordinating very large groups of wide area search munitions. In D. Grundel, R. Murphey, & P. M. Partdalos (Eds.), *Series on Computers and Operations Research, vol. 4, Theory and Algorithms for Cooperative Systems* (pp. 451-480). World Science Publishing Company.

Schumacher, C., Chandler, P. R., Rasmussen, S. J. (2002). *Task allocation for wide area search munitions,* AFRL-VA-WP-TR-2002-301. Wright-Patterson AFB, OH: Air Force Research Laboratory, Air Vehicles Directorate, Control Sciences Division.

Schumacher, C., Chandler, P. R., Rasmussen, S. J., & Walker, D. (2003). *Task allocation for wide area search munitions with variable path length,* AFRL-VA-WP-TR-2003-306. Wright-Patterson AFB, OH: Air Force Research Laboratory, Air Vehicles Directorate, Control Sciences Division.

TRAINING INTERVENTIONS TO REDUCE AIR FORCE PREDATOR MISHAPS

Robert Nullmeyer
Lt Col Robert Herz
711th Human Performance Wing
Mesa Arizona

Gregg Montijo
Crew Training International, Inc.
Memphis Tennessee

The use of unmanned aerial systems (UASs) is expanding rapidly. In military operations, this increased use has been accompanied by relatively high mishap rates compared with rates across more mature manned aircraft. These higher rates led to multiple high-level reviews of unmanned operations, but surprisingly little consensus emerged across reports regarding root causes. To help close this gap, Air Force Predator Class A mishap reports through FY 2008 were analyzed in detail. Mishap rates, counts, and causal factors appeared to shift systematically over time, with an increase in mishap reports citing shortfalls in several skill and knowledge areas in FY 2004-2006. Individual and team Predator training objectives were revisited at the end of 2006 and the content of crew resource management (CRM) training was refocused on improving these key operator skills. In FY 2007-2008, Predator Class A mishap attributed to operator error decreased despite increasing numbers of mishaps overall.

While early attempts to use UASs for military purposes can be traced back to World War II or before (Gambone, 2002), unmanned technology clearly entered the mainstream of combat operations during the recent conflicts in Kosovo, Afghanistan, and Iraq. In these conflicts, UASs were first used as Intelligence, Surveillance, and Reconnaissance/Target Acquisition assets, providing commanders with imagery intelligence, electronic intelligence, and streaming video. Resulting information could be used to monitor enemy movements and conduct battle damage assessment. The Predator system added a strike capability, and similar capabilities were not far behind in other Department of Defense (DoD) UAVs. Across all United States military services, the use of unmanned aerial systems (UASs) is expanding rapidly. Predators accounted for 4% of all Air Force flying hours in FY 2007 and 19% of all Class A mishaps (Air Force Safety Center, 2009). Predator flying hours (and Class A mishap counts) nearly doubled in FY 2008. Despite the rapid rise of flying hours to date, only about one half of requests for UAS surveillance can currently be met, with growth in flying hours being limited by the ability to train enough crews to meet the demand for battlefield surveillance. The Air Force now flies 27 round-the-clock Predator and Reaper orbits in the Central Command area of operation, involving about 450 pilots. Military leaders want 50 orbits to be flown by 2012, requiring 1,100 pilots (Hoffman, 2008). The Quadrennial Defense Review predicted that approximately 45% of the future long-range strike force will be unmanned (Office of the Secretary of Defense, 2006). Emerging roles for UAVs include homeland security (e.g., border patrol), long-duration law enforcement surveillance, and battlefield delivery of critical medical supplies (Bone and Balkcom, 2003).

The rapid rise in UAV employment was unfortunately accompanied by high mishap numbers across all military services. This, in turn, led to several senior reviews to understand the root causes. The Office of the Secretary of Defense published a report on UAV reliability in 2003) that looked at non-weather related mission aborts or cancellations. The "Achilles heels" of UAV platforms appeared to revolve around component quality, redundancy, and maintenance, and concluded that it was critical to improve UAV platforms in these areas because reliability affects affordability, availability, and acceptance. A Defense Science Board Study on Unmanned Aerial vehicles and Uninhabited combat Aerial Vehicles (2004) concluded that UAV programs have not yet expended the resources necessary to fix the root causes leading to mishaps, and that manned-aircraft-like reliability is achievable, but will require substantial additional investment. Tvaryanas, Thompson, and Constable (2005) conducted an in-depth review of UAV mishaps across the United States military services. They reported that, since the inception of the systems in the 1990s through the end of FY 2003, 334 mishaps per 100,000 flying hours had occurred with the Navy/Marine Corps Pioneer, 55 mishaps per 100,000 flying hours had occurred with the Army's Hunter system, and 32 mishaps per 100,000 flying hours occurred with the Air Force's Predator system. For comparison purposes, overall Air Force Class A mishap rates ($1 million damage or fatality) are typically in the low single digit range per 100,000 flying hours (O'Toole, Hughes, & Musselman, 2006).

A recent challenge from the Secretary of Defense to reduce the numbers of preventable mishaps by at least 75% (Rumsfeld, 2006) focused attention on Predator mishap frequencies which accounted for about 20% of all Air Force Class A AIB reports (at least $1 million damage or a fatality) in the past two years (FY 2006 and 2007). It should be noted that most manned aircraft are mature systems, while most UAV programs are relatively early in their life cycles, and mishap rates tend to improve with system maturation.

A consistent picture of the problem to be solved has not yet emerged for UAS mishap reduction. Even at basic levels such as the relative contributions of equipment failure versus human error, different analysts reached widely differing conclusions. The Office of the Secretary of Defense Reliability Study (2003) reported that human error represented 16% of all sources of Predator A (MQ-1) system failures and 2% of Predator B (MQ-9) mishaps and the Defense Science Board (2004) reported that 17% of UAS mishaps were attributable to human error. On the other hand, Tvaryanas and his colleagues (2005) reported that 68% of UAS mishaps involved causal human factors and Williams (2004) reported that 67% of Predator mishaps involve human factors. Some researchers looked at Class A (more than $1 million damage or a fatality), B (more than $200,000 damage, and C (more than $20,000 damage) mishaps (e.g., Tvaryanas, 2006), some considered Class A mishaps only (Williams, 2004), and others did not specify the scope of the mishaps analyzed.

Experience with previous efforts to reduce mishaps in manned aircraft dictates that successful interventions to improve reliability must be based on an accurate understanding of the root causes leading to failure. Several researchers recently documented differing root cause patterns across organizations and platforms. Helmreich, Wilhelm, Klinect, and Merritt (2001) studied threats to safety and the nature of errors in three domestic air carriers in the United States, and observed striking differences among these airlines regarding both threats to safety and operator errors despite obvious commonality with respect to mission and environment. Nullmeyer, Stella, Montijo, and Harden (2005) reported differing mishap root causes across Air Force manned aircraft types. Williams (2004) reported major deviations in root causes across UASs, and Tvaryanas, et al. reported significant differences among root causes depending on the service involved.

Based on rapidly increasing UAS operations in both military and civilian organizations, the emphasis from senior military leaders on reducing UAS mishaps, and the lack of consensus in the literature on causal factors, we felt that root cause analyses with known parameters were needed to assess the role that training interventions could play to reduce mishaps and increase capability for a given platform. Our focus in this paper is on root causes and other characteristics of Air Force Predator Class A mishaps. This focus was chosen in part because Class A mishap reports are more detailed than Class B or Class C mishap reports, and in part because Class A mishap counts have become a highly visible metric of safety and reliability. Based on the patterns that emerged from our analyses, training interventions are proposed to address the areas of greatest potential.

Nullmeyer, Herz, Montijo and Leonik (2007) analyzed findings from all Air Force Predator Class A mishaps that had occurred from the introduction of this system into the Air Force inventory in 1995 through the end of FY 2006 to identify training-related trends. Substantial changes were reported over time regarding annual mishap rates, annual mishap counts, and causal factors. Mishap rates across the past three years were consistently less than one half the combined rate across earlier years. Mishap *counts*, however, steadily increased, as did Predator flying hours. Early mishap reports typically cited mechanical problems and operator station design issues. From 2003 through 2006, 80% of mishaps cited causal human error factors. Equipment interface problems were still cited as causal or major contributing factors in almost half of these mishaps. More specifically, mishap reports from 2003-2006 often cited shortfalls in skill and knowledge (checklist error, task misprioritization, lack of training for task attempted, and inadequate system knowledge), situation awareness (channelized attention), and crew coordination.

Based on the findings of Nullmeyer, et al., crew resource management training was developed for both the Predator formal school and for continuing Crew Resource Management training that was given to mission qualified crews. The focus of the new courseware was having students understand the primary threats to safety in the Predator community and providing techniques to manage the types of operator error that were repeatedly cited in Predator mishap reports. The remainder of this paper updates previous findings regarding human factors trends in Predator mishaps, focusing on publicly accessible information.

Methods

The United States Air Force Judge Advocate General's office maintains an online repository of Accident Investigation Board (AIB) report summaries for Air Force Class A mishaps. This site (http://usaf.aib.law.af.mil) is publicly accessible, lists Class A mishaps by fiscal year across all platforms in the Air Force, and provides one page executive summaries of AIB reports as they are released. These summaries describe the mishap and discuss probable cause. Most, but not all Class A mishaps are analyzed by an AIB. The publicly accessible database accounted for over 90% of all Class A Predator mishaps.

The Air Force Safety Center generates mishap investigation reports for every Class A mishap and provides results at varying levels of granularity. The analyses reported here provided the structure for further analyses of information from three distinct Safety Center sources that were used to guide changes in mishap reduction training. Moving from general to specific, the first was statistical data from the Air Force Safety Center web site (http://afsafety.af.mil). These data include hours flown and numbers of Class A mishaps by fiscal year and by aircraft type. The second data source was safety investigation summaries. These provide a brief narrative of the mishap, and categorized the Predator Class A mishaps as being primarily logistics-, maintenance-, or operations-related. Summaries also provide descriptive data for each mishap such as phase of the mission and time of day in which the mishap occurred, and list conclusions and recommendations. The third Safety Center source used was discussions of human factors from the full mishap investigation reports. Safety Investigation Board (SIB) findings are formally documented as a section of the full mishap report. These findings were reviewed for descriptions of human factors causing or contributing to the mishap. In addition to the board findings, a separate Life Sciences Report is prepared by the Life Sciences Branch of the Air Force Safety Center. The Life Sciences Report provides a chronological mishap narrative and a discussion of every element cited in the human factors database. Interrelationships among the human factors may be addressed. AIB summary reports (http://usaf.aib.law.af.mil) were initially analyzed to generate descriptive trend data regarding mishap frequencies and the general nature of the mishaps (equipment failure or operator problems) over time. Flying hour data were obtained from the Air Force Safety Center web site (http://afsafety.af.mil). Data from both sources were combined to generate mishap rates.

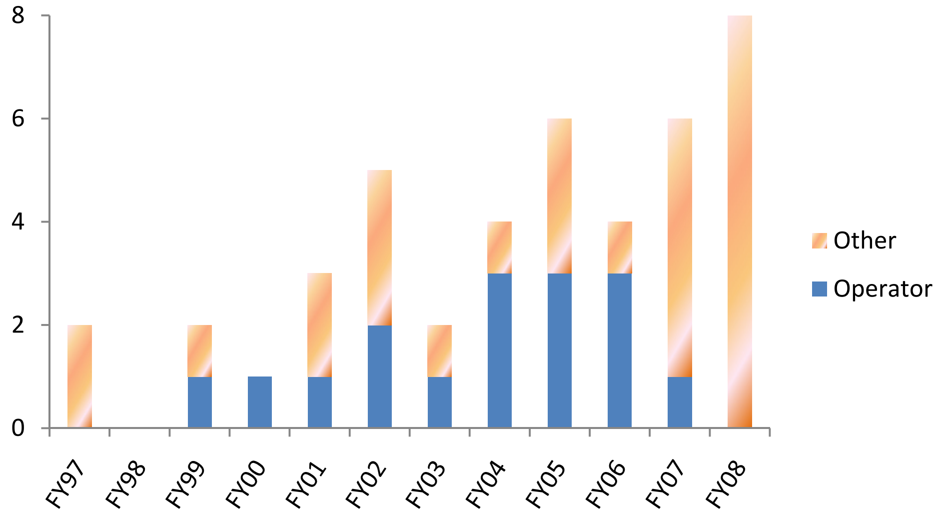
Four training-related problem areas emerged from analyses of the full and more detailed Safety Center mishap reports: (1) situation awareness development and maintenance, (2) task management, (3) decision making, and (4) crew coordination. Legacy Crew Resource Management (CRM) training for both initial qualification training students and recurrent training for qualified operators was refocused on these four areas and on Predator operations. This new training continued to be accomplished in a seminal format, but content and case studies were updated to emphasize Predator-specific threats to safety that are under the control of crews and strategies to mitigate the types of crew error that have led to Class A mishaps.

Results

Predator mishap frequencies per year increased systematically over the history of the system (http://usaf.aib.law.af.mil, 2009) as shown in Figure 1, with fiscal years accounting for over 65% of the variability observed in mishap frequencies (correlation = .81, $p < .001$). Causes in AIB reports are usually stated in terms of a single "clear and convincing" causal issue. We summarized each cause as being either operator error or some other factor, which was almost always some sort of equipment failure. Mishaps in the first five years of the Predator life cycle (FY 1997 - 2001) were most often attributed to equipment failure. Mishaps from each of the next five years (FY 2002– 2006) were attributed primarily to operator error. AIB reports of Class A mishaps in the most recent two years (FY 2007-2008) reflected two major shifts--a large increase in mishaps attributed to equipment problems accompanied by a consistent reduction of mishaps attributed to operator error.

Safety Center Class A reports provide more detailed descriptions that address multiple causal and contributing factors and also address the interactions among mishap factors. In the detailed Safety Center analyses, nine of 15 mishaps from FY 1997-2003 were attributed to equipment factors, and even four of the six operator-error mishaps cited causal equipment interface problems. In total, thirteen of the fifteen mishaps from 1997 through 2003 cited causal equipment factors. In FY 2003-2006, one mishap was attributed primarily to equipment failure and the remaining 14 were attributed to operator (12 mishaps) or maintainer (2 mishaps) error. Further, only three of the 12 mishaps attributed to operations cited equipment as a causal factor. Functional design continued to be cited as a contributing factor, however, in many of these recent mishaps.

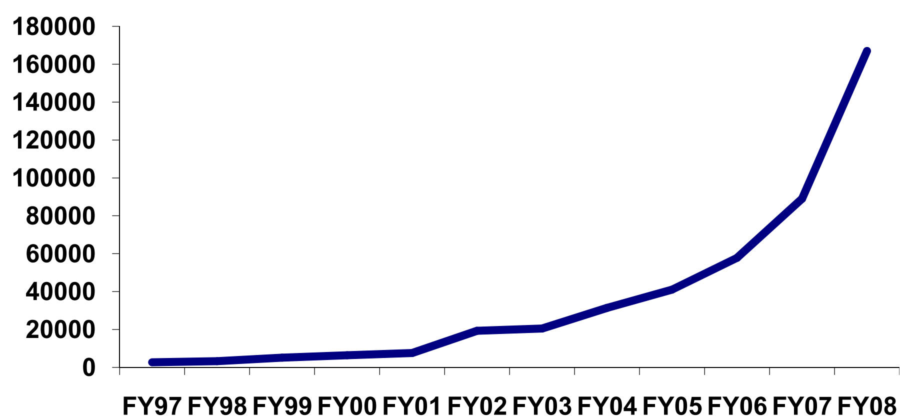
Figure 1: Predator Class A Mishap Frequencies by Fiscal Year (http://usaf.aib.af.mil)



Mishap frequencies before and after training modifications. Training focused on threats to Predator safety and strategies to identify error states and mitigate consequences of operator error was introduced at the end of FY 2006 and continued throughout FY 2007 and 2008. Twenty nine mishaps occurred between1997and 2006, averaging 2.9 per year. Fourteen additional mishaps occurred between 2007 and 2008, averaging seven per year. The AIB reports revealed a substantial drop in mishaps attributed to operator error following the implementation of threat and error training, and a reversal of probable cause from operator error through 2006 to other factors, usually equipment failure, since then (chi square = 7.61, df = 1, $p < .01$). A similar shift was seen in Safety Center mishap reports.

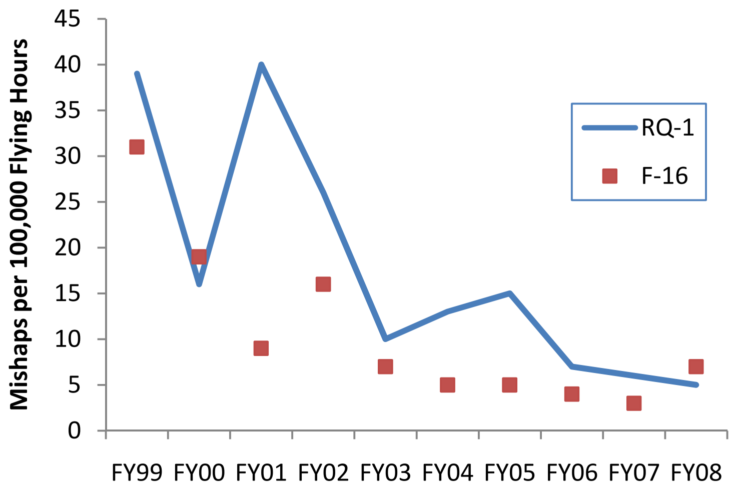
Predator flying hours are shown in Figure 2. The numbers of Predator mishaps clearly need to be interpreted in the context of the accelerating growth of Flying hours are reported on the Air Force Safety Center Web site (http://afsafety.af.mil). Annual flying hours increased from less than 3000 in FY 2000 to almost 80,000 in FY 2007, the latest year reported. Projections call for continuing increases in UAS operations. These changing utilization levels are important to consider when interpreting trends in mishap frequencies over time.

Figure 2. Predator Hours Flown Annually from 2000 – 2008



Mishap rates per 100,000 flying hours help take into account the rapidly growing utilization of Predators and provide a better metric of safety then frequencies. Figure 3 depicts both Predator and historic F-16 mishap rates starting with the first year in which more than 5000 hours were flown annually in either platform. With rapidly increasing operations, mishaps per 100,000 flying hours decreased substantially since the early 2000's despite increasing mishap counts. Comments from some of the early reviewers suggesting unusually high mishap rates in UASs were based on the early years, when mishap rates were high. Over time, Predator mishap rates are following a pattern that is very similar to the rates seen early in the history of the F-16 weapon system. Both systems encountered mechanical and human error problems early in their life cycles. F-16 mishaps are now very close to overall Air Force mishap rates. For F-16 mishaps, the data points represent the time period between 1977 and 1984. Predator Class A mishap rates were about 6 per 100,000 flying hours in 2007 and are projected to be close to 10 per 100,000 flying hours in FY 2008. For comparison purposes, the overall Air Force Class A mishap rate has been slightly less than two mishaps per 100,000 flying hours for the past decade.

Figure 3: MQ-1 (Predator) and F-16 Class A Mishap Rates (with Historic F-16 Rates FY 79-88)



Conclusions

Predator mishap trends reflected systematic and substantial changes over time. The overall direction of these trends depends on the measure used. Mishap frequencies steadily increased over time as have Predator hours flown. Mishap rates per 100,000 flying hours decreased substantially (from 23 Class A mishaps per 100,000 flying hours from fiscal years 1997-2003 to less than 10 in fiscal years 2004-2008). Despite the decrease, Predator mishap rates remain high relative to more mature Air Force weapon systems, but they are similar to the rates seen in the early years of F-16 operations and are dropping quickly.

In FY 2004-2006, a substantial increase was observed in mishaps that cited insufficient operator skills and knowledge. The threat and error management model (Helmreich, et al, 2001) is widely used by air carriers to enhance safety. We believe that it also provides a reasonable structure for improving UAV mishap rates in military operations and ultimately for increasing combat capability. A key part of this approach is to use evidence to structure interventions to alleviate the specific problems that actually plague a particular community. Training is one of several tools that can be used to meet safety and capability objectives, but other changes such as equipment modifications and altered procedures may also be integral parts of an effective overall error mitigation strategy. The bottom line is that the better we understand the real threats to safety, the more successful we are likely be in developing effective strategies to mitigate them. With the recent rise in equipment-related problems, it would be prudent to address the role of human error in equipment maintenance and if warranted, develop threat and error management training for maintainers. Following the effort to refocus CRM training to address known threats to safety, overall mishap rates continued to climb, but the causes cited in AIB reports shifted from human error to equipment failure. Similar patterns were apparent in the more detailed Safety Center mishap reports.

To substantially increase the numbers of Predator operators, the Air Force is currently evaluating alternatives to using experienced pilots to control Predator platforms. Two programs are underway, one looking at the ability of pilots who recently completed undergraduate pilot training, and another assessing the ability of non-pilot candidates to perform the tasks required of Predator pilots. One measure of merit is mishap rates. It is clear that mishap frequencies, rates and causes are all dynamic in the emerging field of UAS operations, and that mishap reports provide a fertile source of insight into where training and operations need to be improved. Our analyses suggest that raw frequencies could be misleading, especially in light of few operator error mishaps in the past two years. Instead, safety analyses for the purpose of assessing crew performance need to focus on mishaps where operator error is a factor.

RFEFERENCES

Air Force Judge Advocate General (2009). *Class A Aerospace Mishaps.* (http://usaf.aib.law.af.mil).

Air Force Safety Center (2009). *Aircraft Statistics*. (http://afsc.af.mil) .

Bone, E. & Bolkcom, C. (2003). *Unmanned Aerial Vehicles: Background and issues for Congress.* 25 April, 2003. www.fas.org/irp/crs/RL31872.pdf.

Defense Science Board (2004). Defense Science Board Study on Unmanned Aerial Vehicles and Uninhabited Combat Aerial Vehicles. Washington DC: Office of the Undersecretary of Defense for Acquisition, Technology, and Logistics.

Gambone, G. (2002). From U.S. civil war to Afghanistan: a shore history of UASs*. DefenseLINK News*, April, 2002, http://www.defenselink.mil/news/Apr2002/n04162002_200204163.html.

Hoffman, M. (2008). Hundreds of predator reaper, pilots needed. *Air Force Times,* October, 2008, http://www.airforcetimes.com/news/2008/09/airforce UAV_pilots_092908w/.

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.), *Applying resource management in organizations: A guide for professonals.* Hillsdale, NJ: Erlbaum.

Nullmeyer, R., Stella, D., Montijo, G.A., and Harden, S.W. (2005). Human factors in air force flight mishaps: Implications for change. *Proceedings of the Interservice/Industry Training, Simulationi, and Education Conference,* Orlando FL.

Nullmeyer, R.T., Herz, R.A., Montijo, G.A., & Leonik, R. (2007) Birds of Prey: Training solutions to human factors problems. *Proceedings of the 28th Interservice/Industry Training Systems and Education Conference.* Orlando, FL.

Office of the Secretary of Defense (2003). Unmanned Aerial Vehicle Reliability Study. February, 2003. www.acq.osd.mil/uas/docs/reliabilitystudy.pdf.

Office of the Secretary of Defense (2006). The Quadrenial Defense Review Report, February 6, 2006. http://www.globalsecurity.org/military/library/policy/dod/qdr-2006-report.pdf.

O'Toole, K.J., Hughes, T., and Musselman, B.T. (2006) USAF Aviation Safety: FY 2005 in Review. http://afsafety.af.mil/SEF/Downloads/FY05_Life_Sciences_Safety_Review(AsMA06).ppt

Rumsfeld, D. (2006) *Reducing Preventable Accidents.* Memorandum for Secretaries of the Military Departments. The Secretary of Defense, Washington DC.

Tvaryanas, A., Thompson, B.T., and Constable, S.H. (2005). U.S. military unmanned aerial vehicle mishaps: assessment of the role of human factors using HFACS. Brooks AFB, TX: 311 Human Systems Wing.

Williams, K.W. (2004). A summary of unmanned aircraft accident/incident data: human factors implications. Report number DOT/FAA/AM-04/24, Oklahoma City, OK: Civil Aerospace Medical Institute. Federal Aviation Administration.

AIR TRAFFIC CONTROL CREW RESOURCE MANAGEMENT: TO FIND TRUTH AND FACILITATE CHANGE

Skip Grieser, M.Ed.
Subsystem Technologies
Denver, Colorado, USA

The Federal Aviation Administration's Air Traffic Organization Office of Safety sponsors a comprehensive and ongoing program of Crew Resource Management (CRM), Human Factors in Air Traffic Control (ATC). CRM begins with a one-day workshop for all field ATC personnel, including management, staff, supervisors, and air traffic controllers. Facilitators present principles and methods in teamwork, individual performance, and Threat and Error Management, and participants discuss and record how they can use those principles and methods in their air traffic operations and safety cultures. Participants identify local safety issues and record their own recommendations, which then are compiled from all the workshops in each facility, and the data is delivered to local management for follow-up action. This is *proactive* data, intended to improve individual, team, and system performance *before* errors and accidents occur. To avoid the regression effect, reinforcement includes recurrent training, team debriefs, periodic articles and publications, DVDs, and CRM posters.

Introduction

In *Megatrends* (1982), John Naisbitt described one major trend as "High Tech, High Touch." Along with improvements in technology and technical standards, people in all fields are searching for ways to address the needs and potentials of their workers, and to improve human performance.

Most every field has its technical side and its human side. On the technical side in aviation, we have better aircraft design, construction, and maintenance. We have better weather detection and avoidance equipment, and we have better navigation and ATC equipment.

After these safety gains in technology, human and system factors now cause or contribute to up to 80 percent of all aviation accidents, and almost all air traffic controller operational errors (losses of required separation), which jeopardize safety. Human factors, and system factors which are created and managed by humans, are the biggest safety gap remaining to be closed.

Background

In 1979, after several fatal airline accidents were caused or contributed to by breakdowns in human and system factors, the National Aeronautics and Space Administration (NASA) sponsored a conference, Resource Management on the Flightdeck. In 1981, United Airlines was the first airline to start CRM training, and the International Civil Aviation Organization now requires all of the world's commercial airlines to deliver CRM training to flight crews.

In 1985, the Seattle Air Route Traffic Control Center developed Controller Awareness and Resource Training, an excellent three-day workshop with fourteen major subjects. Known as CART, it was the first well-known grassroots ATC human factors course. In 1992, with CART as the forerunner, the Federal Aviation Administration (FAA), the National Air Traffic Controllers' Association (NATCA), and consultants from Human Technology, Inc. developed Air Traffic Teamwork Enhancement (ATTE), the FAA's first national ATC human factors course.

ATTE was a three-day workshop that was delivered, over ten years, to only an estimated one-third of the FAA's ATC workforce. Budget and staffing restrictions were obstacles to more complete delivery and, even more so, to recurrent training.

In 1997, the National Research Council of the National Academy of Sciences published *Flight to the Future: Human Factors in Air Traffic Control,* in which a panel of human factors experts raised seven concerns about ATTE. Two of the concerns were that "the program does not demonstrate organizational commitment to the concepts by being budgeted and mandated at the national level and integrated into ongoing training and evaluation activities," and that "the training is designed as a single-event program without provision for annual recurrent training." The panel recommended that the FAA "initiate a systematic effort to reinforce the value of teamwork within its organizational culture," and require an "improved … centrally funded program" which provides "recurrent training, hands-on practice, and reinforcement" at all air traffic facilities.

In 1999, Denver controllers and supervisors – following these recommendations in another grassroots effort – developed ATC CRM. These one-day CRM workshops were delivered to Colorado facilities in 2000, and then CRM grew to a regional program when workshops were delivered to Seattle and Salt Lake City air traffic facilities in 2002 and 2003. In 2004, the FAA Air Traffic Organization's Office of Safety – when it was looking for a one-day human factors workshop that was more deliverable than the three-day ATTE course – discovered the Denver program, and began sponsoring it nationally in 2005.

The CRM program described in this paper is an evolution of the Denver program that has been field-tested, revised, and refined many times. The shorter length of the CRM workshop is only one difference from ATTE. The CRM workshop's content is more ATC-oriented, and more specific to each facility's own operations and culture. It generates more active engagement of participants in determining how they can use the principles and methods in their own operations and culture. Importantly, proactive data is identified by the workforce on local safety issues, and their recorded recommendations are used for taking follow-up action. Finally, there are systematic, ongoing reinforcements designed to avoid the regression effect that commonly follows one-time, "flavor of the month" training events. The training field calls one-time events "spray and pray." Spray it on, and pray that it sticks. It doesn't.

CRM Workshop Content

CRM, as defined in the workshops, is the intentional use of effective human factors principles and methods to improve team, individual, and system performance, and to reduce errors and accidents. As it is in airline CRM, Threat and Error Management (TEM) is the cornerstone of ATC CRM, and one of the three major subjects in the workshop. And because controllers – like members of flight crews, surgical teams, bands, orchestras, and sports teams – operate simultaneously as individuals and as members of a team, teamwork and individual performance are the other two major subjects.

Improving Teamwork

Teamwork is the first of the major subjects in the workshop, so that defense mechanisms are a little more relaxed among wary participants, who may come in wondering whether they are there to be "fixed" – in terms of their individual performance and their actual or potential errors. It's easier for them to first talk about their team. By the time they have done that, and they have seen the non-threatening approach and format of the workshop, then they are ready to consider their individual performance, and threat and error management. The teamwork lesson is divided into four main subtopics, as follows.

CRM Behaviors in an Operational Safety Culture

Six CRM behaviors, adapted from Dr. Robert Helmreich and the University of Texas Human Factors Research Project, are presented. The behaviors are: 1) provide active operational leadership and support, 2) effectively distribute workload and tasks, 3) clearly communicate all operational plans to everyone, 4) make "safety-first" decisions and review them to reinforce safety, 5) brief and plan for known safety risks and threats, and 6) maintain safety culture vigilance, speak up, and listen.

Using definitions of these six behaviors, small groups discuss and record what they already do in these areas – good and not so well – and what they start doing, or could do better. Small group reports to the large group then generate further discussions and ideas.

Organizational Dynamics

Productive organizational change is discussed in terms of supporters, fence-sitters, and resisters. Working for change in areas where you already have influence, or can start having influence, is discussed. The point is made that any group, facility, or team is going to be *exactly* what the people who are in it, make it.

Best Practices

In small group discussions, participants identify "best practice feeds" for delivering a good product to the next controller, in their own local operations. Again, small group reports to the large group generate more discussion and sharing of best practices.

Team Debriefs

The value of having teams debrief after operational ATC sessions is discussed, and debriefs are encouraged to support and reinforce the use of CRM behaviors and best practices. Teams that debrief routinely communicate better, understand each other's expectations, and do improve individual and team performance.

Improving Individual Performance

Because controllers operate simultaneously as individuals and as members of a team, improving individual performance is also addressed. A central focus of this lesson is that everyone, at any level, can perform better if they learn from experts, who develop higher abilities in two critical skills: their abilities to maintain situational awareness and to develop, revise, and execute their plan. Situational awareness and executing the plan are discussed after two supporting tools are presented – a formula for consistency, and a formula for improving commitment, confidence, and control. The lesson concludes with a formula to remove the ambiguities in the goal conflict between protection and production, or, in ATC terms, between safety and capacity.

Consistency

Former football coach George Allen said, "Consistency is the truest measure of success. It requires concentration, determination, and repetition." A group puzzle, solved by the whole class as a fill-in-the-blank guessing game, reveals "the secret to ATC" consistency – "Do the right thing, with every aircraft, at the right time, every time, no matter how many aircraft you have. And if you need help to do that, call for it in time."

Commitment, Confidence, and Control

Dr. Robert Kriegel and Dr. Marilyn Harris Kriegel developed *The C Zone* to improve peak performance under pressure. Attitude is critical, because it leads to commitment, confidence, and control. Raising any one of these will automatically raise the other two, and methods to raise each are presented. Control can be raised by using "CAN-DOs" – specific actions that will help, can be done now, and are in your control. Managing both overloads and underloads are discussed as significant human factors challenges, and workshop participants explore using C Zone methods to maintain a reasonably comfortable balance between challenge and mastery.

Maintain Situational Awareness

Dr. Mica Endsley identified the three major components of situational awareness as perception (what we see and hear), comprehension (what we understand), and projection (what we plan to do). Expert controllers develop higher abilities to maintain situational awareness. Individual and team CAN-Dos – to maintain situational awareness and to raise it back up when it falls – are identified, discussed, and recorded in a large group brainstorming session.

Develop, Revise, and Execute the Plan

For any ATC position, in any tower cab or radar room, with any type of air traffic volume and complexity, in any weather or airport configuration, expert controllers develop higher abilities to develop, revise, and execute their plan. By focusing on this skill, along with situational awareness, anyone at any level can learn from the experts' examples and get better. Novices can become intermediates sooner, intermediates can become experts, and experts can be more consistent. Participants explore developing, revising, and executing the plan in small group discussions that identify and record best practices for working their own positions.

Best practice feeds (from the teamwork lesson) are about "what's good for the next controller." Best practices for working your own position are about "what's good for you." Again, small group discussions are followed by reports to the large group, with discussions and sharing of ideas.

Protection versus Production

Dr. James Reason has maintained that people in hazardous technologies, where people's lives are at stake, must effectively manage the inherent goal conflicts between protection and production. It is inarguable that people have sometimes died when production has been valued over protection. In ATC, "protection means safety" and "production means capacity."

NASA Ames Chief Scientist for Human Factors, Dr. R. Key Dismukes, maintains that the protection versus production goal conflict creates ambiguities in people's minds, especially when under pressure. The resolution is to "disambiguate" that goal conflict. It is not a matter of keeping the goals in balance, which leads toward ambiguity. It's a matter of prioritizing them, keeping safety first, always. Although capacity remains a goal, it is a secondary goal. "Every thing you do to enhance capacity must be safe, or you wouldn't do that thing."

Threat and Error Management

Keeping safety first provides a bridge to the cornerstone and "grand finale" of the workshop, Threat and Error Management (TEM). A model is explored in which workshop participants systematically identify the unsafe acts of individuals and teams, and the local workplace factors that sometimes put them in error-prone conditions.

The Risk Denial Syndrome

Adapted from Dr. Robert Besco, the "risk denial syndrome" makes us vulnerable to error when we circumvent standard procedures and make risky decisions, while thinking that we're gaining an operational advantage, achieving worthy goals, and that "it won't matter." The resolution is to catch ourselves thinking, "it won't matter," and then "do the right thing, with every aircraft, at the right time, every time …"

Internal Risks and External Threats

Internal risks are within the facility. They include quick turnaround schedules, outdated airspace and procedures, and control room distractions. External threats are from outside the facility, and include weather, airline schedules, and problems with adjacent facilities. In a large group brainstorming session, these are explored in terms of identifying local vulnerabilities, ways to eliminate them, and how to countermeasure those that are not, or cannot be, eliminated.

CRM Error Types

Exploring specific local error types is another segment adapted from Dr. Robert Helmreich and the University of Texas Human Factors Research Project. Again, in small groups, workshop participants identify, discuss, and record actual and potential errors from their own operations of five types: procedural, intentional noncompliance, communications, proficiency, and decision-making. For each actual or potential error, participants identify and record ways to prevent them, and to catch and correct them if they still occur. Again, small group reports to the large group generate further discussions and sharing of ideas.

Conclusion: Find Truth, Facilitate Change

"Find Truth, Facilitate Change" is a slogan adopted from Hank Krakowski, the FAA Air Traffic Organization's Chief Operating Officer. The overarching goal of CRM is to promote open and honest dialogues and processes that will empower individuals and teams to take ownership of their local safety cultures, identify local issues and solutions, and take local actions to improve them.

There is a wide range of CRM success stories about resolved safety and separation issues, operational and workload issues, and systems and process issues. They include airspace and procedure revisions, the use of best practices and more regular team debriefs of operational sessions, and the local development of supplemental training and action planning processes.

References

Besco. R. analyzing and preventing knowledge deficiencies in flight crew proficiency and skilled team performance, The Journal of Aviation Psychology, January 1992, Vol. 2, Issue 1, 53-74.

Dismukes, R., Berman, B., Loukopoulos, L. (2007). *The limits of expertise: Rethinking pilot error and the causes of airline accidents.* London: Ashgate Publishing Company.

Endsley, M. (2001). Training for situational awareness. Presentation to the Royal Aeronautical Society. (http://www.raes-hfg.com/reports/22may01-SitAssessment/220501-endsley.pdf)

Helmreich, R. L., Butler, R. E., Taggart, W. R., & Wilhelm, J. A. (1994). The NASA/University of Texas/FAA Line/LOS Checklist: A behavioral marker-based checklist for CRM skills assessment. NASA/UT/FAA Technical Report 94-02. Revised 12/8/95. Austin, TX: The University of Texas.

Helmreich. R. Musson, D. (2000). The University of Texas Threat and Error Management Model: (http://homepage.psy.utexas.edu/homepage/group/HelmreichLAB/publications/pubfiles/Pub248.pdf)

Kriegel, R., Kriegel, M. (1984). *The C Zone: Peak performance under pressure.* New York: Anchor Press-Doubleday.

Naisbitt, J. (1982). From forced technology to high tech/high touch. *Megatrends: Ten new directions transforming our lives.* (39-53). New York: Warner Books, Inc.

National Research Council. (1997). *Flight to the future: Human factors in air traffic control.* Washington, D.C.: National Academy Press.

Reason, J. (1997). *Managing the risks of organizational accidents.* London: Ashgate Publishing Company.

DEVELOPMENT OF METHOD FOR CRM SKILLS ASESSMENT

Hiroka Tsuda, Tomoko Iijima, Fumio Noda and Kohei Funabiki
Japan Aerospace Exploration Agency,
Tokyo, Japan

Crew Resource Management (CRM) is currently considered as one of the most effective methods for avoiding human errors or minimizing their effects. In training, measurement of the level of flight crews' CRM Skills is necessary in order to evaluate objectively which Skills have been adequately learned and which are lacking. The Japan Aerospace Exploration Agency (JAXA) has developed CRM Skills Behavioral Markers and CRM Skills Measurement Methods that can identify a crew's level of CRM Skills by which human errors and threats are managed. A series of simulated-LOFT (line oriented flight simulation training) were conducted to examine the applicability of the method.

While improvements in aircraft systems technology have dramatically reduced aircraft accident rates over the past few decades, at present accident rates have flattened out and so different approaches are required to further reduce accidents in the future. Human factors are now a primary causal factor of fatal accidents, and so addressing these should yield further reductions in the accident rate. After Helmric revealed that most human factors-related incidents are caused by inappropriate crew coordination, importance began to be placed on flight crew CRM training, and the first CRM training programs were started by airlines in United States in the 1980s. In Japan, flight crew CRM training was mandated by the Japan Civil Aviation Bureau (JCAB, 1998), and Japan Airlines began CRM training in 1986.

It is considered that concrete behavioral indicators are necessary for effective CRM training, and so from 1999 to 2002 JAXA has been developing CRM Skills Behavioral Markers with the support of airlines (Japan Air System, All Nippon Airways and Japan Airlines) that take into account the particular behavioral and psychological characteristics of Japanese crew members, which would be suitable for the Japanese flight crews operating in a domestic environment. Here, "CRM Skills" is defined to be the ability to carry out CRM. Fig. 1 shows the CRM Skills proposed by JAXA. These are classified into five clusters with three or four skills elements in each. Each skills element has two or more CRM Skills Behavioral Markers.

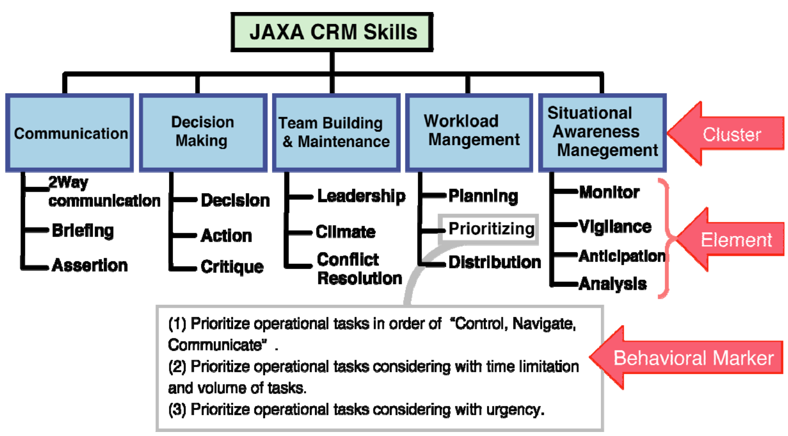


Figure 1. *JAXA proposed CRM Skills.*

Verifying the effect of CRM training is as important as conducting the training. To assess the effect of CRM training and to provide feedback to a training program to reduce possible inadequacies, we have developed a CRM Skills Measurement Method that can identify the extent to which CRM skills have been learned (Iijima *et al.*, 2003). Our proposed method utilizes a subjective rating technique based on the JAXA CRM Skills Behavioral Markers. This paper describes the development of the CRM Skills Measurement Method and the evaluation of its applicability.

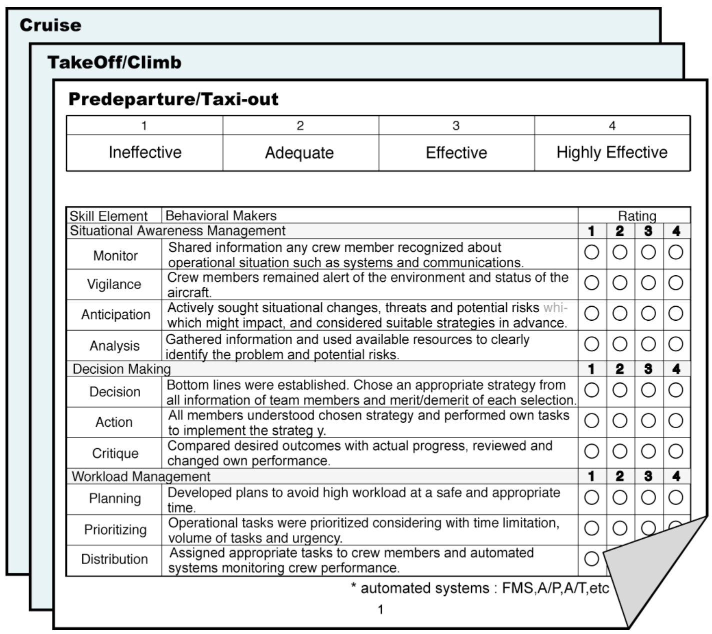
Design of CRM Skills Measurement Method

Development of the CRM Skills Measurement Method consisted of three phases: a preliminary study, design of a prototype rating sheet, and evaluation of the method. The Method includes CRM Skills, a Rating sheet, an Observation sheet, LOFT scenarios, and assessment of inter-rater-reliability.

CRM Skills Rating Sheet and Observation Sheet

To design the CRM Skills Measurement Method, a survey was made of airlines that incorporate CRM Skills into LOFT (Line Oriented Flight Training) and/or LOE (Line Operational Evaluation) (JAL, 2003), and based on the results, an initial version of the Measurement Method, prototype No. 1, was developed for trial purposes. The method used an initial prototype CRM Skills Rating Sheet on which "Raters" (training personnel who evaluate crews' CRM Skills) record scores for the CRM Skills they observe, at that stage on a three-point scale. Raters evaluate only actions that can be clearly observed based on the CRM Skills Behavioral Markers proposed by JAXA. After initial prototyping, the Measurement Method was refined by applying it to sample recorded LOFT sessions, resulting in prototype No. 2.

To improve the second prototype, a preliminary rating experiment was conducted by applying it to a video of a sample LOFT session (the "LOFT Videotapes" below), with two Raters. Feedback from the Raters was then used to produce a third version of the Rating Sheet, part of which is shown in Fig. 2.


Cruise

TakeOff/Climb

Predeparture/Taxi-out

1	2	3	4
Ineffective	Adequate	Effective	Highly Effective

Skill Element	Behavioral Makers	Rating			
Situational Awareness Management		1	2	3	4
Monitor	Shared information any crew member recognized about operational situation such as systems and communications.	○	○	○	○
Vigilance	Crew members remained alert of the environment and status of the aircraft.	○	○	○	○
Anticipation	Actively sought situational changes, threats and potential risks whi- which might impact, and considered suitable strategies in advance.	○	○	○	○
Analysis	Gathered information and used available resources to clearly identify the problem and potential risks.	○	○	○	○
Decision Making		1	2	3	4
Decision	Bottom lines were established. Chose an appropriate strategy from all information of team members and merit/demerit of each selection.	○	○	○	○
Action	All members understood chosen strategy and performed own tasks to implement the strateg y.	○	○	○	○
Critique	Compared desired outcomes with actual progress, reviewed and changed own performance.	○	○	○	○
Workload Management		1	2	3	4
Planning	Developed plans to avoid high workload at a safe and appropriate time.	○	○	○	○
Prioritizing	Operational tasks were prioritized considering with time limitation, volume of tasks and urgency.	○	○	○	○
Distribution	Assigned appropriate tasks to crew members and automated systems monitoring crew performance.	○			

* automated systems : FMS,A/P,A/T,etc

1

Figure 2. *Example of CRM Skills Rating Sheet.*

Scenarios for Simulated LOFT

Scenarios for simulated LOFT were generated based on Seamster *et al.* (1998), but that document specifies the Boeing 737 and uses routings within the United States. To adapt these for Japanese use, the aircraft fleet types were changed to the Boeing 767 and 777, and Japanese routings, air space and airports were substituted. As a result, four types of scenario were created.

A prepared script for the Raters described the event set, weather information, NOTAM, weight & balance, communications with ATC, cabin crew, company radio, and ground staff, and specified which CRM Skills were to be observed.

Selection of Raters

Assessment of inter-rater-reliability is an important issue in the evaluation of flight crews' CRM Skills. In these experiments, Raters were selected based on the following requirements.

(1) Job experience in a CRM training-related department of an airline. Knowledge of CRM skills is indispensable.
(2) Aircrew experience in at least one "glass-cockpit" airplane type.
(3) Ability to understand the proposed CRM Skills Behavioral Markers before the experiments.
(4) Aircrew experience of the aircraft fleet used in the simulated LOFT is not necessary, since aircraft type-specific Standard Operating Procedures (SOPs) are irrelevant to CRM Skills.

Nine Raters (A to I) ware selected from Japanese airlines. Their experience is shown in Table 1. X, Y, and Z are their airlines, and the left column shows the airplane types with which they had crew experience.

Table 1. *Classification of nine Raters by Airline and Airplane Type.*

	X-Airline	Y-Airline	Z-Airline
B747-400	A, B	C	
B767	G	D	
B777	H	F	
Others	E		I

Simulated LOFT Experiments

Five sets (Cases #1–#5) of simulated LOFT sessions using the four scenarios were flown on B777 and B767 flight simulators, and the sessions were recorded to LOFT Videotapes. Using the five LOFT Videotapes, experiments to evaluate the Measurement Method were conducted in the following manner.

(1) The CRM Skills Behavioral Markers, Measurement Method and experiment procedure were explained to the Raters.
(2) Scoring was on a four-point scale: 1 denotes Ineffective, 2 Adequate, 3 Effective, and 4 Highly Effective. '3' is the reference standard.
(3) Only the degree of CRM skills practice is to be evaluated; whether or not a crew follows SOPs is irrelevant.
(4) The skills of the crew itself should be evaluated, not the skills of individual crewmembers.
(5) Comments should be recorded regarding CRM skills that are judged to be better or worse than the reference standard.
(6) A CRM Skills entry may be left blank in the case where Behavioral Markers cannot be observed in the crew.

After watching a video, each Rater completed the Rating Sheet and was then interviewed to determine the reasons for his scorings and to obtain general comments on the Measurement Method.

Results

Overall

Table 2 shows the average of rating (score) and the average of standard deviation (SD) across all Skills calculated for each case. The average rating is the highest for Case #1, and its SD is the second smallest. As already mentioned, Case #5 shows the greatest variance.

Fig. 3 shows a plot of the average rated scores across all cases. It can be seen that Cases #1 and #3, which are based on same Scenario #1, were rated relatively consistently, while Cases #4 and #5 show greater variance of the average scores awarded by the Raters. When looking at the relative scores between the cases rated by each Rater, consistency is observed for 8 out of the 9 Raters (excepting G), excepting Case #5.

Features of each CRM Skills Behavioral Marker

Average rating and Average SD

For each Behavioral Marker, the scores of the nine Raters and five cases were totaled and the average rating and average variance were calculated. While the average rating for any Behavioral Marker was concentrated at the standard score level three (from 2.932 to 3.111), the average variance extended from 0.054 to 0.402. From this, it is understood that there is a difference in ratings between Raters or between cases.

The distributions of average rating and average variance for each CRM Skills Behavioral Marker are plotted in Fig. 4. As expected, the figure shows a strong correlation between average rating and average variance (r=0.505); that is, average variance tends to grow for Behavioral Markers which receive high scores.

Correlations between Behavioral Markers

It was assumed that the evaluated score for a single Behavioral Marker might influence the score for other Behavioral Markers. The correlation coefficients between Behavioral Markers calculated from all the gathered data. were analyzed to examine the extent of this influence.

"Total Team Performance" has a relatively strong correlation with "Leadership", "Climate" and "Assertion". Correlation was observed not only between Behavioral Markers that belong to the same Element, such as "Monitor" versus "Leadership", but also between Markers in different Elements, such as "Leadership" versus "Climate" or "Distribution" versus "Prioritizing".

On the other hand, there was hardly any correlation in the combinations of “Planning” versus “Assertion”, “Critique” versus “Anticipation”, and “Two-Way Communication” versus “Briefing”.

Table 2. *Average Rating and Average of Standard Deviation (SD) for each Case#.*

LOFT No.	Scenario No.	Average Rating	Average SD
Case #1	1	3.100	0.331
Case #2	2	2.983	0.327
Case #3	1	3.035	0.399
Case #4	3	2.916	0.401
Case #5	4	2.982	0.503

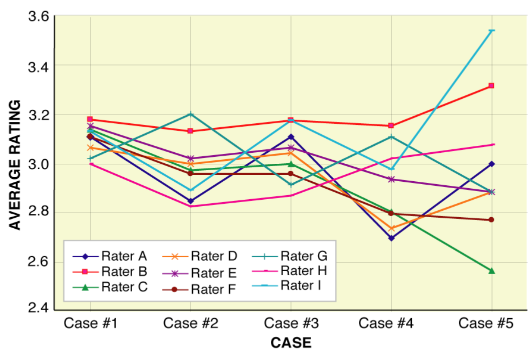


Figure 3. *Average Rating for Each Case.*

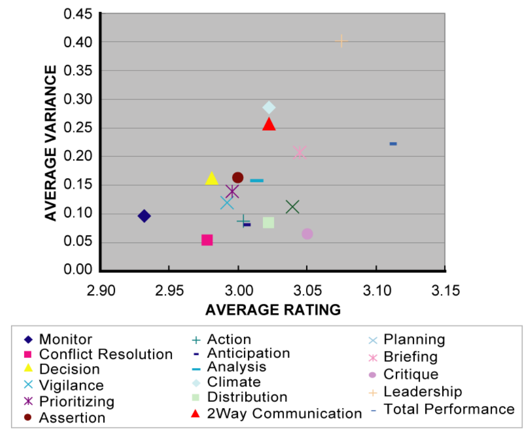


Figure 4. *Scatter Plot of Average Rating VS Average Variance.*

Number of Empty Behavioral Markers Columns

Raters were permitted to leave a column in the CRM Skills sheet blank in the case that the corresponding crew behavior was not observed, or for other reasons. This implies that skills with more empty columns in the Rating Sheet are more difficult to observe.

Relatively high numbers of empty columns were recorded for the Behavioral Markers “Analysis”, “Critique”, and during the Takeoff/Climb and Cruise flight phases. The following narrative comments related to the empty columns were obtained from interviews with Raters.

- Some items were difficult to score because they were not visually prominent in the video record.
- Understood that “Vigilance” is identical to “Anticipation”, only one of these was scored.
- “Critique” was not scored but was included in “Communication” in “Overall”.

Moreover, it is considered that difficulty in identifying transitions between flight phases caused more empty columns during take-off and cruise. For example, some cases contained a missed approach but the timing of the transition is not clearly apparent in the video recordings.

Raters' Aircraft Type Experience

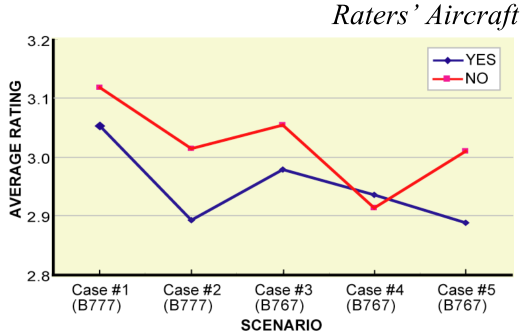


Figure 5. *Average Rating Sorted by Experience YES/NO.*

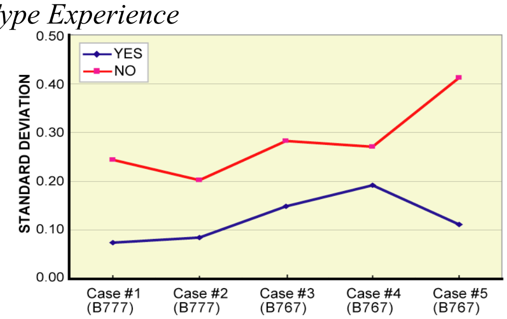


Figure 6. *Average SD Sorted by Experience YES/NO.*

An analysis was conducted to investigate the effect of Raters’ experience with the type of aircraft used in the scenario. The average and standard deviation of ratings in each group (w/ or w/o experience) are plotted in Figs. 5

and 6 respectively. Significant differences at a 5% threshold level were observed for both average and STD between experienced and non-experienced groups (average: p=0.013 and SD: p=0.014). The correlation of differences between cases was strong for averages (r=0.665) but not for SDs (r=0.169).

From these results, it is concluded that a Rater with experience of the aircraft type used in the LOFT tends to award lower scores, and variance between Raters in the experienced group is small.

Raters' Comments

After the Raters had completed the Sheets, they were asked to comment freely on the experiment. The obtained comments are summarized below:

(1) CRM Skills Rating Sheet

- Some CRM Skills Behavioral Markers should be unified into a single rating item to make it easy to evaluate crew behavior.
- "Communication" and "Team Building & Maintenance" may be rated not Overall but rather for each flight phase, as for other skills. Rating a skill "Overall" disguises differences between flight phases.

(2) Background of the Raters

- Rating might be easier if the events sets in the scenarios are known to the Raters in advance, while it is impossible in LOSA (Line Operational Safety Audit).
- Raters' knowledge of the SOPs of the type of aircraft is not essential, but it does help the rating task.
- Some raters with flight inspector experience may easily award low scores (such as 2). On the other hand, other Raters may find it hard to do so because they may imagine that low scores might affect the certification of the crew.

(3) Method of scoring

- Four degrees of rating level is too little/much.
- There seem to be two ways of rating an event related to crew error. When a crewmember makes an error and notices (corrects) it by himself, Raters may either evaluate him negatively for making the error, or may evaluate him positively for correcting it.
- If the rating is done long after viewing the video, Raters are allowed to think about the reasons for the crew's behavior and it if becomes understandable, and then the score tends to be higher.

(4) Others

- There seem to be too many event sets. It was therefore felt that the scenario was designed more to look at technical skills and the evaluation of SOP practice than CRM Skills.
- The experiment provided a good opportunity to know the status of operations and recommended CRM Skills of other companies.

Discussion

Correlation between Behavioral Markers

When the score average variances of the Raters are examined it becomes clear that some Behavioral Markers have large variance. There were some Behavioral Markers for which the Rating Sheet scores were often left blank. The causes of this are considered to be not only differences in individual Raters' judgment, but also the format of the CRM Skills Rating Sheet itself.

At the early stage of this research, we presented Raters with concrete examples of crew behaviors that could be expected to be observed corresponding to CRM Skills Behavioral Markers. In response to this, it was suggested that it might not be clear how to evaluate skills if behaviors other than the examples provided were observed, and that with detailed examples of crew behavior anyone could be a Rater without training. As the result of this feedback, only guidelines on scoring were presented to Raters for this experiment, with neither detailed examples of crew behaviors nor how to guidance on how to evaluate them.

In this experiment, Raters commented that while each Behavioral Marker was to be scored separately in the Rating Sheet, some crew behaviors corresponded to more than one Behavioral Marker and in such cases, the Behavioral Markers should be unified to form a single column. For example, "Vigilance" and "Anticipation", "Monitor" and "Analysis", "Planning" and "Prioritizing", "Conflict Resolution" and "Briefing". These Behavioral Markers were often left blank in Rating Sheet. For a Behavioral Marker which is strongly correlated with another, if its Rating Sheet column is left blank then is possible to guess the score from that of its correlated Behavioral Marker.

It is therefore understood that while there is no need to improve the CRM Skill Behavioral Markers themselves, there is a need to review and restructure the measurement items in the CRM Skills Rating Sheet. However, unification of some Behavioral Markers into a single measurement item makes the evaluation of CRM skills coarser, and is not always necessarily better from the viewpoint of identifying a crew's inadequate CRM Skills.

Standardization of Raters

As already discussed, Raters' type experience with the aircraft in the simulated LOFT affected their scoring behavior. Although Raters were instructed that they should not score execution of SOPs, their knowledge of the aircraft SOPs did in fact influence their ratings. Standardization of Raters should therefore be conducted taking into account their type experience. In this experiment, no limitations or requirements were stipulated on Raters' flight crew backgrounds, and no standardization was conducted prior to the experiment. Although the authors had assumed that that adequately developed CRM Skills Behavioral Markers would require no standardization in advance, the experiment result highlighted the necessity of Rater standardization.

A method for Rater standardization widely_used by world airlines is as follows. Two or more Raters watch a recorded LOFT session together, and then compare their own rating scores with each other and with a Standard Score while discussing. Repeating this procedure minimizes scoring variation between Raters. In the present experiment, some Raters commented that it was very effective to know the opinion of other Raters. However, contrary to this, Case #5 was scored after the greatest amount of discussion but showed the highest variation between Raters' scores. The reason may be to do with the following comments concerning Case #5: "I had became accustomed to the experiments, so it came to be able to evaluated that I thought", and, "By comparing with the past scores of other Raters, I noticed that my own scores had been relatively high, so I reduced my scores in this case." Consequently, it is concluded that Raters without their own firm rating criteria were easily influenced by the opinions of other Raters, and a familiarization with use CRM Skills evaluation method is necessary before the rating session.

For Rater standardization, the method mentioned above seems not to be only effective for standardizing Raters' scoring criteria, but is also effective for familiarization with how to rate before the actual rating session.

Degree of Scoring Level

Many Raters commented that the current four-degree scale of scoring is confusing. The current scale of "1" to "4" gave Raters an impression that "2" means "unacceptable", and it was difficult to decide whether to score a "2" or a "3" if minor deficiencies were observed for a skill. The Rating Sheet is one measurement tool for evaluating crew CRM Skills levels, and its main objective is to extract skills that require improvement. It is thought that current four level scoring scale should be revised to allow Raters to score "2" more easily. However, if the Sheet is used only for LOFT, where it is assured that the score record is immediately discarded after the training session, the current four degrees is perfectly acceptable.

Conclusion

JAXA has developed a CRM Skills Measurement Method to evaluate the effectiveness of flight crew CRM Skills training. The method was developed by means of a survey, interviews and several simulated LOFT experiments. Nine Raters evaluated crew CRM Skills performance using this Measurement Method in LOFT experiments.

Analysis of the ratings and consideration of Raters' comments indicate that the concept of the CRM Skills Measurement Method is sound and that suitable CRM Skills Behavioral Markers are available, but there is room for improvement in the Rating Sheet and in the method by which the rating is carried out. The importance of inter-Rater-reliability was recognized, and insights into CRM Skills Measurement Method were also obtained.

References

Japan Civil Aviation Bureau_(JCAB) (1998). Order of execution of CRM training to aircrew. Flight Standards Order Vol._410, Engineering Department, Ministry of Land, Infrastructure and Transport .

Seamster, T. L., Edens, E. S., McDougall, W. A., and Hamman, W. R. (1998). Observable Crew Behaviors in the Development and Assessment of Line Operational Evaluations (LOE's), FAA OP-US/6821.

Iijima T.,_Noda F., Sudo K., Muraoka K. and Funabiki K._(2003). Development of CRM skills behavioral markers, TR-1465, National Aerospace Laboratory Report.

Flight Crew Technical Service, Flight Operations, Japan Airlines Co., Ltd. (JAL)_(2003). Research Report of LOFT/LOE and CRM skills.

Helmrech, R. L., Klinect, J. R, Wilhelm, J. A and Sexton, J. B. (2001). The Line Operations Safety Audit (LOSA), Proc. of 1st LOSA week.

Acknowledgments

This research is supported by ANA, JALI, JALJ and other the staff of other airlines. The authors wish to thank them all.

A PC Based Methodology for CRM – Corporate Resource Management Practice Training

Cabral, Lisia Maria Espinola da Silva Pacheco
Brazilian Airports Administration (INFRAERO - Empresa Brasileira de Infra-estrutura Aeroportuária) / East Regional Bureau (SRGL – Superintendência Regional do Leste) / Air Navigation Management (GLNA / NAGL – Gerência de Navegação Aérea)
Rio de Janeiro, Rio de Janeiro, Brazil
Ribeiro, Selma Leal de Oliveira
National Institute for Spatial and Aeronautics Development (IDEA - Instituto Nacional para o Desenvolvimento Espacial e Aeronáutico)
Rio de Janeiro, Rio de Janeiro, Brazil
Landau, Luiz
Advanced Methods of Computer Engineering Laboratory (LAMCE – Laboratório Avançado de Métodos Computacionais em Engenharia) / Rio de Janeiro Federal University (UFRJ – Universidade Federal do Rio de Janeiro)
Rio de Janeiro, Rio de Janeiro, Brazil
Gerson, Cunha
Advanced Methods of Computer Engineering Laboratory Laboratory (LAMCE – Laboratório Avançado de Métodos Computacionais em Engenharia) / Rio de Janeiro Federal University (UFRJ – Universidade Federal do Rio de Janeiro)
Rio de Janeiro, Rio de Janeiro, Brazil

On 2003, the National Civil Aviation Agency (ANAC – Agência Nacional de Aviação Civil) elaborated the Civil Aviation Instruction (IAC - Instrução de Aviação Civil) 060-1002A, reviewed on 2005, about the CRM Training, based on the AC – Advisor Circular 120-51 A-E of FAA - Federal Aviation Administration, aiming at giving a standard direction about the subject for civil aviation organizations. The purpose of the study is to use a PC Based Methodology for the CRM Training / 2nd. Phase - Recurrent Practice and Feedback, in a low cost artificial environment differently from the LOFT – Line Oriented Flight Training, which, besides of being a technical Training, also requires an expensive simulator equipment for it. Although the IAC 060-1002 A was elaborated for civil aviation organizations, military institutions also adopted it, therefore the research was developed into a military institution operating helicopters. Future steps of the Methodology validation need to be further developed.

CRM – Cockpit / Crew / Corporate Resource Management Training, as the names indicate, has passed through an evolution, since its origin after some aviation accident (Portland, Tenerife, Dryden etc.) occurrences, consolidating the comprehension about human error contribution to optimize team human performance, so that the newest generation of CRM, known as TEM - Threat and Error Management, not only emphasizes human errors, but also internal and external threats which may affect flight safety as a final result (CABRAL, 2006).

The study here presented was developed to enable the organizations to implement CRM Training / 2nd. Phase - Recurrent Practice and Feedback, by using computer games, based on some previous experiments (BOWERS, SALAS, PRINT & BRANNICK, 1992), in order to fill up high fidelity flight simulators gap in Brazil, mainly of helicopters, once there is a belief that CRM Recurrent Practice and Feedback Phase ought to be realized in this kind of equipment. This is an inadequate point of view, because, although the CRM Training may be implemented as a behavior training in high fidelity flight simulator environments, at the same time of the LOFT Training (USA, 2003), which is a technical Training, both have different purposes and the second one will never substitute the first one.

This article comments the implementation of CRM Recurrent Practice and Feedback Phase using the Microsoft Flight Simulator Software, demonstrating the facilities of computer games application for the CRM Recurrent Practice and Feedback Phase, as a lower cost and an easier acceptance tool, comparing to high fidelity flight simulators environment (CABRAL, 2006).

CRM Brazilian Regulation

According to Civil Aviation Instruction (IAC - Instrução de Aviação Civil) 060-1002A (BRAZIL, 2006), about CRM - Corporate Resource Management Training, based on the AC – Advisor Circular 120-51 A-E (USA, 2004), it consists of three different phases: Initial Indoctrination / Awareness; Recurrent Practice and Feedback; and Continuing Reinforcement.

The 1st. Phase / Initial Indoctrination / Awareness, based on the organizational diagnosis of the institution, aims at pointing out critical situations concerning teamwork, in order to improve attitude towards team performance; the 2nd. Phase / Recurrent Practice and Feedback requires previous implementation of the Initial Indoctrination / Awareness Phase, aiming at reinforcing team attitude in operational routine and emergencies, in order to achieve a change from individual behavior to team behavior; and the 3th. Phase / Continuing Reinforcement, aims at consolidating the CRM Training into the organization culture.

The aviation accident statistics of the last decades indicates a considerable contribution of human error, what outlines the CRM Training as a valuable tool to optimize team performance and increase safety of aviation operation.

Although the IAC 060-1002 A is proper to guide the CRM implementation in civil aviation organizations, it also is adopted by military institutions. Therefore, this article will present a research (CABRAL, 2006), developed in a military institution which operates helicopters, where it was possible to create a PC Based Methodology for the CRM 2nd. Phase / Recurrent Practice and Feedback, using the Microsoft Flight Simulator Software, in order to optimize the interface between pilots and flight engineers related to teamwork performance.

The Study

Based on some american previous experiments (BOWERS, SALAS, PRINT & BRANNICK, 1992), this study subsidize a behavioral and not a technical training realized in a low fidelity aviation environment. It was developed a proper Methodology using the Microsoft Flight Simulator Software / 2004, installed in a laptop, with a mouse and a keyboard as devices, for the2nd. Phase / Recurrent Practice and Feedback.

This research was based on some demands concerning 2nd. Phase / Recurrent Practice and Feedback, such as the lack of: parameters towards internal and external variables; a theoretical base for the CRM concepts; criteria to observe and evaluate team behavior in CRM Recurrent Practice and Feedback Training; systematic procedures to assess the CRM Debriefing. It was considered an occasional sample, which was divided in four groups, each one composed by: a facilitator, a captain, a pilot and a flight engineer. It was required a specific training for the facilitators to apply this Methodology.

The occasional sample of the study sums up 15 participants, who represents 16,67% of the 90 crew members (facilitators, captains, pilots and flight engineers), which can be observed in Figure 1.

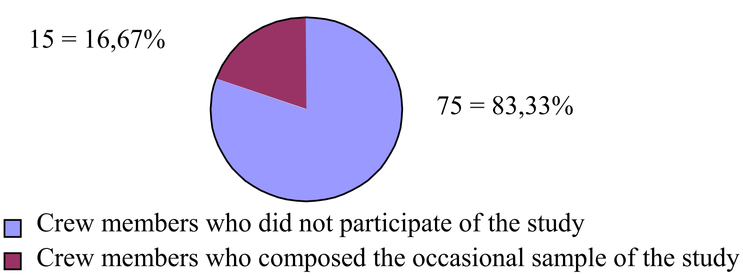


Figure 1: Occasional Sample Percent of the Study

The Methodology

To develop the referred Methodology, some techniques were used, such as: observations, reports and interviews. The result was the implementation of the PC Based Methodology for the CRM 2nd. Phase / Recurrent Practice and Feedback, using the Microsoft Flight Simulator Software. The Methodology application sums up, approximately, five hours, and was divided in three steps, which correspond to the basic moments of the Training - Before Training, During Training and After Training, as follows:

Table 1: *Computer Based Methodology for CRM 2nd. Phase using the Flight Simulator*

Computer-based Methodology for CRM 2nd. Phase Using the Flight Simulator	TIME ≅ 5 h
Before Training	
Facilitator Checklist / ANNEX I and Briefing about the Training	≅ 10 min
CRM Expectations and its Concepts applications Questionnaire (QE-CRM / ANNEX II)	≅ 30 min
Familiarization on Flight Simulator Main Commands (ANNEX III)	≅ 60 min
CRM Refreshment (CRM Refreshment Parameters / ANNEX IV)	≅ 30 min
During Training	
Facilitator Briefing about the flight	≅ 10 min
Flight operation and Videotape (Internal Variables / ANNEX V, External Variables / ANNEX VI and Script / ANNEX VII)	≅ 60 min
After Training	
Debriefing	≅ 60 min
Training Evaluation (Participant Reaction Questionnaire / ANNEX VIII and Facilitator Reaction Questionnaire / ANNEX IX)	≅ 15 min
Validation after Training (CRM Feedback Form for Training Optimization / ANNEX X)	≅ 30 min

The three phases of the Methodology will be commented bellow:

Before Training

Consists of the following components:

-Facilitator Checklist / ANNEX I – Facilitator tasks and responsibilities guide for each CRM Training step.

-Briefing about the Training – Explanations about the purposes of the Training.

-CRM Expectations Questionnaire / ANNEX II – Assessment of the CRM knowledge and motivation.

-Helicopter Commands Parameters / ANNEX III – Familiarization of the main helicopter commands in Flight Simulator Software.

-CRM Refreshment Parameters / ANNEX IV – Review of the main CRM concepts.

During Training

Presents the following tools:

-Facilitator Briefing – Facilitator explanation about the flight task to the crew members.

-Small paper forms written with the following types of variables:

- Internal Variables Parameters / ANNEX V – Different combinations of equipment problems, also classified in crescent levels of complexity (low, medium and high), to be chosen to compose the Script.
- External Variables Combination / ANNEX VI – Different combinations of problems from environment sources, classified in crescent levels of complexity (low, medium and high), to be chosen to compose the Script.

-CRM Script Form/ ANNEX VII – Facilitator guide about the scene description and the parameters of behavior skills (communication / assertiveness, situation awareness, decision making and leadership) assessment.

-Videotape – Record of the whole Training.

After Training

Presents the following tools:

-Debriefing - CRM Video Session conducted by the facilitator, showing the interaction among the participants during the flight experience, emphasizing the behavior skills, based on the CRM concepts.

-Participant Reaction Questionnaire / ANNEX VIII – Assessment of the participant feelings about the Training right after the Debriefing.

-Facilitator Reaction Questionnaire / ANNEX IX – Assessment of the facilitator feelings about the Training right after the Debriefing.

-CRM Feedback Form for Training Optimization / ANNEX X – To be fulfilled by the facilitator right after the Debriefing and the Facilitator Reaction Questionnaire.

When applying the PC Based Methodology (CABRAL, 2006) for CRM Recurrent Practice and Feedback Training, the sample seemed receptive. During Training, the script was composed by a scene involving an one hour flight, from the City of Taubaté, São Paulo, to the City of São Pedro D'Aldeia, Rio de Janeiro (Brazil), about, approximately, 442 km, in Squirrel Helicopter. During the flight, some variables were introduced, either internal (problems with the equipment) or external (problems in the flight environment) ones, which required of the crew members to make arrangements to find the best alternatives for their resolution.

The variables chosen for the scene that composed the script could not be plotted into the Microsoft Flight Simulator, as first purposed, so they were written in small paper forms to be given to the crew members by the facilitator, once at a time. This procedure turned out to be an effective way of perceiving and managing the problems by the crew members. Also, it was addressed the necessity of classifying the variables in crescent levels of complexity - low, medium and high, so that they could be introduced During Training, according to the facilitator assessment related, not only to the flight evolution, but also to the ability and maturity of each sample to deal with them, based on some parameters of behavior skills, such as: communication / assertiveness, situation awareness, decision making and leadership.

Here are some examples of variables used for the CRM Training: winkle generator lights (internal variable / low complexity), light authority pressure inside the aircraft (external variable / low complexity), phone degradation communication (internal variable / medium complexity), hydraulic failure (internal variable / high complexity) etc.

During Training is considered the most important moment of the 2^{nd}. Phase / Recurrent Practice and Feedback, therefore it must be videotaped and registered by the facilitator in the CRM Script Form / ANNEX VII, to make the participants possible to have a feedback about it, in the Debriefing, After Training.

Results

As part of the study results, some aspects collected from the PC Based Methodology (CABRAL, 2006) tools will be mentioned, as follows:

CRM Expectations Questionnaire:

Have you participated of CRM 1^{st}. Phase / Indocrination?

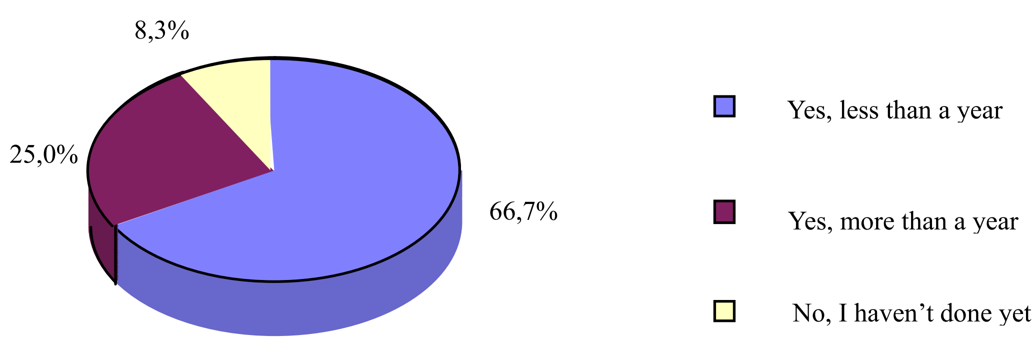


Figure 2: Have you Participated of CRM 1^{st}. Phase / Indocrination?

Among the crew members who participated of the study: 66,7% realized CRM / Indocrination Training less than a year; 25% realized more than a year; and 8,3% did not realize it. Although the majority (66,7%) had already be trained on CRM 1^{st}. Phase before, we must advise that it is required its previous implementation for CRM 2^{nd}. Phase, based on a customized culture organizational diagnosis focused to find aviation operational problems concerning team interaction, which must guide its implementation.

Participant Reaction Questionnaire:

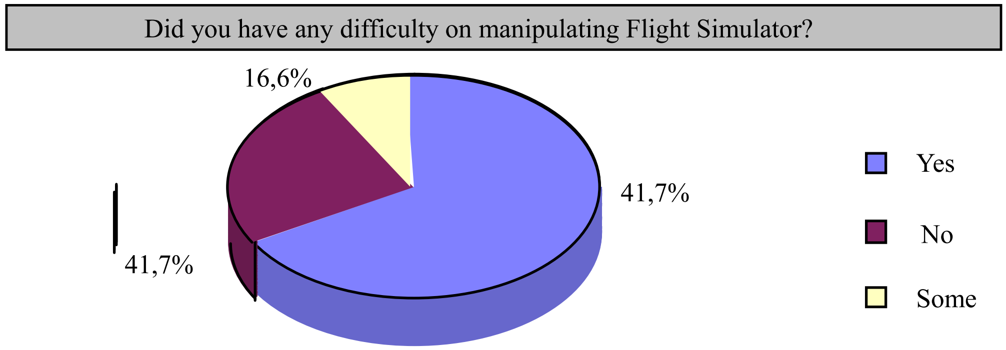


Figure 3: Did You Have any Difficulty on Manipulating Flight Simulator?

Among the participants, half 41,7% had difficulty on manipulating the Flight Simulator, and half 41,7% did not, therefore it was provided the implementation of the Familiarization Flight Before Training / ANNEX III, with the Helicopter Commands Parameters mentioned before.

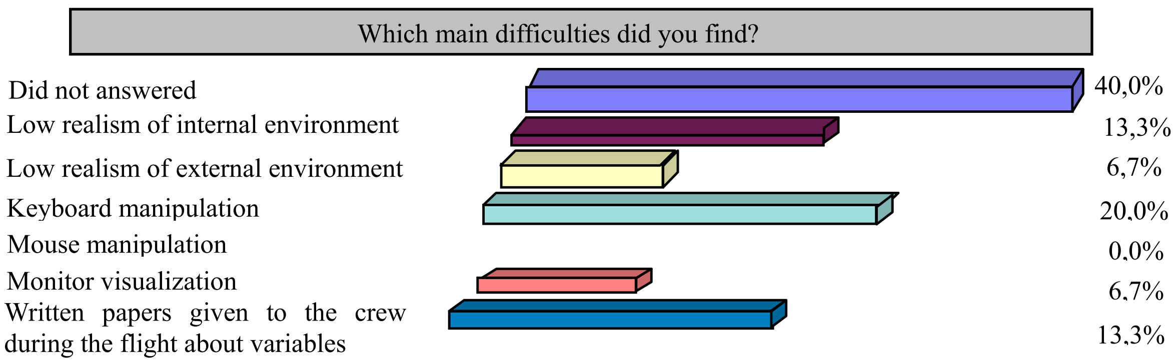


Figure 4: Which Main Difficulties Did you Find?

Although 40,0% of the sample did not answer, among the difficulties presented, 20,0% are related to the keyboard, suggesting to substitute it by a joystick, which constitutes a low cost and a possible initiative.

Facilitator Reaction Questionnaire:

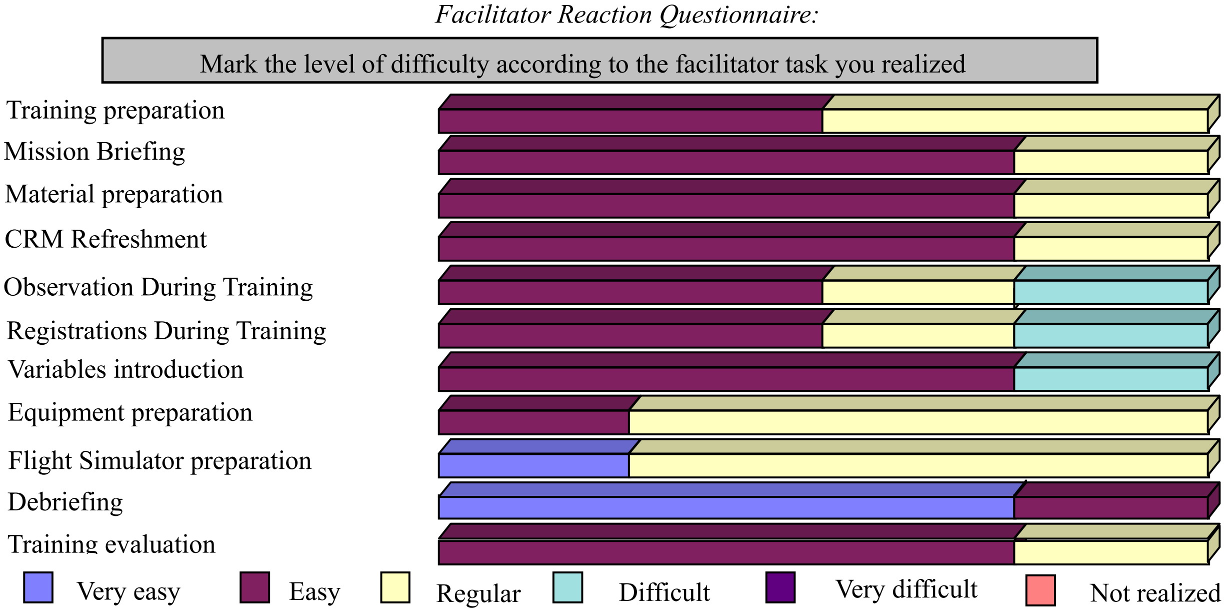


Figure 5: Mark the Level of Difficulty According to the Facilitator Task you Realized

No task was considered very difficult or not realized, but three of them were considered difficult by a minority: observation and registration During Training, and variables introduction. It indicates a high level of requirement for the facilitator During Training, which requires a specific training to prepare the facilitator for an adequate performance. Besides, the facilitator has, also, to plan, create and update continuously Before Training, based on the PC Methodology After Training, which reinforce the need of the facilitator training.

It is appropriate to mention that here was illustrated a summary of the whole research, which may be consulted and deepened by researchers (CABRAL, 2006).

Conclusion

The PC Methodology for CRM 2nd. Phase / Recurrent Practice and Feedback showed a high level of acceptance among the study sample, but it would be better to adopt the use of a joystick for the helicopter controls instead of a keyboard, which is, also, a low cost solution.

The study results were presented not only to the sample, but also to some organization managers, who decided to adopt the PC Based Methodology (CABRAL, 2006) using the Microsoft Flight Simulator Software for CRM Recurrent Practice and Feedback Training.

At the beginning, the idea was to focus the CRM 2nd. Phase / Recurrent Practice and Feedback, only to helicopter flight operations in Microsoft Flight Simulator, based on the script involving the scene once planned, but, when the Methodology was concluded, it revealed that it can also be developed for other types of aircrafts, if the necessary adaptations be implemented.

The main benefit of this PC Methodology is to offer a low cost and an accessible computer game as a resource to implement the CRM Training, in place of an expensive simulator environment, as usual, although it may also be used in this kind of equipment. Future steps of the research may include the Methodology (CABRAL, 2006) validation in military and civil aviation environments and in other types of aircrafts.

The study also shows the need for each organization to elaborate an internal instruction for the CRM Program and to use the Methodology results as an instrument of culture organizational diagnosis for an annual review of CRM 1st. Phase / Indocrination.

Finally, it must be emphasized the need to adapt the Methodology to the use of appropriate games involving other aeronautic segments, besides aviation crews, such as air navigation and maintenance, considering the contribution of professional interculture to operational safety.

Acknowledgments

We're very grateful for all the efforts and contribution of the Aviation Command of Brazilian Army, which made possible the development of this study.

References

Bowers, C., Salas, E., Prince, C. & Brannick, M. (1992). Methods & Designs – Games teams play: A method for investigating team coordination and performance in behavior research methods, instruments and computers, 24 (4) 503-506.

Brazil. (2005). "Instrução de Aviação Civil 060-1002 A. Treinamento em Gerenciamento de Equipes" [Civil Aviation Instruction 060-1002 A Corporate Resource Management – CRM]. Brasília, DF: ANAC - Agência Nacional de Aviação Civil [Brazilian Civil Aviation Agency]. (www.anac.gov.br/legislacao).

Cabral, L. M. E. S. P. (2006). "Uma proposta metodológica para tripulantes de aeronaves baseada em jogos computacionais" [Proposition of a Behavioral Training Methodology for Aircrafts Crew Members Based on Computer Games]. Rio de Janeiro, RJ: UFRJ - Universidade Federal do Rio de Janeiro. (www.minerva.ufrj.com.br).

USA. (2003). AC (Advisory Circular) 120-23C, LOS (Line Operational Simulation), LOFT (Line Oriented Flight Training), SPOT (Purpose Operational Training), LOE (Line Operational Evaluation). Washington, DC: FAA (Federal Aviation Administration). (www.faa.gov).

USA. (2004) AC (Advisor Circular) 120-51 E, CRM - Crew Resource Management Training. Washington, DC:

FAA - Federal Aviation Administration. (www.faa.gov).

PROPOSING ATTITUDE INDICATOR MODIFICATIONS TO AID IN UNUSUAL ATTITUDE RECOVERY

Nathan B. Maertens
University of Illinois at Urbana-Champaign
Savoy, Illinois

Pilots' inability to recover from unusual attitudes (UA) is a major factor in loss of control in-flight (LOCIF) accidents, the largest cause of commercial aviation fatalities (Boeing, 2008). One study found 58% of professional pilots and 72% of general aviation pilots were unable to recover from LOCIF upsets (Regional Aviation News, 2008). Statistics also show that LOCIF is the only fatal aviation accident type not to appreciably decrease over a 21 year period ending in 2002 (Sumwalt, 2003a). A revision of the attitude indicator (AI) is proposed to examine if this would reduce the problem by keeping pilots from flying into UA, accelerate UA identification and enhance recovery. The proposed modifications are: add triangles filled with graduated colors to indicate horizon position, roll and pitch indicators that inform pilots of UAs and corrective procedures, and a thrust indicator that indicates throttle action to maintain adequate energy for aircraft recovery.

Loss of control in-flight (LOCIF) is the primary cause of worldwide aviation fatalities. A LOCIF is classified as a failure or an inability to get an aircraft's wings level and usually results from an unusual attitude (UA) (Schlimm, 2005). An UA condition is defined as greater than 45^o bank, 25^o pitch up, 10^o pitch down or flying at speeds inappropriate for conditions while within the above parameters (Sumwalt, 2003a). Failure to recognize and promptly recover from such an UA can easily result in a plane crash.

Over the ten year span from 1998-2007 LOCIF resulted in the greatest number of aviation fatalities for the commercial jet fleet, 22 fatal accidents and 2051 lives lost (Boeing, 2008). For corporate aircraft accident data from 1982 to 2002 LOCIF is the only mishap type that had not appreciably decreased (Sumwalt, 2003a). General aviation (GA) numbers are equally dismal. In 2000, there were 261 GA LOCIF accidents, 111 of these were fatal with 179 lives lost (NTSB, 2001). This accounted for 14% of all GA mishaps that year. 2001 saw 233 mishaps from LOCIF, 13% of all mishaps that year (NTSB, 2002). Although the LOCIF rate went down slightly, it still caused 111 fatal crashes with 190 lives lost. These numbers have not improved; in 2005 18% of all fatal GA accidents were due to a LOCIF (FAA, 2006).

Several studies have examined how well certification requirements ensure that pilots are capable of recovering from an UA. According to one study 58% of professional pilots and nearly 72% of general aviation pilots were unable to recover from LOCIF upsets (Regional Aviation News, 2008). "As evidenced by our research results, pilots are ill-equipped to deal with loss of control scenarios beyond the accepted limitations of their training requirements during pilot certification and recurrent simulator training" (Regional Aviation News, 2008, p. 1).

Another study corroborated these findings by noting that unusual attitude recovery training for commercial pilots was inadequate (Gawron, Berman, Dimuskes and Peer, 2003). In particular, it was noted that the stress of the unexpected scenario as well as the demand for immediate and correct analysis of the situation and correct action were overwhelming for many of their participants, directly inhibiting their ability to recover the aircraft. According to a director of flight operations for a major U.S. based carrier another key skill that has been reported missing from the aviation industry's efforts to reduce LOCIF is training pilots to be able to "recognize potential upset conditions" (Sumwalt, 2003b, p 14).

Unfortunately, one of the primary instruments designed for pilot orientation, the attitude indicator (AI), has been indicted as part of the problem. Numerous studies have shown that the western AI variant may lead to pilot misinterpretation and subsequently a recovery in the incorrect direction, complicating the problem and robbing the pilot of precious time. This situation is known as a roll reversal. Interestingly, the Soviet AI variant utilizes a different display type and has been shown in multiple studies to be more intuitive and less likely to induce roll reversals (Previc and Ercoline, 2000).

LOCIF consistently costs hundreds of lives and millions of dollars annually. Something must be done to help improve pilots' abilities to recover from UA and prevent these disasters. Some of this problem can be attributed to pilot disorientation, some to instrument misinterpretation and some to pilots failing to identify the developing situation. This paper proposes a comprehensive and intuitive solution to this problem by improving pilot awareness of developing aircraft attitude problems, reducing the opportunity to misinterpret their instruments, providing guidance for an expeditious recovery and in accomplishing these objectives reduce the threat of loss of control in-flight.

Proposed Solution

The solution proposed is to make modifications to the AI that will help improve its interpretation and have it provide corrective guidance in throttle settings as well as in the roll and pitch directions. Three specific modifications are being proposed: a horizon indicator that improves ability to discern horizon location, roll and pitch guidance and throttle guidance. These modifications should be compared with both the Western and Soviet AI displays to determine which combination improves performance best.

Horizon Indicator

The horizon indicator is designed to help pilots more readily discern the horizon's location. Isosceles triangles drawn into the ground and sky of the AI can meet this goal. The apices of these triangles would meet at the horizon line to form an hour glass with their opposing lines residing on their respective 90^o pitch lines. To further elucidate horizon location the inside of the triangles would be colored with a gradient that is dark near the 90^o pitch lines but pale near the horizon line. In this way, regardless of the orientation displayed on the AI, the pilot will always be able to discern the horizon location by following the visible triangle's contours and color gradient to its apex. This intervention seeks to mitigate conditions such as seen in Figure 2 where no horizon line is evident, which can easily lead to a roll reversal.

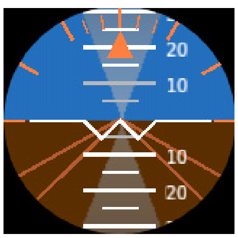


Figure 1. Horizon indicator at horizon.

Figure 2. Extreme unusual attitude without horizon indicator.

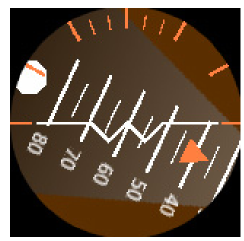


Figure 3. Extreme unusual attitude with horizon indicator.

Roll and Pitch Indicator

The second modification is to add roll and pitch correction indicators. Roll and pitch correction indicators would flank the perimeter of the AI (Figure 4). These indicators would illuminate with a green hue in the direction of correction necessary to recover to straight and level flight. These indicators are designed to illuminate to improve their salience. Not only do they instruct a pilot of correct recovery inputs but also promptly raise awareness of a hazardous situation as it develops, thereby compensating for the cognitive deficit experienced by a task saturated, complacent or distracted pilot.

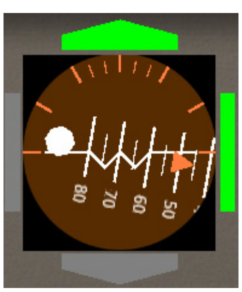


Figure 4. Roll and pitch indicators commanding a right bank and a pull up.

To prevent human misuse of automation (e.g., overtrust) these indicators would only illuminate when the aircraft attitude is approaching a situation that could devolve into danger. Should the pilot fail to recognize these warnings and the attitude become worse, the appropriate roll and pitch indicators would blink to further increase their salience. The turn off threshold should be set at a value less than the turn-on threshold. This gap would prevent issues where a pilot flying on the cusp of the roll indicator's illumination settings would cue the indicator to turn on and off with slight changes in bank. This type of situation could lead to a detriment in safe flight performance as it might result in annoyance, distraction or complacency.

Another item to be considered with the roll guidance indicators is an inverted attitude. In this situation the wings would be straight and level so the bank indicators would not be illuminated. This could be extremely dangerous because if the plane concurrently had a nose low attitude the pilot's roll and pitch indicators would instruct the pilot to pull back on the controls—in effect complicating the situation and flying the aircraft to the ground. To prevent this, the system could be programmed to account for inversions. Should an inversion occur the roll indicators should instruct the pilot to recover to the right because people are biased toward this direction (Wickens, personal communication, September 9, 2008).

Throttle Guidance Indicator

The first two modifications were designed to help pilots more easily interpret their AI and provide guidance to recover to straight and level flight. The final modification examines another key aspect of aircraft control and unusual attitude recovery—the energy state. An attempt to correct an unusual attitude without attending to the current energy state may stall the aircraft, exacerbating the situation. The throttle guidance indicator was designed to assess the aircraft's total energy state and make recommendations on whether more or less throttle input is necessary to recover the aircraft (Figures 5 and 6). To prevent automation misuse the throttle indicator should only be active when the roll and pitch indicators are illuminated. The throttle indicator would be displayed off-center of the AI and move vertically to indicate necessary changes in throttle setting. This display would be high on the AI if the system should decrease its energy (Figure 5); conversely, if an increase in energy state is warranted the indicator would be low on the AI (Figure 6). If the aircraft has the appropriate amount of energy the display would be in the vertical center of the display and covered by the aircraft watermark and 90^0 bank lines; and thus, not be visible when unnecessary.

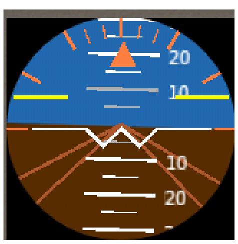


Figure 5. Throttle guidance indicator prompting a decrease in energy state.

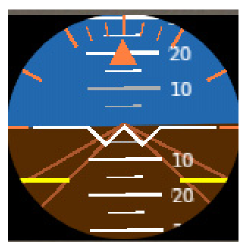


Figure 6. Throttle guidance indicator prompting an increase in energy state.

Theoretical Foundations

Psychological Display Dynamics

First, it should be discussed why visual symbols should be used rather than alternatives (e.g. auditory or tactile recovery cues). Visual information is placed into the visuospatial sketchpad of spatial working memory which people use to help them in mental manipulation, recalling items they see and executing actions (Wickens and Hollands, 2000). In an unusual attitude the pilot is trying to control the aircraft and make the AI "read right," spatial working memory activities; therefore, it is wholly appropriate that the display be pictorial. Additionally, accuracy and speed of recognition is greatest if displayed stimuli are presented in a format commensurate with the unit in memory's representation, indicating that a pictorial display should also accelerate recovery (Wickens and Hollands, 2000). Abiding by the proximity-compatibility principle and co-locating information on energy state, roll and pitch information may also accelerate recovery and ease the effort of an instrument scan. Lastly, in time constrained situations with high cognitive demand, such as an unusual attitude recovery, a command display can be best as they save pilots one extra processing step, thereby accelerating response and reducing error rates (Wickens and Hollands, 2000). This is why the roll, pitch and throttle indicators were chosen to be command displays.

AI Display Type

The discussion of superiority between the Western and Soviet AI has been examined for decades. Previc and Ercoline (2000) conducted a thorough analysis of this literature and made an extremely convincing argument of the superiority of the Soviet display. The Soviet display presents an aircraft symbol that banks right and left against a vertically moving background. This display type is known as an outside-in (O-I) display because it is similar to what an individual outside of an aircraft would see looking at an aircraft. The Western display holds the aircraft stationary and the world shifts. This display is conformal to what a pilot in an aircraft would see when looking out at the horizon and is hence known as an inside-out (I-O) display. Problems with the I-O display center around the fact that people have a tendency to control the moving part and hence enter control inputs based on the world's motion, not the aircraft's—exactly opposite that which is appropriate, leading to roll reversals. It would be interesting to see if the I-O display with these modifications can outperform the O-I display or if the O-I display would be even further enhanced with these modifications.

Horizon Indicator

The primary advantage of the horizon indicator is that the triangles always point toward the horizon and the triangle embedded in the ground always points in the direction (i.e. left or right) in which the controls must be moved to regain level flight from a bank. The color gradient further supports this by capitalizing on the innate human perception of aerial perspective—as objects get further away their color becomes bluer and paler.

This is not the first time that such a modification has been evaluated. Liggett, Reising and Hartsock (1992) attempted this in their background attitude indicator display, except they used a wedge rather than a triangle. Pilots performed best when color shading with a trapezoid pointing to the horizon was used. They theorized that this finding was commensurate with the concept of optical flow fields—their color shading and wedge pattern both functioned to direct the pilot back to the horizon. These cues were thought to be perceived without the necessity of much information processing. In their study they found that pitch lines broke up some of the optical flow and that numbers next to the pitch lines further slowed the pilot's interpretation of the display. In further discussion of this experiment Liggett, Reising and Hartsock (2000) noted a significant double interaction between the wedge and pitch lines with numbers when color shading was present. However, the delay in reaction time, 42msec, was not deemed practically significant by participant matter experts. Furthermore, the pilots were found to prefer the pitch lines with numbers despite their lack of performance enhancement (Liggett, et. al., 2000). Due to preference, the importance of pitch lines and their numbers in indicating absolute orientation and the negligible practical impact on performance, pitch lines and their numbers should be incorporated with this horizon indicator.

Roll and Pitch Guidance Indicator

The roll and pitch indicators were intentionally selected to illuminate on the side in which correction would be necessary rather than on which side they are over-banked/pitched. Research on the Simon effect has shown that stimuli presented to the left or right of a fixation, when stimulus location is irrelevant (i.e., color is the cue), results in faster response if the stimulus location coincides with the assigned response's location (Proctor, Lu, Van Zandt, 1992). Proctor, et. al. also found that response pre-cuing enhances this effect. Although the Simon effect typically discusses horizontal responses, due to a right-left dimensional preference, the Simon effect also works in the vertical direction as well (Proctor, Vu, Nicoletti, 2003). This right-left preference actually serves pilots well as it helps them correct their roll prior to their pitch which is typically appropriate in unusual attitude recovery. Additionally, this display-action compatibility noted by the Simon effect corresponds to performance standards found in instruments with which pilots are already familiar (e.g. instrument landing system needles).

The effectiveness of peripheral cuing for flight guidance during instrument approaches was demonstrated by Hasbrook and Young (1968). In their study they placed peripheral light cues on the yoke of the aircraft that would indicate if the aircraft orientation had strayed greater than 1.5 degrees from straight and level flight. As the magnitude of deviation increased the light cues would blink more rapidly and transition from green to red to provide situational feedback and improve their salience. In this study it was specifically noted that pilots maintained straight and level flight 35% more frequently with solely peripheral cuing (no visible attitude indicator) than they did with their regular instruments. Hasbrook, et. al. credit this to the fact that pilots were able to recognize and correct for their banked situation by peripheral cues even if they were not looking at the attitude indicator. Peripheral cueing resulted in a statistically significant ($p \leq .01$) improvement in unusual bank attitude corrections and no roll reversals; whereas 30% of the participants with their regular attitude indicator had a roll reversal.

Throttle indicator

The movement of the throttle indicator was specifically designed to correspond with the principle of the moving part—"the direction of movement of an indicator on a display should be compatible with the direction of physical movement and the operator's mental model" (Wickens and Hollands, 2000, p. 135). If there is too much energy in the system the pilot would need to pull back on the throttle. This movement of the throttle directly correlates to the throttle indicator moving down on the AI. Adhering to this principle makes the system intuitive and hence more beneficial to the user in cognitively demanding situations, such as unusual attitude recovery.

Summary

This paper sought to propose some modifications to the current attitude indicator used in airframes to help improve safety rates by reducing loss of control in-flight mishaps. Three modifications were proposed: a horizon indicator, roll and pitch guidance indicators and throttle guidance indicators. These modifications were specifically designed to capitalize on multiple experimentally proven concepts to provide better information to the pilot in the dire situation of an unusual attitude. Examining these display types with both the conventional I-O and Soviet O-I attitude indicators may improve pilot performance in preventing, identifying and recovering from unusual attitudes and should be conducted.

Acknowledgments

The genesis for the roll and pitch corrective indicators spawned from the SD command indicator from Small, Fisher, Keller and Wickens (2005) and Wickens, Self, Andre, Reynolds and Small (2007). The idea for the throttle indicator was inspired by the work done by Amelink, Mulder, van Paassen and Flach (2005) in looking at total energy maintenance for aircraft on approach to landing. I would like to thank several people for help in developing this project: Don Talleur for his expertise and guidance with respect to flying and recovery procedures. Ashwin Jadhav for his assistance in translating several Russian articles on attitude indicators and also his aeronautical engineering expertise in developing the throttle indicator. Jonathan Sivier for his help in drawing the illustrations in the figures and in programming an experiment to test these displays. Lastly, my advisor, Terry von Thaden in helping me flesh out these ideas. The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

References

2000 GA accident aircraft data used in annual review. (2001). [Data file]. Available from National Transportation Safety Board web site: http://www.ntsb.gov/

2001 GA accident aircraft data used in annual review. (2002). [Data file]. Available from National Transportation Safety Board web site: http://www.ntsb.gov/

Amelink, M.H.J., Mulder, M., van Paassen, M.M., and Flach, J. (2005). Theoretical foundations for a total energy-based perspective flight-path display. *International Journal of Aviation Psychology*, 15(3), 205-231.

Boeing. (2008, July). Statistical summary of commercial jet airplane accidents, worldwide operations 1959-2007. Seattle WA.

FAA. (2006 March 15). General aviation joint steering committee fatal accident statistics (2005). Retrieved November 20, 2008, from http://www.faa.gov/safety/programs_initiatives/pilot_safety/safer_skies/gajsc/ga_stats2005/

Gawron, Berman, Dimuskes and Peer, (2003). New airline pilots may not receive sufficient training to cope with airplane upsets. *Flight Safety Digest*, 22(7-8), 19-32.

Hasbrook, A.H. and Young, P.E. (1968). Peripheral vision cues: Their effect on pilot performance during instrument landing approaches and recoveries from unusual attitudes. *AM [reports]. United States Office of Aviation Medicine*, Pages 1-17.

Liggett, K.K., Reising, J.M., and Hartsock, D.C. (1992). The use of a background attitude indicator to recover from unusual attitudes. *Proceedings of the Human Factors Society 36th Annual Meeting*, 43-47.

Liggett, K.K., Reising, J.M., and Hartsock, D.C. (2000). Development and evaluation of a background attitude indicator. *International Journal of Aviation Psychology*, 9(1), 49-71.

Nearly 60 Percent of Professional Pilots Fail LOC Tests. (2008, May). *Regional Aviation News*, 6(21).

Previc, F.H. & Ercoline, W.R. (2000). The "Outside-In" attitude display concept revisited. *International Journal of Aviation Psychology*, 9(4), 377-401.

Proctor, R.W., Lu, C. Van Zandt, T. (1992). Enhancement of the Simon effect by response precuing. *Acta Psychologica,* 81(1992), 53-74.

Proctor, R.W., Vu, K.L. and Nicoletti, R. (2003). Does right-left prevalence occur for the Simon effect? *Perception & Psychophysics*, 65(8), 1318-1329.

Schlimm, K. (2005, July 10). Unusual attitude recovery: Reacting quickly in an over-banked situation. AVWeb, Retrieved on July 15, 2008, from http://www.avweb.com/news/airman/190089-1.html

Small, R.L., Fisher, A.M., Keller, J.W. and Wickens, C.D. (2005). A Pilot Spatial Orientation Aiding System. AIAA 2005-7431. p 1239-1253.

Sumwalt, R.L., (2003a), Airline upset recovery training: A line pilot's perspective. *Flight Safety Digest*, 22(7-8), 1-17.

Sumwalt, R.L., (2003b), Preventing the upset from occurring: The first essential step. *Flight Safety Digest*, 22(7-8), 14.

Wickens, C.D. and Hollands, J.G. (2000). Engineering Psychology and Human Performance, Third Edition. Upper Saddle River, NJ, Prentice Hall.

Wickens, C.D., Self, B.P., Andre, T.S., Reynolds, T.J., and Small, R.L. (2007). Unusual attitude recoveries with a spatial disorientation icon. *The International Journal of Aviation Psychology*, 17(2), 153-165.

A COMPATIBILITY ANALYSIS OF ATTITUDE DISPLAY FORMATS

Motonori Yamaguchi
Purdue University
West Lafayette, Indiana
Robert W. Proctor
Purdue University
West Lafayette, Indiana

The present research investigated factors that contribute to the compatibility of attitude display formats with actions taken to control an aircraft. In three experiments, participants performed a speeded response task in which they responded by banking an aircraft according to a nonspatial aspect of lateralized stimuli. The format of attitude display was *horizon-moving* or *aircraft-moving*, and each participant used *normal* and *reversed* controls. These manipulations dissociated influences of three response factors (aircraft, display, hand) on the *stimulus-response compatibility*, or *Simon, effect*. The influences of the three factors on the Simon effect were nearly additive, and their contributions depended on the task contexts. In particular, throughout the three experiments, the major factors in representing responses were the movement directions of the aircraft and the operating hands, and the influence of display format became small as participants directed attention onto the display.

One of the most important factors in interface design is compatibility between the way information is presented in the display and operations that are performed using that information (Proctor & Van Zandt, 2008; Wickens & Holland, 2000). A design principle that is most relevant in the present study is *the principle of moving part* (Roscoe, 1968), according to which the movement direction of display indicators should be consistent with the physical movement and/or operators' mental model of what is indicated. In the conventional design of a glass cockpit, the roll and pitch of an aircraft are indicated by moving the artificial horizon. This format was adopted based on the assumption that the correct format of an attitude indicator is an exact analogue of the pilots' view from the cockpit window (Roscoe, Corl, & Jensen, 1981, p. 343). However, the issue of whether or not this conventional format is really representative of pilots' mental models is a classic one (e.g., Conklin & Lindquist, 1958) that is still a topic of debate (e.g., Previc & Ercoline, 1999). The present paper reports three experiments that examined this issue. More specifically, the current research investigated what factors are relevant to represent the task context and to what extent the relevant factors influence performance.

Stimulus-Response Compatibility and Representation of Task Context

The stimulus-response (S-R) compatibility effect refers to the fact that, in a choice-reaction task, responses are faster and more accurate when stimuli and responses share certain properties, or when they correspond, than when they do not (Proctor & Vu, 2006). This effect is known to be so robust that it is observed even if the S-R correspondence is irrelevant to performing the task. For instance, in a task in which participants are instructed to press a left or right key in response to the color of a stimulus, responses are faster and more accurate when the stimulus occurs on the same side as the location of the correct key than when it occurs on the opposite side. The variation of the S-R compatibility effect that occurs on the basis of task-irrelevant S-R correspondence is termed the *Simon effect* (Simon, 1990). The Simon effect implies that task-irrelevant information is still encoded and affects performance. In turn, it is an indication of how the task context is represented.

The study of task representations in the context of the Simon task can be traced back to Simon, Hinrichs, and Craft (1970). In their experiments, participants were asked to respond by pressing a left or right response key to high- or low-pitch tones that were presented to the left or right ear. Therefore, the task-relevant stimulus feature was the tone pitch, and the side of the ear to which the tone was presented was task-irrelevant. In a condition where the left key was pressed by the left hand and the right key by the right hand, responses were faster if the location of the response hand corresponded with the side to which the tone occurred, yielding a regular Simon effect. However, in a condition where the left key was pressed by the right hand and the right key by the left hand (i.e., when the hands were crossed), responses were faster if the location of the response hand was opposite to the side to which the tone occurred. From these results, Simon et al. inferred that responses were encoded in terms of the key locations, not the hands with which the keys were pressed.

More recently, Hommel (1993) conducted experiments in which high- and low-pitch tones were presented from the left or right speakers, and participants pressed the left or right key in response to the tone pitch. In his experiments, two lights were positioned near the speakers, and a keypress turned on a light that was spatially noncorresponding to the key location (i.e., a left keypress turned on a light on the right, and a right keypress turned on a light on the left). When participants were instructed to press a key, responses were faster if the key location corresponded to the speaker location from which a stimulus was presented, yielding a regular Simon effect. However, when they were instructed to turn on a light, responses were faster if the key location did not correspond to the speaker location, that is, if participants turned on the light that was located at the same side as the speaker location. Thus, in the latter condition, responses were coded in terms of the light location, rather than the key location. Note that the illumination of a light is a distal effect of keypress response. The consequence of one's action, like the illumination of a light as a result of hitting a switch, is called an *action effect*, which Hommel's study has shown to be an important component in representing the task context.

The above studies exemplify the effectiveness of using the Simon task to investigate how people represent a task context. That is, if the Simon effect is observed between a task-irrelevant stimulus feature and a response component of interest, it can be taken as evidence that participants represent the response in terms of that component. The present study uses this paradigm to examine factors that contribute to performance in flight operations with two formats of an attitude indicator.

Present Study

The conventional format of an attitude indicator has been that of horizon-moving, in which the artificial horizon rotates to the left if the aircraft banks to the right, while it rotates to the right if the aircraft banks to the left; on the other hand, the aircraft symbol stays stationary at the center of the display. This format is in accordance with the pilot's perspective of the actual horizon as it is viewed from the cockpit window. However, several researchers suspect that the apparent movement of the horizon may be inconsistent with the pilot's mental representation of the relationship between the horizon and the aircraft (e.g., Previc & Ercoline, 1999; Roscoe et al., 1981). For instance, Patterson et al. (1997) argued that the spatial representation constructed for monitoring the cockpit indicators is different from that constructed for viewing the scene outside the cockpit window. When monitoring the indicators, the spatial representation is based on the coordinate system that is centered at the aircraft, which, as the aircraft changes its attitude, moves in relation to the actual horizon but is stable in relation to the pilot. In contrast, when viewing the outside scene, the spatial representation is based on the coordinate system that is centered at the direction of the gravity, which is stable in relation to the horizon but moves in relation to the pilot when the aircraft changes its attitudes (see also Previc, 1998). The researchers proposed that a format more consistent with the spatial representation for monitoring the attitude indicator is that of an aircraft-moving, in which the artificial horizon stays stationary while the aircraft symbol rotates with the roll of the aircraft.

The aircraft- and horizon-moving displays present physically equivalent information, that is, the relationship between the aircraft's attitude (roll and pitch) and the horizon, but in different ways. Their difference is which display object is actually moving to represent the relationship. According to the *principle of moving part*, the compatibility of the two displays with pilots' operations depends on which aspects of display information enter into the mental representation of the operations. Because the mental representation is not directly observable, one has to use an indirect method to investigate the issue. For this purpose, the Simon task is useful.

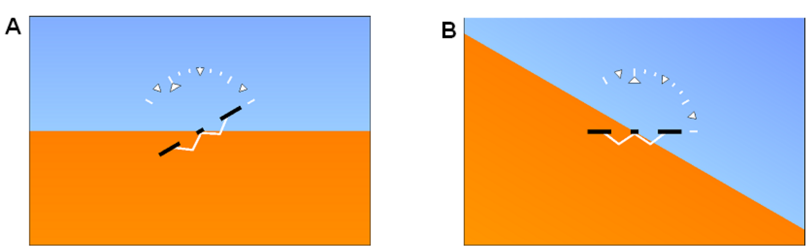


Figure 1. The aircraft-moving (A) and the horizon-moving (B) displays used in Experiments 1-3.

In the present study, participants were asked to bank an aircraft simulated in the computer display with an aircraft-moving format or a horizon-moving format (see Figure 1). Responses were made by turning a flight yoke, and stimuli were auditory tones in Experiment 1 and visual signals in Experiments 2 and 3, both of which contained a spatial feature (left or right). However, the spatial feature of stimuli was irrelevant to performing the task. In the context of these Simon tasks, there are at least three factors that can contribute to task representation. These factors are (a) the movement direction of the operating hands, (b) the movement direction of the display object, and (c) the movement direction of an aircraft simulated on the display. Note that the latter two factors are distal effects of the operating hands, that is, they are action effects. As mentioned, they are known to constitute an action representation.

In the aircraft-moving format, the aircraft symbol rotates to the left or right according to the direction of the yoke input. In the horizon-moving format, the artificial horizon rotates to the left or right but in the opposite direction to the yoke input. Although the roll of the aircraft is simulated differently in the two formats, both displays convey physically equivalent information; the rotational relationship between the aircraft and the horizon. Thus, the aircraft rotates in accordance to the direction of the yoke input in both displays. Therefore, the directions of the operating hands and the aircraft are always consistent. To examine the separate contributions of the three factors to pilots' mental representation, we introduced a manipulation in which the control was either 'normal' or 'reverse'. In the *normal-control* condition, the aircraft banked to the left (right) if the yoke was turned to the left (right). In the *reverse-control* condition, the aircraft banked to the left (right) if the yoke was turned to the right (left). With this manipulation, there were a total of eight trial conditions, which are summarized in Table 1. For the aircraft-moving format with normal-control (the first and second rows), the three response components are all compatible (incompatible) with the stimulus location if one of them is compatible (incompatible). When the control is reversed (the third and forth rows), the compatibility relationship of the operating hand is dissociated from the other two components. Similarly, for the horizon-moving format, the compatibility of the display object is isolated from the other components in the normal-control condition, whereas the compatibility of the aircraft is isolated from other components in the reverse-control condition. Therefore, by comparing the Simon effect for these trial conditions, the extent to which the three factors contribute to mental representation of the task context can be assessed.

Table 1. *Summary of Compatibility Relationships between Stimulus and Three Response Components*.

		Response Component		
Display Format	Control Condition	Aircraft	Display	Hand
Aircraft-Moving	Normal	+	+	+
		–	–	–
	Reverse	+	+	–
		–	–	+
Horizon-moving	Normal	+	–	+
		–	+	–
	Reverse	+	–	–
		–	+	+

Note. Plus (+) and minus (–) signs indicate 'compatible' and 'incompatible', respectively.

General Method

Three experiments are reported. In each experiment, 40 undergraduate students enrolled in Introductory Psychology at Purdue University participated (a total of 120 participants). Participants were randomly assigned to one of the two display formats (aircraft-moving, horizon-moving) and performed two trial blocks between which the control (normal, reverse) varied. The order of the two control conditions was counterbalanced across participants. At the beginning of an experimental session, participants were shown the display that they were about to use during the test trials. They were told that the display represented the roll of an aircraft (the pitch was fixed in the present study) and asked to move the aircraft to familiarize themselves with the display. The attitude indicator was 22 cm in width and 14.6 cm in height (see Figure 1). The viewing distance (unrestricted) was approximately 70 cm. After the familiarization, participants read instructions for the task presented on the computer screen and were told that one block would be the normal-control condition and the other would be the reverse-control condition. Each block began with 12 practice trials, followed by a pause screen. The test trials started when the experimenter pressed a start key. A test block consisted of 156 trials (of which the first 12 trials were considered to be warm-up and thus discarded from the analysis). The experiment was conducted individually in a well-lit cubicle. An experimental session lasted for less than 30 minutes.

Each trial started with the roll angle of 0°. With a 1,000-ms foreperiod, the imperative stimulus was presented. In Experiment 1, the stimuli were high- and low-pitch tones (880-Hz and 440-Hz, 64 dB) presented through headphones to the left or right ear. In Experiments 2 and 3, the stimuli were green and red rectangles (2.2 cm in width and 1.5 cm in height) presented on the left or right above the attitude indicator. Participants wore headphones throughout the session in all experiments.

In Experiments 1 and 2, participants were simply instructed to turn the aircraft to the left or right according to the task-relevant aspect of stimuli. When the aircraft banked 45°, the display was paused until the yoke was returned to the neutral position, after which the roll was automatically set to the initial zero point. In Experiment 3, participants were asked to bank the aircraft to the bank marker at 45° and maintain the roll angle for 1 s (the error window was ±3º from the target position). Therefore, the difference between Experiments 2 and 3 was that the former required participants simply to turn the yoke to the left or right, whereas the latter required more fine control. Note that in both cases, the task instructions were based on the aircraft movement, not the operating hands or the display objects.

Response time (RT) was the interval between stimulus onset and displacement of the yoke from the neutral position approximately 10° to the left or right. A response was considered to be an error when the yoke was turned to a wrong direction beyond this criterion position, but participants were not aware of this response criterion (thus, they could correct before completing a trial, though the response was recorded as an error on that trial). If the eventual response was incorrect, an error message "ERROR" (Experiment 1) or an error tone (440 Hz, 64 dB; Experiments 2 and 3) was presented at the end of a trial.

Results

Trials for which RT was less than 100 ms or greater than 1,500 ms were discarded (< 1% of all trials). Mean RT for correct responses and percentage errors were computed for each participant (we report only the analysis of the RT data due to the space limitation) and submitted to analysis of variances as a function of Correspondence (corresponding vs. noncorresponding; within-subject), Control (normal vs. reverse; within-subject), and Display Format (aircraft-moving vs. horizon-moving; between-subject). Note that the Correspondence variable was coded based on the spatial relationship between stimulus and the correct direction of the aircraft. The Simon effect was computed by subtracting mean RT for the corresponding trials from mean RT for the noncorresponding trials, which is summarized in Figure 2.

Experiment 1: Auditory Stimuli

There was significant interaction between Display Format and Correspondence, $F(1, 38) = 5.11$, $MSE = 331$, $p < .030$, and between Correspondence and Control, $F(1, 38) = 42.08$, $MSE = 360$, $p < .001$. The Simon effect was larger for the aircraft-moving format ($M = 28$ ms) than the horizon-moving format ($M = 15$ ms), and for the normal-control condition ($M = 41$ ms) than for the reverse-control condition ($M = 2$ ms). However, the three-way interaction of the Display Format, Correspondence, and Control was not significant, $F(1, 38) = 2.68$, $MSE = 360$, which indicates little evidence for a violation of additive effects of Display Format and Control on the Simon effect.

From Table 1, the larger Simon effect for the aircraft-moving format than the horizon-moving format implies a contribution of the display motion to the response representation. Similarly, the larger Simon effect for the normal-control than the reversed-control implies a contribution of the operating hands to the response representation. Also, there was still a significant main effect of Correspondence, $F(1, 38) = 57.76$, $MSE = 331$, $p < .001$, which yielded a 21 ms of the Simon effect. This observation implies a contribution of the aircraft to the response representation (because the null effect is expected if there is no contribution of that factor). Therefore, Experiment 1 suggests that all three components contribute additively to the response representation.

Experiment 2: Visual Stimuli

As in Experiment 1, there was a significant interaction between Correspondence and Control, $F(1, 38) = 5.25$, $MSE = 682$, $p < .028$. The Simon effect was larger for the reverse-control condition ($M = 38$ ms) compared to the normal-control condition ($M = 19$ ms), implying a contribution of the operating hands to the response representation. In contrast to Experiment 1, however, the interaction between Correspondence and Display Format was not significant, $F(1, 38) < 1$. That is, the contribution of the display motion to the response representation was very small when visual stimuli were used. However, a main effect of Correspondence was still significant, $F(1, 38) = 32.63$, $MSE = 1{,}007$, $p < .001$, which implies the contribution of the aircraft movement to response representation. The 3-way interaction of Correspondence, Display Format, and Control, $F(1, 38) < 1$, was not significant, as in Experiment 1.

Experiment 3: Fine Control

There was a significant effect of Correspondence, $F(1, 38) = 118.06$, $MSE = 781$, $p < .001$. Thus, the contribution of the aircraft movement to response representation appears to be an important factor when fine control of the aircraft attitude is required. However, the interaction between Correspondence and Control or between Correspondence and Display Format was not significant, *F*s(1, 38) < 1. Thus, in contrast to the preceding experiments, neither the contributions of the operating hands nor the display motion were apparent in Experiment 3. There was a trend of a main effect of Display Format, $F(1, 38) = 4.02$, $MSE = 26{,}574$, $p < .052$, reflecting faster responses for the horizon-move display ($M = 566$ ms) than for the symbol-move display ($M = 617$ ms). This outcome is, however, probably a between-subject error because the effect was not consistent throughout the three experiments (responses were faster for the horizon-moving in Experiment 1 but slower in Experiment 2).

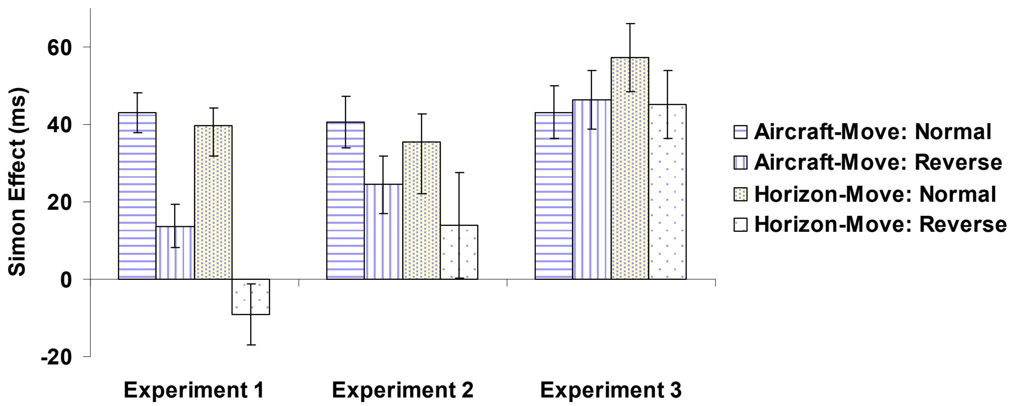


Figure 2. Compatibility effects as a function of display format and control conditions for Experiments 1-3.

Discussion

The study of the display format in attitude indicator has been centered at the compatibility of display information with pilots' mental representations of flight operations. To address this issue, we used the Simon task that has been shown to be effective to investigate the influences of task features on stimulus and response representations. A particular focus was placed on how action (response) is represented in controlling the roll of an aircraft while monitoring an attitude indicator.

The three factors of interest were the movement direction of (a) the aircraft simulated on the display, (b) the display object, and (c) the operating hands. The former two factors are action effects that result from the physical action taken by the operator, which have been shown to be important factors in representing one's actions. To examine the contributions of the two action effects and the physical action to the mental representation of the task context, we dissociated their influences by introducing the normal- and reverse-control conditions. The three experiments also differ in whether the imperative stimuli were auditory or visual and whether a ballistic or fine control of the aircraft was required.

In Experiment 1, where the stimuli were auditory, all three factors influenced the Simon effect, implying that participants represented their responses in terms of the three response components. Because the only difference between the aircraft- and horizon-moving was the movement of the display objects, the most important observation in this experiment is the influence of the display object. That is, the result implies that if the motion of display object is incompatible with the direction that the pilot intends to move, it can interfere with flight operations. Hence, the present experiment supports the advantage of an aircraft-moving format for an attitude indicator, as several researchers have argued (e.g., Patterson et al., 1997; Previc & Ercoline, 1999). However, this conclusion is attenuated by the results of Experiments 2 and 3.

In Experiment 2, where the stimuli were visual, only the aircraft movement and the operating hands were significant factors, and little influence of the display motion was obtained. The use of visual stimuli was likely to have forced participants to pay more attention to the screen, compared to when the stimuli were auditory. At surface, such a manipulation would have increased the influence of the display motion. The results indicate the contrary; the influence of the actual display motion is weakened if participants pay more attention to the display. A likely reason for this outcome is that the display information is interpreted more accurately if participants attend to that information, which is the relationship between the aircraft's attitude and the horizon. If so, Experiments 1 and 2 collectively imply that the advantage of an aircraft-moving format can occur when a sudden change in the flight condition forces the pilot to quickly read the attitude indicator or when the pilot has to quickly shift between multiple displays, but when the pilot continuously monitors the attitude display, the difference between the aircraft- and horizon-moving formats is not influential.

In Experiment 3, participants were asked to roll the aircraft to the bank marker at 45° and maintain the angle for a period of time. Thus, the task required continuous monitoring of the indicator, in contrast to the preceding two experiments. Consistent with the above interpretation, participants represented their actions in terms of the aircraft's movement, and the other two factors were virtually ignored. Thus, when participants had to pay attention to the screen, the display information was correctly interpreted throughout the session.

In conclusion, the three experiments suggest that the two display formats provide equivalent task performance as long as the pilot pays attention continuously to the attitude indicator. However, the advantage of an aircraft-moving format may emerge when the pilot has to read the aircraft's attitude quickly, for example, to recover from an abnormal attitude. It should also be acknowledged that the conclusions are restricted to the type of display used in the present experiments. Whereas the current results are likely to be applicable to a head-down glass cockpit display, which embeds an attitude indicator similar to the one used in the present study, the generalization of the results to different types of displays, such as head-mounted displays and analogue attitude indicators, requires caution. Finally, while the present research relied on a nonpilot population, the validity of the results for the trained pilot population is an important issue for future investigations.

Acknowledgements

The present research was supported in part by Army Research Office MURI Grant W911NF-05-1-0153.

References

Conklin, J. E., & Lindquist, O. H. (1958). Recovery from unusual aircraft attitudes under the influence of vertigo. *Journal of Applied Psychology*, *42*, 136-138.

Hommel, B. (1993). Inverting the Simon effect by intention. *Psychological Research*, *55*, 270-279.

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. *Aviation, Space, and Environmental Medicine*, *68*, 463-471.

Previc, F. H., (1998). The neuropsychology of 3-D space. *Psychological Bulletin*, *124*, 123-164.

Previc, F. H., & Ercoline, W. R. (1999). The "outside-in" attitude display concept revisited. *The International Journal of Aviation Psychology*, *9*, 377-401.

Proctor, R. W., & Van Zandt, T. (2008). *Human factors in simple and complex systems* (2nd ed.). Boca Raron, FL: CRC Press.

Proctor, R. W., & Vu K.-P. L. (2006). *Stimulus-response compatibility principles: Data, theory and application*. Boca Raton, FL: CRC Press

Roscoe, S. N. (1968). Airborne displays for flight and navigation. *Human Factors*, *10*, 321-332.

Roscoe, S. N., Corl, L., & Jensen, E. J. (1981). Flight display dynamics revisited. *Human Factors*, *17*, 341-353.

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (pp. 31-86). Amsterdam: North-Holland.

Simon, J. R., Hinrichs, J. V., & Craft, J. L. (1970). Auditory S-R compatibility: Reaction time as a function of ear-hand correspondence and ear-response-location correspondence. *Journal of Experimental Psychology, 86*, 97-102.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

MODELING THE EFFECTS OF HUD VISUAL PROPERTIES AND CONFIGURATIONS ON A MULTI-DIMENSIONAL MEASURE OF CLUTTER

Sang-Hwan Kim, David B. Kaber, Karl Kaufmann & Theo Veil
Industrial & Systems Engineering, North Carolina State University
Raleigh, NC
Amy L. Alexander & Emily M. Stelzer
Aptima, Inc.
Woburn, MA
Lawrence J. Prinzel
NASA Langley Research Center
Hampton Roads, VA

The objectives of this study were to: validate a multidimensional measure of display clutter for advanced head-up displays (HUDs) incorporating enhanced and synthetic vision systems (EVS/SVS); assess the influence of HUD configuration on pilot perceptions of display clutter and flight performance; and model clutter scores in terms of visual display properties. Eighteen pilots flew a flight simulator in a landing approach using three different sets of HUD configurations (low, medium, or high clutter). Pilot ratings of overall display clutter and its underlying dimensions were recorded along with flight performance measures (deviations from localizer, glideslope, altitude and approach speed). A display image analysis software application was used to measure the visual properties of HUDs. The multidimensional measure of clutter showed internal consistency with high correlations of overall perceived clutter. Calculated clutter scores were sensitive to the various HUD configurations and in agreement with prior display classification (based on the new measure). There was a trend for the extremes of display clutter to cause less stable performance. High clutter displays were associated with cognitive complexity of flight tasks and low clutter was associated with a lack of display information. Multiple linear regression models of perceived clutter were developed based on HUD visual properties.

Future concepts for the National Airspace System integrate cockpit technologies to support flight safety through improved terrain and traffic awareness. A subset of these technologies includes synthetic and enhanced vision systems (SVS and EVS) for pilot use (Bailey et al., 2002). The goal of EVS and SVS displays in the aircraft cockpit is to reduce the incidence of low visibility accidents (Prinzel et al., 2002); however, the design of such display may obscure other important information features when integrated in existing head-down displays (HDDs) and/or HUDs. While these new technologies provide pilots access to information that may not be visible with traditional flight instrumentation, the presentation of this additional information in HUDs may serve to produce visual display clutter (Ververs & Wickens, 1998; Prinzel & Kramer, 2006). Kaber et al. (2008) investigated the effect of HUD features, including EVS, SVS, "highway-in-the sky" (tunnel), traffic collision avoidance system (TCAS) icons and the amount of basic flight information, on experienced pilot perceptions of clutter. Clutter was defined as the presence of irrelevant information (or obscuration of relevant information) in a display. They found that the greater the number of information features and visual density, the higher perceived clutter ratings. Kaber et al. (2008) also conducted a psychological decomposition of the phenomenon of clutter in terms of underlying HUD display qualities. They identified a concise language (set of terms) that expert pilots find useful to assess aviation display clutter, including "redundant/orthogonal," "monochromatic/colorful," "salient/not salient," "safe/unsafe," and "dense/sparse." In addition to this, Alexander et al. (2008) identified bottom-up (data-driven) and top-down (knowledge-driven) factors in pilot perception of display clutter. These studies provided a basis for defining aviation display clutter in terms of display information features and perceived display qualities, as well as developing a multidimensional measure of display clutter. The objectives of the present research were to build upon this initial study by: (1) defining and validating a multidimensional measure of display clutter; (2) assessing the influence of HUD clutter classifications on pilot perceptions of displays and actual flight performance; and (3) modeling subjective HUD clutter scores in terms of objective visual properties, including luminance, target-to-background contrast, feature occlusion, and visual density.

Methodology

Development of Multidimensional Measure of Display Clutter

The set of display descriptor terms selected by pilots in the Kaber et al. (2008) study were used as anchors in a collection of bipolar subjective rating scales covering the underlying dimensions of clutter, including "redundancy (orthogonal/redundant)", "colorfulness (monochromatic/colorful)," "feature salience (salient/not salient)," "feature dynamics (static/dynamic)," "feature variability (monotonous/variable)," and "global density (sparse/dense)." The scales were integrated into an overall clutter index, which required pilots to rank the importance of each dimension for characterizing HUD clutter (in context) and to rate displays on each scale. The ranking of clutter dimensions and the ratings for display were then combined in an overall clutter score (rank-weighted sum of ratings across dimensions). This measurement approach was very similar to the design of measures examining other psychological phenomenon, such as the NASA-Task Load Index (TLX) (Hart & Staveland, 1988) for assessing cognitive workload.

HUD Configurations

Thirty-two different HUD configurations studied by Kaber et al. (2008) were rank-ordered based on measures and predictions of expert pilot ratings of clutter. For the present study, the top 20% of HUDs were classified as "high-clutter," the middle 20% as "medium clutter," and the lowest 20% as "low clutter." From each of these groups, three target displays were selected to represent unique HUD feature sets within each group for a total of nine test displays. Figure 1 shows the nine HUD configurations selected across the three clutter groups.

Participants and Experiment Design

Eighteen current line-pilots with varying levels of flight experience (six with <5 yrs.; six with 5-15 yrs.; six with >15 yrs.) but no HUD experience participated the experiment. They were asked to fly the Integration Flight Deck (IFD) simulator at NASA Langley on a landing approach under low and high workload conditions (no wind or a substantial crosswind, respectively). The IFD was used to present pilots with the standard instrument landing system (ILS) approach to Runway 16R at Reno-Tahoe International airport. Each approach was divided into three segments, including: (1) initial approach fix (IAF) at the PYRAM intersection to glideslope (G/S) intercept; (2) G/S intercept to just inside the final approach fix (FAF; DICEY); and (3) from the end of the preceding segment to decision height (DH), either as published or EVS minimums (as appropriate for the HUD configuration; FAR 91.175c). After completing a training period of basic airwork and an approach under visual conditions, participants completed six test trials. During the test trials, each pilot was presented with a different set of HUD configurations, representing "low," "medium" or "high" levels of clutter. Thus, two between-subject variables (three levels of pilot experience and three levels of display clutter) and two within-subject variables (two levels of flight workload and three segments of flight) were manipulated in the experiment.

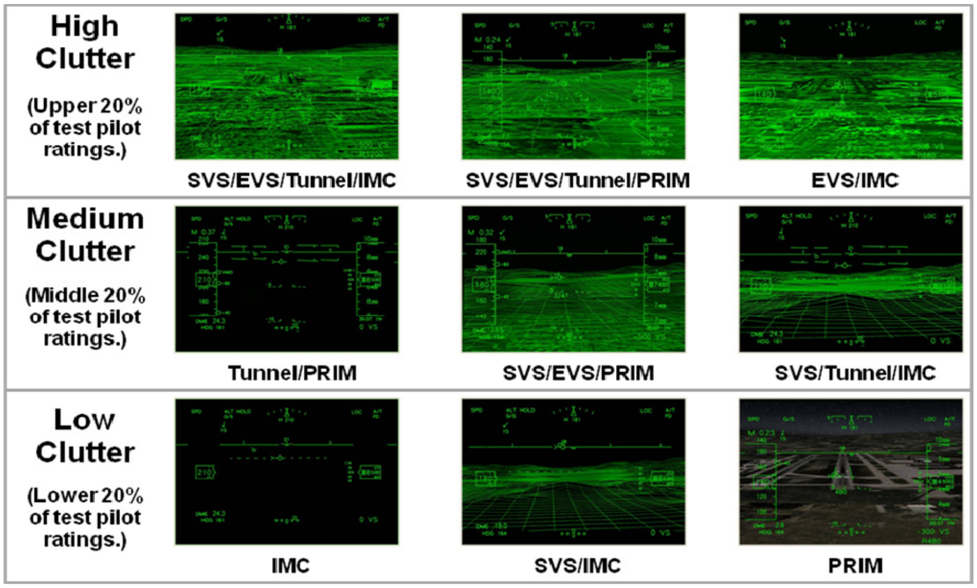


Figure 1. *Nine HUD configurations used in the experiment (IMC = Reduced symbol set for instrument meteorological conditions; PRIM = Complete primary flight display symbol set; and TUNNEL = Highway-in-the-sky guidance).*

Dependent Measures

Pilots subjective ratings of overall perceived display clutter and ratings on the underlying dimensions of clutter were collected at the end of each segment of flight for a trial. We also recorded pilot performance, including localizer (LOC) and glideslope (G/S) deviations and flight altitude and speed control. In addition to this, we developed a display image analysis software application to measure objective visual properties including average contrast of iconic (e.g., airspeed and altitude) to non-iconic (SVS/EVS) features, percent occlusion of iconic features by non-iconic features, and percent overall density for all HUD configurations in advance of the experiments. We also measured the lumens for each HUD in the IFD using a hand-held photometer.

Results

Multidimensional Measure of Display Clutter

Correlation analyses on ratings on the dimensions of clutter with overall perceived display clutter revealed significant positive linear relations for "colorfulness," "dynamics," "variability," and "density". However, "redundancy" and "saliency" were found to be negatively related with overall clutter. Table 1 shows the Pearson coefficients for the subscale and overall clutter ratings and the anchors of the various scales. On this basis, an overall clutter index was formulated by integrating the pilot ratings across all scales (including reversing ratings on "redundancy" and "saliency" (for which maximum scale values were given negative ratings)). Calculated clutter scores were generated by multiplying pilot rankings of dimensions with ratings for each HUD and were found to be highly correlated with overall perceived display clutter ratings (r = 0.77, p<0.0001). These results indicated that the new multidimensional measure of clutter had internal consistency and was valid for describing pilot experiences of clutter.

Influence of HUD Configuration on Pilot Perceptions of Display and Flight Performance

An ANOVA on pilot ratings of display clutter revealed significant effects of pilot flight experience ($F(2,44) = 3.768$, $p= 0.031$) and HUD configuration ($F(2,44)= 5.043$, $p=0.011$). High experience pilots tended to be more sensitive to clutter than the medium and low experience pilots ($p<0.05$).The high clutter display group led to higher ratings than the low clutter group ($p<0.05$). These main effects, however, were moderated by a significant experience by display interaction ($F(4,44)=3.122$, $p=0.024$). The interaction analysis revealed that low experience pilots rated the low clutter displays as more cluttered than the medium and higher clutter displays, perhaps because these displays lacked critical relevant information. Alternatively, medium and high experience pilots provided higher ratings for medium and high clutter displays than low clutter displays (see Figure 2). This finding suggested that display clutter perceptions of higher experience pilots were more consistent with the predefined groups of HUDs in terms of clutter features.

Table 1. *Correlation of subscale with overall clutter ratings and descriptor terms used as scale anchors.*

Subscales	**Correlation with Overall Clutter Ratings**	**Descriptor Terms / Scale Anchor**	
		Lower Clutter	**Higher Clutter**
Redundancy	$r= -0.431$, $p<0.0001$	Orthogonal	Redundant
Colorfulness	$r= 0.237$, $p<0.0001$	Monochromatic	Colorful
Salience	$r= -0.185$, $p<0.0001$	Salient	Not Salient
Dynamics	$r= 0.567$, $p<0.0001$	Static	Dynamic
Variability	$r= 0.474$, $p<0.0001$	Monotonous	Variable
Density	$r= 0.856$, $p<0.0001$	Sparse	Dense

Pilot flight performance was assessed in terms of the degree of variability of flight path control relative to identified targets in various segments of the landing approach. Specifically the kurtosis (degree of centrality) of the distribution of LOC deviations was calculated for all flight segments and the distribution of G/S deviations was calculated in those segments following intercept to evaluate the effects of pilot experience, HUD configuration, and flight task workload. An ANOVA revealed pilot experience to be insignificant in any of the critical performance measures. An interaction of the HUD configuration and segment of flight was highly significant for G/S deviations ($F(2,45)=8.514$, $p=0.001$) and marginally significant for LOC deviations ($F(4,90)=2.130$, $p=0.083$). A main effect of display appeared for G/S deviation ($F(2,45)=3.533$, $p=0.038$) and LOC deviation ($F(2,45)=14.51$, $p<0.001$), but flight segment was only significant for the G/S deviations ($F(1,45)=6.254$, $p=0.014$). There was a clear trend for the extremes of display clutter to cause less stable performance (lower distribution kurtosis). Low clutter displays led to unstable vertical path control, attributable to a lack of critical information. The high clutter displays produced less stable performance, attributable to redundant flight information, and the majority of pilots commented in post-experiment interviews that the displays were “cognitively complex”. In general, the medium clutter HUDs appeared to be superior to low and high clutter displays for performance across all other conditions. Figure 3 shows G/S deviation stability by HUD configuration and flight workload level. Differences in stability across HUD clutter levels were greater for the low workload condition. The same pattern of results also appeared for pilot control of airspeed. Regarding LOC deviations, the more information included in the display, the more stable path control was. This pattern of results also appeared for pilot control of altitude (MSL) on the segment prior to G/S intercept. For both measures (LOC and G/S deviations), pilot control variability appeared to decrease significantly inside the FAF as compared to between the G/S intercept and FAF, possibly due to concentration on vertical path deviations shortly before landing.

Modeling of HUD Clutter Scores in Terms of Visual Properties

Multiple linear regression models of clutter scores were developed based on the objective visual properties of the HUDs measured with the image analysis software. Two separate models were created; one for the low-workload (no crosswind) condition and one for the high-workload (crosswind) condition. It was expected that the crosswind condition would cause greater density of visual features in the HUD (e.g., the crosswind would drive the flight path marker group towards the distally located altitude and speed tapes) and for the pilots to have higher perceptions of clutter. The models for both workload flight conditions were significant (R^2 = 0.33, p<0.0001 for low workload; R^2 = 0.18, p<0.0001 for high workload) for predicting clutter index values and *t*-tests on the lumen, contrast, occlusion, and density parameters. Model parameter tests revealed display lumens, contrast and density to all be significant contributors (p<0.05) to the clutter score; however, occlusion was not significant. Table 2 shows the magnitude and directions of all model predictor terms along with *t*-statistics and significance levels for both the low and the high workload models.

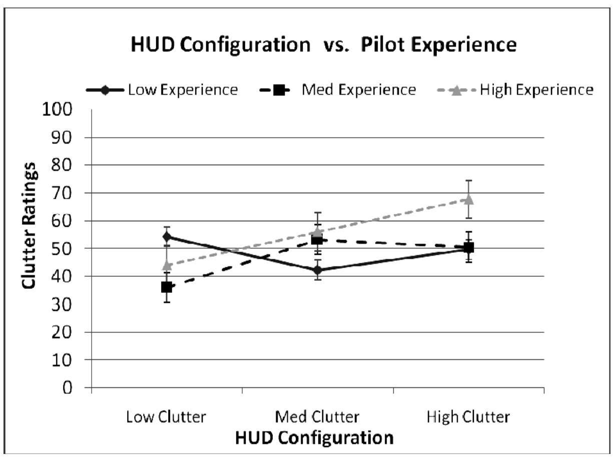


Figure 2. *Clutter ratings by HUD configuration and pilot experience.*

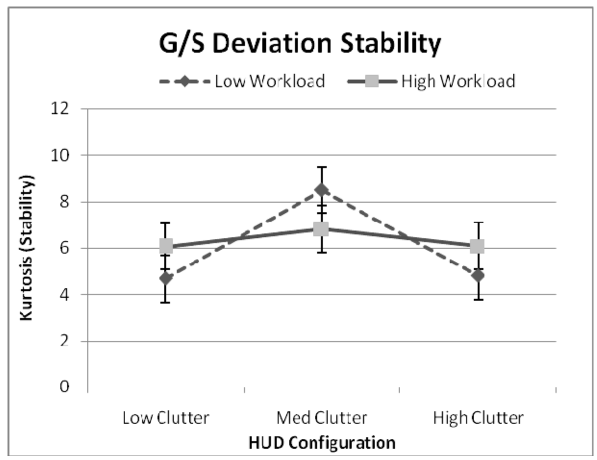


Figure 3. *G/S deviation stability by HUD configuration and flight workload.*

Table 2. *Regression analysis results for model of clutter score in visual display properties.*

Workload	Parameter	Estimate	Error	*t* Value	Pr > |*t*|
Low-workload (no crosswind) condition	Intercept	29.26	3.49	8.39	<0.0001
	Lumen	-12.29	3.05	-4.03	<0.0001
	Contrast	-10.38	2.92	-3.55	0.0005
	Occlusion	1.12	0.40	2.80	0.0057
	Density	15.69	3.91	4.01	<0.0001
High-workload (crosswind) condition	Intercept	39.77	3.54	11.23	<0.0001
	Lumen	-7.75	3.09	-2.51	0.0132
	Contrast	-8.70	2.97	-2.93	0.0039
	Occlusion	0.55	0.41	1.35	0.1785
	Density	9.87	3.97	2.48	0.0141

In general, decreases in display lumens and contrast and increases in occlusion and density caused increases in clutter scores. Average active pixel count (density) and average lumens of the HUDs appeared to account for the greatest amount of variability in the calculated clutter scores. Occlusion was not as strong a contributor as expected from the pilot subjective ratings. It is possible that this was due to a narrow stroke width of the SVS, EVS and tunnel features in the display configurations.

Discussion and Conclusion

The present study achieved the objectives of developing a new multidimensional measure of display clutter, which proved to be sensitive to manipulations of HUD configurations. This measure was correlated with overall ratings of perceived clutter, supporting construct validity. Experiment results indicate that pilot experience plays a role in perceptions of clutter, with high-time pilots being more accurate and consistent in judging the occurrence of clutter. Across workload and performance measures, negative effects of low and high clutter displays were observed, suggesting some optimal amount of HUD information may exist in terms of avoiding information overload while supporting flight path control. Software-based analysis of HUD images yielded visual property measures that were highly predictive of clutter. This indicated that pilot perceptions of clutter in new HUD designs could be projected based in part on low-level display characteristics. Top-down factors, such as the information feature content and task relevance also need to be considered in future extension of such models. One additional caveat of this research is that we did not use a full-motion flight simulator for the experiment; however, few pilots directly commented on the absence of kinesthetic cues or suggested a potential influence on clutter evaluations.

The results of this research, including the multidimensional measure of clutter, and model of perceived clutter in terms of visual display properties, are expected to be applicable for evaluation of a range of NextGen display concepts, beyond EVS/SVS HUDs. One direction of future work would be to apply the new multidimensional subjective measure of clutter for evaluating air traffic management support display technologies for the occurrence of clutter and to assess the reliability of the measurement outcomes.

Acknowledgement

This research was supported by NASA Langley Research Center under Grant No. NNL06AA21A. Randall Bailey and Lawrence Prinzel were the technical monitors. The opinions and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of NASA.

References

Alexander, A. L., Stelzer, E. M., Kim, S-H., & Kaber, D. B. (2008). Bottom-up and top-down contributors to pilot perceptions of display clutter in advanced flight deck technologies. In *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Bailey, R. E., Parrish, R. V., Kramer, L. J., Harrah, S., & Arthur, J. J. (2002). Technical Challenges in the Development of a NASA Synthetic Vision System Concept. In *Proceedings of the North Atlantic Treaty Organization (NATO) Symposium on Enhanced and Synthetic Vision Systems*.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (Vol. 1, pp. 139-183): Elsvier, Amsterdam.

Kaber, D. B., Alexander, A. L., Stelzer, E. M., Kim, S-H., Kaufmann, K. & Hsiang, S. M. (2008). Perceived clutter in advanced cockpit displays: Measurement and modeling with experienced pilots. *Avia. Space & Env. Med.*, 79(11), BOI: 10.3357/ASEM.2319.2008.

Prinzel, L. J., Kramer, L. J., Comstock, J. R., Bailey, R. E., Hughes, M. F., & Parrish, R. V. (2002). NASA synthetic vision EGE flight test. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Prinzel, L. J., & Kramer, L. J. (2006). Synthetic Vision Systems. In W. Karwowski. (Eds.), *International Encyclopedia of Ergonomics and Human Factors* (pp. 1264-1271), 2nd Ed., Talyor & Francis.

Ververs, P. M., & Wickens, C. D. (1998). Head-up displays: Effects of clutter, display intensity, and display location on pilot performance. *International Journal of Aviation Psychology,* 8(4), 377-403

DESIGN OF AN ECOLOGICAL VERTICAL SEPARATION ASSISTANCE COCKPIT DISPLAY

F. M. Heylen, S. B. J. van Dam, M. Mulder, M. M. van Paassen
Faculty of Aerospace Engineering, Delft University of Technology,
Kluyverweg 1, 2629 HS Delft, The Netherlands.

A tactical navigation support tool was designed to effectively deal with conflict situations in the vertical plane, while preserving travel freedom as much as possible. Based on Ecological Interface Design principles, the Vertical Separation Assistance Display is developed as an extension to the existing Vertical Situation Display. Functional information is presented via overlays that show pilots how their vertical maneuvering possibilities are constrained by ownship performance, and by limits imposed by surrounding traffic. A questionnaire-based evaluation shows that the ecological overlays considerably improved pilot traffic awareness in vertical conflict situations.

Airspace congestion and delays force airspace authorities and governments to explore more effective ways to manage air transportation. Novel Air Traffic Management concepts such as the Next-Generation Air Transportation System (FAA, 2008), and Single European Sky ATM Research (SESAR Consortium, 2007) initiatives, advocate the potential benefits of adopting a more flexible approach to ATM. In the future, during cruise flight aircraft may obtain more freedom to optimize their trajectories by allowing 'direct routing' and 'cruise climb'. In order to reduce the workload of the air traffic controller in this situation, the separation task is delegated to the flight deck. The problem of how to assist pilots in this task has attracted great interest in the research community, and several solutions have been proposed in the past decade (Merwin & Wickens, 1996; Johnson et al., 1997; Thomas & Johnson, 2001; Hoekstra, November 2001).

Many of these proposed airborne separation assistance tools provide pilots with *explicit*, ready-to-use automated solutions. This has proven to be effective as far as providing conflict resolution and reducing workload are concerned. However, the use of explicit solutions holds pilots back from exploring solutions other than those presented, and therefore may preclude full exploitation of airspace capacity. Also, the explicit advice often fails to show the 'cognition' behind the automation that deals with the separation problem, and requires cognitive effort from pilots to mentally integrate the different pieces of traffic-related information before they fully understand the conflict situation.

In this paper an alternative airborne self-separation assistance tool for the vertical plane is described. Adopting the principles of Ecological Interface Design (Vicente & Rasmussen, 1992), the Vertical Situation Display is extended with graphical overlays that present functional information regarding how the own aircraft vertical maneuvering possibilities are constrained by the ownship vertical flight performance limits, and by limits imposed by surrounding traffic. The resulting display, the Vertical Separation Assistance Display aims in particular at supporting pilots in maintaining a high level of traffic Situation Awareness (Endsley, 2000).

Ecological Approach

Ecological Interface Design (EID) is an interface design framework that addresses the cognitive interaction between users and complex socio-technical systems, and was originally applied to process control (Vicente & Rasmussen, 1992). Its approach to interface design gives priority to the worker's environment, concentrating on how it imposes constraints on the work. EID principles have been applied to support pilots in various tasks, including an interface for horizontal separation assistance support (Van Dam, Mulder, & Van Paassen, 2008). The Vertical Separation Assistance Display (VSAD) presented in this paper can be considered the 'vertical' complement of this earlier design.

Such an 'ecological' separation assistance tool would aim to visualize the separation problem in such a way that it reflects the cognition needed to cope with the conflict geometry in motion, while at the same time preserving maximum pilot maneuver freedom. EID is a design framework that provides useful tools to achieve these objectives. When adopting its design guidelines, two main questions need to be addressed (Vicente & Rasmussen, 1992). First, how can the content and structure of the work domain be described in a psychologically-relevant way? And second, in which form can this information be effectively communicated to the operator? In this paper, these questions are addressed through a Work-Domain Analysis, followed by an ecological interface design, which aims to visualize the constraints and means-end relationships in the environment in such a way, as to fully take advantage of the human capacity to directly perceive, and act upon cues from the environment.

Work Domain Analysis

The first step of ecological interface design consists of a workspace analysis, using Rasmussen's Abstraction Hierarchy (Rasmussen, 1986). Using the Abstraction Hierarchy (AH), the principal work domain functions and constraints can be identified. The boundaries of the work domain in this study are restricted to the task of self-separation in the vertical plane, during cruise flight. The pilot task consists of the on-board path (re-)planning of climb or descent maneuvers, with the main goal of separating themselves from other traffic in the vicinity.

Minimal separation can be defined using a Protected Zone (PZ), a virtual coin-shaped area, around each aircraft, which is to remain free of other aircraft. General dimensions for the PZ are: 5 NM horizontally, and 1000 ft vertically. A conflict occurs when two aircraft would enter each other's PZ at some instance in the near future, if neither aircraft changes its flight path. Many different ways of detecting a conflict and providing potential resolutions have been proposed; for a review see Kuchar & Yang (Kuchar & Yang, 2000).

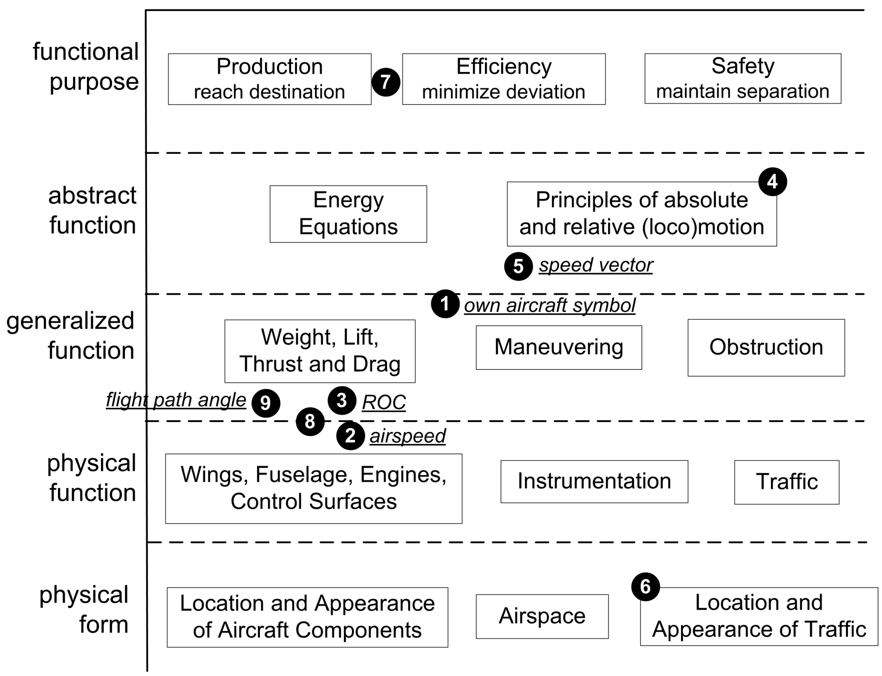


Figure 1: Abstraction Hierarchy for tactical navigation in the vertical plane.

Figure 1 shows the Abstraction Hierarchy that has been developed for the tactical navigation in the vertical plane. The abstraction hierarchy is a stratified hierarchical description of the workspace, defined by means-end relationships between the adjacent levels. Along the vertical axis, the five levels of the AH represent the constraints at decreasing levels of abstraction, starting at the top with the purpose(s) for which the system was designed, all the way down to the spatial topology and appearance of the components that make up the system on the bottom level. Along the horizontal axis, constraints are arranged from internal constraints on the left, to external constraints on the right.

At the functional purpose level, the purposes of the system are defined. As for most transportation systems, three main purposes can be identified at this level, production, efficiency and safety. Here, safety relates to staying within the performance envelope, maintaining separation. The efficiency purpose is to resolve and prevent conflict situations by minor deviations of the planned flight path. The production goal is to fly towards the destination of the programmed flight path. The abstract function level in this case contains the general physical laws that dictate locomotion. The general function level describes how the causal laws at the abstract function level are achieved, independent of the actual implementation of the system. Properties such as weight, lift, thrust and drag all impose internal constraints on aircraft behavior. Obstruction describes other traffic as external constraints. The physical function level describes the various components, and their capabilities and states, and at the physical form level the appearance and location of components, the airspace, and other aircraft are described.

Internal Aircraft Constraints

The internal constraints are defined by minimum and maximum speed and thrust. The minimum velocity is the stall speed. The maximum velocity is the never-exceed speed. Two particular figures of merit related to maximal thrust are the steepest (SC) and fastest climb (FC). SC flight establishes the maximum flight-path angle that an aircraft can achieve. FC occurs when the rate of climb is maximal. In gliding flight, aircraft fly on idle thrust. The minimum and maximum thrust settings yield non-linear contour lines for the flight-path at various airspeeds, Figure 2. These contours depend on aircraft type, configuration, and altitude. In this paper, a model of the Cessna Citation I is used, trimmed at 16,405 ft and 292 kts True Airspeed (TAS), in clean configuration.

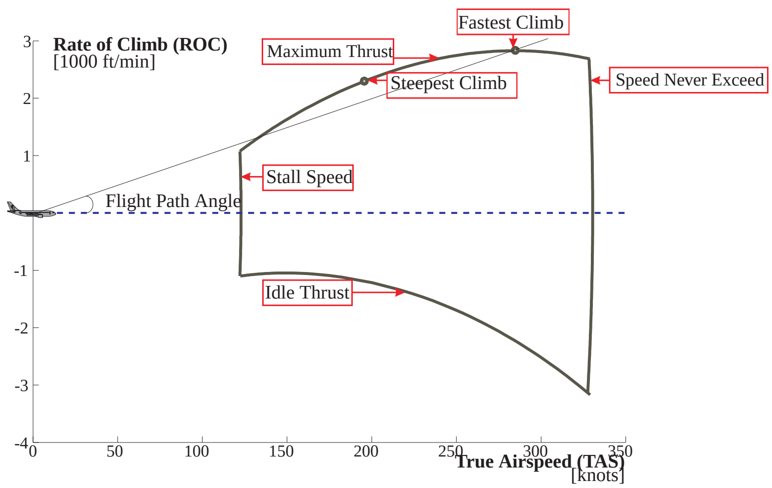


Figure 2: Performance envelope of the Cessna Citation I, in TAS/ROC state space.

External Traffic Constraints

The position and motion of 'traffic' in the vicinity of the own aircraft determine the external constraints on the maneuvering of the own aircraft. A conflict will occur if the speed vector relative to the intruder points in the direction of the intruder aircraft protected zone, Figure 3. This can also be visualized by drawing a beam-shaped area, originating from the ownship position and tangent to the outer sides of the rectangular shape of the protected zone, from hereon called the 'Forbidden Beam Zone' (FBZ). If the tip of the relative velocity vector lies within or moves into this FBZ, separation will eventually be lost. In order to be able to combine the internal and external constraints, the external constraints are translated to the aerodynamic reference frame. In this frame, the conflict geometry is presented from the perspective of the own speed vector, by translating the FBZ over the intruder's speed vector, see Figure 3(c). Then, the pilot should simply move the own aircraft speed vector out of the FBZ to resolve the conflict. If multiple conflicts occur simultaneously, the FBZ's are superimposed after being translated and presented in the absolute velocity plane. This allows pilots to choose a 'global' solution that avoids all FBZ's at once. The combination of the performance overlay and the conflict geometry overlay is called the State Vector Envelope (SVE) (Van Dam et al., 2008).

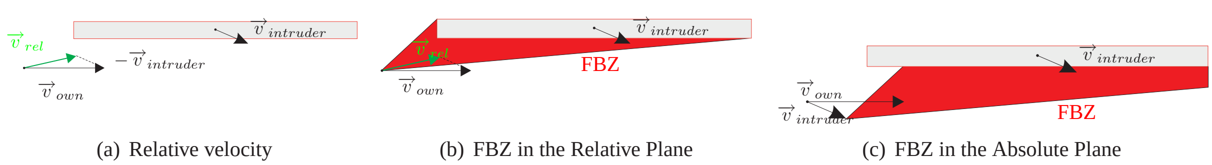


Figure 3: Definition of the Forbidden Beam Zone (FBZ), in the relative and absolute velocity planes.

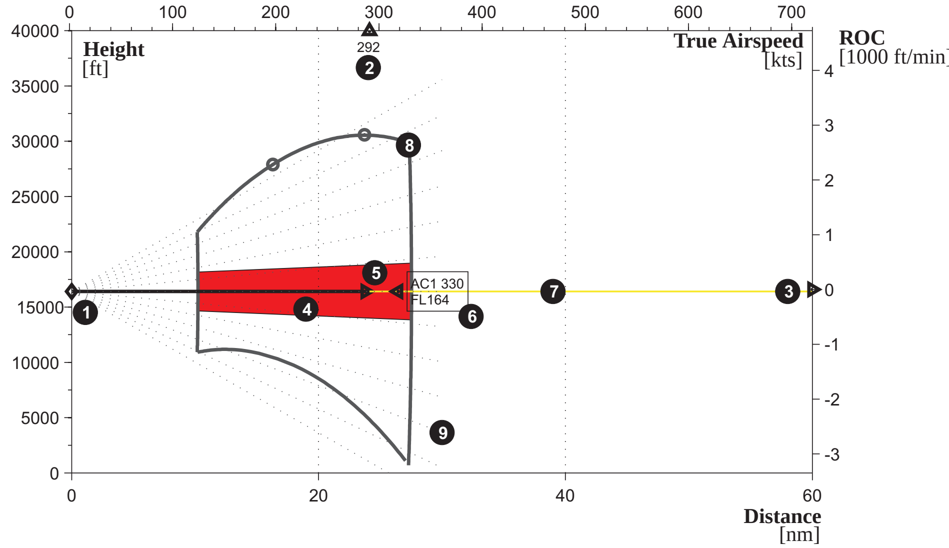


Figure 4: The Vertical Separation Assistance Display (VSAD).

Interface design

The VSAD has been implemented using an existing VSD standard (Prevot & Palmer, 2000), adding layers of functional information identified in the previous section. Since the VSD describes vertical space in terms of distance and height, a transformation of the vertical speed towards height and the horizontal speed towards distance was needed. For this purpose, a horizontal and vertical speed overlay was added on the VSD. The scaling of the speed overlay was based on a prediction time of five minutes, a prediction interval that is frequently used for the detection of conflict situations (Hoekstra, November 2001; Kuchar & Yang, 2000).

Figure 4 shows the VSAD. It integrates the performance envelope of Figure 2 and the conflict geometry visualization of Figure 3(c) in a conventional VSD. Here, ❶ is the own aircraft symbol, ❷ is the speed indicator, ❸ is the ROC indicator, ❹ is the conflict geometry overlay, ❺ is the own speed vector, ❻ shows the intruder aircraft with a label containing callsign, true airspeed and flight level, ❼ shows the own aircraft programmed flight path, ❽ is the performance envelope overlay, transformed to the 5 minute time interval, and ❾ shows potential flight path angle settings in one-degree intervals. These numbers also correspond with the numbers in the abstraction hierarchy, Figure 1. The use of the prediction time means that the performance envelope of the aircraft represents any location the aircraft can reach within that time frame. The speed vector represents a trajectory predictor within the VSAD, based on the current state. Three markers for the Rate of Climb (ROC), airspeed and altitude give the pilot an additional reference to this prediction.

Evaluation

To check whether the Vertical Separation Assistance Display is set to meet its main goal of supporting pilot traffic SA, an evaluation was conducted, with twelve professional airline pilots, with extensive experience with glass cockpits. Pilots were shown movies of 20 to 30 seconds, illustrating dynamically a certain conflict situation in the vertical plane. Using a set of questionnaires before and after the experiment, pilot situation awareness was measured in a systematic fashion. Two display configurations were compared: the Vertical Situation Display (VSD), and the Vertical Separation Assistance Display (VSAD). Ten scenarios were designed that were considered to best represent six ‘typical’ conflict situations. These consisted of opposite maneuvers, parallel maneuvers, overtake maneuvers, situations with multiple intruders, and situations where no conflict is present. For each scenario, between 1 to 5 intruder aircraft were simulated. The overtake scenarios, or in fact, any scenario where traffic was not visible on the VSD, were considered ultimate test-cases for one of the benefits of the VSAD, although here a comparison with the VSD is not possible. It was hypothesized that the traffic SA scores depend neither on the number of intruder aircraft, nor on the conflict situation. Also, the VSAD was hypothesized to significantly improve pilot traffic SA.

Results and Discussion

Regarding their flight strategy, 7 of the 12 pilots indicated that, primarily based on their day-to-day experience, they preferred to resolve a conflict by changing velocity, not altitude. This is contrary to pilots' preferred strategies in the horizontal plane, collected in previous work on the horizontal separation assistance display (Van Dam et al., 2008), where pilots indicated that they preferred heading changes over speed changes. Pilots further commented that during cruise flight it is often impossible to climb higher or fly any faster. Note that this would indeed be shown by the VSAD, through the performance envelope overlay, but none of the scenarios involved cruise flight near maximum altitude.

Rather surprisingly, the answers from the pre- and post-questionnaires indicate that pilots were more appreciative of the performance envelope overlay in the VSAD *before* the dynamic questionnaire. In the post-questionnaire, 4 out of 12 pilots judged the overlay to be 'too theoretical', whereas another 2 pilots found that not all boundaries were necessary. Tentatively, this reflects their preferred flying strategy to resolve conflicts through changing speed only, a strategy for which the aircraft climbing capability, presented through the minimum and maximum thrust contours, would be irrelevant.

Linking of the conflict geometry to the conflicting aircraft was initially thought to be easy if the number of intruder aircraft stays limited. After the dynamic questionnaire, however, 8 out of 12 pilots found it hard to detect which conflict geometry belongs to which intruder aircraft. It can be concluded that, generally, pilots found it easy to attach information presented by the VSAD with data from the PFD. Some pilots (3) found it unnecessary to have any additional links between both displays, 6 other pilots appreciated the speed vector presentation in the VSAD though.

Regarding pilots' overall opinion about their traffic awareness with the VSAD, mixed responses were obtained. Whereas 7 pilots were more or less satisfied, 5 pilots were sceptical about the VSAD; one pilot found it 'too complicated', 4 pilots commented that, in actual flight, they expect to simply lack the time to check all information provided. Note that, in contrast to the decline in pilot appreciation of the VSAD overlays during the experiment, pilots became more supportive about the VSAD as a tool to improve their traffic awareness.

Some pilots commented on the symbology used to show whether intruder aircraft were climbing or descending. They suggested to adopt more TCAS-like symbology, like the use of an 'arrow up' when the intruder aircraft is climbing more than 500ft/min, to be positioned near the intruder label. Similar to TCAS, pilots also recommended to show the difference in height rather than the intruder aircraft flight level in the label. To become better aware of the time-to-conflict, pilots proposed the use of a color scheme: 'yellow', when conflict was more than 3.5 minutes away; 'orange', conflict 2 minutes away; 'red', conflict 1 minute away and prepare for traffic advisory. Subjective pilot SA ratings also indicated that pilots found themselves less aware of the time-before-conflict, and that they had difficulty in understanding what intruder belonged to what FBZ on the VSAD conflict geometry overlay.

Despite the overall lack of appreciation, pilot SA and meta-cognition scores were significantly larger with the VSAD. The averaged SA and meta-cognition scores indicate that pilot SA is higher with the VSAD as compared to the VSD, at all levels of SA and meta-cognition. These effects were all highly-significant ($p < 0.001$), except for the meta-cognition scores at the 'perception' level, where the difference between VSD and VSAD was small and not significant. SA and meta-cognition scores are lowest at the comprehension level, for both displays, but especially for the VSD. The benefits of the VSAD appear in particular at the levels of comprehension and projection, as was hypothesized. The fact that the meta-cognition scores are rather low with the VSD at these levels indicate that pilots often gave the wrong answer to SA queries that regarded a potential conflict's risk level, the time before initiating an escape maneuver, and also the understanding of how many aircraft would cause a potential conflict. Although the scores with the VSAD are higher, on average they do not reach the level of 'fairly sure'. This illustrates that, although the pilots' answers to the SA queries were generally correct with the VSAD, pilots were still unsure about their understanding of the situation. Tentatively, working with the VSAD for a longer time might increase these scores considerably, as the pilots would gain more experience and confidence in using the novel ecological overlays.

What also became clear is that whereas the SA and meta-cognition scores remain more or less the same for the VSAD, they decrease significantly with the VSD when the number of intruder aircraft increases. This causes a significant effect of 'intruder' (total SA: p=0.018; total meta-cognition: p=0.021), and a significant two-way interaction 'display $\times$ intruder' for the SA scores (p=0.006). The interaction was not significant for the meta-cognition scores. This result supports our hypothesis that with the VSAD, pilot SA does not depend on the number of intruder aircraft. In fact, remarkably, the scores with the VSAD are highest for the situations with the largest number of intruders, a non-significant effect, however.

Conclusions and Recommendations

Pilot Situation Awareness scores improve significantly with the ecological overlays presented on the Vertical Separation Assistance Display. These overlays give pilots a better sense of what maneuvers are possible to assure separation from surrounding traffic. Traffic awareness increases in particular at the higher levels of comprehension and projection. Awareness scores did not drop when the number of intruder aircraft increased, nor were they affected by changing conflict situations. The relatively low meta-cognition scores reflect the fact that although pilots were generally correct in answering the situation awareness queries in the questionnaires, they were still rather unsure about their answers. Extensive training with the novel display concepts are expected to increase pilot confidence and appreciation considerably. The evaluation further showed that in particular the conflict geometry overlay needs improvement, as pilots had difficulties in relating its components to the various intruders. Also, part of the display 'space' should be used to show 'what is behind' the own aircraft. Future research should also investigate the influence of maneuvering dynamics on the prediction times. It is recommended to conduct an extensive flight simulator evaluation, where pilots are more actively involved in maintaining safe separation.

References

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* (p. 3-32). Mahwah (NJ): Lawrence Erlbaum Associates.

FAA. (2008). *FAA's NextGen Implementation Plan* (Tech. Rep.). FAA, Washington (DC), USA.

Hoekstra, J. M. (November 2001). *Designing for Safety: The Free Flight Air Traffic Management Concept*. PhD dissertation, National Aerospace Laboratory (NLR) and Delft University of Technology.

Johnson, W. W., Battiste, V., Delzell, S., Holland, S., Belcher, S., & Jordan, K. (1997). Development and Demonstration of a Prototype Free Flight Cockpit Display of Traffic Information. *AIAA and SAE, 1997 World Aviation Congress, Anaheim (CA), Oct. 13-16*(AIAA 1997-5554).

Kuchar, J. K., & Yang, L. C. (2000). A Review of Conflict Detection and Resolution Modelling Methods. *IEEE Transactions on Intelligent Transportation Systems*, *1*(4), 179–189.

Merwin, D. H., & Wickens, C. D. (1996). *Evaluation of Perspective and Coplanar Cockpit Displays of Traffic Information to Support Hazard Awareness in Free Flight* (Tech. Rep. No. ARL-96-5/NASA-96-1). Savoy (IL): Aviation Research Lab: University of Illinois Institute of Aviation.

Prevot, T., & Palmer, E. (2000). Staying Ahead of the Automation: A Vertical Situation Display Can Help. In *AIAA and SAE, 2000 World Aviation Conference, San Diego (CA), Oct. 10-12* (p. 1-9).

Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. New York (NY): Elsevier Science Inc.

SESAR Consortium. (2007). *SESAR – The ATM Target Concept (Deliverable D3)* (Tech. Rep.). EUROCONTROL, Brussels (Belgium).

Thomas, L. C., & Johnson, W. W. (2001). Evaluation of CDTI Dynamic Predictor Display. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting, Minneapolis (MN), Oct. 8-12*, 171-175.

Van Dam, S. B. J., Mulder, M., & Van Paassen, M. M. (2008). Ecological Interface Design of a Tactical Airborne Separation Assistance Tool. *IEEE Transactions on Systems, Man & Cybernetics, Part A*, *38*(6), 1221-1233.

Vicente, K. J., & Rasmussen, J. (1992). Ecological Interface Design: Theoretical Foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, *22*(4), 589-606.

EFFECTS OF VISUAL, SEAT, AND PLATFORM MOTION DURING FLIGHT SIMULATOR AIR TRANSPORT PILOT TRAINING AND EVALUATION

Judith Bürki-Cohen, Andrea L. Sparko, and Young Jin Jo
USDOT/RITA/Volpe National Transportation Systems Center
Cambridge, Massachusetts
Tiauw H. Go
Nanyang Technological University
Singapore, Singapore

Access to affordable and effective flight-simulation training devices (FSTDs) is critical to safely train airline crews in aviating, navigating, communicating, making decisions, and managing flight-deck and crew resources. This paper provides an overview of the Federal Aviation Administration-Volpe Center Flight Simulator Human Factors Program examining the requirements for the qualification and use of FSTDs. We will summarize past research investigating the need for a full hexapod-platform motion system, describe regulatory and industry developments, and report on current activities.

Aviation Training Challenges

As we are writing this paper in early 2009, the global economy has entered a down-turn that may temporarily ease the world-wide pilot and aviation personnel shortages that have plagued the aviation industry in recent years. However, the challenge of training crewmembers with increasingly different backgrounds may remain, as the recent merger of several major airlines would suggest. At the same time, the proportion of certain types of accidents appears to be increasing (Boeing, 2007). One airliner disabled by a bird-strike was able to successfully land in the Hudson River, saving everybody on-board and sparing many that could have been killed on the ground. The crew had been trained under the rules of the Advanced Qualification Program (AQP), an alternative scenario- and proficiency-based, team-building curriculum targeting not only technical and procedural skills, but also cognitive and human-factors awareness [Federal Aviation Administration (FAA), 2005]. The recently published Notice of Proposed Rulemaking regulating traditional air-crew training takes a similar approach (Federal Register, 2009). Scenario-based full-mission training will play a critical role as the United States (US) and the European Community (EC) are overhauling their air-traffic-management (ATM) systems to improve current operations and accommodate future increases in traffic volume and complexity. This requires an update of the existing ATM infrastructure and the introduction of advanced tools to increase the efficiency of the air-transportations system (ATS) both in the air and on the ground without compromising safety.

For training, this all boils down to an increased need for effective and affordable Flight-Simulation Training Devices (FSTDs). FSTDs are the only means to efficiently train crews in aviating, navigating, communicating, making decisions, and managing their resources all at once. In fact, AQP is not possible without access to FSTDs, and the proposed new crewmember-qualification rules mandate it. Airplane mergers increase the need for a team-training approach. The US's Next Generation (NextGen) ATS and the EC's SESAR (Single European Sky ATM Research) may radically change the roles and collaborations of air- and ground crews and automation. FSTDs will have to be able to simulate all these interactions, in addition to increased emphasis on accurate simulation of environmental hazards and Loss of Control (LOC) situations. Given that airlines have limited resources for supporting such training under any circumstances, but especially now, this update on the FAA-Volpe Center Flight Simulator Human Factors Program is provided for the benefit of training developers. We will summarize past research investigating the need for a full hexapod-platform motion system, describe regulatory and industry developments, and report on current activities.

FAA-Volpe Center Flight-Simulator Requirements Research

As the FAA launched the "One Level of Safety" initiative for major and regional airlines and more and more airlines adopted AQP in the mid-nineties, the FAA launched the FAA-Volpe Flight Simulator Human Factors Program to identify potentially unnecessary obstacles to universal access to the benefits of flight-simulator training. The purpose of the program was to systematically examine each requirement for simulators used in zero-flight-time training and evaluation of airline pilots for its contribution to the training value of those simulators. The focus of this examination was the training value added by each requirement. A review of the literature and consultation with many subject matter experts (SMEs), including two FAA industry workshops (proceedings available), revealed that the scientific basis for requiring a hexapod-motion platform for airline pilot training and evaluation was shaky. While simulator motion does have great face value (after all, the airplane moves), decades of research have failed to show an effect of motion on transfer of training to the airplane or on reverse transfer from the airplane to the simulator for pilot evaluation (Bürki-Cohen, Longridge, and Soja, 1998). The program thus initiated a series of carefully controlled studies to answer the following questions: 1) Are there maneuvers in airline-pilot training where platform motion cues, in addition to the visual cues from a wide field-of-view (FOV) out-the-window view and instruments, result in an operationally relevant improvement of transfer between the simulator and the airplane? 2) Do airline pilots need to be trained to avail themselves of motion cues? 3) Are motion cues from a hexapod-platform representative of those experienced in the airplane? 4) Can alternative systems provide onset cues and perception of realism?

To ensure valid and replicable studies, we did everything possible to achieve the statistical power to find an effect of motion. For the first three studies,[1] we tested the extremes, i.e., comparing FAA Level C or D full flight simulators (FFS) with the motion turned on or off. We calibrated all cueing and measurement systems. We counterbalanced across groups any other factors that we could not keep constant on all aspects that could affect their flying skills. We also tested more pilots than in most other studies to randomly balance any differences that we were unable to systematically control or counterbalance. We carefully selected maneuvers where motion cues were most likely to serve an alerting function according to the literature and SMEs, i.e., high-workload maneuvers with unpredictable mechanical or weather disturbances. We measured nearly 80 dependent variables directly from the simulator representing pilot control inputs, flight precision, and simulator motion performance at a high sampling rate (at least 30 Hz). We also asked all participants for their opinions on all aspects of the simulator and the behavior of the pilots flying. We started out with experienced pilots, who according to the literature would be most likely to rely on motion cues (Young, 1967). We kept the motion status and in fact the purpose of the study secret to prevent any bias for or against motion from affecting the results.

We performed three studies and did not find any operationally relevant effects of motion. The first two studies examined the effect of motion on recurrent training and evaluation. Study 1 tested regional airline captains in a Level C FFS of a turboprop "powerhouse" airplane with wing-mounted engines. The simulator had a 60 inch stroke with a 1.7 Hz heave bandwidth. The maneuvers consisted of engine failures with continued and rejected takeoffs (V1 cuts and RTOs) at a quarter mile Runway Visual Range and with 10 knots crosswind. For the V1 cut, we found no differences between the pilots that flew the simulator with motion and those that flew it without motion at "First Look," i.e., when pilots flew the simulator the first time after having flown the airplane for twelve months.[2] For the RTO, the motion pilots showed lower yaw activity, but motion did not affect their heading compared to the no-motion pilots. After additional training (when necessary) and transfer to the simulator with motion of all pilots, pilots having been trained with motion showed lower speed exceedance that may have been achieved with higher pitch standard deviation (STD). They also showed higher yaw activity with more pedal reversals, but without directional control benefit. There were no differences between the two groups for the RTO. In general, all differences were small and operationally irrelevant. However, the data also showed a relatively large discrepancy between the lateral acceleration of the simulator compared to the lateral accelerations from the flight model during the first few seconds following the engine failure. Was the motion of our test simulator just unusually ineffective? With the help of the FAA's National Simulator Program, we obtained data from eight other Level C and D simulators and found that in terms of lateral acceleration, the test simulator used appears to be representative of other FAA qualified FFS.

Given these findings, in collaboration with the National Aeronautics and Space Administration (NASA), we conducted a second study using the FAA-NASA Level D B747-400 simulator with its motion re-engineered so as to provide greatly improved lateral acceleration and somewhat improved heave motion, trading off rotational motion which, according to some studies, is less important than translational motion (see AIAA-2003-5678 for references). For this study, we tested both captains and first officers, and we replaced the rejected takeoff with a V2 cut, arguing that it was a more diagnostic maneuver requiring multi-axis control. We also had pilots return the one-engine-out airplane to the airport for two difficult landings, one a side-step landing with a microburst (SSL) and the

---

[1] Due to the six-page space limit, the link to all studies performed within the FAA-Volpe Center Flight Simulator Human Factors Program is given in the references.

[2] Reported effects have a probability of $p<0.05$ to have occurred by chance. Effects of $0.05<p<0.10$ are reported as trends.

other a hand-flown Precision Instrument Approach and landing with quartering head- and tailwinds (PIA). During First Look, we finally found the V1-cut pedal reaction time advantage with motion that may arise from faster perception of vestibular motion compared to visual motion, but it was less than half a second and had no effect on flight precision. Most importantly, it did not transfer to the simulator with motion; once the pilots that were trained without motion were given motion cues, they responded as fast as the pilots trained with motion. Also at First Look, the motion pilots had lower heading STD with lower root mean square (RMS) pedal and yaw activity, but higher pedal bandwidth. They also had lower pitch STD. For the V2 cut, the motion group had more pedal reversals than the no-motion pilots at First Look, and when all pilots transferred to the simulator with motion, the motion-trained pilots had a 0.8 s slower mean pedal reaction time. There were no differences between the two groups for the V1 cut at transfer. Again, all differences between the two groups were small and operationally irrelevant. The only difference that approached operational significance, according to our SME, was found with the one-engine-out PIA during the approach-fix-to-decision height phase, where the motion-trained group had worse horizontal directional control than the no-motion trained group both before and after transfer to motion (by 0.19 dots higher localizer STD and 0.16 dots higher localizer exceedance). The motion group had also higher STD for heading and bank angle. For the SSL, the motion group had lower pedal bandwidth and landed more softly (by 42 ft/min) but less precisely (by 225ft, but both groups landed well within the landing box). All these latter differences were small.

Having found no operationally relevant advantage of providing motion cues for recurrent training and evaluation with two very different groups of pilots flying a very different airplane, the question remained whether it was perhaps their familiarity with the motion of the real airplane that masked an effect regardless of a heightened sensitivity of experienced pilots to motion cues that was reported in the literature. We thus tested a group of new-hires fresh from ground school in a Level D B717-200 simulator with a 54 inch stroke of platform motion and a 90 degree phase lag at 8.3 Hz. Due to time constraints, we focused on the PIA and the V1 cut. For the First Look V1 cut, we found a trend for a faster pedal response with motion (again, less than half a second), and again, the effect disappeared once all pilots transfer to motion. Also, this trend did not translate in an operationally relevant effect on flight precision. Both during First Look and Transfer to motion, the motion group had steadier column RMS, smaller airspeed exceedance, but higher pitch STD during the V1 cut and higher pedal RMS for the PIA. Again, all effects were small and operationally irrelevant.

In summary, it appears that the answers to the first three research questions are "No." Consulting the literature and SMEs, we had carefully chosen maneuvers within and even somewhat beyond the air-transport-pilot curriculum that should have depended on an alerting function of physical motion, if there were one. In the one maneuver where pilots did respond minimally faster with motion, even pilots that had not experienced motion during training were able to use the motion cue once they received it, indicating motion was not necessary for training. Finally, we obtained the same results even in a simulator that was tuned to provide the best possible motion cues for the maneuvers trained. Our follow-up to Study 1 found large differences between the equations of motion and the real motion for several simulators available for examination, indicating that a good overall safety record had been obtained despite negligible motion cues in at least some devices (see Bürki-Cohen et al., 2001 manuscript for summary of follow-up study).

Regulatory and Industry Impact

The lack of finding an operationally relevant effect of costly motion platforms, even with state-of-the-art motion and visual systems and with demonstrated power to find an effect ($\beta<0.20$, Cohen, 1988), has received a mixed reception. The FAA tried to introduce tighter motion standards in its proposal for Part 60, regulating FSTD qualification and use, but removed the minimum excursions, velocities, and accelerations in the final version based on industry questions regarding their effect on training (see Bürki-Cohen et al., 2005, and FAA, 2008). In general, the experimental rigor and the reliability of the studies were not questioned. Some critics wanted transfer to be tested in the airplane instead of the simulator with motion as a stand-in for the airplane. Aside from safety, cost, and experimental control issues, however the fact that during testing with motion pilots trained without motion performed no differently than pilots trained with motion appears to preclude any conclusion that motion plays a major role in training. Others have questioned whether the results would generalize to other maneuvers, pilot populations, or airplanes. However, while our research began by testing the potentially most motion-reliant maneuvers, pilots, and simulated airplane (i.e., disturbance maneuvers, experienced pilots, wing-mounted powerhouse), our subsequent studies did take into account other aspects that may increase the need for motion. We still found no impact on training effectiveness. We therefore feel that these results would hold for other maneuvers, pilots, and airplanes trained in air-transport operations.

Internationally, a European turboprop airplane manufacturer responding to its world-wide customers' urgent need for access to the benefits of flight-simulator training used this research to specify a high-level "Full Flight Trainer" (FFT X™). The FFT uses the same data package as a Level D simulator to simulate the manufacturer's 74 passenger high-wing twin-engine turboprop airplane providing "motion cueing without a motion base." Motion is simulated via a wide FOV collimated visual system and a dynamic seat providing heave motion-onset cues via electric jacks. Vibration cues are provided via loud-speakers. The simulator thus takes advantage of the fact that humans are very proficient in perceiving motion via multiple perceptual systems, including not only the vestibular, but also the visual, tactile, and proprioceptive sensory systems. The existence of this simulator, and the fact that it was granted training and checking credits similar to a Level B FFS (including recurrent training and checking) by the French National Aviation Authorities (NAA, a member of the Joint Aviation Authorities) critically influenced the drafting of the International Civil Aviation Organization's Doc. 9625 Edition 3 draft: Manual of Criteria for the Qualification of Flight Simulators (Royal Aeronautical Society, 2008). Despite taking a generally strong stance for platform motion in its summary table with device examples, the draft takes a training-task-based approach and provides for an alternative, albeit cumbersome, à la carte approach to developing cost-effective simulators tailored to airlines' needs.

Current Research Activities

The FAA-Volpe Center heard about the FFT while participating in the international working group drafting Doc. 9625 and offered to help with its evaluation in an attempt to answer the final research question, on whether there is an affordable alternative to full platform motion. A first "proof-of-concept" phase concluded with the successful type rating of six pilots, the first time in the world that pilots were type-rated using a simulator without a hexapod-motion base. All pilots were employees of the NAA. Two were designated as "Experienced Pilots" due to the fact that they held a multipilot-crew license and had airline experience. The four "Non-Experienced" pilots held single-pilot licenses, and in one case had flown as little as 563 hours. Those latter pilots underwent a somewhat expanded curriculum in the classroom and in an FFT without seat motion that was used for both groups before using the FFT-X with seat motion. We collected opinion data throughout the course of training, which are reported in Bürki-Cohen et al., 2007. The most interesting comparisons were those collected after participants had flown the actual airplane. In those, instructors reported the pilots to be the "same as typical trainees" with regard to performance, control strategy and technique, workload, and ease of learning. The trainees themselves declared the FFT to be the "same as the airplane" with regard to handling qualities, feel and response of controls, ease of learning, comfort, workload, and overall simulator cues. They said that they used a "somewhat similar" control strategy, and rated the acceptability of the simulator as "satisfactory as is." During a final debriefing session, the trainees and instructors agreed that the transition between the FFT and the airplane had been successful. The NAA decision maker concluded that the strategy of using an FFT to focus on effective stimulation of the pilot, instead of rote simulation of the airplane, had been validated, and authorized the next phase of the FFT evaluation, which resulted in the successful type rating of 16 more pilots and also served as the set-up phase for Phase 3, a systematic comparison of the training values of the FFT with those of the FFS of the same airplane. For this FFT/FFS comparison phase, all pilots were first prepared for type-rating in the FFT. Next, they were divided into an FFT and FFS group, keeping the experience level constant between the two groups. The pilots were then brought into their assigned training device and each flew two take-offs and landings to familiarize themselves with the device. After familiarization, they were trained on V2 cuts followed by PIAs with quartering head and tail winds, and on V1 cuts followed by SSLs with microbursts. Pilots were trained in each scenario three times. After training, they filled out questionnaires. The next day, both groups were brought into the FFS and tested on a V1 cut followed by a PIA with quartering head and tail winds and on a V2 cut followed by an SSL with microburst. After testing, they filled out another questionnaire. Instructors also filled out questionnaires after training and testing. During both training and testing, the two pilots took turns flying the scenarios. This experiment is taking place in France and directed via e-mail and telephone communications with no direct supervision by the experimenters. To date, the data of 7 FFT- and 5 FFS-trained crews have been analyzed and a few interesting trends seem to be emerging. We will first describe the pilots' opinions collected in questionnaires.

This is the first experiment in our program where the motion condition (and thus perhaps the purpose of the experiment) could not be concealed from pilots. In Studies 1 through 3, pilots were tested and trained in the exact same simulator. During training without motion, we still lifted the bridge leading to the simulator and initialized the motion, before washing it out. Therefore, many pilots never realized or realized only after several maneuvers that the physical motion cues were missing. We also found no marked preference for either condition in extensive questionnaires administered to all study participants. However, in the present study, pilots knew exactly whether they entered the FFT via a few steps or the FFS over a bridge. Looking at some "pilot" questionnaires completed by pilots participating in Phase 2 and at questionnaires from two crews participating in the current Phase 3, we noted a preference for the FFS that may have been the expression of a true preference or, alternatively, a bias for the motion device arising from the knowledge that the airplane moves. After rejecting countermeasures such as blindfolding or administering pre-experiment questionnaires to diagnose bias as impractical or potentially inducing additional bias, we settled on removing the detailed questions on simulator cues, leaving only questions on simulator properties assumed to be affected by all cues, namely handling qualities, feel and response of controls, comfort, workload, acceptability, control strategy and technique, and ease of learning.

To date, we have completed preliminary analyses on the last three assessments, using the SAS General Linear Model procedure with the factors Group (FFT vs. FFS trained) and Session (Training vs. Testing). For acceptability, we found no effects of any factors, with Least Squares Means (LSMean) ratings of 3.9 by the FFS-trained group for both sessions and 3.8 vs. 4.1 by the FFT-trained group for training and testing, respectively [all F, including Group by Session interaction, $F(1,65)<1$]. The rating scale ranged from 1 (uncontrollable) to 5 (excellent). For control strategy and technique, we just missed a Group effect [$F(1,187)=2.61$, $p>.10$], but found a trend of a Group by Session interaction [$F(1,187)=3.69$, $.05<p<.10$]. According to a Tukey-Kramer multiple pairwise comparisons test on the LSMeans, this was due to a trend ($p<.10$) of the FFT-trained group to rate the simulator higher during testing than the FFS-trained group [LSMeans 3.34 vs. 3.08 on a scale of 1 (very different from the airplane) to 4 (same as airplane)]. None of the other comparisons were significant, which meant that the transition to the FFS did not increase the rating of the FFT-trained group. Finally, for ease of gaining proficiency, we found an overall effect of Group, with the FFT-trained group rating the simulators slightly lower than the FFS-trained group regardless of session [2.9 vs. 3 on a scale of 1 (very hard) to 4 (very easy), $F(1,65)=5.42$, $p<.05$].[3] In summary, the pilot opinions described to date do not indicate a marked preference for either the FFT or the FFS, despite the fact that in this study, the absence of a motion platform and thus substantial physical motion cues was obvious.

For all of our studies, a first objective analysis looked at the number of success rates, defined as successful take-offs or landings without LOC or abnormal ground contact. To be considered a success, maneuvers also had to be flown within four STDs of the most important performance variables (go-arounds, where a pilot forgoes landing in favor of another try, were eliminated from all analysis). As in all our other studies, unsuccessful maneuvers were rare and did not differ between the two groups in this study. The V1 and V2 cuts to date were flown with a 100% success rate regardless of group. The PIA and SSL success rates for the FFT group were 96 percent. The respective success rates for the motion group were 97 percent for the PIA and 92 percent for the SSL. With respect to the detailed data on pilot control inputs or workload and flight path precision, we have performed some preliminary MANOVAs on the V1 cut and on the first stage of the PIA (approach fix to decision height), the two maneuvers with the most interesting results across the studies. It appears that we have again replicated the effect of platform motion on the pedal reaction time to the V1 cut during training. It took the pilots in the FFT an average of 1.73 seconds to respond to the engine failures, 0.42 seconds longer than it took the pilots in the FFS [$F(1,68)=4.18$, $p<.05$]. This reaction time advantage, however, did not affect the heading STD [$F(1,68)<1$]. Most importantly, it again disappeared after transfer of all pilots to the FFS [$F(1,21)<1$], indicating that even very-low time pilots do not have to be trained to use motion cues. For the first stage of the PIA, the only significant difference discovered so far is a slightly lower RMS wheel response for the FFT-group during training of 6.04 degrees vs. the 7.79 degrees of the FFS-group [$F(1,67)=16.02$, $p<.01$], with no effect on localizer STD [$F(1,67)<1$]. Again, this effect disappeared when all pilots were tested in the FFS.[4]

Conclusions and Future Plans

Pending further data collection and analyses, it appears that the answer to the final research question—whether alternatives to platform-motion systems can provide onset cues and perception of realism—is yes. However, we do not really know what the contribution of the seat-motion system is to training effectiveness aside from added face validity compared to a system with only visual motion cues. We are currently planning to conduct similar operational testing for recurrent training and testing, again using a high-level simulator with the motion

[3] We are always reporting the results based on Type III Sum of Squares (SS).
[4] In this preliminary analysis, we performed the analyses separately for the two sessions and thus do not have any interaction data.

turned off and on, with the intent to pursue skill maintenance of pilots trained with and without motion over a longer period of time than just the experiment itself.

Acknowledgments

The authors thank their FAA Program Manager, Dr. Eleana Edens of the Air Traffic Organization Human Factors Research and Engineering Division, and Dr. Thomas Longridge, Manager of the FAA Flight Standards Service Voluntary Safety Programs, for their critical insight and support of this work. We also thank Dr. Shuang Wu (Computer Sciences Corporation) for her emergency statistics advice. Dr. Stephen Popkin, Director of the Volpe Center Human Factors Research and System Applications Center of Innovation, and Dr. Michael Zuschlag, Engineering Psychologist at the Center, stood by with excellent advice as usual.

References

All references to work conducted within the FAA-Volpe Center Flight Simulator Human Factors Program can be found at http://www.volpe.dot.gov/hf/pubs.html or by contacting the first author at Judith.Burki-Cohen@dot.gov

Boeing Company (2007). Statistical Summary of Commercial Jet Airplane Accidents: Worldwide Operations 1959-2006.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

FAA (2005). Advanced Qualification Program. (http://www.faa.gov/education_research/training/aqp/)

FAA (2008). Code of Federal Regulations Title 14. Part 60. Flight simulation training device initial and continuing qualification and use. (http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=32a30a7ce06986548dd94d1b635acfe9&rgn=div5&view=text&node=14:2.0.1.1.1&idno=14)

Federal Register (2009). Vol. 74, No. 7 / Monday, January 12, 2009 / Proposed Rules. Pages 1288-1453. ([Docket No. FAA–2008–0677; Notice No. 08–07]). (http://edocket.access.gpo.gov/2009/pdf/E8-29584.pdf)

Pasztor, A. & Lunsford, J.L. (2009). Crash probe turns to pilot training. *Flight Safety Information 19FEB09-0*44. (from *Wall Street Journal*).

Royal Aeronautical Society (2008). Manual of Criteria for the Qualification of Flight Simulators (Doc. 9625) (draft, submitted to ICAO Aug. 29, 2008). (http://www.raes-fsg.org.uk/cgi-bin/yabb2i/YaBB.pl)

Young, L. R. (1967). Some effects of motion cues on manual tracking. *Journal of Spacecraft and Rockets, 4 (10),* 1300-1303.

PSYCHOLOGICAL FIDELITY OF SIMULATOR HUMAN PERFORMANCE LIMITATION TRAINING

K. Wolfgang Kallus
Karl-Franzens-University Graz
Graz, Austria

Problem

Fidelity of simulators for training of pilots has to be judged from the final end of the training goal. This conclusion can be derived from the overview of Hays & Singer (1989), which has been published a considerable time ago. Nevertheless, an ongoing debate questions the need of simulator features like motion for the training of pilots – partly without giving attention to the training goals at hand. Especially in the area of threat and error management requirements for the simulators differ markedly from operational recurrence training. For experienced ATPL- pilots we can assume that a high fidelity visual simulation and a proper representation of the avionics and a high fidelity simulation of the flight dynamics might well be sufficient to refresh rare standard situations. From a psychological point of view we would predict that the well elaborated cognitive model of professional pilots with respect to aircraft, its dynamics and the situation will allow to simulate the situations without motion. Pilots are able to add the not simulated aspects from their highly elaborated mental model. On the other hand a broad range of situations in the area of human performance limitations are beyond the experience of pilots or trainees. A proper simulation of the aircraft performance and the perceptions and sensations is necessary to improve performance by simulator training to cope with situations beyond the standard environment. Especially for successful disorientation recovery training it may be necessary to provide the correct physical sensations enable the pilot to learn the correctly timed and executed actions to re-establish safe flight parameters. Perceptual illusions of the vestibular system and problems in vestibular-optic coordination are core elements in the development of a multitude of spatial disorientation phenomena (Bles, 1998; Cheung, 2004; Previc and Ercoline, 2004). A couple of reports have been published, which show convincingly that disorientation recovery training with a motion base simulator improves performance in jet pilots (Cheung, 2004; Kallus & Tropper, 2004) as well as in helicopter pilots (Hays & Singer, 1989) and in pilots of small VFR aircraft (Kallus, Tropper & Boucsein, 2009). These studies all used simulators, which are at least capable to rotate in one axis. Disorientation due to sensory illusion is not only caused by vestibular illusion (like gyro spin or leans, for details see Previc & Ercoline, 2004 or Kallus & Tropper, 2004). Some accidents in the area of spatial disorientation occur primarily due to visual illusions (like the black hole approach or the runway width/slope illusion). For VFR pilots, flight into IMC is one of the most problematic and often fatal causes of disorientation. Unintended flight into IMC due to gradually worsening weather conditions seems also to be a primarily visual problem. The more visually based disorientation situations might not require motion cues during the training, as motion does not seem to play a predominant role in the development of the state of disorientation. An experimental study was designed to evaluate the role of motion cues for different disorientation recovery exercises in the simulator.

Methods

Subjects and experimental conditions

Forty-two pilots with a valid PPL-license participated in the experiment. Age ranged between 20 and 56 years (M = 41.2 years, SD = 8.7). Only pilots without IFR-rating and with less than 500 fight-hours were admitted to the study.
The 42 pilots were randomly assigned to one of three groups: The training-motion group (n=15) received a disorientation recovery training, which was based on the successful procedures of a previous study (Kallus, Tropper & Boucsein, 2009). A second group received an identical training without motion. For the training-no motion group (n=15) the motion function of the simulator was switched off during training sessions. In addition a control group was studied under motion conditions. The control-motion group (n=12) did not receive a specific training, but had to execute free flights and some of the flight maneuvers of the experimental groups (e.g. approaches) under standard conditions to equal the simulator experience. Table 1 summarizes the experimental conditions.

Table 1. *Experimental conditions.*

	simulator phase I	**simulator phase II**	**simulator phase III (test)**
TG_MO (N = 15) training group motion	**familiarization flight** MOTION ACTIVE	**training** MOTION ACTIVE	**test** (5 test profiles) MOTION ACTIVE
TG_noMO (N = 15) training group no motion	**familiarization flight** NO MOTION	**training** NO MOTION	**test** (5 test profiles) MOTION ACTIVE
CG_MO (N = 12) control group motion	**familiarization flight** MOTION ACTIVE	**free flight** **control condition** MOTION ACTIVE	**test** (5 test profiles) MOTION ACTIVE

Procedure

A motion base flight simulator (AIRFOX spatial disorientation trainer DISO by AMST Systemtechnik GmbH, Austria, 2006) was used for training and test. The exercises were performed with a two engine turboprop aircraft model. The experiment took place in three subsequent phases: instruction, training, and test. Instruction and test was identical for all subjects.

The following exercises were used for training:

- Pitch up illusion (by configuration change just after take off under minimal visibility conditions),
- Inadvertent Flight into IMC (climbing to 3000ft under deteriorating weather conditions),
- Unusual approaches (black hole approach and approaches with tilted or narrow runway)
- Unusual attitude recoveries (returning the aircraft to near straight and level flight from an unexpected bank and/or pitch angle)
- spin recoveries and a gyrospin demonstration

The test exercises in phase 3 correspond to the training exercises. Motion was on for all groups during the test exercises

Measures

The study was conducted in a multivariate multilevel assessment approach, only performance data (observation data, instructor ratings, time-measurements, self-assessment) will be reported here. For detailed results on the psychological and physiological state before, during and after the exercises see Kallus, Tropper & Boucsein (2009).

Objective performance data were time to regain safe flight parameters was taken for UAR recoveries and spin recoveries. A blind scoring of performance was conducted for the other profiles using a five point rating scale with objective rating criteria for each of the five categories. These ratings were based on flight recordings using the digital video recording system of the DISO Airfox simulator.

Instructor ratings. The instructor rated the pilots' flight performance immediately after each exercise according to the following six evaluation criteria: allocation of attention, situation awareness, stress resistance, multi tasking, aggressiveness, and overall performance. Ratings used four categories: excellent (4), good (3), fair (2), and unset (1). For each category, five subcategories were available: double minus, minus, middle, plus, and double plus. Thus, the whole scale ranged from 0.6 (unset, double minus) to 4.4 (excellent, double plus). As the unusual attitude recovery sequences were of short duration (average about 13 sec per UAR), the instructor rated the overall performance for each UAR.

Self-ratings of performance using the same rating scale were obtained during a reconstruction interview, which was conducted after the test phase with each pilot.

Statistical analyses

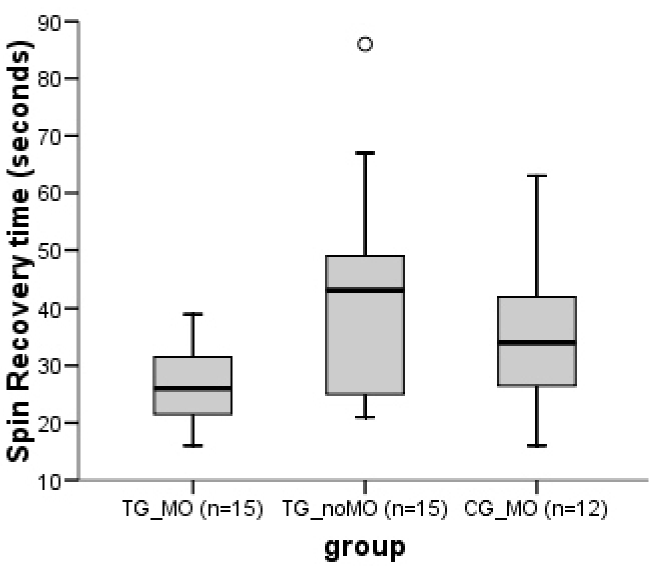
The performance data were analyzed with a multivariate analysis of variance and the instructor ratings were analyzed with a repeated measures analyses of variance for a controlled statistical decision with alpha=0.05 and Bonferroni-Holm adjustment (Holm, 1979) for multiple testing. In a second step a traditional statistical analysis was conducted using analyses of variance procedures for self rated performance.

Results

Objective performance data

The statistical analyses of the performance data from the test phase resulted in clear cut group-effects below the adjusted type-I-error of $\alpha=0.05$. Thus, it can be concluded that training effects could be proved with a type-I-error of 5%.

The motion based training outscored the other two groups in the test profiles, which resulted in a highly significant statistical effect ($F(10,72)=3.06$, $p=0.003$). Univariate analyses show that the positive training effects are most prominent in the profiles “Take-off with Pitch-up Illusion” ($F(2,39) = 6.68$, $p = .003$), “Inadvertent Flight into IMC” ($F(2,39) = 5.14$, $p = .010$), and “Spin Recovery” ($F(2,39) = 4.87$, $p = .013$). Figure 1 depicts the results of the spin recoveries as box-whisker-plots, which show means (bars), interquartile distances (boxes), and the 95% intervals (whiskers). In addition outliers are shown if present (single points).



TG_MO training group motion
TG_noMO training group no motion
CG_MO control group motion

Figure 1. Boxplots for the spin-recovery time for the three experimental groups.

The results of post-hoc tests (Tukey-test) are depicted with stars. For the profiles “spin recovery” the training without motion showed the worst performance indicating a “negative training” effect in this motion oriented profile. Similar results were obtained for the profile “pitch-up illusion”. Even for inadvertent flight into IMC the only significant effect was obtained for the motion-based training, which differs significantly from the control condition. In this profile the training without motion results in an intermediate performance. Additional analyses for the objective data with a repeated measures analysis to check for interaction between training effects and the kind of profile (using profile as repeated measure after standard-normal-transformation and alignment of scoring direction) resulted in a significant interaction term ($F(8,156) = 2.51$, $p = .014$) indicating that the differences in effects for the different profiles are substantial.

Instructor ratings

The repeated measures analysis of variance all in all show corresponding results to the objective performance data with a significant main effect for the training condition ($F(2,39) = 8.47$, $p = .001$).

Self-rating of performance

Differences in performance were also represented subjectively. Significant effects emerged in the performance ratings of NASA TLX ($F(2,39)=5.38$, $p=0.009$). The motion based training resulted in better subjective performance compared to the controls and to the no-motion training. Again the no-motion training does not differ substantially from the control group.

Discussion

The results fit well into a mental training framework of simulator training. Motion oriented test procedures profit a lot from motion cues during training. A profile like spin recovery, which has a complex, partly contra-intuitive recovery procedure showed no training effect with the training based only on visual cues. Motion enhanced performance significantly compared to the no-motion training group. Without motion it might have been impossible (or at least much more difficult) to obtain a proper mental representation of the situation. Considering that VFR-Pilots do not have access to motion simulators a preparation for situations like spin recoveries is not possible during simulation. An acrobatic aircraft trainer is the only option to learn procedures like spin recoveries properly in Europe as long as Disorientation training simulators like the DISO Airfox are not accredited in the pilot's training syllabus. The main reason for this is the generic avionic, which works well – but is not a face valid naturalistic representation of a VFR aircraft. The option to use more generic simulators for specific training purposes has also been claimed by Dahlström et al. (2009). They also argue that the mere reliance on increased photorealistic fidelity of simulation systems can be the wrong path to follow for a couple of training goals. For the training of a couple of no-tech-skills technical fidelity might even distract the attention from the training goals towards technical details of the simulated situations. Our data strengthen the view, that training simulators have to mimic the relevant cues as realistic as possible. Cues outside the focus can be simulated in a very generic way, especially, when the trainees can fill in their correct mental representation. Of course – basic principles of mental training should be met, when technical simulation and metal representations are used in a training paradigm. The data provided with the disorientation trainer DISO AIRFOX show that motion cues during training are crucial for an adequate test performance. The results given in figure 1 rise the problem of possible negative training effects. These effects occur if the simulated training situation results in a response pattern or a mind set, which is dysfunctional in the aircraft. Motion is a basic feature of every aircraft – thus exclusion of motion cues from training might cause problems in the long run. As full flight simulators are unable to simulate extreme (motion-)situations the requirement to include more motion axes into the training seems inevitable especially for pilots, who are at risk of extreme motion situations in their operational environment. This is especially true for helicopter pilots and military pilots.

For trainings of human performance limitations we currently face the paradoxical situation, that JAR-FCL require substantial knowledge of human performance and human performance limitations from CPL and ATPL certified pilots, while PPL licences only have to know the basics (probably without any option to make this knowledge relevant for their decision making in disorientation prone flight situations). To provide extended knowledge to the better educated pilots is useful – but in large commercial aircraft there is a much lower probability of disorientation prone situations. For VFR pilots the knowledge might be life saving, especially if it is transferred into action relevant mental models, which trigger recovery and adequate decision making.

References

Bles, W. (1998). Coriolis effects and motion sickness modelling. *Brain Research Bulletin*, 47(5), 543-549.

Cheung, B. (2004). Spatial Orientation – Nonvisual Spatial Orientation Mechanisms. In: F. Previc, W. Ercoline (Eds.) *Spatial Disorientation in Aviation. Progress in Astronautics and Aeronautics*, Vol. 203, 37-94. American Institute of Aeronautics and Astronautics, Inc. Restoin, Virginia.

Cheung, B. (2004). Spatial Orientation – Nonvisual Illusions in Flight. In: F. Previc, W. Ercoline (Eds.) *Spatial Disorientation in Aviation. Progress in Astronautics and Aeronautics*, Vol. 203, 243-281. American Institute of Aeronautics and Astronautics, Inc. Restoin, Virginia.

Dahlström, N., Dekker, S., van Winsen, R. & Nyce, J. (2009). Fidelity and validity of simulator training. *Theoretical Issues in Ergonomic Science*, in print.

Hays, R. T. & Singer, M. J. (1989). *Simulation fidelity in training system design – bridging the gap between reality and training*. New York: Springer.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65-70.

Kallus, K. W. & Tropper, K. (2004). Evaluation of a Spatial Disorientation Simulator Training for Jet Pilots. *International Journal of Applied Aviation Studies*, 4(1), 45-55.

Kallus, K. W., Tropper, K. & Boucsein, W. (2009). The Importance of Motion Cues in Spatial Disorientation Training for VFR-Pilots. *International Journal of Aviation Psychology*, submitted.

Previc, F. H. & Ercoline, W. R. (2004). Spatial disorientation in aviation: Historical background, concepts, and terminology. In F. H. Previc & W. R. Ercoline (Eds.), *Spatial Disorientation in Aviation* (pp. 1-36). Virginia: American Institute of Aeronautics and Astronautics.

KNOWLEDGE AND SKILL-BASED EVALUATION OF SIMULATED AND LIVE TRAINING – FROM EVALUATION FRAMEWORK TO FIELD APPLICATION

Martin Castor
Jonathan Borgvall
Swedish Air Force Air Combat Simulation Centre
Swedish Defence Research Agency
Stockholm, Sweden

Winston Bennett, Jr.
Warfighter Readiness Research Division
711 Human Performance Wing
Air Force Research Laboratory
Mesa, AZ

In a study, a simulated spin-up exercise and the corresponding large-scale live military flight training exercise was evaluated based on the Alliger et al. augmented taxonomy of Kirkpatrick's training criteria. The data collection was developed and designed to assess the training from reactions to in-simulator knowledge and skill development to operative training effect. The basis for the evaluation was knowledge and skills identified with the Mission Essential Competencies (MEC) process. Using surveys, quantitative and qualitative data from 14 fighter pilots were collected regarding reactions to training, perceived training value and additional training needs. This paper will present the rationale and theoretical framework behind this methodological approach. The main contribution is the description of how the underlying theoretical frameworks have been transformed into measures allowing structured evaluation of training "in the wild".

In July 2008 the Swedish Air Force (SwAF) for the first time participated in the world's largest military live flying exercise – Red Flag (RF). The exercise is managed by the US Air Force at Nellis Air Force Base (Las Vegas, NV) and the flying takes place over the Nevada Test and Training Range (NTTR). Four Red Flag exercises are usually conducted each year and international participation is becoming increasingly common. The exercise is the largest of its kind in the world and by many considered as the most realistic training event available for military aircrew. During this specific RF exercise, RF 08-3, up to 65 aircraft flew two sorties per day for a period of two weeks. The Swedish unit that participated in RF 08-3 was the SwAF rapid reaction unit, which consists of highly motivated and experienced combat ready pilots flying the fourth generation fast-jet JAS39 Gripen. In many ways the SwAF participation in RF 08-3 was considered to be not only an excellent training event but also a war-like criterion of operative performance – a way of confirming developed tactics, techniques, and procedures (TTP) at an individual, team and organizational level.

In order to prepare for the live exercise RF 08-3, the pilots of the SwAF RF 08-3 contingent conducted a series of preparations, in the class-room, live at their squadrons and in a simulator environment. One of the major preparatory events was a simulated spin-up exercise, called RF spin-up, at the Swedish Air Force Air Combat Simulation Centre (FLSC). Several processes of monitoring and documenting the SwAF participation in RF 08-3, ranging from preparations to return to home base, from logistic procedures to fast-jet performance, were undertaken simultaneously. One of them is the work presented here – a training evaluation study monitoring the fast-jet pilots' individual reactions to the spin-up training together with their perceived training value and self-assessed performance readiness before, during and after the exercises.

Training criteria and training evaluation taxonomies

Kirkpatrick's taxonomy (1959a, 1959b, 1960a, 1960b) with four levels of training evaluation criteria has received widespread attention from researchers and practitioners since its original definition. Kirkpatrick's taxonomy originally was a set of practical suggestions, drawn from personal experience, providing a useful heuristic for what can be measured, not necessarily providing strict directions on what should be measured. The taxonomy has been debated by a number of researchers, and flaws in the taxonomy have been highlighted by, for example, Alliger, Tannenbaum, Bennett, Traver and Shotland (1997), who provide an augmented framework.

Bell and Waag (1998) suggests an alternative model for evaluation of simulator training. Even though their model was explicitly focused on simulator training evaluation it was based on Kirkpatrick's training criteria and shows clear similarities with the Alliger et al. augmented taxonomy (1997).

Kraiger, Ford and Salas (1993) describe, as learning is multidimensional, how the result of training should be assessed terms of changes in affective, behaviour (skill-based) and cognitive (knowledge-based) capacities. For the current study, the Alliger et al. augmented taxonomy (1997) was used as the fundament when conceptualizing the surveys. The different levels of the Alliger et al. augmented taxonomy are briefly described below (the corresponding levels from Kirkpatrick's taxonomy and Bell and Waag's simulator training evaluation model within parenthesis):

- Level 1. Reactions (Kirkpatrick level: Reactions, Bell & Waag level: Utility evaluation)
    - Level 1a. Affective reactions: measures assessing to what extent trainees liked and enjoyed the training.
    - Level 1b. Utility reactions: measures assessing the perceived utility value of the training.
- Level 2. Learning (Kirkpatrick level: Learning, Bell & Waag level: Performance improvement [in-simulator learning])
    - Level 2a. Immediate post-training knowledge: measures assessing how much trainees know about the training topic directly after the training.
    - Level 2b. Knowledge retention: measures similar or identical to level 2a measures but administered at a later time than directly after training.
    - Level 2c. Behaviour/skill demonstration: measures assessing the behavioural proficiency within the training, rather than the work environment.
- Level 3. Transfer: measures that assess to what extent the knowledge and skills attained during training actually are usable in the real work environment. (Kirkpatrick level: Behaviour, Bell & Waag levels: 3.a.) transfer to alternative simulation environment, and 3.b.) transfer to operational environment)
- Level 4. Results: measures that assess the organizational impact of training such as, for example, productivity gains, cost savings etc. Measurements on the results level are the most distal from the actual training but by some perceived as the most fundamental when judging training success, as they are linked to they underlying reason why the training was performed. (Kirkpatrick level: Results, Bell & Waag level: extrapolation to combat environment)

Training Evaluation Method

Participants

Fourteen SwAF JAS39 Gripen pilots participated in the study. They constituted the current SwAF rapid reaction unit at the time, which means they were all experienced combat ready pilots. Individual total military flight hours varied from 950 hr to 2500 hr (M = 1705 hr, SD = 520 hr). Based on their experience, all participants were considered as subject matter experts (SMEs), able to provide highly reliable estimates concerning their own training. This was a pre-condition for the methodological approach of the study.

Design of the study

The study was conducted as a within-participant design. The dependent measures were the reactions, and the ratings of perceived training value and additional training needs. Independent measures were the training during the RF spin-up exercise and RF 08-3 respectively. All data collection was based on surveys.

Description of the two exercises

Red Flag spin-up. The RF spin-up exercise was conducted over four days at the SwAF simulation facility FLSC about one month prior to the live exercise RF 08-3. All simulation scenarios were flown over a satellite photo generated geographical database of NTTR. During a number of workshops, pilots of the unit in dialogue with instructors and training designers at FLSC analyzed training needs for the live exercise and identified essential experiences that could be provided at FLSC. Each pilots' levels of tactical execution performance was considered to meet or exceed the requirements for entering the live RF 08-3 exercise. Based on that, for the RF spin-up training, tactical execution was not considered a prioritized training objective. On the other hand, and to enable highest level of tactical execution at RF 08-3, the elements that were identified as needs among the pilots were knowledge closely attached to Nellis AFB and NTTR. Examples of such elements are local flight control procedures, ground operations, communication protocols, training rules and regulations, geographical knowledge and familiarity, and time and fuel management. The goal for the RF spin-up training was the provision of the experiences supporting the development of the knowledge and skills associated to these elements in order to allow a strong focus on tactical execution and highest possible performance at RF 08-3.

During RF spin-up every pilot was exposed to five different scenarios: a familiarization flight over NTTR, a Nellis AFB air traffic control procedures, a four vs. four air-to-air engagement over NTTR, and two large force employments (LFE) where airspace and time management was in focus. All through the week, researchers and SMEs (retired pilots) from the US Air Force Research Lab (AFRL) Warfighter Readiness Research Division (Mesa, AZ) and one pilot from the 65th Aggressor squadron at Nellis AFB supported the training, based on their local knowledge and experience. The aggressor pilot provided briefings on Nellis AFB air traffic control procedures, RF training rules and regulations, and NTTR airspace and target areas. The aggressor pilot also provided feedback on the performed sorties and was available for questions.

Red Flag 08-3. For the SwAF pilots the primary training goals for the exercise was to participate in LFE sorties (sorties with many and usually dissimilar aircraft types) in the air-to-ground role to further develop their capacity to participate in coalition operations. Another goal was to enhance and validate their close air support (CAS) and dynamic targeting (DT) ability and tactics. A number of different aircraft types participated in the exercise such as fighter aircraft in air-to-air and air-to-ground roles, tankers, command and control aircraft, and electronic warfare aircraft. Missions were flown twice per day and the enemy side consisted of special aggressor squadrons flying air-to-air missions to counter the LFEs. Simulated surface-to-air missile platforms provided a challenging ground based air defence (GBAD) "threatening" the aircraft.

Contributing research efforts

Mission Essential Competencies. Under a project agreement between FOI and AFRL, the Mission Essential Competencies (MEC) process (Colegrove & Alliger, 2003; Alliger, Beard, Bennett, Colegrove, & Garrity, 2007; Alliger, Beard, Bennett & Colegrove, in press) have been utilized to identify the essential core of experiences, knowledges and skills necessary for a combat mission ready JAS39 Gripen pilot (Bennett et al., 2006). MEC entails competence descriptors at various levels. For this evaluation a set of 36 knowledge and 50 skill requirements was used as a thorough description of "all the pilots need to know". The MEC knowledge definition is "info or fact that can be accessed quickly under stress" and one example of a knowledge statement for SwAF JAS39 Gripen is "radar warning and threat reactions". The MEC skills are defined as "compiled actions that can be carried out successfully under stress" and one example is "interpret rules of engagement".

Similarities and differences to previous MEC-based training evaluations. In the US/UK Red Skies study (Smith et al., 2007) similar research objectives such as the ones of this study were present. The Red Skies study was a DMO/MTDS (Distributed Mission Operations/Mission Training through Distributed Simulation) research trial, in addition to being a training event. This was not the case for the present study, thus the surveys that were developed and used placed a somewhat different emphasis on what was assessed. In the current study measures of development of knowledge and skills were in focus, whereas the assessment the Red Skies study placed an emphasis on MEC experiences. Further, no observer protocols of performance were used in this study. Given the criterion issues in performance measurement for a knowledge and skill set as extensive as that identified during the MEC process, training effect rather than performance was chosen as the primary construct to assess.

Surveys and data collection

To capture the relevant data from the participants before, during and after RF spin-up and RF 08-3 a battery of five surveys was developed. Table 1 provides an overview of where and when each survey was collected and what it captured.

Demographics survey. A demographics survey, not presented in Table 1, designed to capture background information about the pilots to be able to sort them by previous experience was collected before the RF spin-up training started.

Knowledge and skills addition training needs survey. The rationale for the knowledge and skills additional training needs survey (ATN) was to establish the participating pilots initial performance readiness prior to RF spin-up, and then repeatedly monitor their performance readiness development in terms of additional training needs desired for each of the 36 knowledge and 50 skill requirements. The pilots completed the ATN survey three times:

first and last day of the RF spin-up and last day of RF 08-3 (plan was to collect this data also the first day of RF 08-3 but that data could not be collected due to operational constraints). The participants also provided calibrated values of their previously rated performance readiness at the two last rating occasions (i.e., ATN3, ATN6 and ATN7 in Table 1). Their actual rating from the previous occasion was not presented to them when making this calibration.

Knowledge and skills perceived training value survey. The knowledge and skills perceived training value survey (PTV) assessed the pilots' ratings of perceived training value of each sortie flown at RF spin-up and RF 08-3 right after it was finished (i.e., to what extent did the previous sortie provide training value for the knowledge and skills). In a workshop, pilots, simulator instructors and training designers identified a subset of 28 knowledge and 40 skill requirements (out of the total of 86 from the ATN survey) based on how likely they were anticipated to be developed at RF 08-3. This was to decrease the intrusiveness of the survey since it was used within the training context.

Reactions survey. The reactions survey contained questions concerning both the affective and utility reactions to RF spin-up. It was collected both after the spin-up and after RF 08-3 (in both cases concerning reactions to RF spin-up).

Top 3-Bottom 3 survey. The top 3-bottom 3 survey (T3-B3), not presented in Table 1, was completed once at the end of each day, both during RF spin-up and during RF 08-3. The pilots straightforwardly listed the three best and the three worst events of the day and thus provided a wealth of qualitative data concerning the exercises.

Table 1. *Surveys used for data collection before, during and after RF spin-up and RF 08-3.*

Exercise	Time	Knowledge & skills additional training needs survey (ATN)	Knowledge & skills perceived training value survey (PTV)	Reactions survey (R)
RF spin-up (simulator exercise)	Before	ATN1. Current performance readiness		
	During		PTV1. Perceived training value	
	After	ATN2. Current performance readiness ATN3. Calibrated ATN1		R1. Affective & Utility reactions
RF 08-3 (live exercise)	Before	ATN4. Current performance readiness[a]		
	During		PTV2. Perceived training value	
	After	ATN5. Current performance readiness ATN6. Calibrated ATN1 ATN7. Calibrated ATN2		R2. Affective & Utility reactions

[a] ATN4 data was not collected in this study due to operational constraints at RF 08-3.

It is sometimes claimed that the reliability and validity of subjective ratings of many psychological constructs are insufficient and that it can be difficult to fully determine what has been measured. Doubts about their validity, although sometimes justified, should not be exaggerated. Even if the precision of any single rating is modest, data may still be sufficiently rich of information to be useful. The authors want to highlight the experienced pilot, as an intelligent filter against the complexity of the world, who has the capability to integrate his or her experience into a balanced measure. Note that this capability is dependent upon the nature of the construct that is being assessed, as not all mental processes are introspectively available. However, for assessment of operative field training effects, subjective ratings are useful and in most cases the only practicable way forward. Validity is further increased when ratings are collected before, during and after training, and when calibration ratings of previous status are included as this allows for control of what Golembiewski, Billingsley and Yeager (1967) calls alpha, beta and gamma types of change.

As a note to Bell and Waag's level "extrapolation of transfer to combat environment" the training during an exercise as Red Flag can in many respects provide better training of mission execution than war operations, as these highly realistic exercises allow more comprehensive and concentrated exposure to the full envelope of situations during a mission. In that respect, it constitutes not only an excellent training opportunity but also a representative training criterion.

Experiences from application

Survey data mapping to Alliger et al. augmented taxonomy

The Alliger et al. augmented taxonomy (1997) was used to conceptualize the surveys in order to capture data at all levels and meet expected analysis goals. Table 2 shows how the collected data set was linked to the specific levels.

Table 2. Summary of how data from each survey evaluates training on the different levels of the Alliger et al. (1997) augmented taxonomy.

1. Reactions		2. Learning			3. Transfer	4. Results
1.a. Affective	1.b. Utility	2.a. Immediate Knowledge	2.b. Knowledge retention	2.c. Behavioural/Skill demonstration		
R1 R2	R1 R2	PTV1[a] PTV2[a]	ATN2[b]	ATN5[c]	ΔATN6-ATN7[d]	Post analysis[e]

[a] PTV classified as immediate knowledge based on the assumption that a perceived training value indicates that learning has occurred. [b] Performance readiness status after one week of exposure to RF spin-up training. [c] Performance readiness status after exposure to RF 08-3 live training and performance criterion. [d] The delta between ATN6 and ATN7 is a measure of how much of the total training effect from before RF spin-up to after RF 08-3 that each pilot attributes to RF spin-up. This is based on the fact that ATN5, ATN6 and ATN7 for each knowledge and skill are rated at the same time. [e] Linking ATN training effects to the official SwAF training objectives for RF 08-3, for example to support cost-benefit analyses comparing magnitude of training effects and associated cost levels.

Examples of data

The data from the ATN surveys distinguished which knowledge and skill statements that received the most development during RF spin-up and during RF.

The mean of means from the ATN survey, together with SME mappings of each knowledge and skill to the SwAF training goals enabled a post-hoc analysis concerning the fulfilment of the same. The results of this type of evaluation yield interest at all level of the military hierarchy, decision maker not the least, and provide the often difficult mapping to the Results level.

In the MEC training gap analysis previously conducted for the JAS39 Gripen, a number of experiences for which the pilots desired more exposure were identified. Through SME mappings between these training gaps and the exposure to experiences that lead to reduction of training gaps during RF08-3, indications to what extent RF08-3 addressed the training gaps was extracted from the data.

The PTV survey provided detailed data from each sortie that can be used when analysing the training contribution from each specific sortie. The data also entail information useful when discussing the design of the sorties and when the SwAF in future exercises express expectations and training needs to the RF staff.

Through the T3-B3 survey a large amount of qualitative data concerning the exercises was captured, which corroborates the quantitative data.

Future analyses & methodological enhancements

The Alliger et al. meta-analysis (1997) reported that few studies of training present data and correlations between the different levels of Kirkpatrick's original taxonomy. One of the goals of this study is to analyze the effects of each measure but also to correlate between the levels of the Alliger et al. augmented framework as presented in Table 2.

In order to assess the training effect for the full range of knowledge and skills, which can be described as the lion's share of the pilots' full professional competence, comprehensive and structured assessment approaches such as the one described here are needed when studying training outside of the laboratory. Ultimately a mix of structured self-ratings, instructor ratings and logged performance data from simulator and/or aircraft would be collected. However, for studies of training with high ecological validity there are often issues with ceiling effects in the performance measures, such as number of air-to-air missile hits or bomb accuracy, when studying high performing pilots. This leads on to the almost the philosophical question whether training value and training effect are more representative constructs to evaluate than performance, although performance increases are the end goal of training.

The formulation of organizational goals and training objectives has always been a challenge for both researchers and the operational community. An observation from this study, describing a point observed many times before, is that if training goals were expressed in a more clear and precise form, training evaluations, and in the long run the training itself, could be developed much further.

References

Alliger, G. M., Beard, R., Bennett, W., Jr., & Colegrove, C. M. (in press). Mission essential competencies: An integrative approach to job and work analysis. In M. A. Wilson, G. M. Alliger, W. Bennett, Jr., R. J. Harvey., F. P. Morgeson, K. J. Nilan & E. Salas (Eds.), *The handbook on job and work analysis: The methods, systems, applications, & science of work measurement in organizations.* Mahwah, NJ: Taylor Francis.

Alliger, G. M, Beard, R., Bennett, W., Jr., Colegrove, C. M., & Garrity, M. (2007). Understanding Mission Essential Competencies as a Work Analysis Method (AFRL-HE-AZ-TR-2007-0034). Mesa, AZ: Air Force Research Laboratory.

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. Personnel Psychology, 50, 341-358.

Bell, H. H., & Waag, W. I. (1998). Evaluating the effectiveness of flight simulators for training combat skills: A review. *International Journal of Aviation Psychology, 8*(3), 223-242.

Bennett, W., Jr., Borgvall, J., Lavén, P., Castor, M., Gehr, S. E., Schreiber, B., et al. (2006). International mission training research (IMTR): Competency-based methods for interoperable training, rehearsal and evaluation. In the proceedings of the Simulation and Interoperability Standards Organization (SISO), *European Interoperability Workshop (EuroSIW)*. Stockholm, Sweden: SISO.

Colegrove, C. M., & Alliger, G. M. (2003). Mission essential competencies: Defining combat readiness in a novel way. In proceedings of NATO Research and Technology Organization (RTO) System Analysis and Studies Panel's (SAS) Symposium, *Air Mission Training Through Distributed Simulation (MTDS) – Achieving and Maintaining Readiness (SAS-038).* Brussels, Belgium: NATO RTO.

Golembiewski, R.T., Billingsley, K. R., & Yeager, S. (1976). Measuring change and persistence in human affairs: types of change generated by OD designs. *Journal of Applied Behavioural Science*, 12, 133-157.

Kirkpatrick, D. L. (1959a). Techniques for evaluating training programs. *Journal of ASTD, 13*, 3-9.

Kirkpatrick, D. L. (1959b). Techniques for evaluating training programs: Part 2-Learning. *Journal of ASTD, 13*, 21-26.

Kirkpatrick, D. L. (1960a). Techniques for evaluating training programs: Part 3-Behaviour. *Journal of ASTD, 14*, 13-18.

Kirkpatrick, D. L. (1960b). Techniques for evaluating training programs: Part 4-Results. *Journal of ASTD, 14*, 28-32.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, 78, 311-328.

Smith, E., McIntyre, H., Gehr, S.E., Symons, S., Schreiber, B., & Bennett, W., Jr. (2007). Evaluating the impacts of mission training via distributed simulation on live exercise performance: Results from the US/UK "red skies" study (AFRL-HE-AZ-TR-2206-0004). Mesa, AZ: Air Force Research Laboratory.

DEVELOPMENT OF PROACTIVE SAFETY MANAGEMENT SYSTEM FOR INDUSTRIAL FIELDS BASED ON THE FRAMEWORK OF AVIATION SAFETY REPORTING SYSTEM

Capt. Akira Ishibahsi
Graduate School of Engineering, Tohoku Univ.
Sendai, Japan
Daisuke Karikawa, Toshio Wakabayashi, Makoto Takahashi and Masaharu Kitamura
Graduate School of Engineering, Tohoku Univ.
Sendai, Japan

The safety reporting system such as the Aviation Safety Reporting System (ASRS) draws attention from various industrial fields as an effective safety management method to prevent further accidents. However, it became apparent that the industrial fields often confront the difficulty of the development and effective operation of safety reporting system due to the differences between aviation and other industrial fields. In this study, an effective safety reporting system for practical use in a conventional industrial field has been developed based on ASRS. Although the detailed evaluation of the proposed safety reporting system is still underway, its effectiveness has been strongly implied through the actual utilization as a proactive safety management system of a construction company.

In order to enhance the safety of public transportation, Japanese Ministry of Land, Infrastructure, Transport and Tourism (MLIT) issued a mandatory requirement for all service providers of public transportation to introduce Safety Management System (SMS), based on the lessons learned from the Aviation Safety Reporting System (ASRS). The safety reporting system also draws attention from other industrial field as an effective safety management method to prevent industrial accidents. However, it has become apparent that the industrial fields often confront the difficulty of the development and effective operation of proactive safety management system. In general, the safety reporting system in aviation area is operated by the government with ample budget and human resources. The collection and analyses of safety information are performed by many concerned parties. On the other hand, in most of the conventional industrial fields, a company is required to establish and operate safety reporting system within limited resources from collecting information to formulating safety measures based on the analysis of the acquired information. Research and development activities taking such differences between aviation and other industrial fields into consideration are certainly required to apply SMS developed in aviation field to other conventional industrial domains.

In this study, the essential issues to be considered to develop effective safety reporting system for practical use in a conventional industrial field have been discussed based on the extensive application experiences. The target field of the safety reporting system tackled in this study is not only transportation area but also manufacturing industry, construction, nuclear power plant and space development industries.

Method

In order to apply the safety reporting system to conventional industrial fields with various differences from aviation area, the authors have developed a modified safety reporting system called the Industrial Safety Reporting System (ISRS) based on ASRS. The conceptual scheme of ISRS is summarized in Table 1 as contrasted with one of ASRS. The point is that required human resources for system operation of ISRS should be reduced compared with ASRS because available resources for safety activities are very limited in many industrial fields

Table 1. Conceptual scheme of ASRS and ISRS

ASRS	**Industrial Safety Reporting System**
1) Basic Principle	***1) Basic Principle***
Voluntary, Confidential, Non-punitive	Voluntary, Confidential, Non-punitive, Field-oriented
2) Reporter Protection	***2) Reporter Protection***
Immunity (assured by formal documents) Confidentiality and De-identification	Non-punitive to the reporter Separation from personal evaluation De-identification after completion of hearings
3) Transparency of the System	***3) Transparency of the System***
Clarification of Responsible official Clarification of Objectives and Operation	Clarification of Responsible official Clarification of Objectives and Operation Clarification and Publicity of Submitting Procedure
4)Organization and Operation	***4)Organization and Operation***
Secretariat Report processing, Root cause analysis, Callback, Countermeasures and Recommendation, Publication	Secretariat, Project team Report processing, Root cause analysis, Hearings, Risk assessment, Countermeasures, Publication of case list Improved reporting format to simplify data entry work
5)Feedback	***5)Feedback***
Statistical report, Callback, Quarter report, Quick actions etc.	Statistical Report (1 page), Poster, Case list, Notice letter
6) Promotion	***6) Promotion***
Certificate of receipt, Appreciation letter	Kick-off campaign, Safety forum or Workshop Incentive awards (from¥1000-¥10,000), nickname of system and slogan (“Near-incident is a present by angels”)

compared with ones in aviation. For example, reporting format of ISRS are designed to simplify data entry works into the incident database so that staffs without domain knowledge can perform the works easily and smoothly. In addition, because the know-how to analyze near-incident reports and to develop countermeasures based on the result of analysis is not so common in conventional industrial companies, effective analytical methods of collected reports should be provided as a part of the framework of safety reporting system. Promotion activities are quite important for developing shared understanding of non-punitive and non-disparage principle of a safety reporting system which can significantly affect its feasibility in a company. One of the effective ways for such kind of promotion activities is to give a friendly nickname or a catch-phrase to the safety reporting system such as "A near-incident is a present by angels".

Application

ISRS has been applied for safety enhancement activities of Company T. The domain of Company T is maintenance for electrical power plants. Company T has 1,310 employees and also has 4,000 employees of subcontractors. For enhancing safety in the company, one of the authors has worked with them for introduction of ISRS from the year of 2006. This chapter summarized the process and result of the application of ISRS in an actual field as one of the examples of the ISRS application to the conventional industrial domains.

Introduction of ISRS

Before introduction of ISRS, Company T had the original safety reporting system gathering employee's safety report by a questionnaire format. However, the original safety reporting system failed to gather useful information and ISRS has been introduced as the renewed company's safety management system since April 2006. The overview of the introducing process is described in the following.

Table 2. Criteria of Risk Quantification in ISRS

Evaluation Points	Risk Quantification (Calculation of Risk Score)
1) The number of past reports about similar near-incidents	The risk score is added 1 point for each past report. (The max additional score is 6 points.)
2) Frequency of the work involving the risk	Daily: +4 points. Every work period: +2 points. Rare: +1 point.
3) Anticipated damage	Virtually-undamaged: +1 point light injury: +5 points Fatal & serious injury: +10 points

Establish secretariat and project team. A group manager has been assigned as a responsible official. Two newly-employed part time stuffs were appointed as administrators of the near-incident database. In addition, project team was organized in each regional office as back-up team for the secretariat. The secretariat has wide and important roles as listed below.

receive and analyze reports / conduct hearing investigations / conduct the risk assessment of reported near-incidents / develop countermeasures / give feedback to company members / produce case list

Improve the near-incident database. New near-incident database system has been installed. The permission to access the database has been given to everyone in the company. Management works of the database has been performed by the secretariat.

Improve the reporting format. Reporting format has been changed to a new format with high compatibility with the near-incidents database system. It contributed to reduce the burden of data entry works into the database.

Give a Nickname to ISRS. In order to familiarize company members with the proposed safety reporting system, the slogan "A near-incident is a present from angels" was decided. The reporting format was also named "Experience note of angel's present".

Introduce the incentive award. In order to collect as many as experiences of near-incidents from workers, the institution of the incentive awards has been introduced. A reporter has been given the reward from 1,000 yen to 10000 yen based on the contribution of her/his report to the enhancement of safety.

Analysis and Utilization

As a part of ISRS, the methods for analysis of collected reports and for utilization of the result of analysis have been provided to the company. This is a very important point because the know-how for analysis and utilization of the safety reports is not so common in the conventional industrial fields. The standard process for analysis and utilization in ISRS is described as follow:

1. Case analysis. In order to reveal the root cause and background factors of the near-incidents, the collected reports are analyzed by the secretariat from the view point of human factors. M-SHEL model and Variation Tree Analysis (VTA) are utilized as standard analytical tools in Company T

Table 3. Risk Level

Risk Level	Risk Score (cf. Table. 2)
Level 1 (negligibly-small)	1 - 3points
Level 2 (acceptable)	4 - 7points
Level 3 (should be reduced)	8 - 11points
Level 4 (should be immediately reduced)	12 - 20 points

Table 4. Quick reference matrix of countermeasures effectiveness

Type of Countermeasures	Effectiveness
Remove the safety risk	7 points (Ex. Using safety belt in high-place work、Put up safety net, improvement of construction method
Decrease the safety risk by improving manner of operation	4 points (Ex. assignment of safety observer)
Cautionary notices about the safety risk	2 points (Ex. heads-up at daily meeting)

2. Risk assessment. As it is important to quantify the risk level of the collected near-incidents for prioritizing countermeasures, the ISRS provide the convenient criteria for quantification of potential safety risks as described in Table 2 and Table 3.

3. Hearing investigation. The collected reports do not always contain enough information to extract lessons from the occurred near-incident. In such a case, hearing investigation is performed by the secretariat. A member of the secretariat visits a reporter and conducts the interview confidentially. After the completion of the hearing investigation, the collected report is de-indentified.

4. Countermeasures (Preliminarily Evaluation of Effectiveness). For realizing steady enhancement of safety, the effectiveness of countermeasures should be evaluated explicitly. ISRS provides a quick reference matrix for simplified evaluation of countermeasures as shown in table 4. For example, if a countermeasure removing the objective safety risk is taken, 7 points are subtracted from the risk score calculated based on table 2. Although the score given by the evaluation scheme provided here is just an approximated figure, it helps intuitive understanding of important facts concerning human factors such as “cautionary notices about the safety risk are hardly effective to prevent accidents”.

5. Feedback. The results of analysis of the collected reports have been fed back to company members in the form of a periodic report, a bulletin board, a letter ruling and so on. All the member of the company also can access safety information obtained from the safety reports through the near-incident database. In addition, a case list is quite important to spread safety information horizontally throughout the company. Our case list contains not only facts and direct causes of an occurred incident but also background factors, original risk level calculated by means shown in table 2 and 3, adopted countermeasures and remaining risk level. The additional information can support frontline workers to utilize the case list as a reference to prevent similar incidents by themselves.

Result

Company T has applied ISRS as a company's safety management system for two years. In the last two years, 1555 reports have been submitted and 155 cases have published as a case list. The case list has been utilized as an effective reference in the safety meeting of each work place every day. The detailed evaluation of ISRS is still underway, but the fact that ISRS has been positively accepted in Company T have strongly implied the effectiveness of ISRS

Conclusion

For the proactive safety management in conventional industrial field, the present study has proposed a safety reporting system called the Industrial Safety Reporting System (ISRS) based on ASRS. ISRS has been adapted for practical use in conventional industrial fields taking their features and constraints into consideration. Although the detailed evaluation of ISRS is still underway, the authors believe its effectiveness based on the achievements of ISRS as the safety reporting system of a construction company in the past two years. The further study to develop objective and quantitative safety measure is still going on to elaborate the proposed framework.

References

NASA ASRS Office (2000). Aviation Safety Reporting System Program Overview

Association of Air Transport Engineering and Research (ATEC) Tokyo (2007). Safety Management System in Air-Transport System（SMS） Seminar Text

Capt. Akira Ishibashi (2008). Information-Sharing on Factors that induce Human Error in Aviation. Kitakyusyu, JAPAN: *In Proceedings of the 1st East Asian Ergonomics Federation Symposium*

TEAM ERRORS IN AIR TRAFFIC CONTROL: ANALYSIS BASED ON VOLUNTARY REPORTS

Renli Lu

Civil Aviation Safety Academy, Beijing, China

Yan Zhou

Air Traffic Management Bureau of East China, Shanghai, China

Mu Zhou

Civil Aviation Safety Academy, Beijing, China

By using the framework of team errors taxonomy(Sasou & Reason, 1999）and Threat and Error Management (TEM), operation errors reported in the ATC voluntary safety report system are analysed. Team errors have been proved in this research as a main risk source of ATC, a sophisticated social-technical system. Team errors contributing to 31% of the total reporting ATC errors. According to the results of sample statistics and analysis of typical cases of team errors, the main types and distributions of ATC team errors have been identified, as well as the framework and importance of sequencing of the team PSFs（Performance Shaping Factors）.

Introduction

In aviation operation system directly related with aviation safety, to reduce human error, multi-staff working posts are established to realize cross-check and coordination (e.g. two-pilot crew in cockpit). But the human error in team work situation still continuously happened. Team error is a special type of human error which defined by Reason (1990). When an operation team tries to accomplish common task, but sometimes team lose functions of identifying, specifying and correcting human errors. The human error of individual made in a team processes is not effectively detected, indicated or corrected, can be defined as team error.

Sasou and Reason（1999）has conducted team error analysis on events that happened in the nuclear, aviation and shipping industries. Following their framework, this research conduct team error researches in air traffic operation context.

Regarding research data, the most common used data source in air traffic control human error research basically concentrates on flight accident/incident database, such as work of Pape and his colleagues(Pape et al, 2001）. However, even though data from accident/incident formal reports are very detailed, the reliability of research conclusion is quite restricted due to the limited number of reports. Air Traffic Management system in Chinese Civil Aviation system (44 directly subordinated operation facilities) normally reports no more than 10

ATC incidents each year. From 1998 to 2007, only 42 ATC incidents were reported in ten years.

Aviation Voluntary Reporting System is a successful safety approach in international aviation community (Such as ASRS). It encourages operation staffs actively report potential problems or mistakes in working, created a safety information sharing mechanism by communication and discussion. Proactive and just culture is important foundation to ensure the success of this mechanism. Voluntary report is important data source for aviation safety research, especially in human factor research. Even though there could be some limits to make it the research database, such as information not valid, report bias, etc, which make report information difficult to used in relative frequency of occurrence and trend analysis (ICAO, 2006). But voluntary report can still be counted as a reliable data source, and the ignorance of punishment normally will encourage reporters more comprehensively report the process and background of operation mistakes, therefore promote reliability and accountability of data.

According to Civil Aviation Administration of China(CAAC) regulations, ATC operation uses teamwork mode. Each control sector runs “dual-post” rule. Two controllers will collaboratively work according to standard control process, and jointly take responsibility of the safety and efficiency in airspace operation control. Therefore CAAC ATC operation is an ideal environment in conducting team error research. The data of this research come from the voluntary report system established in one ATC center under CAAC ATMB.

Team errors definition and taxonomy

According to team error framework（Sasou & Reason, 1999）, team-error can be categorized as:

- Individual Errors: The errors are made by individuals, the other controllers do not direct relate with the occurrence of the errors.
- Shared Errors: The entire team makes understanding error, furthermore makes wrong decision or adopts wrong activity.

Or team errors can be categorized from decision information prospective:

- Independent Errors: The errors are made when all available information is correct.
- Dependent Errors: Some of the information is incorrect, inappropriate or biased which causes the controller behaves inappropriately under specific condition.

The correction of team error can be categorized as 3 key phases:

- Fail to Detect: the first phase of error correction. If team members fail to identify the occurrence of error, there will be no chance to correct it;
- Fail to indicate: Some team members detected the error, but for some reason they fail to indicate it to the responsible controller. This is the second phase of error correction.
- Fail to correct: Team members detected and indicated the occurrence of error. But due to inappropriate ways, methods or channels of communication, they failed to urge the responsible controller effectively correct the errors.

At least one of the 3 above stated phases appeared team error really happen, as shown in Figure 1.

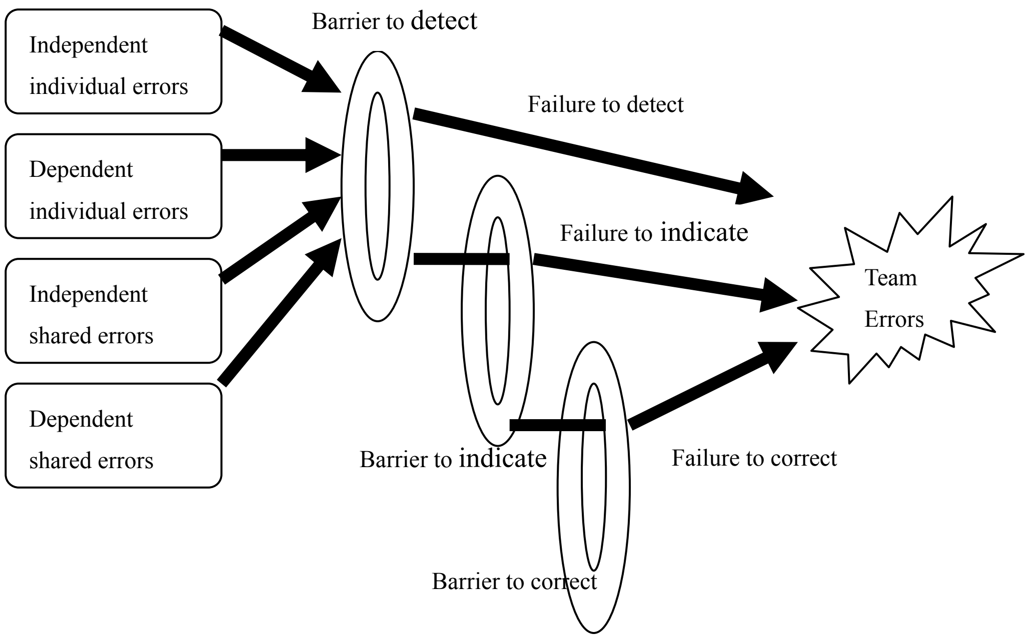


Figure 1. Team Error Process (Sasou & Reason, 1999).

Threats in air traffic control

Threat and error management framework is the theory foundation of Line Operation Safety Audit(LOSA) in cocpit and Normal Operation Safety Survey（NOSS） in ATC. ICAO categorizes ATC threats into 4 types (ICAO, 2008), as shown in Table 1. One of the important differences between threat in voluntary report and threat detected by NOSS is that threats in voluntary are those result in consequences (errors or undesired states). Therefore, the threats from "Other Controllers" in team operation condition are the major reason of team error.

Table 1. *Threats Categories in ATC (ICAO, 2008).*

ATSP internal	ATSP external	Airborne	Environmental
Equipment	Airport layout	Pilots	Weather
Workplace factors	Navigation aids	Aircraft performance	Geographical
Procedures	Airspace infrastructure/design	R/T communication	environment
Other controllers*	Adjacent units	Traffic	

Note. Threats from other controllers are most relevant to team errors.

Team PSFs

The identification and understanding process of human being relies heavily on environment factors.

Generally speaking, error is the result of combinational effect from various affecting factors. These factors can be defined as PSF (Performance Shaping Factors). When team members detected but failed to indicate or correct individual or share error, the reason normally can be found from effects of human-to-human internal relationship or working environment. Team PSFs can be identified as factors arising from a group of people working together on a common project or task.

Some of the common Team PSFs are: deficiency in communication, excessive belief (on one assumption or precondition), deficiency in resource/task management, inappropriate authority gradient, excessive professional courtesy, over-trusting, air of confidence, friendship, organization factors, etc.

Methodology

Voluntary reports provide a lot of real cases for human errors analysis. To ensure the integration and comparability of analysis, the voluntary reports from one ATC center in 2004 are selected as basic research sample in this research. ATC operation unrelated errors (such as report of study, reflection of work, summary of new regulations or new operation methods, etc.); duplicated reports (reports on same occurrence from various units) and invalid reports (information incomplete or hard to understand) are deleted. Altogether 370 valid reports are collected as analysis samples.

In analysis, two ATC experts, who understand the team error and TEM principles, made independent selection and analysis of samples, discussed about discrepancy during selection and analysis, and ultimately achieved consistent conclusion as the analysis result.

CASE

En-route control center. Radar control circumstance. Flight A estimated flied over position X at 13:30, on flight level 7500 metres and expected to pass position Y at flight level 5100 metres (handover flight level). Flight B estimated cross position Y at 13:31, on flight level 5400 metres. Two flights were flying on opposite direction. After Flight A reported passing X, principle controller instructed Flight A descent to flight level 5100 metres. Assistant Controller detected the error and reminded principle controller that Flight B was in adjacent which meant the clearance was not acceptable. But he did not clearly suggest that two flights should maintain vetical seperation until across each other. After being indicated, the principle controller just instructed flight A expedited descending. As it is still difficult to make a cross with acceptable separation, he instruct Flight A turn left on heading 360 to deviate from planed route. In the process, the assistant controller did not take any further action. When two aircrafts crossed each other, the separation of two flights was less than minimum.

Conclusion: Independent individual error; error be detected, indicated, but not corrected.

Team PSF: (1) Principle controller’s over-confidence; (2) Assistant controller’s excessive professional courtesy; (3) authority gradient is too flat in this team.

Results and Discussion

In comprehensive sample distribution (see table2), 45% of voluntary reports are related with operational error. Among these errors, 15% of samples are cases that other controllers indicated timely and errors are corrected timely. 14% samples are cases that team corrected timely on unsafe occurrences caused by pilots (e.g. change altitude or course without permission, take wrong flight route, forget to follow instructions, etc.), or team indicated timely on flight conflicts in adjacent ATC centers. This proved the effectiveness of team work mode.

Table 2. *Comprehensive Distribute of Voluntary Reports.*

Type	Sub-type		Number of Reports	ratio
Reports without Operational Errors	notifying of abnormal circumstance, experiences of resolve complicated flight conflicts, dealing with unusual situation, discussion of handling measures for special operational scenario, etc.		N_1=204	55%
Reports with Operational Errors	Flight Errors	Detected and Indicated Pilots' Errors	N_{21}=52	14%
	ATC Error	Detected errors made by themsevles and corrected timely	N_{22}=7	2%
		Detected errors made by Other Controllers and corrected timely	N_{23}=56	15%
		Team Error	**N_{24}=51**	**14%**
Total			370	100%

Among all reports, 14% reports (n=51) are related with team errors. To emphasize, errors related with collaboration, supervision and correction inside the team are 14% of all effective reports; 31% of all reports with operational errors (number of team errors related reports divided by number of reports with operational errors). According to distribute of voluntary reports, after ATC errors do occurred, the cases that team can not effectively detect, indicate and correct are 44.7%(number of team errors related reports divided by number of all reports with ATC errors). The error management capability will still be an important area of future ATC operation risk control.

By using Threats and Error Management (TEM) research framework (ICAO, 2008), we conducted operation threat analysis on sample reports. The analysis result is illustrated in Table3. The threats of internal collaboration inside operation team are 24.31 % (n=132) of all threats, therefore it becomes into the biggest individual source of threats in ATC operation. This explains that in ATC operation, the key source of threats to safety is Human, including those other controllers or team members who willing to assist and cooperate.

Table 3. *TEM Threat Analysis of Voluntary Reports.*

Type	Sub Type	Number of Threats	Ratio		Explanation
Airborne	R/T Communication	39	7.18%	31.3%	Controller-Pilot VHF radio communication errors
	Air Traffic	49	9.02%		Overloaded traffic or special flight requirements
	Aircraft Performance	21	3.87%		Aircraft performance limitation

	Pilot	61	11.23%		Threats caused by pilot's mistakes or flaws in collaboration.
ATSP external	Airport Layout	3	0.55%	23.02%	Threats caused by unreasonable airport design
	Airspace Structure/Design	12	2.21%		Threats caused by unreasonable airspace structure
	Other ATC Units	110	20.26%		Threats caused by bad coordination with, or special restriction or requirements from other ATC units (including Air Force ATC units).
Environmental	Geographical Environment	5	0.92%	10.5%	Threats caused by special Terrain
	Environmental Interference	11	2.03%		Noise interference in working environment or in R/T communication
	Weather	41	7.55%		Threats caused by bad weather condition such as thunder storm.
ATSP internal	**Other Controllers***	**132**	**24.31%**	35.17%	**Threats caused by cooperation or coordination among controllers**
	Equipments	41	7.55%		Threats caused by ATC Equipment maintenance
	Procedures	18	3.31%		Threats caused by unreasonable working procedures
Total		543	100%		

Note. The threats from other controllers are most relevant to team errors.

51 reports related with team errors are categorized and analyzed according to error types and phases of occurrences. The analysis result is listed in table 4. From the phase of occurrence prospective, even though "fail to correct" are most unlikely happened phase, where 8% (n=4) are of this type. When errors have been detected and indicated but still fail to correct show the most typically flaws of team culture.

From error type prospective, shared errors are 39% (n=20) of all team human errors, which is less than individual errors (61%, n=31). But shared error is caused by concurrent mistakes in understanding, decision making and acting among team members, therefore hard to be detected and corrected. Dependent errors are less commonly seen, with a ratio of 24% (n=12) in all team errors. However, since dependent error is caused by insufficient or wrong operational information, the flaw of information increases the likeliness of errors. Therefore, the dependent shared errors, 12% of all errors, are the error type that brings in high probability of safety hazard.

Table 4. *Statistic Analysis Result of Team Error Categorization from Voluntary Report Samples.*

	Fail to Detect	Fail to Indicate	Fail to Correct	Total	Ratio
Independent Individual Errors	16	6	3	25	49%
Dependent Individual Errors	5	1	0	6	12%
Independent Shared Errors	11	2	1	14	27%
Dependent Shared Errors	6	0	0	6	12%
Total	38	9	4	51	100%

Ratio	74%	18%	8%	100%	

Team PSFs Sample Categorized Statistic Analysis

By making analysis of 51 team errors cases in voluntary report samples, the analysis result is listed in table 5 by sequence of occurring times.

Table 5. *Categorization and Ratio of Team PSFs in Voluntary Report Samples.*

Team PSFs	Times of Occurrring	Ratio(PSFs/Sample Number)
Deficiency in Communication	24	47.1%
Excessive Belief	15	29.4%
Organization Factors	11	21.6%
Deficiency in Resource/Task Management	9	17.6%
Over-trusting	8	15.7%
Excessive Professional Courtesy	4	7.8%
Inappropriate Authority Gradient	3	5.9%
Air of Confidence	3	5.9%
Friendship	2	3.9%

Note. Since maybe more than 1 team PSF is counted in one report, the sum of PSP ratio doesn't equal to 100%.

Combine with the content of sample reports, the statistic analysis result is explained as follow:

- **Deficiency in Communication.** It is the most common PSF in team errors. The apparently identified behaviors include: fail to make mutual communications among team members when disagreements exist on ATC clearances; information is not timely relayed; unwilling to prove suspicions; take actions for granted due to over trust or over self-confident. Excludes these, fail to communicate timely or incomplete communication are also factors bring in threats. Fail to communicate timely means hesitate to make communication; hence the best time to correct errors is delayed. Incomplete communication means information is not completely delivered, so intention is not clearly defined, or designate target in information transfer is obscure, therefore brings in misunderstandings.
- **Excessive Belief.** Excessive belief is always related to one presumption or precondition. For example, in one radar display error (the flight route shown on radar display offset from correct position), to correct the offset, the controller give instruction to a flight who is actually on right route to fly offset more than 100km, the suspicion raised by crew is ignored.
- **Organization factors.** Organization factors are related with on-site management and working mechanism or procedures in ATC operation center, such as unreasonable shifting schedule. In one report, three new controllers who just recently obtain ATC license are put on same ATC sector. The team could not treat with a special event due to lack of experiences. Such organization factors can also include operation procedures, staff training, ATC control environment, etc. Currently the ATC controller's training provided in China is

still concentrate on individual skill training on radar simulators. This training mode actually intensifies controller's intention to make individual decision during work. How to indicate errors? And how to advise when principle controller refuse to take? These are all count on effective teamwork training subjects.

- **Lack of resource/task management**. It is important to apply effective coordination procedures or assisting tools, such as using consistent record or identification methods to help indicate necessary activity information. For example, after runway lights were shut down in one airport, the controller fail to make stripe note as procedure required, which caused an incident that controller give clearance to a flight landed on runway in night condition but forgot to turn runway light on.
- **Over-trusting**. It is related with trusts exist inside the team. In one report, the assistant controller generated suspicion to crew's clearance readback, so indicated to principle controller. The principle controller confirmed crew's readback was correct. The assistant controller did not insist to confirm from air crew again. But the fact revealed later that the principle controller actually misunderstood pilot's readback.
- **Excessive Professional Courtesy.** It always causes hesitation in making indication and correction among team members, therefore missed the best time to indicate or correct. It could also behave as a too soft way of making indication or correction, so team members can not make correction timely.
- **Inappropriate Authority Gradient.** In one report, the principle and assistant controller can not reach agreement on how to handle a flight confliction. Principle controller was just recently got his license, so felt hesitate on his own plan. The assistant controller, however feel confident about his plan. So before a mutual agreement was reached, the principle controller adopted assistant controller's plan. When unexpected situation occurred in flight, the principle controller fail to show flexibility in handling urgent situation, he changed mind to follow his original controlling plan, this finally causes failure in maintain minimum separation. Besides lack of communication, excessive authority gradient is one of significant PSFs in this occurrence. In another report, the principle and assistant controller used very long time in discussing before reached mutual agreement, which delayed to issue clearance. Inappropriate authority gradient is an important background in the occurrence.
- **Air of Confidence.** In many cases is caused by excessive self-estimated thoughts among team members. In one report, thunderstorm was reported in the airspace and detour was a necessity. A flight requested to enter adjacent ATC area at position 25 miles west offset from normal hand-over position. The coordinate controller didn't take this seriously, and failed to relay the request to adjacent ATC Center (the coordinate controller was busy in other coordination issues). Only when the flight almost reached the boarder between these two ATC centers, this coordination controller started to make phone calls to inform next area. Unfortunately there was artificial precipitation enhancement by rockets firing operation, and adjacent ATC refused to take over the deviated flight at offset position. The flight had to turn back to its original control area.
- **Friendship**. In one report, the principle controller prefered Plan A which was complicated but can alleviate work load of coordination controller. In the meantime, coordination controller coordinated with adjacent airspace control center according to Plan B, which he believed will be able to alleviate work load of principle controller. However, during coordination, the principle controller was confused and could not decide which plan to follow, so the most appropriate time to act was missed.
-

Conclusion

Team error is an important risk source in Air Traffic Control, a sophisticated social technology system. This research applied TEM model, team error taxonomy as well as case study, reached the following conclusion by adopting statistic analysis on one ATC center's voluntary report system whose reports were received in the year of 2004:

1. Among all voluntary reports, there are 24.31% of identified threats from voluntary reports are generated by other controllers and 31% of reported errors are belonging to team errors. In ATC operation, human being is important operation risk source, which including those team members who are willing to cooperate.
2. Team errors related reports are 14% of all reports. When ATC related human errors occurred, 44.7% of cases (voluntary reports contain ATC errors) failed to detect, indicate and correct errors timely. To improve team's error management capability is still an important area of ATC operation risk control.
3. From the point of team error occuring process, even though errors have been detected and indicated, still 8% of these errors failed to be effectively corrected by the team. As to types of team errors, the dependent shared error which is 12% of all reported team errors will bring in more serious hazards to ATC operational safety.
4. Team error PSFs are ranked according to significances and ratio of occurrence. It is identified that 'deficiency in communication' is the most common inductive factor of team error. 47.1% of all reported team errors have the background of 'deficiency in communication'.

References

International civil aviation Organization (ICAO). (2006). *Safety Mangement Manual* (Doc 9859).

International Civil Aviation Organization (ICAO). (2008). T*hreat and Error Management (TEM) in Air Traffic Control* (Cir 314).

Pape, M .A., Wiegmann, D .A., Shappell.S. (2001). Air traffic control (ATC) related accidents and incidents: a human factors analysis. *Presented at the 11th International Symposium on Aviation Psychology* .Columbus, OH: The Ohio State University.

Reason J. (1990). *Human Errors*. Cambridge University Press.

Sasou K, Reason J. (1999). Team errors: definition and taxonomy. *Reliability Engineering and System Safety*. 65, 1-9.

Acknowledgements

This paper was funded by the National Natural Science Foundation of China (NSFC) under Contract No. 60776805. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSFC.

REVIEW OF SAFETY REPORTS INVOLVING ELECTRONIC FLIGHT BAGS

Divya C. Chandra and Andrew Kendra
United States Department of Transportation
John A. Volpe National Transportation Systems Center
Cambridge, MA

Electronic Flight Bags (EFBs) are a relatively new device used by pilots. Even so, 37 safety-related events involving EFBs were identified from the public online Aviation Safety Reporting System (ASRS) database as of June 2008. In addition, two accident reports from the National Transportation Safety Board (NTSB) cite EFB as a contributing factor. Underlying EFB issues were ascribed to each ASRS report by the authors based on subject matter expertise. Pilots reported issues such as configuration of the chart display and difficulty using the EFB when they were newly implemented. Both NTSB reports identified use of an EFB for calculation of landing distance as a contributing factor in the accidents. The NTSB reports identify areas for improvement in the evaluation of EFB software as well as training and procedures. This report provides further recommendations for improving EFB guidance materials to mitigate safety issues.

The Electronic Flight Bag (EFB) industry has grown rapidly since the Federal Aviation Administration (FAA) issued Advisory Circular (AC) 120-76A in March of 2003 (FAA, 2003). A recent review of EFB products shows the diversity of implementations that are being purchased and deployed by all types of operators (Yeh and Chandra, 2007). Their benefits include better access to aircraft operating documents, just-in-time flight performance calculations by the flight crew, and reduction of paper charts and documents in the flight deck. For more information on what functions EFBs support, see Shamo (2000) and Hirschman (2009).

The purpose of this report is to examine what, if any, safety impacts EFBs are having as the industry matures and units are deployed more widely. To accomplish this task, safety-related reports pertaining to EFBs were obtained from the public online Aviation Safety Reporting System (ASRS) database managed by the National Aeronautics and Space Administration (NASA). These reports were analyzed to understand what impact the EFB had in the event. In addition, we review two National Transportation Safety Board (NTSB) reports that cite the EFB as a contributing factor in aircraft accidents.

ASRS reports are useful for identifying human factors areas of interest. However, there are limitations to the data. In particular, these are subjective self-reports that were submitted voluntarily. The reporters are not trained observers and may have difficulty in observing their own situation and performance. Also, the ASRS website states that in many cases, the reports have not been corroborated by the FAA or NTSB and therefore the report database cannot be used to infer the prevalence of a particular problem within the National Airspace System.

In this report, we discuss how the EFB-related reports were identified and then describe the overall set of reports that were obtained. We present descriptive observations about the data, interpret the safety reports to determine what EFB human factors issues were encountered, and summarize the EFB-specific issues from the NTSB reports. We conclude with recommendations for improving guidance to mitigate these issues.

Identifying EFB-Related Safety Reports

Safety reports were collected from the online ASRS database in June 2008 (http://asrs.arc.nasa.gov). The NTSB reports were obtained from the online database (http://www.ntsb.gov). Other databases that may contain EFB-related safety reports, such as those kept by airlines, are not publicly accessible and were not searched.

In order to find EFB-related reports in the ASRS database, a key word search was conducted on the full narrative and synopsis of the report. This task was complicated because there is no standard terminology in use for EFB systems and applications. The following search terms were used to identify the relevant reports: *EFB*, *Onboard Performance Computer, Tablet PC, Tablet, Paperless, Electronic Chart, Laptop,* and *APLC* (an abbreviation used at one airline for “Airport Performance Laptop Computer”). In order to locate the more recent accident report on the NTSB website, the term “Onboard Performance Computer” (OPC) was used. The resulting accident report, from 2007, references an older report from 2000 in which the EFB is referred to as the APLC.

Spurious reports were often returned from the ASRS search and these were manually removed from the search results. In some cases the unrelated reports were easily identified, such as references to passenger laptops or medicinal tablets, but other reports were reviewed carefully to determine whether the EFB was actually a factor in the situation. Cases where the EFB was used normally during incidents that were set in motion by other factors were not considered relevant. For example, if a report mentioned that the first officer was using the EFB, then stowed it to listen to air traffic control communications, then no problem with the EFB was documented per se, and the case was dropped from the set. Other cases were excluded for a variety of reasons. For example, in one case, a Part 91 operator's laptop-based moving map display failed to function at a critical time. This case was eventually discarded because Part 91 moving map displays are not addressed under AC 120-76A (FAA, 2003) and because this type of map display is similar to hand-held or installed GPS displays, which are generally not classified as EFBs.

Analysis

Some information was copied directly from the ASRS report into a spreadsheet for analysis, but other information was constructed by the authors based on their subject matter expertise. For example, the authors classified the outcome or anomaly that occurred and judged whether the EFB was a primary or contributing cause for the outcome or anomaly. Table 1 shows what information was copied, extracted, or interpreted about each event.

Table 1. *Information extracted or constructed for each relevant safety report.*

Information Copied Directly From ASRS Report	Information Extracted or Interpreted from ASRS Report	Interpretations of the Event Constructed by Authors
• Case Number • Year • Operating Regulation (e.g., Part 91) • Operator Type (e.g., Corporate) • Synopsis • Callback Interview • Flight Conditions (e.g., visual or instrument) • Light (e.g., nighttime, daytime) • Other Environmental Conditions • Aircraft Make/Model	• Relevant airport (e.g., origin, destination) • Phase of flight (e.g., arrival, climb) • EFB application in use (e.g., electronic charts) • Outcome/Anomaly (e.g., altitude deviation) • Interesting quotes • Search term(s) used to find the report • Description of EFBs in use (e.g., how many, what type)	• Summarized the EFB issue • Categorized the EFB issue • Determined whether the EFB was a *primary* or *contributing* factor to the outcome/anomaly.

In order to identify the outcome or anomaly, the authors tried to determine why the reporter considered the event as being serious enough to warrant filing an ASRS report. In general, the answer to this question was the "outcome." The outcomes and anomalies found in the ASRS reports are typically an actual violation or a "near violation" (i.e., a violation that almost occurred) of a requirement such as an altitude clearance, or published heading for a departure or arrival procedure. Filing a voluntary ASRS report grants the reporter a level of immunity for the violation as detailed in Advisory Circular (AC) 00-46D (FAA, 1997). There was only one such outcome/anomaly in each of the reports.

In order to distinguish between primary and contributing factors for the event, we reviewed the reporter's narrative carefully to determine the order of events and the self-reported actions and difficulties. The primary factor was the one without which the event was likely not to have occurred at all. Contributing factors tended to complicate or exacerbate the situation. In some cases, the narrative clearly identifies what the reporter considered to be the primary factor in the event or there was only one factor in the event (e.g., an expired database on the EFB). In most cases, however, there was more than one factor and the authors attempted to prioritize the factors. The use of these terms, "primary" and "contributing" factors, is consistent with language used in NTSB reports.

In addition, the description and classification of the EFB issue encountered was based on the authors' judgment. This was the most subjective part of the analysis. In some cases there was enough information to judge what the issue was with regard to the EFB (e.g., the reporter mentioned that he/she was unable to read the screen in bright sunlight). In other cases there was not enough information to identify the exact problem. For example, if there were difficulties accessing information, it may have been because there was a software bug, or because the user training was insufficient, or because the EFB design was problematic. Here the EFB issue was classified more generally, to acknowledge that the underlying issue may not be well understood.

Another subjective aspect of the analysis was in determining the list of relevant EFB issues. The final list presented here was constructed iteratively; where there were enough similar cases, the issue was called out on its own, but if the events were relatively unique, they were placed in a "Miscellaneous EFB Operation" category.

Results and Discussion

Thirty-seven relevant reports were identified from the ASRS database. Descriptive statistics about these events are presented first for the information copied directly from the ASRS reports. Next is a discussion of the incident interpretations and the EFB issues that were encountered. A brief description of the two accident reports from the NTSB involving EFBs is provided after a discussion of ASRS events.

EFB-Related ASRS Events

Relatively few EFB-related ASRS reports were filed each year from 1995 through 2005 (zero to five events each year) but many more reports, 13 of the 37, were filed in 2006. The increased number of reports in 2006 may reflect the fact that EFBs were implemented more widely in the recent past. All types of operators are represented in the reports, but air carriers were involved in a relatively low proportion of the reports, just nine of the 37. More recent reports are largely from corporate operators. This may reflect the fact that while a small number of air carriers have been using EFBs for many years recent purchases have largely been from corporate operators. Weather and ambient lighting conditions do not appear to play a part in EFB-related safety reports. Most of the events (33) occurred in visual flight conditions. Reports were filed for both day (24) and night (8) conditions.

Eighteen of the 37 events occurred on departure. Twelve of these occurred during initial climb out, an especially busy time in the flight. Three events that occurred preflight (on the ground) were related to errors in computing weight and balance or flight performance and in two events the problem was an expired database.

Many pilots reported problems using the charts during the approach phase. The charting application was in use for 28 of the 37 reports. Note that the more recent reports tend to be related to the charts application, while older reports tend to be related to the flight performance calculations function. This also reflects the market trend that corporate operators, who purchased EFBs more recently, use them primarily for the charting function, while air carrier users, who purchased EFBs some time ago, use them primarily for flight calculations.

Four of the reports were filed for events that occurred while flying the same location and procedure, specifically, the Teterboro, New Jersey (TEB) departure procedure known as the TEB 5 departure. This procedure provides separation between departures from Teterboro and arrivals into Newark International Airport, which serves the New York and New Jersey metropolitan area. The TEB 5 departure is a complex procedure that imposes a high level of workload regardless of whether an EFB is used or not because there is little margin for pilot error (see NASA, 2007 and FAA, 2008b). We cannot determine whether the use of an EFB is an *additional* risk factor under these high workload conditions without more information than is available in the ASRS report.

Table 2 below provides a summary of the outcomes, and information about whether the EFB was a primary or contributing factor in the event. The most common outcome or anomaly was a deviation in track, heading, or speed; these occurred in 22 of the reports. A runway incursion occurred in four reports. However, the EFB was only a contributing factor, not the primary factor that caused the runway incursion in these four events. Other outcomes included incorrect weight and balance computations in three reports, use of expired databases, altitude confusion, deviation from a company policy, an aborted takeoff, and a tail strike upon rotation.

Table 2. *Outcomes and whether the EFB was a primary or contributing factor.*

Outcome	Total	EFB Primary Factor	EFB Contributing Factor
Spatial Deviation	22	12	10
Runway Incursion	4	—	4
Incorrect weight and balance computation	3	2	1
Expired database	2	2	—
Altitude confusion without violation	2	—	2
Deviation from company policy	1	—	1
Aborted takeoff	1	1	—
Incorrect take-off speed, tail strike on rotation	1	1	—
Altitude deviation during declared emergency	1	—	1
Total	**37**	**18**	**19**

The EFB was judged by the authors to be the primary cause for outcome in roughly half of the cases (18), and a contributing cause in the other half (19). Other factors for the outcomes in these reports included time pressure, fatigue, problems with the Flight Management Computer, and last minute changes to the aircraft clearance. Sometimes when the EFB was found to be a primary cause for the outcome, these other factors were also present as contributing causes. When the EFB was determined to be a contributing factor for the outcome, one of these other factors was typically the primary factor.

The EFB issues encountered in these reports are summarized in Table 3 below, along with examples and a list of related sections from Chandra, Yeh, Riley, and Mangold (2003), which is a primary resource document about EFB human factors considerations for the FAA and industry. Although Chandra et al. (2003) contains material that is related to the issues seen in the ASRS reports, that report does not necessarily address the specific situations that occurred. The purpose of the reference to Chandra et al. in Table 3 is to identify sections that could be updated by incorporating the issues encountered in the ASRS reports as examples. In particular, some of the EFB-related incidents cut across issues in Chandra et al. and the links between these topics could be illustrated more clearly. Note that the number of issues reported in Table 3 is greater than the total number of reports because more than one EFB issue was encountered in some of the reports.

Table 3. *EFB issues encountered, examples, and related references.*

Issue	Description	Cases	Related Section(s) from Chandra, et al. (2003)
Display Configuration	Related to zooming and panning to configure the display for readability. Information may be missed because it is out of view, or workload may be increased because of the task of configuring the display.	14	2.1.1 Workload 6.2.5 Zooming and Panning 6.2.11 De-cluttering and Display Configuration
New to EFB	The EFB is new to the crew.	10	2.1.1 Workload 2.3.3 Documentation for Part 91 operators
Miscellaneous EFB Operation Issues	EFB is difficult to use for a variety of reasons (e.g., stowed away, sluggish response in cold environment, big/heavy for the flight deck).	7	2.2.2 Stowage 2.2.3 Use of Unsecured EFB Systems 2.5.3 Display
Screen Legibility	Screen is hard to use under different lighting conditions.	5	2.1.5 Lighting-Legibility
EFB Inoperative	EFB or application is not available for use (e.g., EFB in sleep mode or rebooting).	3	2.4.5 Multitasking 2.4.9 Display of System Status
Chart Selection	Difficulty in selecting the required chart at the appropriate time (e.g., due to distraction, or turbulence).	3	6.2.6 Chart procedures 6.2.9 Access to Individual Charts
Software bug	Failure of the software to operate as expected.	3	No applicable section.
Flight Deck Procedures	Related to crew procedures for using the EFB(s) (e.g., sharing/cross-checking information).	2	2.3.1 Part 121, Part 125, and Part 135 Operations EFB Policy
Database Expired	Issue in maintenance or crew verification of database currency.	2	2.4.15 Ensuring Integrity of EFB Data 2.4.16 Updating EFB Data
Separated Information	Difficulty of accessing related information	2	2.4.18 Links to Related Material
Data entry	Difficulty with data entry function.	2	5.1.1 Default Values 5.1.2 Data-entry Screening and Error Messages

The EFB issue encountered most often in this set of ASRS reports was related to zooming and display configuration of electronic charts. In order to read detailed information on the chart the pilot has to zoom in, but in several events this resulted in the pilot missing important information that was off the screen. If the display is not zoomed in, small text can be misread. In addition, the display configuration tasks of zooming and scrolling create workload, and this workload contributed to pilot errors in some cases. One interesting comment from many of the reports was that the pilots would have preferred to have paper printouts of the charts for use during approaches and departures. Paper printouts at these times may be especially useful because hand-held EFBs must be stowed for safety during landing and takeoff (FAA, 2003).

The second most common EFB issue was that, in ten of the reports, pilots indicated that the EFB was a new device for them. This appeared to be a factor in the level of workload and pilot performance. Interestingly, of these ten reports, one was from a Part 121 carrier (from 1999) and one was from a chartered Part 121 flight (in 2007). All of the others were Part 91 or 135, either corporate, private, or charter flights. AC 120-76A (FAA, 2003) requires Part 121 operators to be trained on EFBs, but it does not apply to Part 91 operators. Requirements for Part 91 operators are more lenient (FAA, 2007) and they probably receive less training on EFBs than the Part 121 crews.

NTSB Accident Reports Involving EFBs

On July 31, 1997, a Federal Express (FedEx) MD-11 aircraft crashed while landing late at night in visual conditions at Newark International Airport in Newark, New Jersey (NTSB, 2000). Two crew members and three passengers escaped with minor injuries, but the aircraft was a total loss valued at $112 million. On December 8, 2005, Southwest Airlines (SWA) flight arriving from Baltimore ran off the departure end of runway 31C at Chicago Midway International Airport in Chicago, Illinois at nighttime in instrument meteorological conditions (NTSB, 2007). The Boeing 737-700 aircraft rolled through two fences and onto an adjacent roadway where it struck an automobile before coming to a stop. A child in the automobile was killed, and there were injured passengers both in the automobile and airplane.

The EFB was a contributing, not a primary, factor in both these events. Both accidents involved use of the EFB to calculate landing distance. The EFBs had been in use for some time at both SWA and FedEx and the accident crews were experienced with their use and related procedures.

In the FedEx accident, the NTSB found that the crew misinterpreted landing distances provided by the EFB such that they developed an unnecessary sense of urgency to touch down early and initiate maximum braking immediately. If the crew had correctly interpreted the EFB data, they would have known that there was actually an additional 900-ft stopping margin in the calculation. In response to NTSB recommendations from this accident, the FAA issued Flight Standards Information Bulletin for Air Transportation 02-03, which has since been updated to the InFO Safety Bulletin 0831 (FAA, 2008a). The bulletin reminds inspectors to review and ensure adequacy of training and procedures regarding use of EFB and interpretation of the data generated, including landing distance data.

In the SWA accident, the programming and design of the Onboard Performance Computer (OPC) was a factor. The OPC did not show two inherent assumptions that were critical to the pilot's decision to land. First, the pilots assumed that landing distance calculations were based on the value they entered for the tailwind component (8-knot), but the software actually showed landing distance based on a 5-knot limit for poor runway conditions allowed by company policy. The software highlighted the entered (8-knot) tailwind component on the display without indicating that the stopping margin was not based on that entry. Second, the OPC calculations incorporated the use of reverse thrust for this model of aircraft, but not for two other models that the pilots flew interchangeably. The pilots of the accident aircraft believed that the stopping margins they were shown were conservative because they thought that the reverse thrust was not entered into the calculations, but in fact, there was no such margin. The airline's guidance to pilots on these differences has since been clarified.

The NTSB report on the SWA accident correctly points out that guidance in Chandra et al. (2003) states only that the output of the performance calculations should be displayed in a manner that is understood easily and accurately, and that users of the EFB should be aware of an assumptions upon which the flight performance calculations are based (Section 5.1.6 Use of Performance Calculation Output, Chandra et al, 2003). The NTSB report provides specific suggestions for expanding these recommendations to ensure that critical assumptions are presented as clearly on the EFB as they are on paper (NTSB, 2007, pp. 48-49).

Summary and Conclusions

In this review, 37 incident reports related to use of EFBs were identified from the online ASRS database. Two NTSB accident reports involving EFBs were also identified. Descriptive statistics were computed for the ASRS events, and the authors reviewed the events in order to understand the EFB issues that were encountered. The most common EFB issue encountered in the ASRS events was related to display configuration for using electronic charts. Another important issue appears to be related to the introduction of the EFB to new users. The two NTSB reports cite the EFB as a contributing factor in accidents where landing distance calculations were a factor, even though crews were experienced with the EFB. The reports emphasize the need for proper design of the flight performance calculation software for EFBs, and proper review of crew training and procedures for the use of the EFB.

The results of this research can be used by regulatory authorities such as the FAA to update human factors guidance for evaluating and approving EFBs (e.g., Chandra, et al. 2003). In addition, these results can be used by EFB operators to anticipate issues that need special consideration. EFB manufacturers and designers may also find this report informative.

References

Chandra, D. C., Yeh, M., Riley, V., & S.J. Mangold (2003). Human factors considerations in the design and evaluation of Electronic Flight Bags (EFBs), Version 2. (DOT-VNTSC-FAA-03-07.) USDOT Volpe Center: Cambridge, MA. Available online at http://www.volpe.dot.gov.

Federal Aviation Administration, Advisory Circular (AC) 00-46D, February 26, 1997. Aviation Safety Reporting Program.

Federal Aviation Administration, Advisory Circular (AC) 120-76A, March 17, 2003. Guidelines for the certification, airworthiness, and operational approval of electronic flight bag computing devices.

Federal Aviation Administration, Advisory Circular (AC) 91-78, July 20, 2007. Use of Class 1 or Class 2 Electronic Flight Bag (EFB).

Federal Aviation Administration (FAA) Information for Operators (InFO) 08031 on Weight and Balance Control Methods: Auxiliary Performance Computer (16 May 2008).

Federal Aviation Administration (FAA) Information for Operators (InFO) 08036 Teterboro, New Jersey, TEB-5 SID (3 June 2008).

Hirschman, D. (2009, February). Avionics: Electronic Flight Bags. *AOPA Pilot*, 52(2), 95-98.

National Aeronautics and Space Administration (NASA) Aviation Safety Reporting System Callback newsletter, No. 334, October 2007. High density airspace and traffic mix.

National Transportation Safety Board (2007). Runway Overrun and Collision, Southwest Airlines Flight 1248, Boeing 737-7H4, N471WN, Chicago Midway International Airport, Chicago, Illinois, December 8, 2005. Aircraft Accident Report NTSB/AAR-07/06, PB2007-910407).Washington, DC: Author.

National Transportation Safety Board (2000). Crash During Landing, Federal Express, Inc. McDonnell Douglas MD-11, N611FE Newark International Airport Newark, New Jersey, July 31, 1997. Aircraft Accident Report NTSB/AAR-00/02. Washington, DC: Author.

Shamo, M. (2000). What is an electronic flight bag, and what is it doing in my cockpit? *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics.* Toulouse, France: Cépaduès-Éditions.

Yeh, M. and D.C. Chandra. (2007). Electronic Flight Bag (EFB): 2007 Industry Review. (DOT-VNTSC-FAA-07-04). USDOT Volpe Center: Cambridge, MA. Available online at www.volpe.dot.gov.

Acknowledgements

This paper was prepared by the Behavioral Safety Research and Development Division at the Volpe National Transportation Systems Center. It is based on a more comprehensive technical report in preparation. This research was completed with funding from the FAA Human Factors Research and Engineering Group (AJP-61) in support of the Aircraft Certification Service Avionics Branch (AIR-130) and the Technical Programs and Continued Airworthiness Branch (AIR-120). We thank the FAA program manager Tom McCloy and FAA technical sponsor Colleen Donovan for their assistance. Thanks also to Michelle Yeh and Stephen Popkin of the Volpe Center for feedback on this report. The views expressed herein are those of the authors and do not necessarily reflect the views of the Volpe National Transportation Systems Center, the Research and Innovative Technology Administration, or the United States Department of Transportation.

NT-SEEV: A model of attention capture and noticing on the Flight Deck

Dr. Christopher Wickens, Angelia Sebok, Julie Bzostek
Alion Science and Technology
Boulder, CO
Kelly Steelman-Allen, Jason McCarley,
University of Illinois Human Factors Division, Champaign, IL
Nadine Sarter
University of Michigan Dept of IOR. Ann Arbor, MI

N-SEEV is a model that predicts the noticeability of events that occur in the context of routine task-driven scanning across large scale visual environments such as the flight deck or ATC work station. The model is an extension of the SEEV (salience, effort, expectancy, value) model, incorporating the influence of attentional set and allowing the possibility of a dynamic environment. The model was validated against two empirical data sets. In a study of pilot scanning across a high fidelity automated 747 cockpit, the SEEV component of the model predicted the distribution of attention with a correlation of 0.85. In a lower fidelity study of pilot noticing of the onset of critical cockpit events (flight mode annunciators) the model predicted differences in noticing time and accuracy with correlations (across conditions) above 0.95. Other properties of the model are described

Failures of pilots or controllers to notice critical events, such as low altitude warnings, have been responsible for serious accidents with numerous fatalities (Wiener, 1977; NTSB, 2006). To be effective, visual and aural alerts must capture attention (SAE- ARP, 2007). Guidelines for the design of alerts typically emphasize the importance of display characteristics such as flashing, color, brightness, etc. But even a signal that is highly salient, however, can often go unnoticed, as illustrated by the phenomena of change blindness (Simons & Levin, 1997; Rensink, 2002; Stelzer & Wickens, 2006: Martens, 2007) and inattentional blindness (Mack & Rock, 1998;), effects which demonstrate the **failure** of events (e.g., changes) in the environment to capture attention. We can thus characterize human performance, as reflected in **noticing time (NT)** and the **probability of noticing** (or its inverse, **miss rate**), as falling along a continuum between rapid attentional capture and inattentional blindness. Effective cockpit design must be supported by models that can predict the noticeability of visual alerts, and therefore a assure that noticeability increases with the importance of the alert.

Such noticeability is governed by two classes of factors: bottom up and top-down. Bottom up factors are defined primarily by event **salience**. For example repeated flashes are more salient than single onsets, and these more salient than single offsets. Foveal events are more salient than peripheral ones. Top down factors can be subdivided into those related to expectancy and value. In addition, the effectiveness of both bottom up and top-down factors is modulated by workload, with high workload diminishing noticeability vis attentional tunneling (Wickens & Alexander, 2009). While all of these factors have been clearly identified to effect noticeability in isolation, there have been few studies examining their influence in combination, and there appears to be no valid computational model, that can predict their interaction, in a manner that might be useful for flight deck certification. The purpose of the research we describe here is to validate a computational model of noticing, N-SEEV, against two sets of existing data, and to demonstrate how the model can be applied to certification of visual warnings in safety-critical environments such as the cockpit or the ATC work station. We note that the model described and validated here has been subsequently used to **predict** the detection rate of off-nominal or unexpected events, in highly realistic flight simulations, and as projected to occur in NextGen technology and procedures, as described in detail in Gore et al (2009; see also Hooey et al, 2009)

Method: The Model

The N-SEEV model, described in more detail in Steelman-Allen et al (submitted) has two components, N and SEEV. The SEEV component describes the steady state allocation of visual attention (e.g., scan path) across any large scale environment such as the cockpit (Wickens McCarley, et al., 2008; Wickens & McCarley, 2008). Attention is directed to **<u>S</u>alient** events and locations, is inhibited in its movement by the **<u>E</u>ffort** needed to undertake long scans or head movements, and is directed to areas where there is **<u>E</u>xpected** to be information (e.g., high bandwidth areas), particularly if that information is **Valuable.** The first letter of each of the above terms, defines SEEV. As examples, a red flashing alert will have a high salience. The effort required for a pilot to scan an overhead panel will be greater than scanning a HUD indicator. The expectancy of change on the ADI in turbulence will be higher than the change of a heading indicator, or, in particular, a fuel gauge. The value of sampling displays that support aviating (e.g., angle of attack, pitch, speed above stall speed) is greater than those that support communication (e.g, radio frequencies).

The four elements of SEEV are additive, although in the current application, Salience and Effort are lumped into a single parameter. The model drives gaze around a simulated environment in a probabilistic (i.e., Monte-Carlo) fashion, such that the probability of the scan moving from one location to the next, is proportional to its relative "attentional attractiveness" compared to all other areas, as this attractiveness is defined by SEEV. This model of steady state scanning then predicts a distribution of locations of gaze at the moment when the to-be-noticed-event (TBNE) occurs, and therefore, it predicts distribution of the degree of *retinal eccentricity* of the TBNE. Such eccentricity is shown to be a powerful modulator of noticeablity. Our meta-analytic review of the small amount of existing human factors literature on the topic revealed the eccentricity effect on miss rate to be approximately 8% per 10 degrees, although this eccentricity cost is greatly modulated by clutter and expectancy. The N component of N-SEEV then predicts the time to notice the TBNE as a function of eccentricity, expectancy & value of the TBNE, and its salience, as the latter is quantified by a model derived from Itti & Koch, (2000).

Results: Model Validation

A set of data from three experiments was first employed to fit the parameters of N-SEEV, and then the model was validated against the scanning and noticing time data from two experiments. In the parameter fitting phase, we examined a set of scanning data from single-pilot general aviation simulations in which self-separation responsibilities were supported by a cockpit display of traffic information (CDTI) (Wickens, et al., 2002); and also augmented by a data link display (Helleberg & Wickens, 2003; Wickens, Goh, et al., 2003). Using these data we assured that the current version of SEEV, a Monte-Carlo simulation, predicted pilot scanning with the same precision as the analytic (equation) version of the SEEV model used in those three studies. Our results indicated a favorable fit, with the current Monte-Carlo model predicting the distribution of fixations across areas of interest in the three studies, with correlations of 0.93, 0.96, and 0.94.

Following parameter fitting, the first formal model validation was carried out against the pilot scanning data collected by Sarter, Mumaw & Wickens (2007), describing the scan data of 21 commercial aircraft pilots as they flew realistic missions using the flight management system in a high fidelity Boeing 747 simulator, across phases of take-off, departure, cruise, descent and final approach.. We applied the model to the cockpit image shown in figure 1[1], and populated the different areas of interest with estimates of bandwidth (frequency of change, driving expectancy) and value (importance of the task served by the AOI), that would characteristize the automated cockpit. Effort was inherent in the display layout, with greater effort characterizing areas that were farther apart. Other than adjusting AOI parameters for expectancy (bandwidth) and value, the other model parameters were set to the same values as were established in the model fitting exercise described above from the GA cockpit

[1] This cockpit model was created under NASA NRA# NNX08AE87A; See Gore et al., 2009

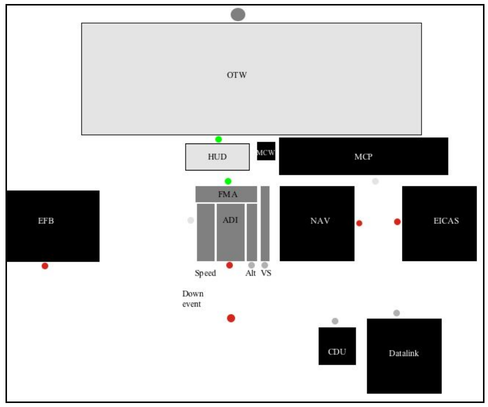


Figure 1: Cockpit image employed for automated cockpit validation model runs. Note that the printed labels on the image were not present during model runs. The dots in the figure represent different possible locations of a TBNE.

Figures 2a and 2b depict the scatter plot of model-predicted versus simulation-obtained percentage dwell time (attentional interest) across the different active areas of interest for the four flight phases of takeoff, cruise, descent and final approach whose point shapes are defined in the legend. The different areas of interest within a phase are not identified within the figures. Figure 2a includes the local areas within and around the primary flight display (altimeter, vertical speed, ADI, airspeed, nav display). Figure 2b includes larger areas more globally, including those specific to automation (primary flight display, navigation display, control display unit, mode control panel and outside world). Both figures reflect reasonably strong correlations between predicted and obtained attentional interest of r= 0.85 (local) and r= 0.88 (global). While both correlations are less than those observed in the model fitting exercise described above, we note that this is the first time that the SEEV model has been applied to the automated cockpit, so we had no firm basis on which to estimate expectancy and value parameters for the different FMS-specific AOIs.

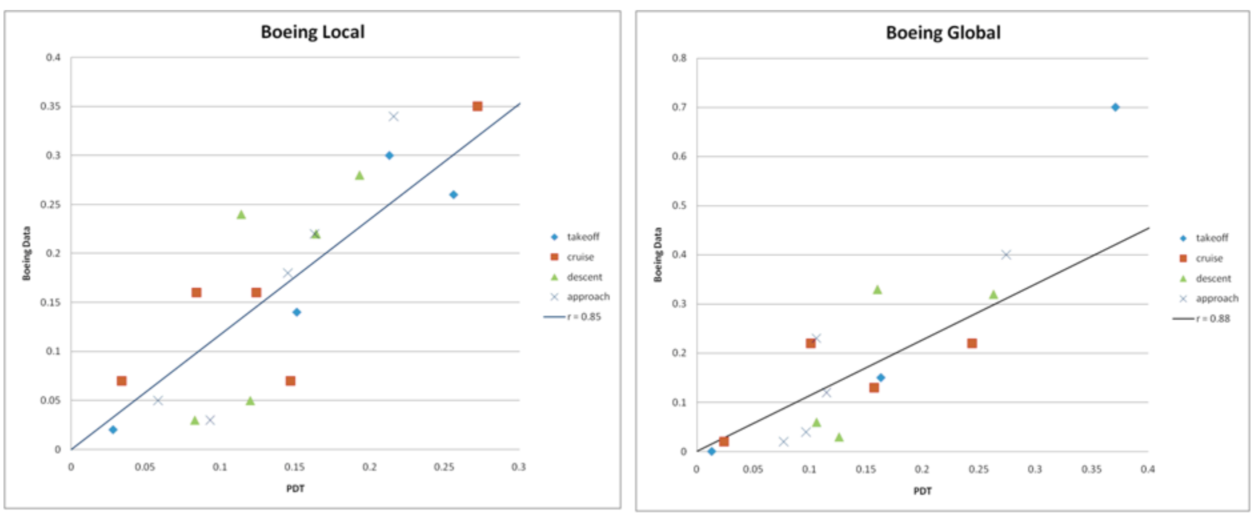


Figure 2a (left): Results of model validation runs from automated cockpit. Predicted scan (X axis) versus obtained scan (Y axis) for **local** areas of interest (r=0.85). All points represented by a common symbol type are measured in a single flight phase. Different points within this symbol type represent different areas of interest. Figure 2b (right): Results of model validation runs from automated cockpit. Predicted scan (X axis) versus obtained scan (Y axis) for **global** areas of interest (r=0.88).

Our second validation study employed the data collected in a simulation of flight management annunciator (FMA) noticing data carried out by Nikolic, Orr and Sarter (2004). The investigators examined the properties of the

FMA that led pilots often to fail to notice the onset of a green box, surrounding the FMA, that indicated an important change in FMS flight mode (i.e., in the way that the automation was flying the aircraft).

This simulation was less realistic than that employed in the Boeing cockpit study, but offered the unique attributes that noticing time and miss rate for events whose salience was varied were measured across two different levels of eccentricity. Our model simulation of the experimental paradigm used by Nikolic et al is shown in figure 3. Participants in the original experiment were asked to play a highly demanding game of Tetris (whose display is shown on the left side of the image), requiring the heavy focus of attention, corresponding to that imposed by high visual engagement with the primary flight display. In parallel, they were to monitor the right side of the image and to report any onset of a green box. Figure 3 depicts an example of the image that we provided to the model for validation. Heavy participant attention is focused on the Tetris box to the left (bandwidth = 0.8, value = 0.8 on a scale of 0- 1.0). The two green boxes between the circles to the right are the locations where a green box onset occurred in the near (35 degrees eccentricity from the Tetris game) and far (45 degrees) condition. Because they were of value to be noticed, they were assigned a value parameter of 0.2. The remaining AOI's represent "clutter displays" which, when present, were green, and rendered dynamic, (to mimic the rotation of dials in the original experiment) by providing them with 0.20 bandwidth settings. In a non-clutter condition, these other instruments were colored pale yellow and had near 0 bandwidth. All clutter AOI's had the value parameter set to 0.

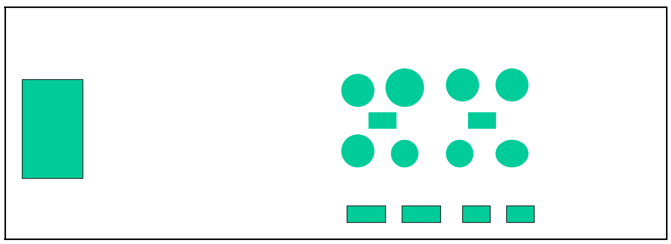


Figure 3: Cockpit image employed to validate noticing predictions, from Nikolic, et al., 2004.

We exercised the model in the near and far conditions, crossed with clutter vs no-clutter conditions. Each model run generates a single output of the number of scans to notice; thus across repeated model runs, a distribution of scans-until-noticing is created. These distributions, not unlike real data, tend to be positively skewed. We made two further assumptions: (1) that participants make approximately two scans per second, so that the conversion of scans-to-notice to noticing time is division by two. (2) Not all event onsets trigger a detection. We made a plausible assumption that since the green box remained illuminated for 10 seconds, any model run for which the TBNE fixation was not achieved within 10 seconds (20 fixations) was considered a "miss". Hence we were able to compute a miss rate. Figure 4a presents the scatter plot of model-predicted versus experimenter-obtained RT (for those responses less than 10 seconds), and figure 4b presents equivalent data for miss rate. Four conditions are represented in each figure: low and high clutter, crossed with near and far display locations.(Note that only that only three points are visible in each graph since, for each, there is a case of two data points directly overlaying each other). Both sets of data show a very high (r > 0.97) correlation, and both also show the approximate values of predicted RT and miss rate, to be within the same range as those of the obtained data.

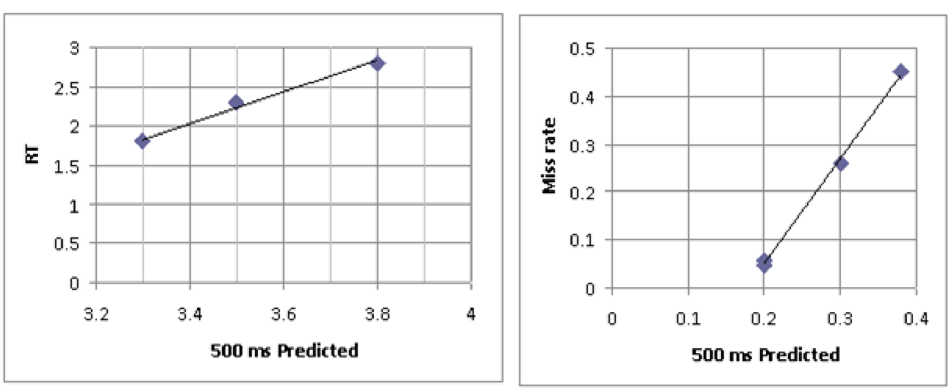


Figure 4a: Noticing predictions and observations across four conditions from Nikolic, et al. (2004) for noticing (or response) time; Figure 4b: Noticing predictions and observations across four conditions from Nikolic, et al. (2004) for miss rate, assuming all runs with fixation time on the green box greater than 10 seconds were misses.

Discussion and Conclusion

The current version of the N-SEEV model appears to do a competent job of integrating both bottom-up (salience and effort) and top-down (expectancy and value) parameters to predict either or both noticing time and event detection rate in a complex visual environment. The model was effectively validated on two sets of experimental data, one predicting scanning, and the other predicting noticing time as inferred from scanning.

In conclusion, we also note the added capabilities of the N-SEEV model, not validated here against empirical data:

- The effects of visual workload can be simulated by increasing the bandwidth at an AOI in question. For example the difference between low and high turbulence can be simulated by increasing the bandwidth at the ADI from 0.2 to 0.8.
- The effects of cognitive workload can be simulated by narrowing the field of view within the model, to mimic the well-known effects of cognitive load on attentional tunneling.
- The effects of attentional set (e.g,. greater for a red than an amber or white alert) are achieved by adjusting a color-tuning parameter within the model.
- The effects of human skepticism in false-alert-prone can be rendered by lowering the expectancy of the alert – essentially mimicking a lower expectancy for a **valid** alert (Dixon & Wickens, 2005). These and other features offer promise for a valid computational model of this vitally important cognitive process.
- The effect of surprise (low expectancy) can be simulated by setting bandwith of the TBNE to 0. As such this mimics the response to off-nominal events. Gore et al (2009) describes how model-predictions of off-nominal events are validated against flight simulation data obtained from a meta-analysis (see also Hooey et al, 2009).

Finally, there is a potential to input the scanning output of the current model into further models of situation awareness (Wickens McCarley et al, 2008; Wickens Sebok et al 2008) and, indeed, into overall pilot performance models (Gore et al, 2009).

Acknowledgements

This research was supported by grant # NASA Research Opportunities in Aeronautics (ROA) 2007, NNX07AV97A, awarded to the University of Illinois. The authors wish to acknowledge Dr. Jeffrey Mulligan of NASA Ames for his support of this work.

References

Helleberg, J. & Wickens, C. D. (2003). Effects of data link modality and display redundancy on pilot performance: An attentional perspective. *The International Journal of Aviation Psychology*, 13(3), 189-210.

Gore, B. F., Hooey, B. L., Wickens, C. D., & Scott-Nash, S. (2009). A Computational Implementation of a Human Attention Guiding Mechanism in MIDAS v5. To appear in *Proceedings of the 13th International Conference on Human-Computer Interaction*.

Gore, B. F., Hooey, B. L., Salud, E., Wickens, C. D., Sebok, A., Hutchins, S., Koenecke, C., & Bzostek, J. (2009). *Identification Of NextGen Air Traffic Control and Pilot Performance Parameters for Human Performance Model Development in the Transitional Airspace.* (Final Report: NRA # NNX08AE87A). San Jose, CA: San Jose State University Research Foundation.

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 of Aviation Psychology. Wright State U: Dayton, OH

Itti, L. & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489-1506.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

Martens, M (2007). The failure to act upon important information. PhD dissertation. Free University, Netherlands.

Nikolic, M. I., Orr, J. M., & Sarter, N. B. (2004). Why Pilots Miss the Green Box: How Display Context Undermines Attention Capture. *The International Journal of Aviation Psychology*, *14*(1), 39–52.

Rensink, R.A. (2002). Change detection. *Annual Review of Psychology, 53*, 245-277.

SAE-ARP (2007). Society of Automobile Engineering Aerospace Recommended Practice ARP4102/4. Flight Deck Alerting Systems. Warrendale Pa:. SAE International.

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.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Science, 1*(7), 261-267

Steelman-Allen, K, McCarley, J, and Wickens, C. (submitted). A salience model of event noticing. Submitted to 2009 Annual meeting of the Human Factors & Ergonomics Society.

Steltzer E. M. & Wickens, C. D. (2006). Pilots strategically compensate for display enlargements in surveillance and flight control tasks*. Human Factors.48*(1), 166-181.

Wickens, C.D & Alexander, A.L. (2009) Attentional tunneling and task management. *International Journal of Aviation Psychology, 19(2), 1-17.*

Wickens, C. D., & McCarley, J. M. (2008). *Applied Attention Theory.* Boca Raton, FL: CRC 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., Helleberg, J., & Xu, X. (2002). Pilot maneuver choice and workload in free flight. *Human Factors, 44*(2), 171-188.

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

Wickens Sebok et al (2007) Modeling advanced flight deck displays. *Human Factors and Ergonomics Society Annual Meeting Proceedings*. Santa Monica, CA: HFES.

Wickens, C. D., Helleberg, J., & Xu, X. (2002). Pilot maneuver choice and workload in free flight. *Human Factors, 44*(2), 171-188.

CONTRIBUTION OF HIGH-FREQUENCY EEG FEATURES TO PHYSIOGICALLY-BASED OPERATOR WORKLOAD ESTIMATION

Jason W. Monnin
Air Force Research Laboratory
Wright-Patterson Air Force Base, OH
Justin R. Estepp
Air Force Research Laboratory
Wright-Patterson Air Force Base, OH

Physiologically-based cognitive workload assessment can discriminate changing levels of operator functional state in complex task environments. In this paradigm, electroencephalography (EEG) is a commonly used physiological measure. Spectral power in clinical frequency bands is used to derive features to train an artificial neural network (ANN) classifier to recognize changes in cognitive workload. Recent research has suggested that power in high frequency bands may be influenced by electromyographic artifact. In a previous study, nineteen channels of EEG were recorded (from 10 participants) during a complex uninhabited air vehicle (UAV) control simulation in which task difficulty was manipulated to induce changes in cognitive demand. In offline analysis, an ANN classifier was trained using feature sets which included and excluded features from high frequency bands. Excluding the high frequency bands reduced classification accuracy, suggesting that, while potentially of electromyographic origin, these features are still important features in physiology-based cognitive workload assessment.

Current physiologically-based operator workload estimation methods have demonstrated very high classification accuracy. Physiological data used for workload estimation typically includes electroencephalogram (EEG), in addition to electrical eye and heart activity (Wilson and Russell, 2007). However, in several recent works, it has been suggested that scalp EEG is largely contaminated by electrical musical activity from facial muscle groups (Goncharova et al., 2003; Whitham et al., 2007; Yong et al, 2008). While the exact nature of this artifact, with respect to the frequency range of contamination, is somewhat disputed (Yong et al., 2008), it is widely accepted that high-frequency components of EEG above approximately 20 Hz, especially in lateral electrode sites (furthest from the midline, are the most heavily influenced by muscle activity. A further compound of this potential problem is that, while the detection of muscle artifact through electromyogram (EMG) signals can be reliable and robust (for one example of many, see van de Velde et al., 1998), the ability to remove muscle artifact from the EEG signal is not trivial, given that both EMG and EEG share similar statistical properties and overlapping frequency content (Djuwari et al, 2005). As discovered by Fatourechi et al. (2007) in a meta-analysis of brain-computer interface (BCI) literature, the presence of muscle artifact in EEG is largely ignored by researchers.

While the artifact correction problem remains difficult, it is still necessary to be able to quantify the effect of muscle contamination in EEG-based systems. The term "artifact," when taken in the strictest sense, would tend to suggest the presence of an undesirable signal, leading to poor signal-to-noise ratio (SNR) in an available dataset. The question remains, however, whether muscle contamination of EEG is truly detrimental in any given application, or whether it could perhaps be of some benefit (given the specific nature of the application in question). The purpose of this work is to understand what effect high frequency EEG components, presumably primarily influenced by muscle artifact, have on the ability to accurately assess cognitive workload in a dynamic simulation environment. Based on the analysis presented here, the removal of high-frequency EEG components negatively impacts the accuracy of being able to discriminate between varying cognitive tasks. These results suggest that, while the source of muscle artifact may very well contaminate underlying EEG components, these high-frequency features of the EEG signal are useful in cognitive workload assessment. In addition, this work tends to agree with others who have shown increased EMG activity in a variety of muscle groups during periods of high cognitive demand (Whitham et al, 2008) and stress (Lundberg et al, 1994).

Methods

Unmanned Aerial Vehicle (UAV) Simulation Task

Ten participants (7 male, 3 female) ranging from 20 to 24 years of age (mean of 22.6, standard deviation of 4.3), after providing signed, informed consent, were trained as operators in an Unmanned Aerial Vehicle (UAV) simulation task. In this task participants were asked to simultaneously monitor multiple UAVs during four unique task conditions. Since the nature of the work presented here can be sufficiently described by considering a 2-class pattern recognition problem, data from only 2 of the 4 task conditions are presented in this analysis. Both of these task conditions consisted of 4 UAVs during a target identification and weapons pairing task. In order to increase the cognitive demand on the operator between the first and second task conditions, the air speed of the UAVs was increased in the difficult task condition, as previously used by Wilson & Russell (2007) in a similar test-bed environment. To guard against possible task learning effects during data collection trials, each participant was trained to asymptotic performance in the simulation. Asymptotic performance was defined by a minimum performance threshold of 90% target placement.

After asymptotic performance was reached, the participants returned on a separate day for a series of trials with physiological data collection. While a variety of physiological measures were recorded, for the purpose of the analysis, only EEG data are reported. EEG was recorded from 19 electrodes placed on the scalp in accordance to the International 10-20 System for Electrode Placement. The data was processed via a MicroAmps (SAM Technology, Inc.; San Francisco, CA) amplifier/filter and software package with a sampling frequency of 256 Hz, bandpass filtered from 0.5 Hz to 100 Hz. Vertical and horizontal eye artifact were corrected using an embedded linear-regression (offline) technique, where vertical electrooculogram (VEOG) and horizontal electrooculogram (HEOG) were used as representative artifact signals in the regression. Post-hoc from data collection, the data from 2 of the 10 participants were excluded from data analysis due to poor data quality.

Artificial Neural Network Training and Feature Reduction

An artificial neural network (ANN) classifier (Wilson and Russell, 2007) was trained using a feature set composed of data solely from the electroencephalogram (EEG). The EEG data was post-processed into frequency band power (via a windowed FFT of 10 seconds, with an overlap of 9 seconds) in the five traditional frequency bands (delta, 1-3 Hz; theta 4-7 Hz; alpha 8-12 Hz; beta 13-31 Hz; and gamma, 31-43 Hz) for each of the 19 electrode sites, yielding a feature set of 95 features. In order to assess the effects of the high-frequency bands on classification accuracy, the ANN was trained and tested using three combinations of frequency bands included in the feature set. The first feature set contained all 95 features, the second dataset contained the lower three frequency bands (delta, theta, and alpha, for a total of 57 features), and the third dataset included only the higher two frequency bands (beta and gamma, for a total of 18 features). A total of two-thirds of the available data was used for training the ANN (of which 25% was used as a separate validation set, independent of the training set), and the remaining one-third of the data was used as the test set. A top-down feature reduction technique was implemented using ranked saliency (calculated using a partial derivative saliency technique, as described in Greene et al. (2000)). The lowest-ranked feature, according to its calculated saliency, was removed between iterations of ANN training. The smallest feature set that yielded the highest classification accuracy on the test data was considered to be the optimally-trained network.

In order to avoid over-training and over-generalization to the provided training set, each network was verified to ensure the accuracy on the validation set (independent of the training set) was above 95%. For each network where classification accuracy was reported (the optimally-trained network), feature count totals (how many times a particular feature was used in the optimally-trained network) were compiled for each 10-20 electrode location (for all features included in a particular training set) in order to see the effects of electrode location in each training set paradigm.

Results

Mean classification accuracy (across participants) for each of the feature sets is shown in Figure 1. Using all 95 features (all bands) yielded a classification accuracy of 90.96%, which is similar to other accuracies reported in physiology-based cognitive workload studies (see Wilson and Russell (2007) for a brief overview). In the feature set containing only the three lower frequency bands (delta, theta and alpha), the classification accuracy was 76.32%. Using a feature set of the highest two frequency bands (beta and gamma), classification accuracy was

87.79%. In each feature set, the mean accuracies reported are well above chance (in a two-class pattern recognition problem chance would be 50%, like a coin-flip to determine the correct class, which indicates that the EEG features used in the training sets are sensitive to changes in cognitive workload between the two tasks).

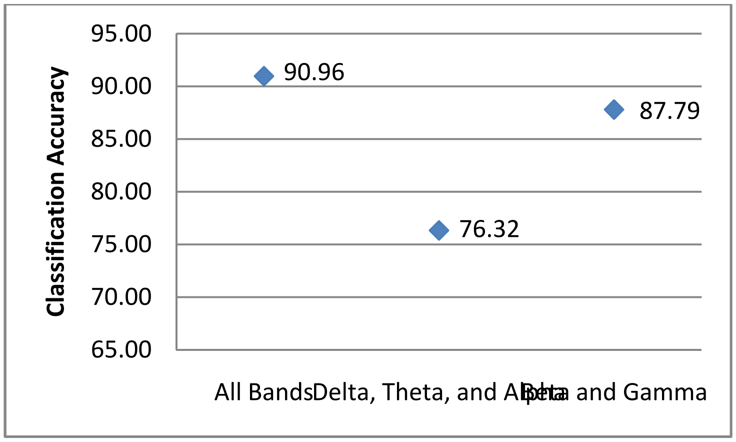


Figure 1. Mean classification accuracies (across all participants) for the three feature sets. Note the significant decrease between using all frequency bands and only using the lowest three frequency bands, as well as the increase in accuracy between using delta, theta and alpha bands verses beta and gamma.

Feature count (collapsed across 10-20 electrode sites) is shown in Figure 2. Especially in the training set where only the beta and gamma (high frequency band) features were used to train the ANN, lateral sites show up as containing the most often used features (in the optimally-trained networks) in this case, with the highest percentage originating from electrode sites: O_1, O_2 and T_4. In contrast to the all frequency band condition, the percentage of features from the medial electrode sites showed an increase in the feature set containing the delta, theta and alpha bands. In the training set containing the beta and gamma (high frequency) bands, the feature count distribution appears to be very similar to that of the all bands feature set. Lateral sites are used more often than the central electrode sites; the only centralized site with a percentage similar to the lateral sites is C_3.

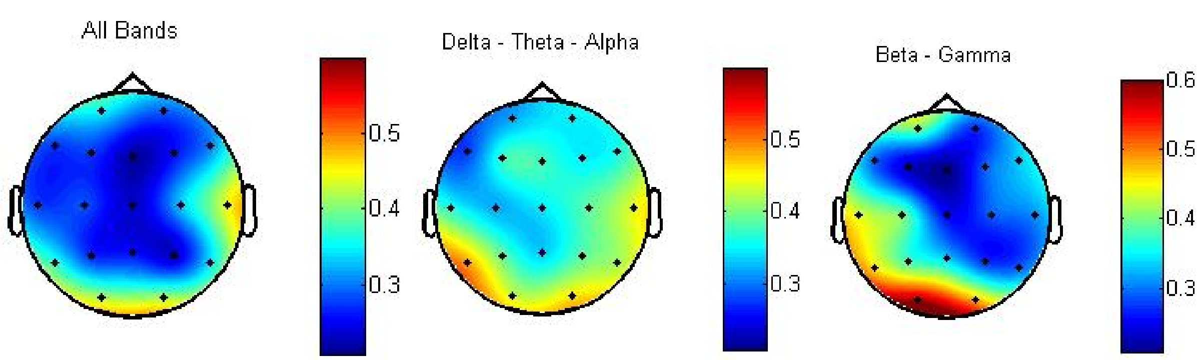


Figure 2. Feature topography using All Bands (95 features), Delta – Theta – Alpha bands (57 features), and Beta – Gamma bands (18 features), mapped to a 10-20 electrode map. Each 10-20 location shows the percentage of features from that location that were used in the optimally-trained network.

Discussion

The results obtained in the analysis agree very well with previous findings; namely, that high-frequency features in lateral electrode sites are often some of the more important features in cognitive workload classification (Wilson and Fisher, 1995). Training the ANN with the beta and gamma bands yielded higher classification accuracy over the delta, theta and alpha feature set; the difference was slightly over 11%, which demonstrates the dependency of classification accuracy on high-frequency features. However, the all bands condition (the full 95 features set) yielded the highest classification accuracy of all three feature sets. This suggests that, although higher classification is primarily driven by the high-frequency band features, the lower frequency bands still contain useful information needed to obtain the best possible classification accuracy.

The two feature sets that obtained the highest classification accuracy (the all bands condition, and the beta and gamma bands condition) also showed the highest feature selection from the lateral electrode sites on the scalp. In contrast, the delta, theta and alpha bands feature set shows that features from medial electrode sites were selected more often than either case where the high-frequency features were included in the training set. Yet, classification accuracy drops when the high-frequency features are not included. This suggests features originating from lateral sites are more likely to increase the overall classification of the ANN.

Given that the higher-frequency features are most useful in obtaining the optimally-trained network, these features are also most susceptible to muscle artifact. In addition, lateral sites are most likely to be contaminated with muscle as well, due to their location on the scalp. Since EEG above 20 Hz is largely contaminated by muscle, it suggests the all bands feature set and the beta-gamma bands feature set are contaminated with muscle as well.

Based on these results, it would appear that, specifically for the purpose of cognitive workload detection, the high-frequency features derived from EEG data are highly salient in correctly assessing levels of workload. While it is largely assumed that these features are highly driven by facial muscle artifact (especially when taken into consideration that the most likely candidate features from the high frequency bands were also from lateral electrode sites), the EEG data itself was not corrected or assessed for the presence of EMG artifact. This, in itself, suggests that EMG may be affected in a predictable manner during periods of high cognitive demand or stress, as previously mentioned in Whitham et al (2008) and Lundberg et al (1994), respectively. Therefore, this “artifact” may actually serve a useful purpose in cognitive workload classification.

References

Djuwari, D., Kumar, D.K., and Palaniswami, M. (2005). Limitations of ICA for artifact removal. *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference* (pp. 4685-4688). Shanghai, China: IEEE Engineering in Medicine and Biology Society.

Fatourechi, M., Bashashati, A., Ward, R.K. and Birch, G.E. (2007). EMG and EOG artifacts in brain computer interface systems: a survey. *Clinical Neurophysiology*, 118, 480-494.

Goncharova, I.I., McFarland, D.J., Vaughan, T.M., and Wolfpaw, J.R. (2003). EMG contamination of EEG: spectral and topographical characteristics. *Clinical Neurophysiology*, 114, 1580-1593.

Greene, K.A., Bauer, K.W., Wilson, G.F., Russell, C.A., Rogers, S.K. and Kabrisky, M. (2000). Selection of psychophysiological features for classifying air traffic controller workload in neural networks. *Smart Engineering System Design*, 2, 315-330.

Lundberg, U., Kadefors, R., Melin, B., Palmerud, G., Hassmen, P., Engstrom, M. and Dohns, I.E. (1994). Psychophysiological stress and EMG activity of the trapezius muscle. *International Journal of Behavioral Medicine*, 1 (4), 354-370.

van de Velde, M., van Erp, G., and Cluitmans, P.J.M. (1998). Detection of muscle artifact in the normal human awake EEG. *Electroencephalography and Clinical Neurophysiology*, 107, 149-158.

Whitham, E.M., Lewis, T., Pope, K.J., Fitzgibbon, S.P., Clark, C.R., Loveless, S., DeLosAngeles, D., Wallace, A.K., Broberg, M., and Willoughby, J.O. (2008). Thinking activates EMG in scalp electral recordings. *Clinical Neurophysiology*, 119, 116-1175.

Whitham, E.M., Pope, K.J., Fitzgibbon, S.P., Lewis, T., Clark, C.R., Loveless, S., et al. (2007). Scalp electrical recording during paralysis: Quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG. *Clinical Neurophysiology, 118*, 1877-1888.

Wilson, G.F., Fisher, F. (1995). Cognitive task classification based upon topographic EEG data. *Biological Psychology*, 40, 239-250.

Wilson, G.F., Russell, C.A. (2007). Performance Enhancement in an Uninhabited Air Vehicle Task Using Pyschophysiologically Determined Adaptive Aiding. *Human Factors*, 49, 1005-1018.

Yong, X., Ward, R.K. and Birch, G.E. (2008). Facial EMG contamination of EEG signals: characteristics and effects of spatial filtering. *Proceedings of the IEEE ISCCSP 2008* (pp. 729-734). Malta: IEEE ISCCSP.

TEAM WORKLOAD: A CONSTRUCT WORTH PURSUING?

Gregory J. Funke[1], Benjamin A. Knott[2], Scott M. Galster[3], and Rebecca D. Brown[4]
[1]Consortium Research Fellows Program, Alexandria, VA, USA
[2]Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA
[3]Aptima Inc., Fairborn, OH, USA
[4]Ball Aerospace & Technologies Corp., Boulder, CO, USA

To date, little research has been conducted on the workload experienced by teams. Attempts to evaluate team workload have typically relied upon validated measures of individual workload, but this approach may not adequately capture the drivers of team workload. Data from three previous research experiments were reexamined using hierarchical multiple regression in the present analysis. Each of the experiments assessed individual and team workload, though the measures employed differed in each. The goal was to investigate whether addition of team workload measures improved prediction of team performance after variance associated with experimental manipulations and individual workload had been removed. Results indicated that inclusion of team workload measures produced an increase of 5-15% in the variance explained. This suggests that workload associated with team processes is not adequately reflected in measures of individual workload. Researchers investigating workload in team settings are recommended to consider inclusion of a team workload measure in their experiments.

The mental workload experienced by an individual operator performing a task has received a significant degree of scientific inquiry, particularly within the last half century (Tsang & Vidulich, 2006). The result has been a proliferation of theories, methods, and metrics designed to evaluate individual workload. Workload assessment continues to remain a vital area of human performance research, providing understanding and prediction of human performance in complex systems (Parasuraman, Sheridan, & Wickens, 2008).

By contrast, however, research on *team workload* has been virtually overlooked (Bowers & Jentsch, 2005), even though teams have been acknowledged as an integral part of the military, industrial, medical, and public service sectors (Bowers, Braun, & Morgan, 1997). A comprehensive, validated theoretical framework for the construct 'team workload,' which includes a description of its relation to individual workload, has not yet been articulated. While some interesting work has been initiated (e.g., Bowers et al., 1997), progress in the area has been quite slow.

One potential explanation for these circumstances is that attempts to assess team workload have been rooted in measures of individual workload. Most research in the area has been conducted using existing individual workload measures , most frequently the NASA Task Load Index (TLX; Hart & Staveland, 1988), which have been modified by alteration of the instruction set, item set, or both to make them more applicable to teams (e.g., Bowers & Jentsch, 2005). This approach, however, is not without its difficulties.

First, consideration must be given to how the data from a team of individuals should be combined and interpreted (Bowers et al., 1997). Several different methods have been advanced, including the averaged workload of all team members, the lowest workload value obtained, and the highest workload value obtained (Bowers & Jentsch, 2005), though it should be noted that there are no theoretically-compelling reasons to adopt any of these. Second, it is perhaps unreasonable to presume that the psychometric properties of these measures, determined at the level of the individual, will be unchanged by transitioning to a team level of analysis. Lastly, several researchers have reported finding no differences in participant workload ratings using a modified team-TLX in response to manipulations of task demands (e.g., Bowers, Urban, & Morgan, 1992; Thornton, Braun, Bowers, & Morgan, 1992; Uban, Bowers, Monday, & Morgan, 1995), leading Bowers, Braun, and Morgan (1997) to suggest that individual workload measures may not be sensitive enough to the sources of team demands to adequately capture team workload.

Given the complexities of measurement and the slow progress of research in the area, a reasonable conclusion may be to question the utility of a team workload construct. If team workload is difficult to assess and provides little additional explanatory power beyond what is already provided by measures of individual workload, then perhaps it is defensible to focus research efforts elsewhere. Conversely, if individual measures of workload are insensitive to team demands, then measures of team workload may provide a more appropriate estimate of workload, and further research in the area is warranted.

The goal of the current experiment was to examine the additional explanatory power measures of team workload may afford beyond that provided by individual measures. To explore this issue, data-sets from several previous experiments were assembled for further analysis. These data were tested using hierarchical multiple regression to determine the degree of additional variance explained by team workload, beyond that accounted for by individual workload, as it was regressed onto a measure of team performance. It was initially hypothesized that the inclusion of a team workload factor in the analysis would significantly increase the variance explained by the regression model.

Method

Data sets from three previous experiments were assembled for further analysis. These experiments featured several communalities. First, each included a complex task which required participants to work together as a team in order to meet scenario goals. The simulated environment utilized in each experiment was Aptima, Inc.'s Distributed Dynamic Decision-making (DDD) simulation (MacMillan, Entin, Hess, & Paley, 2004). DDD provides a scriptable, low-to-moderate fidelity, team-in-the-loop simulated environment. DDD has successfully been used to simulate team command and control tasks and to study realistic and complex team processes in a variety of military and civilian research projects (MacMillan et al., 2004).

In each experiment, task feedback was provided to participants following a trial in the form of a 'team score' which reflected how well the team had achieved the goals of the scenario. This score was scaled so it could range from 0-100; a score of 0 indicated that the team had not met any of the goals of the scenario, and a score of 100 indicated that the team had met all goals perfectly. Team scores were presented to participants after all team members had completed the individual and team workload measures.

Secondly, across experiments, participants completed the same measure of individual workload, the NASA-TLX (Hart & Staveland, 1988), a standard measure of workload that is widely used in human performance research (Wickens & Hollands, 2000). The NASA-TLX provides a global index of task workload on a scale of 0 to 100 and identifies the relative contributions of six sources of workload: mental demand, temporal demand, physical demand, performance, effort, and frustration. In all experiments, participants completed the TLX immediately following a trial.

Thirdly, each experiment included a different, measure of team workload. It should be noted that the focus of these experiments was not team workload and that the included measures were of peripheral interest to the research questions investigated in each. Participants completed a team workload measure immediately following each trial.

A brief description of the experimental task and the team workload measure employed in each data set are included below.

Data Set 1: Schwartz (2008)

Schwartz (2008) examined the relationship between team decision-making structure (centralized, decentralized) and learning (trials) on team performance in a scenario designed to simulate an air strike against hostile forces. Participants completed the experiment in teams of five. Each of the ten participants completed the TLX and team workload measure 23 times in each (decision-making) condition.

Team workload measure. Utilizing a process similar to that proposed by Dickinson and McIntyre (1997), Schwartz (2008) assessed team workload using the Consensus-TLX (C-TLX), a consensus-scored version of the NASA-TLX. Team members worked collaboratively to assign a score reflecting the team's overall perceived workload on each of the six dimension of the NASA-TLX. Participants were told to discuss each trial in conjunction with the six rating scales and to assign a single value to each scale that best reflected the team's workload. Scales of the C-TLX were scored in the same fashion as those of the NASA-TLX (i.e., 0-100).

Data Set 2: Funke, Bennett, Nelson, & Galster (2007)

Funke et al. (2007) investigated the effects of UCAV-control (direct, supervisory) and team-collaboration (standard, whiteboard) conditions on team performance in a simulated suppression of enemy air defense (SEAD) mission. Participants were assigned to one of six team positions; positions differed from each other in their role in the scenario and in the simulated capabilities of their assets. Each of the six participants completed the TLX and team workload measure 12 times in each condition.

Team workload measure. Funke et al. (2007) assessed team workload using the Modified-TLX (M-TLX; Pharmer, Cropper, McKneely, & Williams, 2004). The M-TLX combines the standard six rating scales of the NASA-TLX with an additional five items; the supplementary items addressed sources of team workload and included communication demand, monitoring demand, control demand, coordination demand, and leadership demand. The team workload items of the M-TLX are scored in the same fashion as the NASA-TLX (i.e., item ratings may range from 0 to 100) and a global team workload score is calculated by computing the mean of those five items.

Data Set 3: Funke, Russell, Knott, & Miller (2009)

Funke et al. (2009) examined the impact of task demand (high, low) and collaborative tool availability (with and without access to a resource display, and with and without access to collaborative tools) on team performance in a simulated air defense task. Participants completed the experiment in teams of five. Each of the 105 participants completed the TLX and team workload measure twice in each experimental condition.

Team workload measure. Funke et al. (2009) utilized the Team Workload Assessment Scale (TWAS; Galster & Knott, 2007), a new, unvalidated measure specifically created to assess team workload. The TWAS provides a global index of team workload ranging from 0 to 100 and identifies the relative contributions of ten sources of team workload: temporal demand, physical demand, mental demand, task engagement, coordination demand, task difficulty, control demand, communication demand, team focus, and environmental interference.

Results

Correlations

Table 1displays the mean team performance score, mean individual (NASA-TLX) and team workload ratings, and the correlations between them for each data set. Examination of the correlations suggests that, across data sets, individual and team workload measures were moderately correlated with team performance such that increases in workload were associated with decrements in team performance. The observed correlations also indicate that individual and team workload ratings were quite similar, and raises concerns about the relative contributions of each to predictions of team performance.

One potential explanation is that the measures of team workload included in this analysis do not provide additional information beyond what is provided by measures of individual workload (i.e., individual and team measures of workload are essentially equivalent, and capture the same variance). Alternatively, individual and team workload may strongly covary because comparable processes moderate ratings of each. To address these

Table 1. *Means and correlations for each data set.*

					Correlations		
Source	Team Workload Measure	Team Score (Mean)	NASA-TLX Workload (Mean)	Team Workload (Mean)	Team Score & NASA-TLX	Team Score & Team Workload	NASA-TLX & Team Workload
Schwartz (2008)	C-TLX	49.67	53.87	58.81	.00	-.23*	.83*
Funke et al. (2007)	M-TLX	70.73	49.99	58.35	-.43*	-.61*	.67*
Funke et al. (2009)	TWAS	91.39	48.66	31.32	-.32*	-.35*	.83*

Note. NASA-TLX = NASA-Task Load Index (Hart & Staveland, 1988); C-TLX = Consensus-TLX (Schwartz, 2008); M-TLX = Modified-TLX (Pharmer, Cropper, McKneely, & Williams, 2004); TWAS = Team Workload Assessment Scale (Galster & Knott, 2007).
* $p < .05$.

possibilities, separate hierarchical multiple regression analyses were conducted for each of the three data sets included in this experiment.

Hierarchical Multiple Regression

The goal of these analyses was to examine the incremental increases in variance accounted for by the regression models after variance associated with experimental manipulations and individual workload had been removed. Analyses were conducted in four steps; team score served as the criterion in each analysis. For these regressions, effect-coded vectors for the team, trial, and experimental conditions of each data set were created using the method recommended by Pedhazur (1997). Interaction terms were constructed as product vectors of those task variables. A linear global individual workload term was computed from the mean workload ratings of all team members across the six subscales of the NASA-TLX. Similarly, a linear global team workload term was constructed as the mean workload ratings of all team members across the subscales of the team workload measure employed in each data set (i.e., the C-TLX, M-TLX, and TWAS).

In each regression analysis, the first step entered consisted of the effect coded vectors of the team, trial, task variables, and task variable interactions. These variables were entered first to initially partition variance associated with the experimental manipulations in each data set. The second and third steps entered were the linear individual workload and team workload terms, respectively. This order of entry provided a test of the increase in variance accounted for by the team workload measure after variance associated with individual workload had been removed. The final step in each regression was the entry of all subscales of the NASA-TLX and the relevant team workload measure. The purpose of the final step was to further clarify the potential drivers of individual and team workload in each sample.

Summaries of the results for each regression equation are displayed in Table 2[†]. All subsequently reported regression coefficients (β's) are standardized. Regression coefficients related to experimental manipulations (i.e., those related to team, trial, and experimental conditions) are not reported here as they generally coincide with previously published results. Interested readers are directed to those publications for further information.

Data Set 1: Schwartz (2008). The results of the hierarchical multiple regression analysis indicated that inclusion of the individual and team workload vectors (steps 2 and 3) did not significantly increase the variance explained by the model. However, addition of the NASA-TLX and the C-TLX workload subscales (step 4) substantially increased the variance accounted for. Statistically significant predictors in the final model included the C-TLX performance subscale ($\beta = -.538$, $p < .05$) and a trend toward the C-TLX effort subscale ($\beta = -.333$, $p < .10$). In this experiment, C-TLX subscale scores were more predictive of team scores than were global individual and team workload

Table 2. *Summary statistics for hierarchical multiple regression analyses of task variables (team, trial, experimental conditions), individual workload, team workload, and workload subscales onto team score.*

		Schwartz (2008)				*Funke et al.(2007)*				*Funke et al. (2009)*			
Step	Predictors	R^2	ΔR^2	*df* for Δ	*F* for Δ	R^2	ΔR^2	*df* for Δ	*F* for Δ	R^2	ΔR^2	*df* for Δ	*F* for Δ
1	Task variables	.65	.65	46, 43	1.72*	.58	.58	15, 32	2.95*	.57	.57	42, 293	9.36*
2	Task variables Ind. workload	.67	.02	1, 42	2.58	.59	.01	1, 31	.46	.68	.11	1, 292	98.17*
3	Task variables Ind. workload Team workload	.69	.02	1, 41	2.57	.64	.05	1, 30	4.07*	.70	.02	1, 291	15.90*
4	Task variables Ind. workload Team workload Workload subscales	.85	.16	10, 31	3.23*	.74	.10	9, 21	.99	.77	.07	14, 277	5.70*

Note. * $p < .05$.

[†]Copies of team workload measures and complete summaries of all regression equations may be obtained by contacting the lead author. Email: Gregory.Funke@wpafb.af.mil

estimates. Overall, increases in consensus-scored estimates of team performance and effort were associated with poor team performance.

Data Set 2: Funke et al. (2007). Results of the regression analysis indicated that inclusion of the individual workload vector did not significantly improve the variance explained by the model, but addition of the team workload vector did (albeit the gain was relatively modest). The last step of the regression analysis, which entered the NASA-TLX and M-TLX subscales, did not significantly increase the variance accounted for. The only statistically significant predictor in the final regression model (excluding those associated with task variables) was the global team workload vector ($\beta = -.360$, $p < .05$). As demands on team processes increased, team performance generally decreased.

Data Set 3: Funke et al. (2009). In this data set, each step of the multiple regression contributed significantly to the variance accounted for in the model. Inclusion of the team workload and workload subscale vectors (Steps 3 and 4) resulted in modest increases in the variance explained. Statistically significant predictors in the final regression model included the NASA-TLX mental demand subscale ($\beta = -.402$, $p < .05$), and the TWAS task engagement and environmental interference subscales (β's = -.176 and -.319, respectively, both $p < .05$). Generally, team performance decreased as the mental demands placed on individual team members increased. Team performance was also negatively impacted by team-wide difficulties maintaining task engagement and interference related to the task environment.

Discussion

The purpose of the current research was to explore the potential utility of 'team workload' as an initial stage in the development of more comprehensive theoretical and empirical models of the construct. This was achieved by reanalyzing data from three experiments using hierarchical multiple regression to examine the incremental increases in variance associated with team performance explained in the regression models by the inclusion of measures of individual and team workload. Broadly, the results of the regression analyses support the value of including measures of team workload in team research, though the experimental 'gains,' in terms of variance explained, may be small to moderate.

The results of the multiple regression analyses indicated that, in the experiments sampled, the averaged individual and team workload vectors did not contribute to the prediction of team performance in a uniform fashion. Across the experiments sampled, dissimilar and non-overlapping patterns of statistically significant predictors emerged. This is somewhat unsurprising given the substantial differences in manipulated variables, experimental tasks, and team workload measures originally employed in each data set. It is also worth mentioning again that none of the experiments included in this research were focused on assessment of team workload. Nonetheless, the observed regression results do support the utility of team workload assessment. Across data sets, the global team workload rating (and frequently, the associated subscale ratings) consistently emerged as a significant predictor of team performance.

Measures of Team Workload

The results of the current study do not definitively answer questions concerning the appropriateness of using individual workload measures to assess team workload. The results do support suggestions by Bowers et al. (1997) that measures such as the NASA-TLX may be insufficiently sensitive to sources of team workload. As such, researchers investigating workload in team settings may wish to consider inclusion of a team workload measure in their experiments.

However, it is worth reiterating that a validated measure of team workload has not yet emerged from the research community (Bowers & Jentsch, 2005). The team workload measures sampled in the current study were selected because of their availability, rather than because of their sound psychometric properties. These measures showcase the potential of a validated team workload measure for understanding team performance; they do not provide a means to circumvent the research required for this endeavor. In developing a metric of team workload, the key question researchers must answer is one of construct validity: can 'team workload' be adequately assessed using a more sensitive measure of individual workload, which is then aggregated across team members, or is team workload a new construct which is related to, but distinct from, individual workload and which requires a new measure to address.

Conclusions

The current study reaffirms the need for a validated theory and measure of team workload. Significant research and expertise are still required to resolve the difficult theoretical issues surrounding the construct 'team workload.' In addition, further research concerning the issues of sensitivity and appropriateness of individual workload measures for team workload assessment are certainly warranted.

Acknowledgements

The authors would like to thank Daniel Schwartz for generously allowing the use of his dissertation data in the present study.

References

Bowers, C., Braun, C., & Morgan, B. (1997). Team workload: Its meaning and measurement. In M.T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 85-108). Mahwah, NJ: Erlbaum.

Bowers, C.A., & Jentsch, F. (2005). Team workload. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendrick (Eds.), *Handbook of human factors and ergonomic methods* (pp. 57-1 to 57-3). Boca Raton, FL: CRC Press, LLC.

Bowers, C.A., Urban, J.M., & Morgan, B.B. (1992). *The study of crew coordination and performance in hierarchical team decision making* (Tech. Rep. No. 92-1). Orlando, FL: University of Central Florida, Team Performance Laboratory.

Dickinson, T.L, & McIntyre, R.M. (1997). A conceptual framework for teamwork measurement. In M.T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 19-44). Mahwah, NJ: Erlbaum.

Funke, G.J., Bennett, A.M., Nelson, W.T., Galster, S.M. (2007). Collaboration technologies and the supervisory control of UCAVs in tactical C2: Effects on performance and workload. *Proceedings of the International Symposium on Aviation Psychology, 14*, 216-223.

Galster, S.M., & Knott, B.A. (2007). *The team workload assessment scale (TWAS)*. Unpublished manuscript.

Hart, S.G., & Staveland, L.E. (1988). Development of a multi-dimensional workload scale: Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North-Holland.

MacMillan, J., Entin, E.B., Hess, K.P., & Paley, M.J. (2004). Measuring performance in a scaled world: Lessons learned from the Distributed Dynamic Decisionmaking (DDD) synthetic team task. In S.G. Schiflett, L.R. Elliott, E. Salas, & M.D. Coovert (Eds.), *Scaled worlds: Development, validation, and applications* (pp. 154-180). Brookfield, VT: Ashgate.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*, 140-160.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research*. New York: Harcourt College Publishers.

Pharmer, J., Cropper, K., McKneely, J., & Williams, E. (2004). *Tactical Tomahawk weapon control system V6 land attack combat system prototype human-computer interface: Test report for FY 03 fleet operability test* (Tech. Rep. No. 3184). San Diego, CA: Naval Air Systems Command, Space and Naval Warfare Systems Center.

Funke, G.J., Russell, S.M., Knott, B.A., & Miller, B. (2009). What can a multimodal measure of stress tell us about team collaborative tools? *Proceedings of the International Symposium on Aviation Psychology, 15*, 236-241.

Schwartz, D. (2008). *Exploring team dynamics: The evolution of coordination in a complex command and control environment*. Unpublished doctoral dissertation, Wright State University, Dayton, Ohio.

Thornton, C., Braun, C., Bowers, C., & Morgan, B. (1992). Automation effects in the cockpit: A low-fidelity investigation. *Proceedings of the Human Factors and Ergonomics Society, 36*, 30-34.

Tsang, P., & Vidulich, M. (2006). Workload and situation awareness. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed., pp. 243–268). New York: Wiley.

Urban, J.M., Bowers, C.A., Monday, S.D., & Morgan, B.B., Jr. (1995). Workload, team structure, and communication in effective team performance. *Military Psychology, 7*, 123-135.

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

CONTROL-FORCE INPUTS OBTAINED FROM PILOTS AND NONPILOTS (FLIGHT ATTENDANTS): COMPARISON WITH ESTABLISHED HANDBOOK DISTRIBUTIONS OF PERFORMANCE

Dennis B. Beringer
FAA Civil Aerospace Medical Institute
Oklahoma City, OK

Earlier reports in this series (Beringer, 2006-2008) have reported the force that pilots and nonpilots could exert on flight controls. This paper presents a comparison of some well-known tables of human strength with the values from recent samples of women and men pilots and nonpilots in an attempt to determine how closely those distributions fit tabled distributions of human strength. Findings suggest that some other samples may be used to approximate the difficult-to-sample Part 121 female pilots if the data are treated properly. Specifically, yoke-activation tasks for the female pilots could be reasonably well approximated by fractional performance values of male pilots. It was also determined that some older data obtained from a narrower sample of participants (both in age and gender) were not especially good for estimating present performance of the more diverse population of present-day certified pilots. Percentile values are provided for the lower values of the sampled groups.

Previous publications have reviewed the Code of Federal Regulations [14CFR, Parts 25.143(c) and 23.143(c)] that specify the maximum forces the test pilot can experience in the flight controls during the testing required for certification of an aircraft (Beringer, 2006) and presented some preliminary data comparing the abilities of current pilots and nonpilots with both the values contained in the regulations and with documented abilities of earlier-sampled populations (Beringer, Ball, & Haworth, 2007). The primary emphasis was on determining what proportion of the samples could produce forces at or above those allowed by the CFR in manual control systems (those data will be referenced as Sample 1). A further sample (Sample 2) of pilots and nonpilots from a Part 121 (scheduled commercial carrier) operator was preliminarily reported by Beringer (2008), detailing data for 35 additional individuals. That paper discussed how the application of force on a single manipuland or control was greatly reduced when the operator was required to apply force with both a hand and a foot simultaneously, and presented data regarding the force that could be applied to both rotary and pushbutton seatbelt releases. The purposes of the present article are to (1) trace the sources of often-used anthopometric data, (2) compare data from these sources with some of the recently collected data, and (3) to determine to what extent samples other than Part 121 female pilots can be used to estimate performance of that group, whether they be of male pilots or of nonpilots, given that performance of the female population is likely to be the limiting factor for how many individuals will be able to perform force-exertion tasks at any given level.

Popular Anthropometric Data and Their Sources

Many sources of anthropometric data have been generated in the last 50 years, and a large proportion of those were formed prior to 1980. If we examine the sources, we will find that many date anywhere from 25 to 50 years ago. For example, the tabled values found in CFR Part 23.143(c) and CFR Part 25.143(c) are believed to have been derived from data for 5th to 95th percentile males applying for military service between 1955 and 1957, collected at Wright-Patterson AFB (personal communication, Dr. Richard Jones). This recollection appears to be supported by references in Morgan, Chapanis, Cook, and Lund (1963) (reprinted in Van Cott & Kinkaide [1972]) to the source of the strength data relevant to flight controls reported therein as "Unpublished data, Anthropology Branch, Aerospace Medical Research Laboratories." Additionally, two sets of data summarized in Van Cott and Kinkaide (Table 11-107, Maximal static leg thrust exerted on a fixed pedal by seated males) present data from Rees and Graham (1952) (sample of 20 men) and Rohmert (1966) (sample of 60 men). The summaries in the Van Cott and Kinkaide edition of the handbook were based upon 194 references from the 1950s up through the publication date, thus representing a comprehensive sample of the data collected to that date. Moving forward chronologically, the popular HumanScale 4 (Diffrient, Tilley, & Harman, 1981) dates back 27 years, and the NASA Anthropometric Source Book predates that by 3 years (1978). While it is true that there are some compilations of data with more recent dates, they are just that: compilations of data from earlier studies and assessments.

Although tables of values are, in many cases, simply reprinted/repeated from earlier sources, some have been modified. Ahlstrom and Longo (2003) took the table of arm, hand, and thumb-finger strength (5^{th} percentile male data), Figure 23, from MIL-STD-1472F and reduced all of the contained values to 80% of their original values, presenting it as their Exhibit 14.5.2.1 (page 14-44). The justification, presented on page 14-43 of the document, was: "Since the experimental conditions used to collect the source data yielded maximum possible exertion values for young men, these values are were [*sic*] too high for design purpose. For design, one does not want to deliberately or consistently require maximum exertions. Thus these source values were reduced by 20% before applying them as design criteria." While it is certainly reasonable to expect a downward shift of the distribution of applicable forces with an increased age range (see Stoll et al., 2002, for strength loss as a function of aging), no specific rationale is given for the choice of precisely 20% as the reduction factor. Thus, we have recommended force-application levels for design purposes that are a fractional proportion of earlier tabled values, but without a clear tie of the amount of the reduction to a specific empirically defined reduction factor. While it might be possible that one could take Stoll's data providing profiles of strength loss by age, make an assessment of the distribution of ages in the target population, and then rectify the original data for young men by that factor, this will not be attempted here. Table 1 is partially derived from Beringer (2008; Table 4) and depicts the sample sizes and age ranges for the various groups in each sample that will be compared with extant data.

Table 1. Sample compositions showing group, sample size, median age, and age range. Groups of fewer than 4 individuals have been omitted. Data from Karim are included, as raw data from that study were used to generate percentiles in Figure 2.

			Age		
Sample	**Group**	**n**	**Mean**	**Median**	**Range**
Karim et al. (1972)	Female Part 91 pilots	25	35.4	34	18 to 58
Beringer Sample 1	Male Part 121 pilots	32	49.7	49.5	38 to 58
	Female Part 91 pilots	12	45.7	48.1	21 to 64
	Female nonpilots	12	49.5	50.5	17 to 71
Beringer Sample 2	Female Part 121 pilots	11	40.8	39	32 to 54
	Male Part 121 pilots	6	39.5	38.5	32 to 52
	Female nonpilots (flight attendants)	10	39.9	38.5	24 to 57
	Male nonpilots (flight attendants)	6	32.5	34	22 to 47

COMPARISONS WITH TABLED DATA

Depending upon how much of the population one wishes to accommodate, one may choose one of the lower percentile values from the known tables of human strength. Frequently these tables will present the 5^{th}, 50^{th}, and 95^{th} percentile values. Some, however (i.e., Diffrient et al., 1981), present the 2.5^{th} as the low point, with overlapping distributions for men and women portrayed with two common sets of values (weak men and average women in one; average men and strong women in the other). While it would be possible to thus provide for success of either 97.5% of 95% of the subpopulation from the pilot group (women), the overall success for all pilots would be higher than either of those values due to higher force-application success rates by the men at those same levels. The following figure (Figure 1) provides tabled data for yoke-force input by samples of men pilots and nonpilots. Some sources did not contain

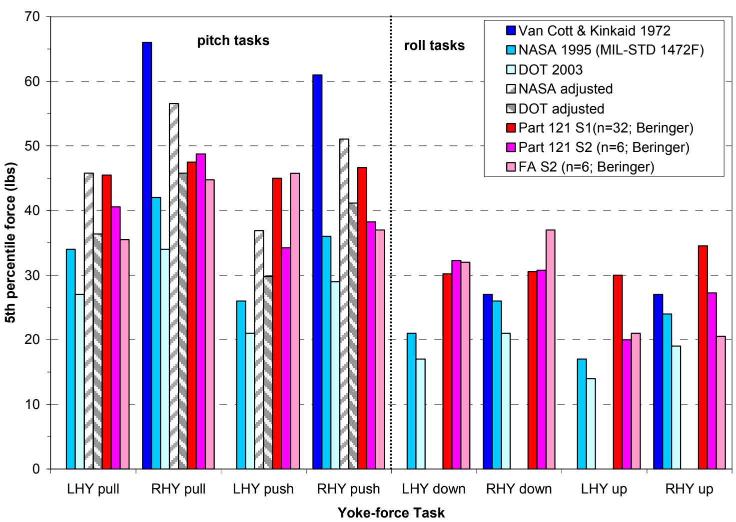


Figure 1. 5^{th} percentile tabled force values and recent samples for men.

the data for yoke-force application, and a derivation factor had to be determined. For example, the Laubach chapter in the NASA source book (1978) presented data for force exertion on a vertical handgrip in various locations near a seated operator, but none of them were a close enough match to the position of the vertical part of the control. However, two of the sources contained data that appeared to be a location match for side-stick data, and the data from Van Cott and Kinkade contained both yoke-push/pull data and side-stick-push/pull data for appropriate elbow angles. Multiplication factors were determined for deriving applied yoke force from applied side-stick force (left-hand push, stick to yoke, 1.3469; left-hand pull, stick to yoke, 1.4186), and the data in Figure 1 labeled as "adjusted" are the resulting values.

Relating data for female pilots to other samples/populations

Given that it is difficult to obtain large samples of women Part 121 pilots due to (1) their comparatively small number relative to men Part 121 pilots and (2) their unavailability due to flight operations being conducted, an attractive alternative would be to use more accessible samples that could somehow be related to the target population. The most obvious choice would be another sample of women with demographics, other than piloting, that were similar. Thus, let us first look at how the 5th percentile values for the four recent samples, pilots and nonpilots, compared amongst themselves and with percentile values derived from the raw data for women pilots in Karim et al. (1972) data (Figure 2). One can see a relative consistency across the samples from the Beringer assessments, with a relatively small but consistent difference, excepting in sample 2 yoke push, in favor of the pilot participants. However, the values from Karim et al. are considerably larger for the yoke push and pull tasks when compared with the other four samples. Aileron (yoke rotation) forces appear to be comparable. One contributing factor for this differential may be that Karim used a hard wooden seat on the test platform that may have allowed participants to use it as a brace more effectively than the padded Cessna seat used in the Beringer assessments. That the flight attendants' force performances (Women FA) were lowest on the hand-up aileron-roll tasks and lower than their own hand-down performances is consistent with other data showing a reduction in the applicable force for hand-up tasks, as compared with hand-down tasks. Otherwise, all recent samples appeared to be relatively close in their 5th percentile levels of yoke-force application.

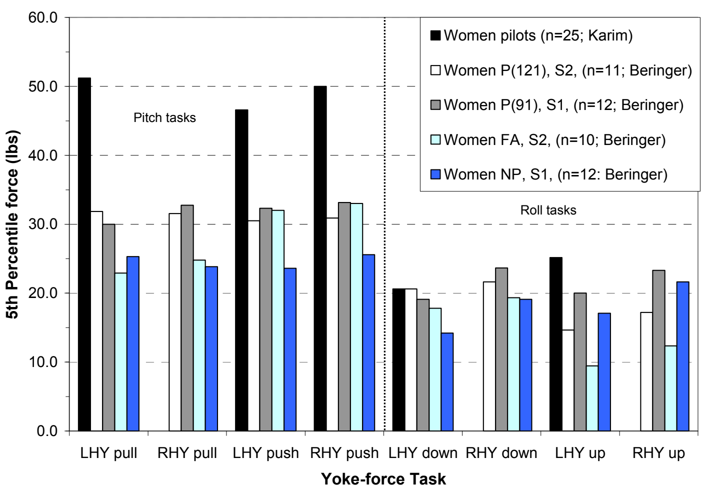


Figure 2. Comparison of recent women pilot samples' calculated percentiles with those derived from Karim et al. (1972).

Humanscale 4 (Diffrient et al., 1981), as mentioned earlier, presents its summarized low-end force data for the 2.5th percentile rather than the 5th percentile and, as such, comparisons with performance data had to be reported separately from the other sources reporting 5th percentile data. Figure 3 depicts the 2.5th percentile Sample 2 data for women and those from Humanscale. The data in Humanscale for the conditions closest to the yoke-manipulation task being investigated consistently underestimate the women's values from the sample by an average of 32.6% for the pitch axis and 35% for the roll axis.

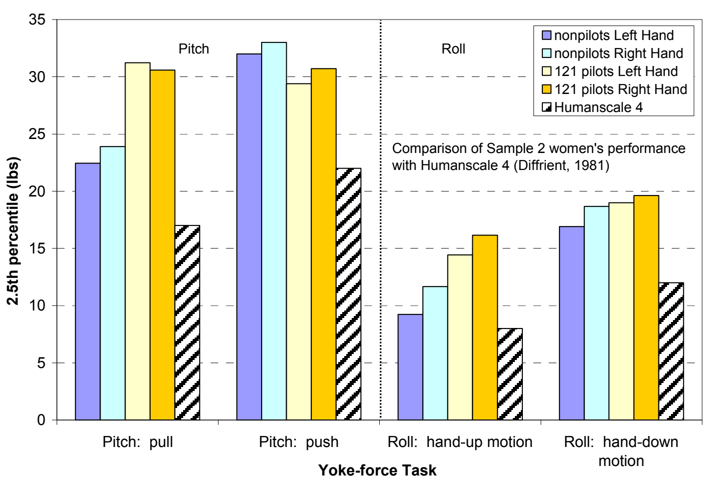


Figure 3. Comparison of women's 2.5th percentile data from Sample 2 with those from Humanscale 4.

Prediction of women's averages from men's

The use of data for males as a baseline can take two forms. First, one can assess the present state of male performance and compare it with tabled values for males, as shown in Figure 1. If the fit were good, then one could use tabled values to represent current men's performance. Alternately, one could just use the new data for males and bias it accordingly. There are existing recommendations as to how to bias the men's data to represent performance expected from the women. For example, Ahlstrom & Longo (2003), in their section 14.5.2.3 (page 14-46) recommend the following reductions of the men's values to apply them to women: "a. For upper extremities, females strength is 56.5% of men. b. For lower extremities, female strength is 64.2% of men. c. For trunk extremities, female strength is 66.0% of men." NASA (1995) presents similar data in that publication's Figure 4.9.3-5, but presents both means and ranges for the differences in total body strength, upper extremities strength, lower extremities strength, trunk strength, and dynamic strength. The three values recommended in Ahlstrom et al. appear to be reproduced directly from the NASA document. Upon closer examination, all of these data appear to have been derived from Laubach (1976; page 85). Other sources have presented the general rule-of-thumb value as 67% (sometimes simply represented as "two-thirds"), undoubtedly derived from Laubach's mean difference for dynamic strength characteristics.

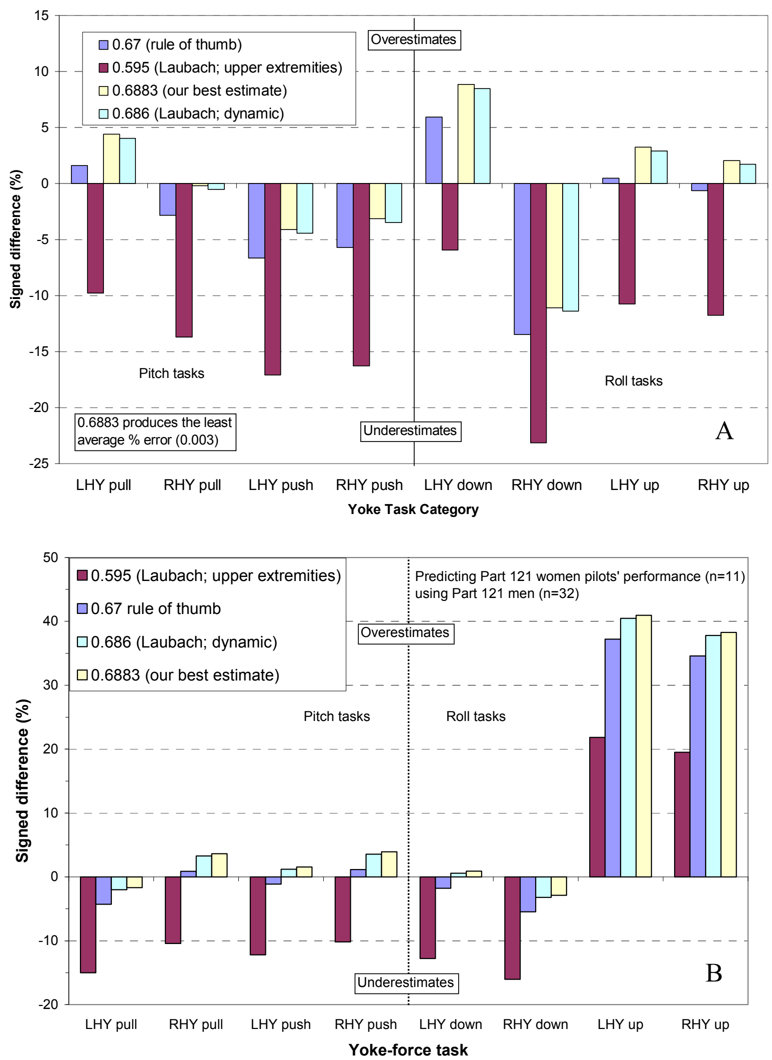


Figure 4. Estimation of Part 91 (A) and Part 121 (B) women pilots' 5th percentile performances as fractional portions of Part 121 male pilots' 5th percentile performances.

The largest sample of male Part 121 pilots (32) was used as the basis for prediction, and the data were restricted to tasks that were common to both major sampling efforts to make comparisons direct and straightforward. Figure 4 depicts the percentage of error for the yoke-activation tasks when using various fractions of male average performance for prediction of female Part 91 pilots' performance. Included are Laubach's value for the upper extremities (.595), Laubach's value for dynamic tasks (.686), a general rule of thumb (.67), and our best estimate for minimizing average error across tasks (.6883). Clearly, next to our tailored value, the estimate for dynamics tasks (.686) produces the best average prediction across these specific tasks (0.003 % average error). Across the first six tasks for the female Part 121 pilots, Laubach's factor for dynamic force input (0.57% average error) appeared to be slightly better than our best estimate (0.91 % average error) from the Part 91 pilots' fit when applied to the Part 121 pilots' data. Although the overall error was less using Laubach's upper extremities factor (0.595), it did so at the expense of having none of the estimates accurate to less than 10% error (increasing the error on the first 6 to balance the overestimations on the last two hand-up roll tasks).

Summary of lower percentiles, Samples 1 and 2

Tables 2 and 3 present summaries of the lower percentile values obtained for the participants in both samples, excluding any groups within a sample that consisted of 3 or fewer. These tables are somewhat more conservative than some of the sources (i.e., Karim et al. for pitch tasks), but are not as conservative as others (Humanscale 4). As such, they may be a reasonable compromise for selecting force levels for the activation of aircraft controls that will allow the majority of users to operate them without difficulty. Three cut-off points in the distributions are provided so that the practitioner will have a little more choice (2.5%, 5%, 10%) than that usually afforded by other tables that provide the 50th percentile and one value in each tail.

Table 2. Women's momentary force-application percentiles (lbs) from field data collapsed across samples 1 and 2.

Control	Direction of movement	Hand/foot used	2.5 %		5 %		10 %	
			Pilots	Nonpilots	Pilots	Nonpilots	Pilots	Nonpilots
Yoke	Pull	Left	30.0	22.0	30.1	22.1	31.1	24.2
		Right	29.8	21.6	30.4	23.2	33.8	27.0
	Push	Left	28.7	23.3	29.4	29.2	33.2	32.0
		Right	28.5	25.3	30.6	31.1	31.3	33.0
	Up	Left	14.7	9.5	15.4	10.1	18.4	12.3
		Right	17.4	12.6	19.6	14.2	22.1	18.2
	Down	Left	18.8	16.0	20.0	16.1	20.0	18.0
		Right	18.9	19.1	20.6	20.1	25.8	21.2
Foot Pedal	Push	Left	103.7	52.5	119.8	54.3	127.4	79.2
		Right	112.8	77.4	120.8	85.5	131.2	96.9

Table 3. Men's momentary force-application percentiles (lbs) from field data, collapsed across sample 1 and 2.

Control	Direction of movement	Hand/foot used	2.5 %		5 %		10 %	
			Pilots	Nonpilots	Pilots	Nonpilots	Pilots	Nonpilots
Yoke	Pull	Left	35.7	32.6	39.2	33.2	51.8	34.4
		Right	41.6	33.7	47.1	35.3	51.7	38.6
	Push	Left	40.8	30.0	44.6	31.9	45.7	35.8
		Right	37.8	33.3	44.0	33.6	48.0	34.2
	Up	Left	22.5	20.2	27.3	21.4	32.0	23.8
		Right	26.8	20.3	29.6	20.6	33.8	21.2
	Down	Left	28.7	31.2	29.9	32.4	30.0	34.8
		Right	29.8	36.2	33.4	36.3	35.0	36.6
Foot Pedal	Push	Left	165.0	179.8	174.6	108.5	209.3	182.0
		Right	155.8	177.8	166.1	178.5	181.5	180.0

CONCLUSIONS

The values obtained in this series of samples of pilot and nonpilot performance suggest that some tabled values from the often-referenced sources of anthropometric data may be overestimates of presently obtainable performance for the target groups of interest, male and female pilots engaged in Part 91 (general aviation) and Part 121 (scheduled commercial carrier) operations. It is also apparent that some specific points on the distributions are comparatively higher values than previously documented and suggest stronger performance than previously suggested. As such, it is recommended that one take the conservative approach when using any of these values to set design limits, using the lesser of the collective values or the median of several sources if in doubt. The data also suggest that some predictions of female performance as a fractional measure of male performance can be accurate, whereas other specific tasks may, for various reasons, not be as amenable to estimation. Ultimately, it is recommended that the practitioner carefully examine the conditions surrounding and mechanisms employed in the execution of any force application to controls to determine what may best suit the particular application and, if in doubt, seek additional data specific to the application.

ACKNOWLEDGMENTS

The author thanks the sponsor, ANM-111, Transport Airplane Directorate, and specifically, Loran Haworth, for his assistance during the initial study conception and the members of the Flight Controls TCRG for their assistance. Special thanks to Ronnie Minnick and Rick DeWeese in AAM-630 (Protection and Survival) and their machine shop for support of the measurement system used for the recent samples. Additional thanks for both American Airlines and to Frontier Airlines (Jeannie Davison and Marc Rusthoven) for making arrangements for and providing assistance during onsite data collections.

REFERENCES

Ahlstrom, V. & Longo, K. (2003). Human Factors Design Standard (HFDS) for acquisition of commercial off-the-shelf, nonpdevelopmental items, and developmental systems. Technical Report DOT/FAA/CT-03/05, HF-STD-001. Springfield, VA: National Technical Information Service.

Beringer, D.B. (2006). Anthropometric standards on the flight deck: Origins of control-force-exertion limits and comparisons with recent surveys of human performance limitations. In *Proceedings of the 50th Annual meeting of the Human Factors & Ergonomics Society*, 116-120.

Beringer, D.B., Ball, J.D., and Haworth, L.A. (2007). Control-force-exertion limits and comparisons with pilot and nonpilot populations. In *Proceedings of the 2007 International Symposium on Aviation Psychology*, 31-37.

Beringer, D.B. (2008). An updating of data regarding the forces pilots can apply in the cockpit, Part II: Yoke, rudder, stick, and seatbelt-release forces. In *Proceedings of the 51st Annual meeting of the Human Factors & Ergonomics Society*, 64-68.

Diffrient, N., Tilley, A. R., and Harman, D. (1981). Humanscale – A portfolio of Information: 4 - Human Strength. A Project of Henry Dreyfuss Associates. Cambridge, MA: The MIT Press. ISBN: 0-262-04059-X.

FAA (1996). Human Factors Design Guide. Atlantic City, NJ: FAA William J. Hughes Technical Center.

Karim, B., Bergey, K.H., Chandler, R.F., Hasbrook, A.H., Purswell, J.L., and Snow, C.C. (1972). A preliminary study of maximal control force capability of female pilots. Springfield, VA: NTIS, TR FAA-AM-72-27.

Laubach, L.L. (1976). Muscular strength of women and men: A comparative study. Springfield, VA: NTIS, TR AMRL-TR-75-32.

Meyer, L.G., Pokorski, B.E., and Ortel, J.L. (1996). Muscular strength and anthropometric characteristics of male and female naval aviation candidates. Pensacola, FL: Naval Aerospace Medical Research Laboratory, Technical Report NAMRL-1396.

MIL-STD-1472F (1999). Department of Defense Design Criteria Standard: Human Engineering. Springfield, VA: NTIS, MIL-STD-1472F.

Morgan, C.T., Chapanis, A., Cook, J.S., & Lund, M.W. (1963). *Human engineering guide to equipment design.* New York: McGraw-Hill.

NASA (1978). Anthropometric Source Book (3 volumes). Springfield, VA: NTIS, NASA Reference Publication 1024.

NASA (1995). Man-systems Integration Standards, Revision B. Springfield, VA: NTIS, NASA-STD-3000.

Rees, J.E. & Graham, N.E. (1952). The effect of backrest position on the push which can be exerted on an isometric foot-pedal. *Journal of Anatomy*, 1952, **86**, 310. Cited in VanCott and Kinkaide (1972).

Romert, W. (1966). *Maximalkraefte von Maennern im Bewegungsraum der Arme und Beine*. Koeln, Germany: Westdeutscher Verlag. Cited in VanCott and Kinkaide (1972).

Stoll, T., Huber, E., Seifert, B., Stucki, G., and Michel, B.A. (2002). Isometric Muscle Strength Measurement. New York: Thieme.

Van Cott, H.P and Kinkade, R.G. (1972). Human Engineering Guide to Equipment Design (Revised Edition). Washington, D.C.: U.S. Government Printing Office.

SPEECH SYNTHESIS FOR DATA LINK:
A STUDY OF OVERALL QUALITY AND COMPREHENSION EFFORT

Martine Godfroy, San Jose State University Foundation
Durand R. Begault, Human Systems Integration Division (ARC-TH)
Elizabeth M. Wenzel, Human Systems Integration Division (ARC-TH)
NASA Ames Research Center
Moffett Field, CA 94035
Contact: martine.godfroy-1@nasa.gov

This study investigated subjective preference for synthesized "spoken data link" messages to provide initial design guidance for communication displays in the context of NextGen (Next Generation Air Transport System) operations. Ratings of Overall Quality and Comprehension Effort were obtained as a function of voice type, synthesized speech rate, and sentence prosody. Rank-order data analyses showed that both Overall Quality and Comprehension Effort were affected by speech rate: under the "Fast Rate" condition (vs. "Default Rate"), Overall Quality decreases and Comprehension Effort increases. However, the introduction of "Prosodic Emphasis" (pitch and level changes for specific phrases) in Fast Rate sentences produced a relative improvement in both comprehension and quality ratings. For both speaking rates, the introduction of "Prosodic Emphasis" resulted in higher quality ratings and lower comprehension effort ratings. The data suggest that faster speaking rates, which may improve message throughput in a display, may be viable when combined with prosodic emphasis.

The overall objective of our work is to investigate human performance costs and benefits of voice communications interfaces in NextGen flight decks, with the particular goal of improving shared situational awareness and minimizing human error and workload. This study investigates perceived Overall Quality and Comprehension Effort of synthesized speech systems in an aviation context of "spoken data link" and "party line" messages. The increased operational autonomy of fight crews in the NextGen will potentially result in higher overall workload and greater dependence on the visual modality to interact with flight instrumentation and data. To alleviate these problems, future systems may be designed to advantage by using synthetic speech displays to convey, e.g., shared situation awareness regarding flight status and trajectory between cockpit to cockpit and cockpit to ground. Such displays may be combined with other design strategies such as 3-D audio presentation to achieve more intelligible, discriminable, and identifiable communications (Begault, 1999).

While human performance analyses predict clear benefits to both workload and capacity in a data link-dominated airspace system (Leiden, et al., 2003), previous studies (Smith, et al., 2001) suggest that voice communications are preferred for interactive tasks involving emergency situations, position reports, and some clearance requests. A study conducted for the Federal Aviation Administration (FAA) examined the use of pre-recorded speech versus text-based data link messages, and whether or not speech displays reduced head-down time and improved response time (Rehmann, 1996). The results of that study indicated advantages of speech for reducing head-down time, but the slow cadence of the speakers was found to be slower than reading (in that system, actual voices were recorded digitally and played from a computer.) Speech was preferred in most cases over text by pilots, and workload was reportedly decreased for confirming messages.

The primary advantage of text-to-speech synthesis is that arbitrary messages can be communicated without pre-recording. A further advantage of more advanced systems is the ability to manipulate the manner of speaking in real time; in particular to change the *prosody* of the spoken utterances to convey meaning outside of the syntactic context. For example, current speech synthesis-recognition systems, known as "interactive-voice-response systems," used in telephonic commerce are able to establish specific personas that can express characteristics like urgency or compassion that enhance semantic content. The current study used the "Rvoice" (Rhetorical Systems, Ltd) commercial text-to-speech synthesis system, which has high speech quality combined with the ability to parametrically vary a number of speech characteristics related to prosody and speaking rate. By manipulating prosody, messages can be formed to add emphasis to important words, such as call signs, and to elevate the perceived urgency of a message, such as when speaking in a "raised" versus normal voice.

A speech synthesizer can also change the rate of spoken words without the loss of intelligibility that an actual speaker would have. Brungart, Simpson, and Nandini (2007) investigated the use of "time-compressed" speech for military aviation applications; they concluded "it is often possible to accelerate a speech signal by a factor of two or more without much sacrifice in intelligibility." The advantage of increasing the spoken word rate is that information transfer can occur more quickly, increasing the amount of information that can be received per unit time.

Given the potential advantages for future flight decks of varying prosody and speaking rate in spoken data link messages, this study sought to determine if these manipulations might impact the subjective Comprehension Effort and Overall Quality of the communications. Subjective Comprehension Effort refers to the perceived effort necessary to understand a message. Sound Quality refers to features of a sound that contribute to the subjective impression made on a listener, with reference to the suitability of the sound for a particular set of design goals, and is meant particularly to account for aspects of communication systems that are not quantifiable by intelligibility measurements. It is well understood that the requirements for speech intelligibility will likely be satisfied by use of an appropriate signal-noise ratio. Recent work in the area of perceptual audio evaluation (Bech and Zacharov, 2006) has shown that subjective data can elucidate a variety of stimulus factors that contribute to high quality audio system design. Telephony research (engendered in various quality testing standards of the International Telecommunications Union-ITU) has long recognized that a particular system may produce an acceptable level of intelligibility while having poor overall quality. Our research presumes that poor audio quality may work against the system's long-term use in an everyday situation.

Experimental Method

This experiment was designed to determine preferences for a given type of synthesized voice (gender, timbre) as well as the effects of speech rate and sentence prosody on the perception of overall quality and on the degree of comprehension effort. The experiment generally followed the ITU recommendations for speech quality experiments (ITU P.800, "Methods for subjective determination of transmission quality") and for synthesized speech (ITU P.85, "A method for subjective performance assessment of the quality of speech voice output devices"). Subjective scales for evaluation per ITU P.800.1 ("Mean opinion score (MOS) terminology") were used to rate "overall quality" and "comprehension effort".

Participants

Eight undergraduate students (4 male and 4 female) from the San Jose State University, CA, aged 19 to 45, took part in the experiment. The participants were all volunteers and were naïve regarding the content and objective of the experiment. All reported that they had normal hearing.

Experimental design

A stimulus corpus of sixty different sentences was composed based on a transcript of Air Traffic Control communications sent by the tower to the crew from a study previously conducted at NASA. The following are examples:

*"**Asiana 2 1 4** turn right heading to 1 3 0 and proceed direct Molen"*
*"**Air France 0 8 4** contact tower 3 2 0 point 5 good day"*
*"**Korean 0 2 3** turn left and maintain 8000 accept approach clearance in 5 nautical miles"*

The 60 sentences were randomly assigned with a voice, a rate, and a prosody manipulation. The three possible voices were American English-speaking voices that are provided as options with the standard Rvoice text-to-speech software package. Rvoice allows manipulation of a number of speech parameters using a scripting language of XML tags imbedded in the text-to-speech string.

The prosody manipulation was accomplished by using (1) the default settings for each voice or (2) increasing the pitch and loudness of the aircraft identifier phrase contained in each sentence (bold face words in the example sentences above) to provide "prosodic emphasis." Manipulation of internal Rvoice software parameters resulted in an increase in the fundamental frequency ("pitch") of the spoken voice by approximately 40 Hz and a loudness increase of approximately 4 dB, depending on the specific phrase. The result was an audible "raising" of the voice, both in terms of pitch and level. The speech rate manipulation was accomplished by using (1) the default settings or (2) an increase in the speaking rate of the overall phrase by a factor of about 1.2. The result was that of an articulate but faster than normal speaking voice. Note that the pitch could be manipulated independently of the rate (unlike, e.g., an analog tape recorder under fast playback). Prosodic emphasis and speech rate manipulations were applied to three different synthetic voices, two male American English speakers and one female American English speaker.

The experimental variables are summarized as follows:

- VOICES: 1 female (Rvoice ID: F019), 2 males (Rvoice IDs: M002, M009)
- PROSODY: 2 prosody levels (No Prosody: NP, Prosody: P)
- RATE: 2 speed levels (Normal: N, Fast: F)

The sentences were synthesized and then digitally mixed with a binaural recording of a 747 flight deck simulator cruise phase background noise, calibrated to a level of approximately 70 dB SPL. The signal-to-noise ratio of the sentences was +10 dB RMS relative to the background noise.

Using the CRC-SEAQ (Communications Research Centre Canada System for the Evaluation of Audio Quality) subjective test module, we designed two separate sessions of 60 trials, one for evaluation of the Overall Quality and one for Comprehension Effort. Two different 5 level scales were used for each session:

-Session 1 (Overall Quality): 1 (bad) to 5 (excellent)

-Session 2 (Comprehension Effort): 1 (excellent, no effort required) to 5 (bad, high effort required)

Sample screenshots for the Overall Quality and Comprehension Effort trials are shown in Figure 1. A trial consisted of the successive presentation of 3 different sentences, noted A, B and C on the experiment screen. No reference was presented. The task of the participants was to rate the 3 different sentences one relative to the others by adjusting a cursor associated with the 5-point scale that was always visible on the left of the screen. The listeners were not allowed to assign ratings until they had listened to each sentence at least once. However, they were allowed to listen to any of the sentences as many additional times as they wished. During a particular trial, the sentences assigned to the A, B and C stimuli from the list of 60 possible sentences were always different sentences. Nevertheless, the random assignment of speaking voice, prosody level, and speaking rate to the A, B and C stimuli was such that the listeners may have heard the same voice, prosody level, or speaking rate more than once. Over the course of all 60 trials in each session, the listeners heard each voice, prosody level, or speaking rate the same total number of times. All listeners heard the same randomized list for both the Overall Quality and Comprehension Effort sessions. Prior to each session, the listeners were given written instructions and the experimenter talked with them to make sure they understood the task.

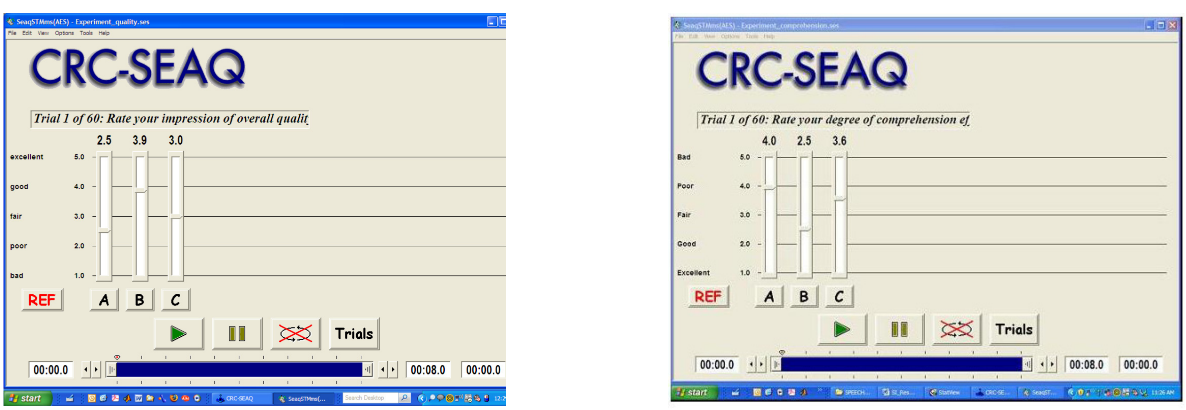


Figure 1: CRC-SEAQ interface for the Overall Quality (left) and Comprehension Effort (right) sessions. The letters A, B and C represent the three successive sentence stimuli. The score estimate is relative, i.e. each sentence stimulus is rated in relation to the two others. No reference was used in this experiment. Note that the scale for Comprehension Effort is inverted compared to the scale for Overall Quality.

Results

Descriptive analyses of the data were conducted with SPSS® to assess whether the data were normally distributed. One sample Kolmogorov-Smirnov test showed that the distributions for both Overall Quality and Comprehension Effort were normal (OQ: Z=.91, p=.37; CE: Z=.61, p=.84). The data for Overall Quality showed some asymmetry (skewness=-.43, SE=.24) and a flat distribution (kurtosis=-.47, SE=.48). Data for Comprehension Effort were closer to a normal distribution in terms of skewness (skewness=-.03, SE=.24), although the kurtosis value again indicated a flat distribution (Kurtosis =-.56, SE=.48). Because of the qualitative nature of the data, non-parametric analyses were preferred. Friedman tests were performed to determine the main effect of the condition of presentation (N=4). A posteriori Wilcoxon tests were used to further analyze the results to determine the specific effects of Speed and Prosody. The significance level was set at *p*=.05.

Overall Quality

A Friedman test performed on the 4 repeated-measures conditions (Non-Prosody/Normal, Non-Prosody/Fast, Prosody/Normal and Prosody/Fast) showed statistical differences between the reported ratings ($\chi^2_3 = 25.28$, *p*<.0001). As expected, a faster speech rate led to a decrease in perceived Overall Quality, both in the Non-Prosody and in the Prosody condition (NPN-NPF: Wilcoxon Z=-3.68, *p*=.0002, PN-PF: Wilcoxon Z=-2.82, *p*=.004). More interestingly, introduction of Prosody in the sentence did not significantly affect the perception of quality when the speech rate is normal (NPN-PN: Wilcoxon Z= -1.35, *p*=.17), but significantly compensated for the deleterious effect of increased speed (NPF-PF: Wilcoxon Z= -3.06, *p*=.002).

Table 1 summarizes the observations and provides supplementary data as a function of subject gender (SM vs. SF) and voice gender (VM vs. VF). It can be seen that the Overall Quality scores are not modified by this a posteriori categorization. The only minor difference concerns the marginal significance (p=.05) of the differences between the NPN/PN conditions for male subjects, who appear to be more sensitive than female subjects to the effect of prosody That is, male subjects produced lower ratings compared to female subjects in the Non-Prosody condition.

Table 1. *Friedman mean rank for perceived Overall Quality (scale from 1= bad to 5=excellent) as a function of the 4 conditions of presentation of the sentences*

Conditions	Total		Subject M		Subject F		Voice M		Voice F	
NPN	2.83	*p*=.17	2.50	*p*=.05	3.16	*p*=.87	2.93	*p*=.25	2.62	*p*=.52
PN	3.25		3.25		3.25		3.31		3.12	
NPF	1.48	*p*=.002	1.66	*p*=.03	1.29	*p*=.02	1.53	*p*=.02	1.37	*p*=.04
PF	2.44		2.58		2.29		2.21		2.87	
χ^2_3	*p*<.0001		*p*=.02		*p*=.0004		*p*=.0004		*p*=.02	

Comprehension effort

Similar to the Overall Quality evaluation, we observed a significant effect of condition on the rating of Comprehension Effort ($\chi^2_3 = 33.54$, *p*<.0001). An increase in speech rate increased the perceived Comprehension Effort, both under Prosody (Wilcoxon Z=-3.41, *p*<.001) and Non-Prosody conditions (Wilcoxon Z=-3.88, *p*<.0001). Once again, the role of prosody was not statistically significant under normal rate speech production (Wilcoxon Z=-1.35, *p*=.17), but was shown to be relevant (Comprehension Effort decreased) when sentences are produced at higher speed (Wilcoxon Z=-3.34, *p*<.001).

Table 2 summarizes the results for Comprehension Effort and provides supplementary data as a function of subject gender (SM vs. SF) and voice gender (VM vs. VF). Similar to the Overall Quality scores, ratings of Comprehension Effort were not modified by this a posteriori categorization. The faster speech rate always produced a greater Comprehension Effort (*p* at least <.01), and the introduction of prosody components in the sentence contributed to a reduction of the Comprehension Effort when the perception was altered by an accelerated presentation.

Table 2: *Perceived Comprehension Effort (scale from 1= excellent to 5=bad) as a function of the 4 conditions of presentation of the sentences.*

Conditions	Total		Subject M		Subject F		Voice M		Voice F	
NPN	2.02	*p*=.17	1.87	*p*=.96	2.16	*p*=.09	2.15	*p*=.06	1.75	*p*=.77
PN	1.58		1.58		1.58		1.50		1.75	
NPF	3.56	*p*=.001	3.62	*p*=.02	3.50	*p*=.01	3.56	*p*=.006	3.56	*p*=.06
PF	2.83		2.91		2.75		2.78		2.93	
χ^2_3	*p*<.0001		*p*=.0002		*p*=.002		*P*<.0001		*p*=.007	

Overall Quality and Comprehension Effort

Figure 2 plots the Friedman mean ranks for Overall Quality and Comprehension Effort as a function of prosody and rate collapsed across subject gender and voice gender since these categories had little or no impact on the results.

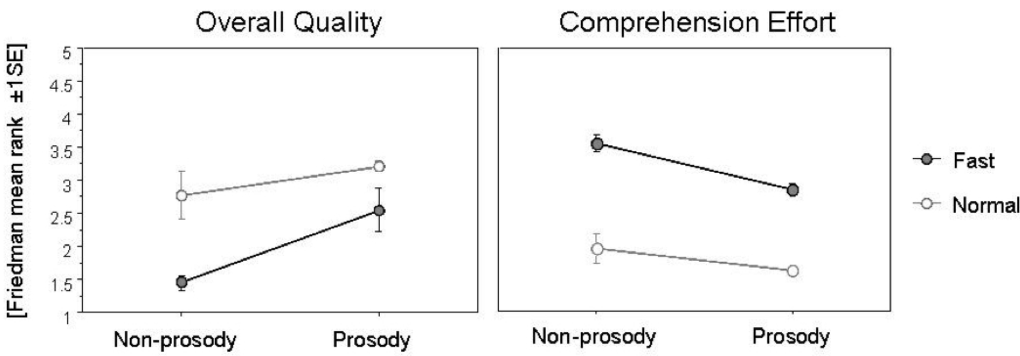


Figure 2: Left: Friedman Mean rank for Overall Quality (1 =bad to 5=excellent) as a function of prosody & rate. Right: Friedman mean rank for Comprehension Effort (1 =excellent to 5=bad) as a function of prosody & rate.

Overall, analysis of the data for Overall Quality and Comprehension Effort revealed that they were very similar (Z=-.02, *p*=.97). When split between the different conditions, the two measures were shown to be highly and inversely correlated as seen in Figure 3.

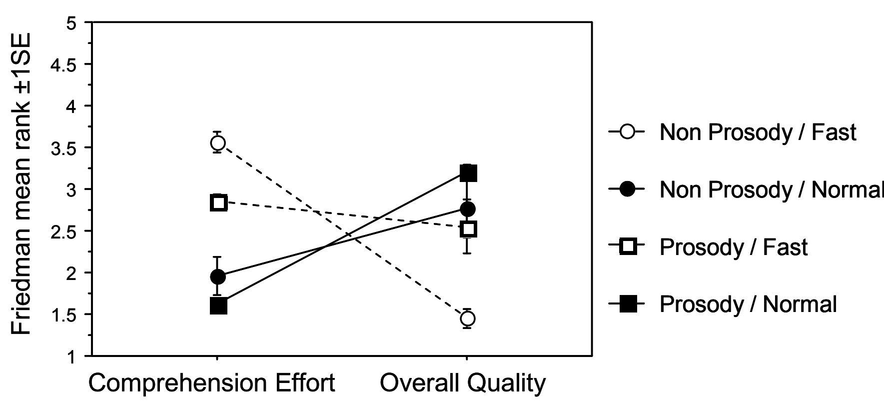


Figure 3: Overall Quality (1 =bad to 5=excellent) and Comprehension Effort (1 =excellent to 5=bad) for the synthesized voices are plotted as a function of the four conditions of presentation of the sentences (rate by prosody). Note that the inversion of the scale leads to the pattern of interaction between the two conditions.

The inverse relationship resulted from the inverse numerical order of the two rating scales. Indeed, fast presentation of the sentences was associated with the highest degree of Comprehension Effort and the lowest perception of

Overall Quality. Similarly, the addition of prosodic emphasis produced the lowest degree of Comprehension Effort combined with the highest level of Overall Quality.

Conclusions and Discussion

In general, that data showed that both the gender of the listener (OQ: Z=-.56, *p*=.57; CE: Z=-.31, *p*=.75) and the gender of the stimulus voice (OQ: Z=-.33, *p*=.73; CE: Z=-.16, *p*=.86) had no significant impact on perceived overall quality or comprehension effort. This is perhaps not surprising since previous work on audio displays (e.g., Brungart, et al, 2007) have found gender to significantly impact intelligibility in multiple talker situations as opposed to the single talker stimuli used here.

As expected, a faster speech rate led to a decrease in perceived Overall Quality and an increase in the perceived degree of Comprehension Effort, both in the Non-Prosody and the Prosody conditions. Introduction of prosodic emphasis in the messages did not significantly affect the perception of quality or comprehension effort when the speech rate was normal. While the combination of prosody and a normal speaking rate produced the highest average ratings of Overall Quality and the lowest ratings of Comprehension Effort, the data were not statistically different from ratings of the normal speaking rate without prosody. However, prosodic emphasis significantly compensated for the deleterious effect of increased speaking rate, by both increasing perceived Overall Quality and decreasing Comprehension Effort. The data indicate that the subjective acceptability of the messages when prosody was combined with the faster rate remained relatively high (above the theoretical mean). The results suggest that if faster message throughput is required in a speech display system, possible negative effects may be offset by the use of prosodic emphasis to enhance meaning in the message.

Future investigations will be concerned with the determination of ceiling effects, i.e. the determination of an acceptability threshold for speaking rate. The impact of prosody and spatial separation on Overall Quality and Comprehension Effort for synthesized speech messages in a background of competing messages will also be examined. We also plan to correlate the results of our subjective impression data with objective measures such as intelligibility.

Acknowledgements

This study was supported by the Integrated Intelligent Flight Deck project within NASA's Aviation Safety Program.

References

Bech, S. & Zacharov, N. (2006). *Perceptual Audio Evaluation: Theory, Method and Application.* West Sussex, England: John Wiley & Sons, Ltd.

Begault, D. R. (1999). Virtual acoustic displays for teleconferencing: intelligibility advantage for "telephone-grade" audio, *Journal of the Audio Engineering Society*, *47,* 824-828.

Brungart, D. S., Simpson, B. D., & Iyer, Nandini (2007). Maximizing information transfer in auditory speech displays. In *Proceedings of the 19th International Congress on Acoustics.* Madrid, Spain.

CRC SEAQ: Communications Research Centre of Canada System for the Evaluation of Audio Quality. (http://www.crc.ca/en/html/aas/home/products/products#CRC_SEAQ)

Leiden, K., Kopardekar, P. & Green, S. (2003). Controller workload analysis methodology to predict increases in airspace capacity. *AIAA Aviation Technology, Integration, and Operations Forum,* Reston, VA.

Rehmann, A. J. (1996). Airborne Data Link Study Report. DOT/FAA/CT-TN95/62, Springfield, VA: National Technical Information Service.

Smith, N., Brown, J. A., Polson, P. and Moses, J. (2001) "An assessment of flight crew experiences with FANS-1 controller-pilot data link communication in the South Pacific." *4th USA/Europe Air Traffic Management R&D Seminar,* Santa Fe, NM.

INTEGRATED MULTIMODAL COMMUNICATIONS MANAGEMENT FOR AIRBORNE COMMAND AND CONTROL

Dianne K. Popik, Victor S. Finomore and Douglas S. Brungart
711 Human Performance Wing, Air Force Research Laboratory
Wright Patterson AFB, Ohio

Increasing the communication efficiency and accuracy associated with Command and Control (C2) operations is crucial in many aerospace applications. This communication intensive environment imposes a high workload on Air Traffic Controllers and other C2 personnel who rely heavily on a variety of communication tools to efficiently plan, direct, coordinate, and control assets during missions. The C2 task is further complicated by the suboptimal integration of the various communication media utilized in the operational environment. The fielded communication tool suites have serious limitations and are not poised to meet the needs of advancing technology. A multidisciplinary research team in the Battlespace Acoustics Branch of the Air Force Research Laboratory has developed the Multi-Modal Communications (MMC) tool suite which is specifically designed to increase communication effectiveness, provide efficient voice communication retrieval, navigation of saved data as well as reducing the perceived mental workload of the operators.

Challenges in Command and Control Communications

Command and Control communications are challenging for many reasons. This time critical communication intensive environment imposes a high workload on the operators, who typically monitor and transmit on as many as eight or more simultaneous voice channels in addition to other forms of communication such as text chat, phone and email. Reducing the workload and increasing the communication efficiency associated with C2 operations is of extreme importance to achieve mission success.

Voice communications pose a unique problem due to the transient nature of verbal transmissions. The recipient has one chance to extract the crucial information or be forced to request a "repeat" which adds to the radio traffic. Additionally, multiple operators can speak at the same time thus reducing the intelligibility of essential messages. Not only is missed communication a problem during the mission, but the operators must take detailed notes to capture the information received. Later, they must rely on manually transcribed audio recordings of the mission for training and debriefing.

Currently, C2 centers are not required to standardize the collaboration tools available and used by the operators. Therefore, it is possible to have very different collaboration tools in various C2 facilities. Since multiple collaboration tools are needed to accomplish the mission, combinations of these collaboration tools are often kludged systems which are meant to assist the operator with the management of these individual systems. The incompatibility of these multiple modes of communication coupled with the variable combination of systems makes it difficult to combine them into a functioning collaborative tool suite suitable for all C2 environments. Presently, the integration of existing communication technologies has not yielded optimal results. The fielded communication tool suites have serious limitations and are not poised to meet the needs of advancing technology.

In addition to the lack of standardization and the incompatibility of the communications systems, there currently is very little capacity to capture, save and review these various communications. For instance, verbal radio communications are transient and perishable and reviewing verbal communications during a mission is impossible at the present time. A multidisciplinary research team in the Battlespace Acoustics Branch of the Air Force Research Laboratory has developed the Multi-Modal Communications (MMC) tool suite which is specifically designed to address these issues.

Current Collaborative Tools

Operators surveyed (Berry et al., 2006) in various C2 centers listed the collaborative tools most commonly used as phone, Chat/text messaging (mIRC), email, radio, secure telephone unit/secure terminal equipment (STU/STE), Voice over Internet Protocol (VoIP), Information Work Space (IWS), and video teleconference (VTC). Operators admitted there were too many different collaboration tools available yet each was limited in its own way. Operators stated that phone, text chat and VoIP were somewhat effective pending their availability. Of these collaborative tools, most operators preferred text chat. Perhaps this preference is due in part to the poor audio tools available making them an undesirable form of communication or it may be due to the convenience of a written record of communications that text chat provides coupled with the overall system stability of the chat functions which were less likely to crash.

Information Work Space

Information Work Space combines a few collaborative tools, but has limitations in sharing, posting, accessing, filing and attaching documents as well as data loss when IWS frequently crashes (Berry et al., 2006). The resulting work following a crash entails rebuilding chat rooms and regaining situation awareness (SA). Another drawback of IWS is its limited audio capability due to the simplex system similar to a walkie-talkie where one party speaks at a time. A simplex system limits the ability to communicate naturally by prohibiting interjections and not surprisingly is an unpopular form of communication. IWS chat functions were more often used than voice functions with operators using on average five chat windows at a time, while a few operators had as many as 14 chat windows open.

Text Messaging/Chat

Text Messaging/Chat was commonly used especially when IWS was unavailable (Berry et al., 2006). When this occurred the operators would switch to mIRC which is an Internet Relay Chat (IRC) client for Microsoft Windows. Despite mIRC's popularity, the operators noted it is difficult to reconfigure following a system crash. In this setting, mIRC was used only temporarily and once IWS was returned to functioning status, mIRC communications were duplicated and placed into IWS again. A common complaint about the chat tool was the inability to copy large amounts of data into a chat room. Brief departures from the chat room also caused a loss of SA since it was not possible to tag the last entry read. The operators would be forced to spend precious time re-reading text to ascertain where they previously left the stream of communication.

Other Collaborative Tools

Email offered a limited capacity in such a dynamic and high tempo environment (Berry & Lindberg, 2009). Phone and STU/STE communication was inconvenient since it required operators to remove their headset in order to hold the phone to the ear. VoIP functioned similarly to telephone use but heard through the headset. VTC was available in certain facilities in the building and not at individual terminals. Radio was often used; however, the operators rarely wore both ear cups in order to hear conversation around them.

Purpose of Multi-Modal Communications Tool Suite

The Multi-Modal Communications (MMC) tool suite was designed to offer C2 operators a combined versatile and intuitive interface which would alleviate the workload and errors associated with an intensive communication environment. This integrated Communications Management tool will improve communications by streamlining the cumbersome and varied forms of communications and give the C2 operator access to the complete spectrum of communications in a single tool. Voice and chat communications will be seamlessly integrated into a single digital communication system over one headset (phone, chat, voice, radio) for internal and external communications.

MMC is an integrated net-centric architecture which will distribute, monitor, archive, and retrieve analog voice transmissions (radio communications), VoIP communications, and text messages across distributed operators on the GIG.

MMC records, archives and displays the verbal and text communications to the operator for real-time playback during the missions reducing workload while enabling the C2 operator to be more effective and efficient. This eliminates the perishable nature of radio communication and allows the operator to focus on the task instead of remembering and writing down information. The MMC tool also employs virtual audio display technology to spatialize the multiple audio signals to aid in the intelligibility of the radio communication. The combination of these technologies has led to the design of a communication interface that will improve the performance of operators confronted with monitoring a high volume of radio communication.

Features of Multi-Modal Communications Tool Suite

Audio Recording

The MMC tool captures the radio communication as text and records each transmission as an audio file. Operators have the ability to play back the original radio transmission by clicking on the desired line of transcribed text. Each audio file or radio transmission is time stamped for easy reference and documentation.

Speech-to-Text Transcription

The speech-to-text transcription feature captures incoming speech and transcribes it into text. This allows all voice traffic to be captured and recorded as a text log. The operator now has the ability to read what was spoken and review previous voice transmissions. Since all radio communications are logged, the operator is easily able to search for keywords during the mission or use the text for debriefing or training purposes.

Spatial Audio

The spatial audio feature (also called 3D Audio) allows users to spatialize each of their monitored radio channels such that the audio signals appear to originate from different azimuth locations. MMC allows the operator to place the radio channels in one of nine spatial locations and to change that location anytime during the mission. This flexibility of configuration allows the operator to organize and more efficiently monitor multiple radio channels. Several studies have shown that the spatialization of speech can improve the intelligibility of communication, lower the perceived mental workload associated with monitoring simultaneous streams of communication, and decrease the negative effects of noise during communication (Bolia, 2003; McAnally, Bolia, Martin, Eberle, & Brungart, 2002; Nelson, Bolia, Ericson, & McKinley, 1998; Nelson, et al.1999; Ricard & Meirs, 1994).

Instant Replay

The instant replay feature provides immediate access to the last message transmitted. By pressing the 'replay' button, the last fifteen seconds of the radio transmission will be replayed, allowing the operator to instantly review or clarify the last transmission. Additionally, the replay feature isolates that particular channel by temporarily muting all other channels.

Isolate

The isolate feature mutes all other radio channels and only plays the specified channel, allowing the operator to focus their attention solely on that channel. The operator may now more effectively direct their attention to critical situations by muting less critical radio transmissions.

Transmit

The push-to-talk feature allows the operator to speak and be heard by other operators monitoring the specified channel. The push-to-talk is activated by mouse clicking the 'Talk' button and holding it down while speaking. Other operators listening to that channel, hear the communication in real-time while the spoken communication is transcribed into the display window.

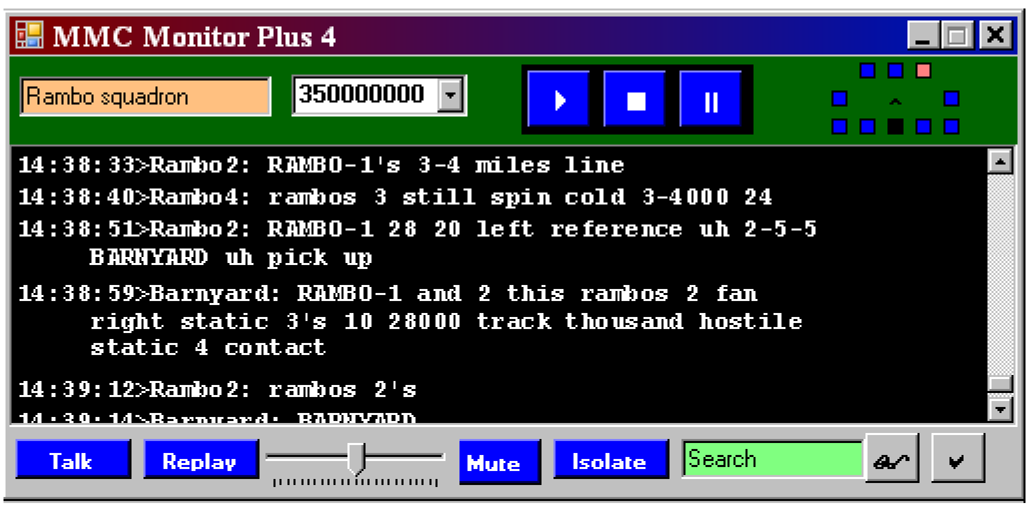


Figure 1. Multi-Modal Communications tool suite display of a single radio communications window.

Future Capabilities

Since the operators prefer the use of chat messaging while communicating; chat capabilities are currently being implemented into the MMC. This chat feature will be similar to current chat clients in that users will sign into secured chat rooms to monitor and transmit text messages. This chat function will have the capacity to convert text to speech. Operators will have the option to speak their messages to be transcribed in the chat window as well as hear text messages in a synthesized voice. Additional features will enable the operator to review logged communications in faster than real-time to be able to review large chucks of past communication in a speeded form in order to be brought up-to-date.

Research and Survey Data

Throughout the development of the MMC tool, 11 subject matter experts (SME) were shown MMC and asked to participate in a usability questionnaire and interview. The SMEs had diverse C2 operations experience ranging from 1 to 27 years (M = 11.57), thus providing comprehensive feedback on the features and design of MMC. This feedback led to the redesign of the interface making it more operator-friendly and congruent with the operator's needs. The SMEs indicated that MMC would improve job performance by allowing them to accomplish tasks more quickly and efficiently. MMC was unanimously supported by all of the SMEs involved as an essential tool in the field and a valuable tool for training and debriefing. Feedback also included comments indicating the spatial audio feature would enhance talker identification and speech intelligibility while the transcription and playback features would help reduce miscommunications and the need for call backs. Overall the SMEs signified that MMC was intuitive and easy to use.

In addition to collecting questionnaire data, performance measures were also collected to compare operators' ability to detect and respond to critical messages via standard radio communication alone (no 3D audio, no voice-to-text and no replay) and with the MMC (3D audio, voice-to-text, and replay). Operators monitored six radio channels for ten minutes for the presence of a critical message which identified a hostile entity along with information pertaining to their location (Viper 1, Hostile-North Lead Group, 55 miles). Their task was to repeat that information back on the correct radio channel. There were six radio transmissions per min with the occurrence of one critical message per min on each of the channels. Nine paid participants from the General Dynamics research pool at Wright-Patterson Air Force Base in the Battlespace Acoustics Branch served as participants in this study. Three separate 2 (Condition) × 2 (Trails) Within-Analysis of Variance was performed for measures of Correct Detection, Response Accuracy, and Response Time. The data revealed that participants were better at detecting critical messages in the MMC condition (M_{CD} = 72. 31%) then in the radio condition (M_{CD} = 50.18%), F (1, 8) = 23.23, $p < .01$. The ANOVA on the response accuracy also showed that MMC condition (M_A = 94.27%) was greater than the radio condition (M_A = 82.56%), F (1, 8) = 7.20, $p < .05$. Analysis of response time revealed that it took longer to reply when using MMC (M_T = 11.59 s) then the standard radio (M_T = 7.65 s), F (1, 8) = 5.39, $p < .05$. Thus it seems the addition of the voice-to-text capability in the MMC algorithm improves overall performance in detection and response accuracy but may increase response time, presumable because listeners who are uncertain were waiting for the transcribed text before making a response. It is important to note that there was on average a 5 to 8 sec delay in the voice-to-text transcription thus inherently increasing the time it took participants relying on the text to reply. The transcription latency has since been decreased to less than a second thus continuing investigation of the performance of the MMC is in progress. None of the other main effect of interactions in these analyses reached significance, $p > .05$ in all cases.

Conclusions

MMC has integrated several stand-alone features known to improve communications in order to create a network-centric communication management suite. The data collection and feedback from subject matter experts (SME) indicates that MMC has the potential to improve the communication effectiveness of operators in intense communication environments. Although MMC currently does not meet all of the needs of an operator we fully expect the empirical tests, both in the lab and field studies, to show that MMC does in fact improve performance during communication monitoring tasks. We also expect these tests to highlight other features to further develop in the MMC thus creating a fully functional multi-modal communication suite.

Acknowledgements

We would like to gratefully acknowledge the technical contributions of Ron Dallman, John Stewart, Christopher Pitstick, and Joel Miller, the MMC project would not exist today without their diligent work. The Multi-Modal Communications team would also like to thank the contributions from David Williamson, Robin Snyder, and Timothy Barry for their work involving speech transcription. Additionally, we would like to thank the subject matter experts who donated their time and expertise to the further refinement of this system.

References

Berry, D. et al. (2006). Current State of Human Systems Integration (HSI) within the Air and Space Operations Center (AOC) (AFRL-HE-WP-TR-2006-0155). Washington, DC: U.S. Government Printing Office.

Berry, D. & Lindberg, R. (2009). Report on Current State of Human System Integration (HSI) and Facilities at KEEN EDGE 2007 (AFRL-RH-WP-TR-2009-#### [in print as of 12 March 2009]). Washington, DC: U.S. Government Printing Office.

Bolia, R. S. & Nelson, W. T. (2003). Spatial audio displays for target acquisition and speech communications. In L.J. Hettinger & M. W. Hass (Eds.), *Virtual and Adaptive Interfaces* (pp. 187-197). Mahwah, NJ: Lawrence Erlbaum Associates.

Ericson, M. A., Brungart, D. S., & Simpson, B. D. (2004). Factors that influence intelligibility in multitalker speech displays. *International Journal of Aviation Psychology, 14*, 313-334.

McAnally, K. I., Bolia, R. S., Martin, R. M., Eberle, G., & Brungart, D., S. (2002). Segregation of multiple talkers in the vertical plane: Implications for the design of multiple talker display. *Proceedings of the Human Factors and Ergonomics Society, 46,* 588-591.

Nelson, W. T., Bolia, R. S., Ericson, M. A., & McKinley, R. L. (1998). Monitoring the simultaneous presentation of spatialized speech signals in a virtual acoustic environment. *Proceedings of the 1998 IMAGE Conference* (pp. 159-166). Chandler, AZ: The IMAGE society, Inc.

Nelson, W. T., Bolia, R. S., Ericson, M. A., & McKinley, R. L. (1999). Spatial audio displays for speech communications: A comparison of free field and virtual acoustic environments. *Proceedings of the Human Factors and Ergonomics Society,43,* 1202-1206.

Popik, D.K. & Finomore, V.S. (2009). Multi-Modal Communications (MMC) (AFRL-RH-WP-TR-2009-0013). Washington, DC: U.S. Government Printing Office.

Ricard, G. L., & Meirs, S. L (1994). Intelligibility and localization of speech from virtual directions, *Human Factors, 36*, 120-128.

CONCEPTUALIZING SPATIAL RELATIONS IN FLIGHT TRAINING

Edwin Hutchins & Will Newsome
University of California San Diego
La Jolla, CA
Christina Middleton
Flight Crew Systems, Inc.
Carlsbad, CA

In the aviation human factors literature, situation awareness (SA) is usually described as arising from disembodied mental processes. Action has virtually no role in current theories of SA. This disembodied view is out of step with contemporary theories that take cognitive processes to be distributed, situated, and above all, embodied. This shift in theory suggests that SA ought to be an embodied phenomenon, and given the highly spatial nature of SA, it would be quite surprising to discover that the body did not play a key role in the construction, elaboration, and maintenance of SA. In this paper we examine the construction of elements of SA in ongoing flight training conducted in a light jet. We show that flight instructors and students make extensive use of their bodies and the relations of their bodies to surrounding space while constructing, remembering, and reasoning about the situation of the airplane.

When pilots transition to a new airplane, they must learn how to think about dynamic flight trajectories in ways that match the performance of the airplane. Awareness of spatial relationships between the airplane and its surroundings is a key element of situation awareness (SA). Contemporary aviation human factors seems to take SA to be a purely mental construct. Endsley (2000) provides a general definition of situation awareness: “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.” For Endsley and many others, this is a one-way process of information input. There is no consideration of the role that bodily activity plays in perception, comprehension, and imagination of the future. For example, Bowers, Jentsch, & Salas (2003) claim that situation awareness is in large part collecting instrument information to construct a “mental picture” of the situation. They also discuss coordination, decision-making, adaptability, and performance monitoring without explicitly mentioning the body. The DoD’s Aviation Safety Improvements Task Force (2009) acknowledges the body as a locus of disease, fatigue and reaction to environmental conditions. Of course the body is given a role in perception and action. However, with respect to communication, reasoning, and conceptualization (for example in constructing the spatial aspects of situation awareness), the body is given no role at all. Crew Resources Management training (O'Connor, Campbell, Newon, Melton, Salas, & Wilson, 2008; Seamster, Boehm-Davis, Holt, & Shultz, 1998) treats both decision-making and situational awareness as “mental factors” and includes no consideration of the role of the body or action in constructing them. Banbury, Dudfield, Hoermann, & Soll (2007) provide a rare exception, noting the importance of non-verbal communication in the construction of situation awareness. The design of pilot interfaces sometimes confronts the reality of the presence of a body. For example, situation awareness can be improved by adding cues in haptic and auditory channels (Lam, Mulder, & van Paassen, 2007; Wickens, Small, Andre, Bagnall, & Brenaman, 2008; Curry, Estrada, Grandizio, & Erickson, 2008).

We believe that the failure to attend to bodily action in the creation of SA is a serious omission because the body is an important resource in these processes and because activities that interfere with the employment of the body in these processes degrade performance. A growing body of research shows that real-world meaning making is multimodal, involving the coordination of verbal and non-verbal behavior with the elements of a shared culturally meaningful setting for action (Goodwin, 1994; Goodwin, 2000; Hollan, Hutchins, & Kirsh, 2000; Hutchins, 1995).

Video data collected in flight instruction in light business jets show how flight instructors and students use their bodies to conceptualize and communicate about the spatial situation of the airplane. Gestural resources in multimodal communication are likely to be especially important when instructor and student do not share native language or culture (Hutchins, Nomura, & Holder, 2006; Nomura & Hutchins, 2007).

Methods

The setting

Through an agreement with a flight school located in Southern California we made video and audio recordings of flight training. In this paper we analyze four training flights conducted in two different light corporate jets, the Cessna Citation I (Model CE-500), and the CitationJet (Model CE-525). The training curriculum included four days of ground school covering the aircraft systems and performance, flight maneuvers, cockpit resource management, and instrument approach flight profiles. The ground school was followed by three days of in-aircraft flight training, and finally, a pass/fail check ride administered by an FAA Designated Pilots Examiner (DPE).

Student # 1 was a 30 year old male ab-initio cadet for a major airline based in Korea. He came to the US earlier for his elementary flight training where he accumulated 256 hours of flight time flying small single-engine and multi-engine piston aircraft. That flight training was conducted at another flight school and was unrelated to this training event. He had a small amount of experience in the Boeing 737 simulator. He held a Commercial Pilot Certificate with Multi-Engine and Instrument Ratings. Student # 2 was a 28 year old French woman who had been working as a flight instructor in Normandy for about one year. She held a Commercial Pilot Certificate with Multi-Engine and Instrument Ratings, as well as a Flight Instructor Certificate. Her total aircraft time was approximately 1935 hours and she had no prior jet experience. Student # 3 was a 60 year old American male, recently retired from the position of captain flying the B737 for a US-based airline. As a career airline pilot, he accumulated approximately 20,000 hours of flight time, and held Airline Transport Pilot and Flight Instructor certificates with ratings in numerous big jets. Two flight instructors, one 38 year old male (Instructor M) and one 34 year old female (Instructor F) conducted the observed training. Instructor F holds an Airline Transport Pilot Certificate and holds two jet type ratings, CE-500 and CE-525(S). Instructor M holds an Airline Transport Pilot Certificate and the following type ratings: AV-L39, CE-500, CE-525(S), CE-510(S), HS-125, LRJET, and WW24. Both instructors hold Multi-Engine Flight Instructor Certificates.

During all four training flights, the students occupied the left (Captain's) pilot seat and the instructor occupied the right (First Officer's) pilot seat. The researcher/observer operating video and audio equipment was also on board for all the flights. On some flights other students and the second flight instructor were also on board. The flights were planned as a standard first flight in a jet. The lesson plan included extended pre-takeoff checks, normal takeoff and climb followed by air work including steep turns, stall recoveries, unusual attitude recovery, then vectors to an instrument approach utilizing an Instrument Landing Systems (ILS). For some flights additional approaches were briefed and flown. The students hand-flew the entire flight with the exception of momentarily handing off controls of the aircraft to the instructor during transitions from one maneuver to the next.

All training flights were approximately 2 hours in duration. The role of the flight instructor was to train the student to perform the flight maneuvers and approaches to ATP standards. In addition to teaching and correcting errors, the instructor also acted as a co-pilot and performed duties such as communicating and coordinating with ATC, running checklists, briefing approaches, tuning navaids, and checking weather. The training fights were conducted under IFR in VFR conditions, with few exceptions when low-level costal stratus mandated actual IFR flight. The student and instructor communicated to each other using the airplane intercom system and both pilots wore standard corporate headsets. The instructor communicated with ATC over the push-to-talk system. The cross-cultural and cross-linguistic aspects of working with foreign students are not typical of the majority of flight training conducted in the US, but we expect it to become much more common as ab-initio programs expand abroad. Overall language differences between students 1 and 2 with their instructors were evident while student #3 was a native English speaker and had no difficulties communicating with both instructors.

Because of the dynamic environment of in-aircraft jet training some variables cannot be controlled, such as weather, ATC workload, the congestion at the airport, more or less chatter on the frequency, other airplanes practicing maneuvers and approaches that lead to more traffic calls, and mechanical problems with the airplane. Although the weather was generally mild throughout all training flights, sometimes IFR clearances or deviations from the original flight plan were necessary. For example, during the flight with student #1 and instructor F, the student had just completed the stall series when the instructor noticed that while it was possible to select the landing gear down and locked, it appeared not to lock in the "up" position. This unanticipated hydraulic problem put an end to the practice air work. It was decided to return to the airport with the gear down and locked. Thus, the approach vectors and ILS approach phases were normal except that the timing of gear extension was disrupted by the fact that the gear was already down.

Data Collection Procedures

The video data were collected with an apparatus we call the "HatCam" which consists of a small 150° field-of-view camera mounted in the brim of a ball-cap and feeding its video signal to a digital video cassette recorder. The HatCam was worn by the instructor and provides a good image of the cockpit environment as shown in figure 1. To record the audio, lapel microphones were clipped to the shirts of both the instructor and the student. Cockpit noise is low enough that the participants can speak in normal voice levels. Air Traffic Control radio frequency was routed to an overhead speaker. The audio signal was recorded onto a digital audio recorder. After the flight the video data was digitized and the audio data was synchronized and added as a track on the video.

Data Analysis Procedures

The analysis began with careful review of the videos. We selectively transcribed the videos, noting all instances of constructions of representations of spatial relations by either instructor or student. To define the boundaries of events, we focused on the representation of relationships of the airplane to spatial or conceptual entities that are elements of situation awareness. We produced a selective transcript table that included all such events. To generate quantitative measures, we defined a set of attributes of the events and then coded every event for the chosen attributes. We coded the primary author of the representation, the referent(s) of the representation, and the resources used to create the representation. Referents were coded into one of two main categories: **performance targets,** which include headings, altitudes, speeds, vertical speeds, and aircraft attitudes, and **geographic features,** which include terrain, landmarks, waypoints, airways, localizer, glideslope, runway environment, and runway and taxiway centerlines. We also coded the resources that were used to create the representation. The attributes here were **verbal** resources, which includes spoken words; **non-verbal** resources, which includes gestures, body orientation, eye gaze and head movement and consequential actions (Segal, 1994); **displays**, which includes flight instruments and documents that were coordinated with the representation; and any **objects** or conditions outside the airplane that were coordinated with the representation. We also coded gestures as **iconic** (representing a spatial concept) and **indexical** (directing attention to objects or events). We noted and coded representations that were initiated by either pilot. We segmented the flights into phases: taxi-out, preflight, takeoff & climb, air work, approach vectors, ILS approach, landing, and taxi-in. We computed intermediate sums for the attributes across each phase of flight separately. It should be kept in mind that attributes of events are not mutually exclusive categories. For example, the vast majority of events make use of both verbal and non-verbal resources.

Figure 1. In this view from the HatCam, the instructor shows the student his pitch target for rotation during the takeoff.

Results

Frequency Counts

Table 1 shows the participants in each flight, the aircraft used, the number of spontaneously created representations of spatial relations, and the duration of the flight in minutes. On average, the participants create more than 4 spatial representations per minute.

Table 1. *The flights.*

Flight #	Student #	Instructor	Aircraft	Representations	Duration
1	1	F	CE-500	189	64
2	2	M	CE-525	412	87
3	3	M	CE-500	240	75
4	3	F	CE-500	312	40
Total				1153	266

Additionally, every video contained a section or sections where either the video or audio feed was lost temporarily. Therefore, the number of spatial representations is undoubtedly underreported within this study. We estimate that these events account for fewer than half of the communicative events generated in the flight overall.

At the beginning of the paper we noted that real-world meaning making tends to be multimodal in the sense that more than one expressive mode is utilized. Examining our data we see that 727 of the 1153 representations spontaneously created by instructor and student utilize a combination of verbal and non-verbal resources. This constitutes 63% of events and demonstrates the importance of *multimodality* of conceptualization in this activity. Every event that utilizes non-verbal resources relies on the fact that the two actors' bodies are co-present in a culturally meaningful space. In our corpus, that is 772 of 1153 events. This is 67% of the events and this large fraction indicates how fully *embodied* the activity is. 910 out of 1153 or 78% of the events incorporate an object or event that is present in the visible environment as an element of the representation. This demonstrates how profoundly *situated* the activity is. The majority of the spontaneously generated representations of spatial relations in this flight are multimodal, embodied, and situated.

Instructor/Student comparisons

Across all phases of flight, instructors make more than 3 times as many representations of spatial relations as students 877/266. This is probably driven by three factors: 1) The instructor role comes with an expectation of creating more representations of everything, 2) the demands of flying the airplane on cognitive resources make it more difficult for students to create representations, and 3) the creation of verbal representations is more costly in cognitive resources for the two foreign students than for the one American student.

Resource limitations

The ratio of student to instructor production of spatial representations was higher while the plane was on the ground vs. in the air. Across all flights, students created 37% of the total spatial representations while on the ground vs. 24% while in the air. It is not surprising that the student's rates of production of representations that use the body decreased from the pre-flight to the flying phases as student's hands were occupied controlling the airplane when hand-flying. However, there was also a sharp decrease in verbal representations as well. In fact, in flight student #1 incorporated verbal resources in the creation of spatial representations 13 times and incorporated non-verbal resources 17 times. Thus, even though the hands were occupied, the student still used his non-verbal resources more than verbal resources in flight. The typical non-verbal event here was a consequential action (setting the heading bug) rather than a gesture. We believe that this is due to the increased workload for the student in flight and to the fact that composing a meaningful action is cognitively less expensive for this student than composing an utterance in English.

Limitations on cognitive resources need not be a one-way causal route from body to mental. Ebbatson, Harris, & Jarvis (2007)investigated pilots attempting to assess crosswind components from the information provided by ATC. They report, "the mental arithmetic associated with calculating the runway crosswind impaired flying performance." Competition among tasks for cognitive resources was an issue for instructor F as well as for student #1. While on the final approach, the approach controller handed the airplane off to the tower controller. While the instructor was waiting for a pause in the tower radio traffic to check in with the tower, the student asked her if 600 feet per minute is a good descent rate to track the glideslope. It's a good question, but because she was waiting for a chance to speak, she did not want to begin a verbal exchange with the student. She pointed to the glideslope indication (on G/S), then to the vertical speed (-600 fpm), then back to the glideslope indication. In the post flight review of the video, the instructor commented that the student choosing to ask her a question while she was waiting to check in, with the tower controller reveals that student's lack of general situation awareness. The reduction in the rate of production of speech in flight might be more than simply a matter of resource limitations. It could also be that with the hands occupied on the yoke and thrust levers, talk is less likely. Rauscher, Krauss, and Chen (1996) found that "preventing speakers from gesturing adversely affected their ability to produce fluent speech when the content was spatial." Other studies show that restricting gesture reduces verbal abilities (Frick-Hornby & Guttentag, 1998; Rime, Schiaratura, Hupet, & Ghysselinckx, 1984).

Relations among gesture, space, and talk

By far the most common type of event observed in this activity arises when the instructor combines a verbalization with a gesture to a flight instrument. In these cases, the meaning of the event is established by the mutually constitutive relations among talk, gesture, and local space (Hutchins & Palen, 1997). A clear example of this happened during student #1's takeoff roll. The instructor called, "V1, and Rotate!" After a two-second pause in which the student applied some back pressure to the yoke, but not enough to lift the nose wheel off the runway, the instructor said, "Pull back. Ten degrees nose up." While saying, "Ten degrees nose up" she pointed to the student's attitude indicator to show him where to look to see the desired pitch attitude (See figure 1). The meaning of this indexical gesture, the pointing as a director of attention, is established by its coupling to the environment (Goodwin, 2000) in this case by its placement on the attitude indicator. The prevalence of indexical gestures that direct attention in and around the airplane is not surprising. Knowing where to look, when to look, and what to see when looking, are key skills in flying any airplane.

Discussion

The value of our two-airplane, three-student, two-instructor analysis is not to make claims about the differences among the cases. Rather, we are impressed by the overwhelming similarity across the cases. No matter the gender or nationality of the instructor or student, no matter the level of flying experience (250 hrs vs. 2000 hrs vs. 20000 hrs), no matter the airplane used, no matter the sort of flight (first flight vs. second, where the second includes a V1 cut and more instrument approaches), all participants (instructors and students) made extensive use of their bodies in constructing and reasoning about SA. Of course, there may be differences across these independent variables in the fine details of the way the body is used, but that is not something we can demonstrate with the data in hand. Everyone believes that SA is an important factor of every flight for every pilot worldwide. We agree. But we also believe that the establishment, expression, and maintenance of SA are embodied cognitive processes. This shows up also in the observations made in the Boeing sponsored project conducted by the first author of this paper. That project, has made in-flight observations of 70 segments of revenue flight with five different airlines based in 4 nations (Japan, New Zealand, Brazil, Mexico), operating in four languages (Japanese, English, Portuguese, Spanish), flown by a total of 64 pilots. Video data on an additional 26 pilots from the participating airlines flying a total of 50 hrs in high-fidelity flight simulators has been collected and analyzed. Those data show a similar pattern of use of the body to construct SA (Hutchins & Nomura, in press; Hutchins, Nomura, & Holder, 2006; Nomura & Hutchins, 2007). Still, embodiment does not yet have a central role in the understanding of SA. We suspect that one reason for this is that the uses of the body are largely unconscious – both in production and in interpretation. As such it may seem unlikely that interventions could change people's behavior much. Furthermore, making the role of the body visible to analysis requires special equipment and techniques.

We are fully aware of the limitation of a study based on a small sample of complex events. In spite of the uncontrolled sources of variability noted in the methods section, the growing literature in embodied, situated, distributed cognition leads us to believe that many of our observations will generalize to other settings in which two or more actors jointly engage in consequential activity. Our analysis demonstrates that this type of flight training is profoundly embodied, multimodal, and situated. We believe that this is a central fact about flight instruction as it occurs in actual flight. Much of flight training concerns the domestication of attention: knowing where to look and what to see when looking there. This is an embodied skill. The representations we observed were tightly integrated with the resources provided by bodies located side-by-side in a complex material setting. These representations fluidly integrated observable with imagined aspects of spatial situation. The meanings of the representations emerged from the mutual elaboration of bodily motion, talk, and local space. This process of mutual elaboration supported the disambiguation of partial representations – including those that may have been due to limited competence in the English language. The observed representations of spatial situation integrated multiple, overlapping, fluidly shifting frames of reference.

In the near future, we are interested in tracking changes in these patterns through the training process. Will a student generate more representations of spatial situation as competence increases? Will the kinds of representations used by student or instructor change? Are certain gestures consistent among different instructors? We are also interested in the relationship between flight instruction as it occurs in an actual airplane and flight instruction in a high-fidelity simulator. Our observations here lead us to believe that with two students in the pilot seats and the instructor at a simulator operator station behind the students, the composition of representations of spatial relations will be very different. We expect to be able to do a comparison study in a high-fidelity simulator in the coming year.

Our preliminary findings lead us to believe that it will not be possible to understand teamwork, situation awareness, decision making, or communication without attending to how people use their bodies in the flight deck.

Acknowledgements
Funding was provided by the National Science Foundation (Award #0729013 DHB: A Multiscale Framework for Analyzing Activity Dynamics) and by the Boeing Commercial Airplane Company.

References

Aviation Safety Improvements Task Force: Human Factors Working Group. (2009). *DoD Human Factors.* Retrieved March 6, 2009, from http://www.safetycenter.navy.mil/hfacs/
Banbury, S., Dudfield, H., Hoermann, H.-J., & Soll, H. (2007). FASA: Development and validation of a novel measure to assess the effectiveness of commercial airline pilot situation awareness training. *The international journal of aviation psychology , 17* (2), 131-152.
Bowers, C., Jentsch, F., & Salas, E. (2003). CRM Training in the new environment: Challenges for flight crew training after September 11, 2001. *International Journal of Applied Aviation Studies , 3* (1), 61-78.
Curry, I., Estrada, A., Grandizio, C., & Erickson, B. (2008). Drift cues from a tactile belt to augment standard helicopter instruments. *International Journal of Applied Aviation Studies , 8* (1), 74-87.
Ebbatson, M., Harris, D., & Jarvis, S. (2007). Crosswind landings in general aviation: a modified method of reporting wind information to the pilot. *The international journal of aviation psychology , 17* (4), 353-370.
Endsley, M. (2000). Theoretical underpinnings of situation awareness: a critical review. In M. Endsley, & D. Garland (Eds.), *Situation awareness analysis and measurement.* Mahwah, NJ: Lawrence Erlbaum Associates.
Frick-Hornby, D., & Guttentag, R. (1998). The effects of restricting hand gesture production on lexical retrieval and free recall. *The American Journal of Psychology , 111* (1), 43-62.
Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics , 32*, 1489-1522.
Goodwin, C. (1994). Professional Vision. *American Anthropologist , 96* (3), 606 - 633.
Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on human-computer interaction: Special issue on human-computer interaction in the new millennium , 7* (2), 174-196.
Hutchins, E. (1995). *Cognition in the Wild.* Cambridge, MA: MIT Press.
Hutchins, E., & Nomura, S. (in press). The multi-author construction of multmodal utterances. In C. Goodwin, J. Streek, & C. Lebaron (Eds.), *Multimodality.* Hillsdale, NJ: Cambridge University Press.
Hutchins, E., & Palen, L. (1997). Constructing meaning from space, gesture, and speech. In L. Resnick, R. Saljo, C. Pontecorvo, & B. Burge (Eds.), *Discourse, tools, and reasoning: essays on situated cognition* (pp. 23-40). Heidelberg, Germany: Springer-Verlag.
Hutchins, E., Nomura, S., & Holder, B. (2006). The ecology of language practices in worldwide airline flight deck operations: the case of Japanese airlines. *Proceedings of the international conference on human-computer interaction in aeronautics*, (pp. 90-96). Seattle, WA.
Lam, T., Mulder, M., & van Paassen, M. (2007). Haptic Interface For UAV Collision Avoidance. *The international journal of aviation psychology , 17* (2), 167-195.
Nomura, S., & Hutchins, E. (2007). The multimodal production of common ground understandings in intercultural flight training. *Proc. of the 14th International Symposium on Aviation Psychology*, (pp. 475-480). Dayton, OH.
O'Connor, P., Campbell, J., Newon, J., Melton, J., Salas, E., & Wilson, K. (2008). Crew Resource Management Training Effectiveness: A Meta-Analysis and Some Critical Needs. *The International Journal of Aviation Psychology , 18* (4), 353-368.
Rauscher, F., Krauss, R., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science , 7* (4), 226-231.
Rime, B., Schiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion , 8* (4), 311-325.
Seamster, T., Boehm-Davis, D., Holt, R., & Shultz, K. (1998). *Developing Advanced Crew Resource Management (ACRM) Training: a training manual.* Office of the Chief Scientific and Technical Advisor for Human Factors. Washington DC: FAA.
Segal, L. (1994). Actions speak louder than words: how pilots use non-verbal information for crew communications. *Proc. of the Human Factors and Ergonomics Society Annual Meeting: Aerospace Systems*, (pp. 21-25).
Wickens, C., Small, R., Andre, T., Bagnall, T., & Brenaman, C. (2008). Multisensory enhancement of command displays for unusual attitude recovery. *International journal of aviation psychology , 18* (3), 255-267.

ARE WE GETTING THE MESSAGE ACROSS? HUMAN FACTORS AND SYSTEM SAFETY EDUCATION – WHAT IMPACT HAS IT HAD?

Sue Burdekin
University of New South Wales at the Australian Defence Force
Canberra, Australia

Education in human factors and systems safety has been incorporated into aviation degree programs at university level for many years. However, there has been little research to measure empirically the impact that this education has had on safety outcomes in the field when the students have completed their degrees. A study is presently being conducted involving graduates from the University of NSW at the Australian Defence Force Academy in Australia who had obtained the degree of Bachelor of Technology in Aviation. Following their graduation, these students had been posted to various flying squadrons within the Australian Defence Force in an operational capacity. Research methods included attitude and knowledge surveys, as well as a statistical and qualitative comparison of academic results and military flying training performance with a control group of other pilots who had not completed this education or had joined the Australian Defence Force in a direct entry capacity. The preliminary results are very encouraging.

Many universities offering a bachelors program in aviation have incorporated the study of human factors and systems safety into their degree. Although pilots are regularly tested in terms of piloting skills and emergency responses it has been more challenging to determine if education and training courses in human factors and system safety have had an impact on the performance of pilots when they become operationally active. Measures of attitudes towards education and training are frequently taken immediately after courses are delivered and whilst student satisfaction is an important component of learning, these measures do not provide evidence of the application of this knowledge in a working environment. In order to tap into how these courses have affected individuals a more comprehensive approach to evaluation is needed.

Kirkpatrick and Kirkpatrick (2006) first developed the 'four level model' of training program evaluation in 1959 and it is used widely in the education and industrial sectors. The model focuses on evaluating student's *reaction*, *learning*, *behaviour* and *results*. A meta-analysis of the relations among training criteria (Alliger, Tannenbaum, Bennett, Traver & Shortland, 1997) found that student reactions could be further augmented into *affective reactions* and *utility judgements.* The latter refers to job relevance, that is, can the trainee provide evidence as to how the training has influenced the way they now approach or perform their work? The measure of *Learning* can also be further broken down into *immediate knowledge* and *knowledge retention*. *Behaviour* refers to on the job performance, that is, has the education and training transferred from the classroom to the workplace, for it is application to the job that, in most cases, defines training success (Allinger et al, 1997). *Results* are defined by organisational impact, for example, productivity gains, customer satisfaction, and could be extended to employee morale. In operational aviation it has been difficult to determine the impact that human factors and system safety education and training has had on pilots because either comprehensive testing was considered to be too difficult or those organisations who have comprehensively evaluated their personnel have not made the results public.

In 2001 a new undergraduate degree program was offered to military officer cadets who had successfully completed 15 hours of flight screening and were selected to undergo training as pilots within the Australian Defence Force (ADF). This degree, the Bachelor of Technology in Aviation (BTech(Avn)) is based on aeronautical engineering together with compulsory, core educational courses in aviation human factors and system safety. Additionally, the students complete a major research project and many choose to conduct experimental research into a human factors related topic. The officer cadets enrolled in this degree spend 26 months at the University of New South Wales at the Australian Defence Force Academy (UNSW@ADFA) concurrently undergoing officer training and then go on to complete the flying training component of their degree prior to graduation (Harrap, Burdekin & Lewis, 2007). Upon graduation they are assigned, according to their operational aptitudes and abilities, to various flying squadrons from airlift to fast jet capabilities where they work alongside pilots who may not have received any education or training in human factors and system safety or who may have had minimal exposure to these concepts.

The aim of the present study was two fold. Firstly to evaluate the BTech(Avn) program and secondly to determine the impact that human factors and system safety education has had on graduate pilots within the ADF flying squadrons.

Method

Participants

Postgraduate students from the UNSW@ADFA BTech(Avn) degree and the ADF Advanced Flying Training School were invited to partake in the study. These graduates are now actively flying in a variety of operational squadrons within the ADF including fast jets, heavy lift transport, maritime and rotary wing platforms. An equal amount of ADF pilots currently flying in the same operational squadrons but, who have not completed the BTech(Avn) degree were also invited to participate and these pilots acted as a control group.

Design

The study was guided by the Kirkpatrick (2006) approach to education and training evaluation and further broken down to compare results by temporal distance. Four levels of review were designed into the evaluation materials – *reaction, learning, behaviour and results* with reaction and learning containing two further levels of distinction – affective reactions and utility judgements; and immediate knowledge and knowledge retention. To measure affective reactions towards the human factors and system safety education, participants were asked to complete a survey entitled the Course Experience Questionnaire (CEQ) which is used in Australia to link institutional performance with Government funding. The CEQ is designed to measure student satisfaction with a degree program post graduation. In this study these results would be compared with the average satisfaction levels that students expressed immediately after completing courses in human factors and system safety subjects. To measure utility judgements participants were asked to complete the Flight Management Attitude Questionnaire (FMAQ), which is widely used by civil aviation organisations and the ADF to measure safety culture; and, a questionnaire which gave participants the opportunity to specifically rate their satisfaction with, and comment on, specific aspects of the BTech(Avn) program.

A knowledge audit was designed to assess how much human factors and system safety knowledge was retained by the participant and how he/she applied this knowledge in the workplace. These results would be compared to the average results achieved for these questions whilst the experimental participants were still at university. In order to evaluate behaviour, the flying supervisors of both the BTech(Avn) graduates and the control group of pilots were asked to comment on the behaviour and performance of the participants using a modified NOTECHS assessment. To measure results the author obtained information relating to any commendations, prizes or awards that the participants had received and questioned the flying supervisors concerning the participant's performance and contribution to the squadron.

Procedure

The study was divided into two stages. Stage one involved the evaluation of the BTech(Avn) program and two questionnaires – the CEQ and the BTech(Avn) Evaluation Survey – were distributed to the BTech(Avn) graduate participants only. To ensure anonymity and to encourage critical evaluation an independent Internet survey company was used.

Stage two of the study was mainly concerned with determining the impact that human factors and system safety education has had on pilots and involved the researcher visiting the ADF squadrons where former BTech(Avn) students are now based. The participating graduates and the control participants were asked to complete a knowledge audit which tested their awareness of human factors and system safety information. Furthermore, the questions asked participants to describe how they would apply or expect these concepts to be applied in their workplace. Both groups were then asked to complete the FMAQ. No names were requested on either document and the only distinguishing feature was the separation of experimental and control groups responses.

The operational flying supervisors of these graduates and control pilots were asked to report their workplace assessment of the participants using the modified NOTECHS questionnaire. Data was also gathered on any prizes and awards that participants had obtained during their short careers as military pilots.

Results

The results so far have been very encouraging. To date 69% of BTech(Avn)graduates have submitted returns for stage one of the study. The researcher is expecting this response to increase, as several graduates are currently serving overseas and have indicated they were unable to respond due to the present operational tempo. Table 1 shows the evaluation rating of the BTech(Avn) degree program.

Table 1. *Course Experience Questionnaire Summary*

Please select a rating based on our satisfaction with the Bachelor of Technology in Aviation program						
	Strongly disagree				Strongly Agree	Rating Average
The staff put a lot of time into commenting on my work	0.0%	3.8%	30.8%	53.8%	11.5%	3.73
The teaching staff normally gave me helpful feedback on how I was going	0.0%	0.0%	26.9%	62.9%	3.8%	3.77
The program helped me develop my ability to work as a team member	0.0%	7.7%	38.5%	50.0%	3.8%	3.50
The teaching staff of this program motivated me to do my best work	0.0%	11.5%	38.5%	50.0%	0.0%	3.38
The program sharpened my analytical skills	0.0%	3.8%	19..2%	57.7%	19..2%	3.92
My lecturers were extremely good at explaining things	0.0%	4.0%	20.0%	64.0%	12.0%	3.84
The teaching staff worked hard to make their subjects interesting	0.0%	0.0%	15.4%	57.7%	26.9%	4.12
The program developed my problem-solving skills	0.0%	11.5%	26.9%	57.7%	3.8%	3.54
The staff made a real effort to understand difficulties I might be having with my work	0.0%	3.8%	42.3%	50.0%	3.8%	3.54
The program improved my skills in written communication	0.0%	11.5%	19.2%	61.5%	7.7%	3.65
As a result of my program, I feel confident about tackling unfamiliar problems	0.0%	7.7%	50.0%	34.6%	7.7%	3.42
My program helped me to develop the ability to plan my own work	0.0%	3.8%	30.8%	57.7%	7.7%	3.68
Overall, I was satisfied with the quality of this program	0.0%	0.0%	15.4%	61.5%	23.1%	4.08

When given the opportunity to elaborate on various aspects of the human factors and systems safety educational subjects they had studied during their degree respondents gave insightful answers, a summary of which follows:

Figure 1. *Question: Can you nominate a particular lecture or piece of human factors or system safety information that has influenced or impacted upon your work or flying and describe how?*

"I've had situations where I've recognised the "holes in the cheese" line up, and subsequently raised concerns, all due to what I learnt in this degree."

"The courses allowed you to be aware of safety and human factors issues so that you may realise a potential problem that you may not have otherwise realised was developing."

"The understanding of the impact of fatigue has been particularly relevant for me. It is a constant factor and understanding that it has an insidious nature has helped me make decisions to not fly when it has been appropriate to do so. I also think that having a good understanding of the Reason Model has made me more aware of potential risks and allowed me to take appropriate action."

"The fact that the culture of an organisation can have such a large impact on aviation safety and is generally the root course of problems is something I think about often. I still look out for systemic problems in my work place regarding procedures and products."

"I remember being taught about visual illusions at night, in particular, the "mothball effect" - how some pilots don't flare because they fixate on the lights on the runway. I had a similar situation where I just stared at the runway, and struggled to keep the aircraft lined up on finals on a particularly dark night (all we could see were the lights). I remembered what I was taught, and went against what "felt" natural, and didn't overcorrect the aircraft. I then told myself not to fixate on the runway and we had a safe landing."

"Despite a massive push on safety from all levels, aviation safety is rarely a perfect system with many problems observed at many levels. The big safety focus from the courses gives me the confidence and ammunition to argue against particular procedures, problems etc."

Another view from a participant indicates that he/she may not have yet realised that safety is an individual responsibility.

Figure 2. *An alternative opinion*

"[The courses] are useful for a bigger picture view of aviation safety but at the junior level have not been particularly applicable. The higher level aviation safety decisions are generally made without input from members as junior as myself. However I think the training will be useful in the future when graduates of our program are more able to influence decisions made in regards to aviation safety."

Discussion

Stage one of the study has thus far provided some extremely useful data. The satisfaction – affective reaction - rating averages for every question asked on both the program CEQ and the human factors and system safety course questionnaire have been in the upper quartile. These results compare favourably to other university programs. However, the answers to the open response questions provide a more comprehensive indication of how the information studied during these courses has impacted upon the careers of the pilot graduates – utility judgements. The comments that were highlighted in this preliminary report include: sound knowledge of the organisational incident and accident; the insidious nature of fatigue; organisational and safety culture issues; and, human information processing and limitations. It is particularly rewarding to hear that at least one graduate believes he/she is sufficiently educated in the subject of aviation safety to feel confident enough to argue a safety issue with a superior officer in a military environment.

Stage one data collection is on-going due to many potential participants currently serving in humanitarian efforts, border protection duties and war zones overseas. Stage two involves visits to military bases around Australia in order to brief pilots and supervisors, and administer surveys. This process is complete and the data are currently being analysed.

Expected Outcomes

It is anticipated that there will be several outcomes from this research. Firstly, the results will indicate the level of transfer of education and training from the classroom to the cockpit and this result will impact upon and shape the future development of content and method of delivery for the UNSW@ADFA BTech(Avn) program. Secondly, the study will compare the attitudes, knowledge, behaviour and results of pilots who have received education and training in aviation human factors and systems safety with those of a control group of ADF pilots who have not. The data from this study will indicate whether human factors and system safety education and training has assisted in developing in ADF pilots a healthy attitude towards and understanding of safety issues, and a greater appreciation and application of safety concepts and knowledge in their flying skills and workplace performance. Evidence of the development of these attributes would help to attract support for future education and training in this field.

References

Alliger G.E., Tannenbaum S.I., Bennett W., Traver H., & Shotland A. (1997) A Meta-analysis of the relations among training criteria. *Personnel Psychology,* 50, 341-358.

Harrap M., Burdekin S. & Lewis R., (2007) The development of the Bachelor of Technology in Aviation degree program at the Australian Defence Force Academy. In *Proceeding of the Australian Association of Engineering Education Conference.* Melbourne. AAEE

Kirkpatrick D.L., & Kirkpatrick J.D., (2006) Evaluating training programs: the four levels 3rd Ed. Berrett-Koehler Publishers Inc., San Francisco.

Kraiger K., Ford K.J., Salas E (1993) Application of cognitive, skill-based and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology.* 78, 311-328.

Acknowledgement

This study is supported by a Department of Defence UNSW@ADFA internal research grant. The author would like to thank GPCAPT Christopher Sawade, Officer Commanding Air Training Wing and LTCDR Rick Sellers, Deputy Director Human Factors, Defence Directorate of Aviation and Air Force Safety for their assistance.

RESULTS FROM THE FIRST FAA INDUSTRY TRAINING STANDARDS (FITS) COMMERCIAL PILOT TRAINING COURSE – A STUDENT'S PERSPECTIVE

Paul A. Craig, Wendy S. Beckman, Mark Callender, Steve Gossett, Wayne Dornan
Middle Tennessee State University
Murfreesboro, Tennessee

In January 2008, students in the Professional Pilot program at Middle Tennessee State University (MTSU) began training using an FAA Industry Training Standards (FITS) Commercial Pilot curriculum. The course was accepted as a scenario-based and competency-based curriculum by the FAA as a "Special Curricula." Students entered the FITS Commercial Pilot course having completed their Private Pilot Certificate and Instrument Rating. Some of the students had completed a FITS accepted and FAA approved combined Private/Instrument course, while others had completed separate Private Pilot and Instrument courses. The first thirty-three students completed the course with lower total flight times than students have historically experienced in conventional training paradigms. Although the students completed the course with less flight time, they nevertheless met FAA standards and passed the FAA Commercial Pilot Practical Exam on their first attempt 88% of the time. This paper presents an analysis of student impressions of this new training methodology.

Continuing Research

In 2004, the Federal Aviation Administration (FAA) implemented the FAA-Industry Training Standards (FITS) program (Federal Aviation Administration, 2004; Glista, 2003), based on recommendations from the 1998 FAA "SAFER SKIES" initiative. Flight training curricula developed using the FITS tenets place major emphasis on: aeronautical decision making skills, risk management, situational awareness, and single pilot resource management using real-time flight scenarios (Ayers, 2006; Glista, 2003). Since 2004, Middle Tennessee State University (MTSU) has been involved in the implementation and testing of FITS-approved Private and Instrument curricula (Craig, et al, 2005a, 2005b; Dornan, et al, 2007b, 2007, 2006; Beckman, et al, 2008). The next step in the development of a completely scenario-based flight training program at MTSU was the creation, approval, and implementation of the first-ever FITS-approved Commercial Pilot Training Course (FAA, 2007). In January of 2008, the first cohort of MTSU students were enrolled in this Training Course. This paper documents an evaluation of the course by the first students to complete training using the FITS-approved Commercial Pilot course.

Commercial Course Requirements

Traditional training for obtaining a Commercial Pilot Certificate is conducted in one of two ways. It is either done in compliance with the laws of the Code of Federal Regulations Part 61 or Part 141. Part 61.129 requires that a Commercial Pilot applicant must have logged at least 250 hours of flight time. Appendix D of Part 141 outlines the requirements for a Commercial Pilot applicant at an FAA Approved School. Appendix D first requires the applicant to have logged at least 35 hours of flight time toward the Private Pilot Certificate, plus at least 35 hours toward the Instrument Rating, plus an additional 120 hours toward the Commercial Pilot Certificate – a total of at least 190 hours. The FITS Commercial Pilot course that was approved at MTSU does not have a minimum flight time requirement. Students progress through the course and ultimately complete the course when they meet the flight proficiency standards, regardless of how many or how few flight hours they accumulate while in the course. This competency-based system is a departure from the traditional flight hour based system and is made possible by the Special Curricula rule of Part 141.57. That law states that schools that were previously FAA approved can, "conduct a special course of airman training for which a curriculum is not prescribed in the appendixes of this part, if the applicant shows that the training course contains features that could achieve a level of pilot proficiency equivalent to that achieved by a training course prescribed in the appendixes of this part or the requirements of part 61 of this chapter." MTSU sought this approval and was granted permission to conduct the FITS-approved Commercial Pilot course. One of the features of the course was the replacement of an hours-based requirement with a proficiency-based requirement.

Methodology

On January 18, 2008, the authors received approval to collect data pertaining to the new Commercial Pilot course from the MTSU Institutional Review Board (IRB #08-151). The project, titled "Evaluation of the FAA/Industry Training Standards (FITS) Commercial Pilot Curriculum Implementation at the MTSU Flight School," authorized the authors to examine student records and collect survey data from the students who complete the FITS Commercial Pilot Course. The survey instrument used in the study was first validated by Graduate Faculty of the MTSU Aerospace Department and the MTSU Office of Research and Sponsored Programs. The survey had two parts. The first was a series of Likert-scale type questions; the second, a series of open-ended questions. The surveys were sent to the students, via email attachment, after they completed the FITS Commercial Pilot course and had passed the Commercial Pilot Practical Test which was administered by an FAA Designated Pilot Examiner. The students were asked to return the completed survey via email attachment. In the early fall of 2008, thirty-three survey requests were made, and twenty surveys were returned. When the surveys were returned they were immediately coded with a number and the name removed. No link is now possible between the students and their survey responses.

Results

A survey question asked of the students was, "How many flight hours did you have on the day you passed the Commercial Pilot Practical Test?" Seventeen students responded to that question. For those seventeen students, the average flight time logged when they became Commercial Pilots was 182.7 hours.. The students were also asked about their flight training prior to beginning the FITS Commercial Pilot course. Ten students had started the FITS Commercial Pilot course immediately following the completion of the FITS combination Private and Instrument course that was taught at MTSU. For students who moved directly from one FITS approved course (Private/Instrument) to the next (Commercial), the average flight time logged when they became Commercial Pilots was 155.2 hours. Another ten students had not completed any FITS training before beginning the FITS Commercial Pilot course. For the students who had no prior FITS training experience before beginning the FITS Commercial Pilot course, the average flight time when they became Commercial Pilots was 217.4 hours.

Students were next asked about their use of flight training devices (FTD) during their Commercial Pilot training. The average FTD time among the students who completed the FITS Commercial Pilot course was 13.5 hours, although the course does not specifically require any FTD time. Any FTD time that was logged was for student proficiency and was considered an extra, outside the course. The range was wide, with one student logging 24.7 FTD hours while two students logged no FTD time at all. The variance was mainly attributed to the fact that instructors utilized the FTDs based on their own preference and FTD availability.

While we were naturally interested in the flight time of students completing the curriculum, the primary goal of the curriculum was to produce a pilot well prepared for the challenges they will face after they become Commercial Pilots and operate in the actual flight environment. Several of the Likert-scale type questions were aimed at this evaluating how well the curriculum performed in this area. To the statement, "I have a high confidence level that the FAA Industry Training Standards (FITS) Commercial Pilot course has prepared me well for the next steps in my professional career" ninety-five percent of the students responded with either "agree" or "strongly agree." Only 5% (one student) entered the response of "disagree." To the statement, "I feel that I am now ready for a Professional Pilot job" 80% responded "agree" or "strongly agree." To the statement, "After the FITS Commercial Pilot course, I am now more confident in my aeronautical decision making skills," 95% of the students entered that they either "agree" or "strongly agree."

One of the unique features of the FITS Commercial Course is that it incorporates instrument training and instrument flights, including a solo instrument flight. Traditional Commercial courses have little or no instrument training and as a result, many practical test applicants are no longer instrument proficient on the day they become Commercial Pilots. The authors saw this as a negative effect and attempted to remedy this problem through the design of the FITS Commercial Pilot course. When students were asked to respond to the statement, "I feel that using the FITS Commercial Pilot course helped me maintain my IFR proficiency as opposed to training with VFR maneuvers alone" 100% responded that they "agree" or "strongly agree."

Another goal of the authors was to assess the satisfaction level that the students had with the course. The curriculum design specifically targeted areas that traditionally are problems for students in training, such as lesson cancellations due to weather or equipment availability, and repetitious lessons. Several survey items attempted to determine if the curriculum was effective in targeting these problem areas. To the statement, “The FITS Commercial Pilot course went smoothly and was relatively easy,” 80% responded that they “agree” or “strongly agree.” Twenty percent made the response that they were “neutral.” The students were asked to respond to the statement, “Compared to other training I have had, the FITS Commercial course was enjoyable.” Ninety-five percent of the students indicated “agree” or “strongly agree.” Five percent (one student) said they were neutral on that statement.

A unique feature of the FITS Commercial Pilot course is its flexibility. The lessons throughout the course can “shuffle.” The course has fours “strands” – VFR, Commercial Maneuvers, IFR and Complex, but students and instructors can switch back and forth between the strands to meet the needs of the training environment. If a VFR flight is planned, but IFR conditions are present, the instructor can switch and conduct a lesson from the IFR strand. This greatly reduced training delays and helped the students maintain proficiency. Traditional maneuvers-based, time-based syllabi require the student to complete all of one stage before moving on to the next set of lessons. This practice, however tends to increase lesson cancellations and training delays. When students were asked to comment on the statement, “The ability to “shuffle” lessons between strands was beneficial,” 100% said either agreed or strongly agreed with the statement.

With less average flight time needed to complete the course, less money is needed to become a Commercial Pilot. Since the average logged flight time was less than what otherwise would have been required, the area of cost savings was also an interest to the authors. When the students were asked to respond to the statement, “The cost for me to get the Commercial Pilot Certificate was less because I used the FITS Commercial Pilot course,” 79% responded “agree” or “strongly agree.” Three students responded “neutral,” one student responded “strongly disagree” and one student did not respond to that statement. One student, who had first completed the FITS combination Private and Instrument course, immediately followed by the FITS Commercial course, had logged 133 total flight hours by the day he passed the Commercial Pilot Practical Test. He reported to one of the researchers that, “The FITS Commercial saved me $6,000.”

Open Ended Questions

The survey instrument also had a series of open-ended questions that allowed the students to further and more freely express their opinions. The researchers believed that students who had just completed the FITS Commercial Pilot Syllabus were in the best position to offer suggestions for improvement. The researchers also felt that having the students respond via email would produce more complete responses because students are so familiar with communicating by computer rather than pen and paper. The students did not hold back with their opinions on the open-ended questions. The following tables include representative responses to the open-ended questions.

Table 1: *Student Responses to “What did you enjoy most about the FITS Commercial Pilot course?*

Question: What did you enjoy most about the FITS Commercial Pilot course?
Representative Responses: > “I enjoyed the fact that I could jump around between lessons and not have to wait for a strand check to be completed before moving on. While waiting for a strand check I could begin another strand.” >”The ability to shuffle around lessons was not only beneficial, it kept me cautious about what I was supposed to be doing. In some cases I had only a couple minutes to plan and file a flight, followed by thirty minutes for preflight inspection.” > “I really enjoyed how quick the process went. Allowing me to move at a faster pace, and transition to other lessons without the stage checks right off the top really allowed me to finish my commercial license a lot quicker, and was more enjoyable.” > ”The scenarios, while sometimes very corny, were also more fun than the basic Commercial Syllabi that are available. I happened to also have a copy of the (traditional) Commercial and flights in the FITS Commercial seemed much more enjoyable then the comparable Part 61 lessons. Also I believe the solo IFR lessons are an excellent building block.” >”It was a bit more laid back and not as demanding as the private/instrument combined. Some of the scenarios were also pretty interesting, and going solo IFR was a big confidence booster.” > ”The real-life scenarios which were presented throughout each lesson allowed the application of regulations and piloting skills which were being taught, giving me a better understanding of the material. I especially enjoyed the flexibility with which I was able to complete the course.” > “I liked how the scenarios were based on realistic situations and simulated real world pressure.”

> “The opportunity to fly IMC frequently and fly IMC solo.” > “I enjoyed the dual flights where it was just flying without learning maneuvers. It seemed more like I was flying for hire than learning.” > “The scenario based lessons, and the solo IFR flights and the 500 nm cross country.” > “The ‘missions’ made FITS a little more down to earth and gave a good sense of things that commercial pilots may actually do.” > “I enjoyed the scenario based lessons, as they helped me to realize what I could do with my commercial license, as well as the responsibilities I would assume as a commercial pilot.” > ‘The solo IFR flight.” > “My favorite part was the ability to choose which ‘string’ to work on. I also thought that the different scenarios were highly enjoyable. Instead of just flying and doing maneuvers, you actually got a taste of what real-world flying will be like. Flying the IFR cross country was also fun.”

Table 2: *Student Responses to, “What did you enjoy the least about the FITS Commercial Pilot course?”*

Question: What did you enjoy the least about the FITS Commercial Pilot course?
Representative Responses: > “I did not like the switching between the PA-28 and DA-40. These are two aircraft that handle completely differently. After I got proper landings in the PA-28, I had to transfer my skills back to the DA-40 to pass the VFR strand check. This is greatly due to the shuffling of the lessons. Flying both the PA-28 and DA-40 on a daily basis did help my piloting skill; however, it was a pain at the time.” > “I don’t have any negative remarks about the FITS Commercial Pilot course.” > “It seemed that the first couple of dual missions in the VFR strand dragged on. I understand their purpose in teaching the Commercial exceptions in Part 119 but I believe that in most cases with almost all students that the lessons could be combined or a few of them taken out and perhaps replaced with more solo or cross country flights.” > “Some lessons I felt were pointless, and while the jumping back and forth saved time, I thought it made the strand checks somewhat more difficult.” > “The complex aircraft strand and commercial maneuvers strand should be required to be at the end of the course so the student doesn’t have to re-fly the same flights to regain proficiency for the checkride. > “The maneuvers were repetitive.” > “I do not think there are enough solo flights built into the syllabus.” > “VFR navigation planning for the 500nm XC.” > “The time restraints placed on students during flight planning. I understand why they were used and it helped me with my ability to quickly and accurately plan a cross country but it was tough at times.” > “I enjoyed all of it.” > “Nothing, everything went great!”

Table 3: *Student Responses to, “Do you have any other comments about the FITS Commercial Pilot course that were not covered in previous questions?”*

Question: Do you have any other comments about the FITS Commercial Pilot course that were not covered in previous questions?
Representative Responses: > “Overall, Commercial FITS is a great program. Much like Private/Instrument FITS, however, there are still a few bugs to work out. I feel more cross-countries should be added to the syllabus. Also, I feel there was not enough time in the Piper Arrow (Complex Airplane). The student is in the aircraft just long enough to learn maneuvers and move on to the checkride. Perhaps making the Arrow a mandatory aircraft on a cross-country or two will not only allow the pilot to become more familiar with the aircraft but also allow the student to become used to flying more than just a DA-40 more than 20 miles from the airport.” > “I believed the FITS syllabus was very well put together (both commercial and private/instrument). I am grateful that I received most of my pilot training in just under two years with thousands of dollars saved. I do not feel incompetent as a pilot, nor do I feel that I was “cheated” that I did not accumulate many flight hours during

my training. I feel that I have a good firm grasp on the aerospace industry and my place within it.”
> “I really enjoyed the realism in the scenarios. I believe to become a successful and safe professional pilot, us as students need to be in stressful situations when we have certain deadlines to meet.”
> “The FITS Commercial course is a rather enjoyable experience. As I stated before some of the scenarios are a bit of a stretch but altogether the syllabus has great potential for experience and it is definitely a plus being able to pocket the money that would otherwise go towards a cross country lab. Finding out you got your Commercial in the same amount of time as someone got their Private is also pretty cool.”
> “The course was well organized, easy to follow, and easy on the budget. I would recommend this course to anyone.”
> “I really did enjoy it a lot more than the fits private instrument combined.”
> “I enjoyed my training very much, and with the right instructor, I believe this syllabus is excellent.”
> “I believe that motivated individuals will be able to successfully and efficiently complete this course while saving valuable time and money. Great Job to all those involved in its implementation.”
> “The scenarios were good real life applications. The maneuvers seemed pointless. Maybe if you incorporated scenarios in which you would need them in real life pilots would understand why to get them down (besides obviously needing to pass the checkride).
> “The staff at MTSU did a great job on providing students a cheaper route to obtain pilot ratings and certificates.”
> “You are doing a great job.”

Discussion and Recommendations

The average reported flight time of 182.7 hours at obtainment of their Commercial Certificate for students completing the FITS Commercial curriculum compares favorable with the 250 hours required by Part 61 and the 190 hours required by Part 141. Obviously, the reduction in cost to students by the lowering of average flight time required is a benefit. However, it should be noted that a reduction in total flight hours for the Commercial Pilot training was not the primary goal of the curriculum writers and researchers of this project. The primary goal was to more effectively train Commercial Pilots for actual flight operations with the use of scenario-based methods – any reduction in total flight time is an additional advantage of the project. It should also be noted here that no attempt is being made to claim that the difference between the Commercial flight time averages of students with previous FITS training (155.2 hours) and students that had not experienced FITS training previous to the Commercial Curriculum (217.4 hours) is due to the FITS training method alone. Many students who had not previously been trained using FITS curricula had “built time” before enrolling at MTSU. This inflated their total time and accounts for much of the higher flight time average for those students.

As previously indicated, the primary teaching methodology of the FITS Commercial Pilot course is the use of scenarios in training. At the conclusion of this course, the students took the standard FAA Commercial Pilot Practical Test. This Practical Test, during the time these students completed the course, was predominantly a maneuvers-based test. This created somewhat of a disconnect between the teaching method and the testing method. At the time of this writing, the Commercial Pilot Practical Test Standards are under revision. The FAA has indicated that future tests will utilize the scenario method, but the students and instructors in this project did indicate some frustrations about the differences between training and testing. Regardless of the disconnect issue, 29 of the first 33 students to complete the syllabus, passed the FAA Commercial Pilot Practical Test on the first attempt. The remaining four students passed on their second attempt. This means that 87.8% of the students passed on their first try. This percentage is comparable to the pass rate experienced by students who trained at MTSU in previous years using the traditional training methods.

The authors believe that both the students’ survey responses and comments are sending an important message. Overall the responses from the students about the FITS Commercial Pilot course was positive. The students expressed enjoying the FITS Commercial Pilot syllabus, commenting specifically on the flexible scheduling that the shuffle feature allows, the realism of the scenarios, and the time and cost savings. Many students indicated that even though the IFR portion was not part of the Commercial Pilot Practical Exam, that they nevertheless saw the value in gaining IFR experience and built confidence with IFR flights, especially the solo IFR flights. There were also areas for improvement that were identified by the students. Some students commented on the Commercial Maneuvers portion of the course. A few students believed that there should be more lessons and practice in a complex airplane, and that the complex airplane lessons should be held to the last because the Commercial Practical test must be accomplished in a complex airplane. Students also indicated that more solo flights, especially solo cross country flights, were needed. Several students commented on their flight instructors and the need for instructors to be completely prepared to teach using the scenario-based methods. Study of the implementation of this training course will continue, and researchers and curriculum writers will take these comments and recommendations into account when revising the syllabus for future use.

References

Ayers, F. H. (2006). The application of scenario based recurrent training to teach single pilot resource management (SRM) under the FAA Industry Training Standards (FITS) Program. *Journal of Aviation and Aerospace Research, 3*, 13-25.

Beckman, W.S., Callender, M., Gossett, S., Dornan, W.A., Craig, P.A. (2008). Moving FITS Training From the Laboratory to the Flight Line. *Collegiate Aviation Review,* 26(1), 1-8.

Craig, P. A., Bertrand, J. E., Dornan, W., Gossett, S., & Thorsby, K. K. (2005a). *Ab Initio training in the glass cockpit era: New technology meets new pilots.* Paper presented at the International Symposium on Aviation Psychology.

Craig, P. A., Bertrand, J. E., Dornan, W., Gossett, S., & Thorsby, K. K. (2005b). Scenario-based Private/Instrument syllabus versus traditional maneuver-based syllabi: A preliminary descriptive analysis. *Collegiate Aviation Review, 23*, 23-27.

Dornan, W. A., Beckman, W., Gossett, S., & Craig, P. A. (2007b). A FITS Scenario-Based Training Program Enhances GPS Pilot Proficiency in the General Aviation Pilot. *Collegiate Aviation Review, 25*(1), 19-27.

Dornan, W. A., Beckman, W., Gossett, S., Craig, P. A., & Mosey, P. (2007). *The Implementation Of The FAA Industry Training Program In Technically Advanced Aircraft (TAA): Lessons Learned.* Paper presented at the International Symposium on Aviation Psychology, Dayton, Ohio.

Dornan, W. A., Craig, P. A., Beckman, W., & Gossett, S. (2007). *Incorporating a Novel Approach using the Technically Advanced Aircraft.* Paper presented at the 78th Annual Aerospace Medical Association Meeting.

Dornan, W. A., Craig, P. A., Gossett, S., & Beckman, W. (2006). Best evidence for the FAA Industry Training Standards program for pilot training in technically advanced aircraft. *Collegiate Aviation Review, 24*(1), 58-66.

Dornan, W. A., Craig, P. A., Gossett, S., Beckman, W., & Mosey, P. (2007). *FAA Industry Training Standards (FITS) Program for Pilot Training: An MTSU Progress Report.* Paper presented at the National Transportation Aviation Symposium.

Dornan, W. A., Gossett, S., Craig, P. A., & Beckman, W. (2006). *Best evidence on FITS.* Paper presented at the National Training Aircraft Symposium, Daytona Beach, Florida.

Federal Aviation Administration. (2004). *FAA-Industry Training Standards (FITS) Program 2004*. Retrieved, from the World Wide Web: http://www.faa.gov/education_research/training/fits

Federal Aviation Administration. (2007, July). *FITS Generic Commercial Airplane Single Engine Land Syllabus*. Retrieved, from the World Wide Web: http://www.faa.gov/education_research/training/fits/training/generic/media/commercial_asel.pdf

Glista, T. (2003, March/April). FAA/Industry Training Standards (FITS): Times (and training requirements) are a changing. *FAA Aviation News,* 1-4.

FAMILY FACTORS INFLUENCING FEMALE AEROSPACE STUDENT'S CHOICE OF MAJOR

Gail Zlotky and Wendy S. Beckman
Middle Tennessee State University
Murfreesboro, Tennessee

In an effort to understand the factors that influence female student choice of an aviation career, a qualitative study was undertaken. Female Aerospace students at Middle Tennessee State University were interviewed to determine if there were common factors that encouraged them to pursue education in aviation. A content analysis of the interviews was performed, and the most commonly cited factor was having parents who were supportive of education. It was noteworthy that the majority of interviewed students did not have parents who specifically encouraged aviation as a course of study, but instead were open to their daughter pursuing a career field that interested them. It was also found that most female students did not have family connections to aviation, but were the first in their families to pursue an aviation career. The continued existence of stereotypes regarding male and female roles in the aviation workplace was also confirmed.

The attraction of female students into the traditionally male-dominated fields of science, math, engineering, and technology (STEM) has become a national priority, as evidenced by the America COMPETES (Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science) Act, which was signed into law in August of 2007 (America COMPETES, 2007). Efforts are underway by industry, government, and academia to increase the representation of females in all of these areas, and aviation is no exception (Chavanne, 2008). Some indications point to improvement in this area. For example, Ison found that the number of women pilots in the United States has increased over the past ten years, that there are now more women enrolling in collegiate aviation programs than ever before, and that there are more female faculty members involved in aviation education than there has been historically (Ison, 2008).

Even given this focus of attention and the improvements indicated above, there is still much room for improvement in this area. An examination of the statistics available on the Federal Aviation Administration's (FAA) website brings clarity to the scope of the problem (Federal Aviation Administration, 2007a, 2007b). For example, there were only 5, 349 female certificated Airline Transport Pilots in the United States in 2007, out of a total of 143,953 Airline Transport Pilots. This means that only 3.7% of the pilots at the highest level of certification are women. The Commercially rated pilot numbers are slightly better, but even then only 6.2% of the Commercial pilots are female; similarly, only 6.8% of the nation's Certified Flight Instructors are women. The maintenance area is even worse, with only 2% of the certificated A&P Mechanics being women. One area that is slightly higher is certificated Dispatchers, where 16.2% of those certified are women; but even this is obviously far below the level of parity with men.

To address the issue of increasing female representation into traditionally male fields, numerous studies have been done. Some of these have focused on arousing initial interest in these fields, while others have focused on retention of students once they are enrolled in a traditionally male program (Turney et al, 2002). Many of these efforts have come to the conclusion that the effect of long-held stereotypes cannot be ignored, but must instead be consciously addressed. Indeed, although recent studies have found that girls are now performing just as well in STEM high school preparatory courses as boys, female students still face the task of overcoming "…stereotypes long held by parents, teachers, and even girls themselves that boys are more suited to math-heavy studies and professions…" (Cavanagh, 2008,1).

The objective of this study was to identify the factors that either encouraged or discouraged female students from enrolling in the Aerospace Department at Middle Tennessee State University (MTSU). The impetus for conducting this study began during the 2007 academic year, when a survey of all graduating seniors in the Aerospace Department at Middle Tennessee State University revealed that only 16.67% of the class was female. Of those female students, 29% indicated that gender stereotypes had influenced their decision-making process when choosing aviation as a career field. This was of concern, as these students were the success stories – the graduating seniors. If nearly a third of the successful female students had experienced the impact of gender stereotypes, what effect did these stereotypes have on students as they chose their academic major? The researchers decided it would be useful to understand what factors led female students to enroll in the Aerospace Department at MTSU, as a first step in identifying a course of action to increase female enrollment. This study attempted to replicate, in limited aspects, a study of female undergraduate engineering students which was done at the University of Oklahoma from 2003-2005, in which efforts were made to identify the factors that effected students' decisions to move into an engineering department (Walden & Foor, 2008). Instead of looking at all potential factors, the decision was made to limit the investigation to the influence of family members and friends on the decision-making process of selecting an undergraduate major.

Methodology

Approval was granted from the MTSU Institutional Review Board (IRB) on June 26, 2008, to conduct a human research subject interview study of female aerospace students (MTSU IRB # 08-340). Twenty female Aerospace students across all five Aerospace concentrations (Professional Pilot, Dispatch/Scheduling, Maintenance, Administration, and Technology) and of each class standing were interviewed by the researchers. The students responded to either a request made in an Aerospace class or to a sign posted in the Aerospace Department hallway, requesting female Aerospace student volunteers to participate in a short interview. The format of the interview was open-ended, with the following six questions asked of each interviewee:

1. Tell me about your relationship with your mother through your childhood and adolescent years.
2. Tell me about your relationship with your father through your childhood and adolescent years.

3. Has there been another person in your life besides your parents who has been very influential in your growth and development? If so, who, and what is their relationship to you?
4. What do you think influenced your choice to pursue aviation as a career?
5. Did any particular person suggest to you that aviation might be a good career choice? If so, who? What was your relationship to that person?
6. Has there been any important person in your life that did not support your career choice? If so, who? What was your relationship to that person?

The responses of the interviewed students were tape recorded so they could be reviewed for a content analysis. Notes were also taken by the interviewers, to supplement the data provided by the audio tape.

Findings

The first focus area of the interview was the investigation of each student's relationship with their mother throughout their childhood. Content analysis revealed that eleven participants responded with the adjective "close" in their description of their relationship with their mother, six participants indicated within their responses the adjective "supportive", and three indicated the adjective "good" (note, many participants used multiple adjectives in their descriptions). Only two students indicated that their relationship with their mother was not positive, with one indicating that the relationship was "rough" and the other indicating that they "fought a lot". Eight participants commented specifically in their response to the first question that their mother had made clear to them the importance of obtaining a college education throughout their childhood years. Only one mother worked in an aviation-related field.

The next area of interest was the student's relationship with their father throughout their childhood. Five participants responded with the adjective "close" in describing this relationship, while seven others indicated that either their father was "not involved in their lives", or had worked long hours or been involved in other activities to such an extent during their childhood that they had hardly knew him. Seven participants specifically commented on the fact that their father expected them to obtain a college education.

Twelve of the interviewees indicated that there were people outside their immediate family who had an impact on their growth and development, including their choice of college major. The identified people included extended family (grandparents, aunts, uncles), teachers, godparents, and family friends. While most of these influencing people were indicated to be positive influences, there were three participants that indicated a negative influence was experienced from these individuals.

Questions number 4 and 5 overlapped somewhat, as for many students, what caused them to become interested in aviation turned out to be a particular person who suggested to them an aviation career. Three students indicated that their father had specifically mentioned aviation to them as a possible career, two indicated that their mother had done so, and one student indicated that both of her parents worked in aviation. One had a supervisor at a job mention the possibility of an aviation career to them, and one played soccer for a coach that was an airline pilot. Several mentioned loving to fly commercially when vacationing as a child, while several others indicated that they had just “always had an interest” in aviation. There were several who responded that they were not sure what caused them to become interested in aviation. Overall, exactly half of the interviewees had experienced a particular person in their life who suggested that aviation might be a good career path for them to follow.

The final question was designed to determine if students encountered any resistance to their decision to major in an aviation field. Ten of the interviewees indicated that they had indeed experienced this negative influence. Three students indicated this resistance was from their mother, two indicated it was from their father, two indicated it was from their extended family (grandmother, mother’s family), two indicated it was from their friends, and one indicated it was from acquaintances. In particular, many of the students who were interviewed talked about the reaction of new acquaintances who learned of their intended major. While some indicated that these new acquaintances (typically other college-aged students) thought it was “neat” and “cool” that they were pursuing an aviation degree, exactly half indicated they often receive negative comments as well. One of the most predominant comments that the interviewees received and perceived as negative was, “Oh, so you’re going to be a flight attendant?” The students did not largely seem to interpret these reactions as being a “negative influence” on their career choice, but instead seemed to view them as a humorous side note.

Discussion

Several themes emerged from the content analysis of the participant interview tapes. First, with few exceptions, the female students characterized their relationship with their mother as being positive. In addition to the general support of their mothers, in many cases the students’ mothers made clear the value of education to their daughters throughout their childhood and adolescence. The female student’s relationships with their fathers were predominantly not described as being as close as the relationship with their mothers, but fathers were also described in many cases as having been supportive of educational endeavors. However, an area of concern to the researchers was whether or not the interviewees were completely candid about their family relationships during the interview. Given the overwhelming positive responses that were received regarding particularly their relationship with their mother, it was wondered if the students were just giving what they perceived to be the “correct” answer to the question. For future research, having female students complete an anonymous survey investigating specifically family relationships and the influence of those relationships may provide more insightful responses.

Second, the majority of interviewees indicated that there had been someone outside of their immediate family that was influential to their growth and development. In some instances, this person turned out to be an individual who had suggested aviation as a career, while in other cases it was just someone who was supportive of their educational efforts. While ten of the students interviewed did have a particular person suggest to them that aviation might be a potential career choice, only six interviewees actually knew someone who worked in the aviation industry prior to enrolling at MTSU. So, for 70% of the students, their college careers were started with very little knowledge of the field of aviation. This number is almost exactly the same as what was found for all Aerospace students in a previous study (Beckman & Barber, 2007). This means that a lack of specific knowledge of the field does not seem to be a factor that discourages young women any more than it discourages young men.

Third, and perhaps most interesting, is the revelation of how predominant negative stereotypes of women in the aviation field still are. While the majority of interviewees indicated that the people influencing their growth and development were supportive influences in their life and had stressed the importance of education, half also reported that a number of these same influential people had negative reactions to their choice of aviation as a major. In addition, while most students seemed to view the casual negative comments of acquaintances with amusement, it was interesting to discover that there is still this level of gender-based stereotypical attitudes prevalent among today's college-age students. This finding indicates that, as a society, even in the younger generations, we are still not past the historical view of "male" versus "female" work role stereotypes. This indicates that current efforts to make girls aware of their career options from the youngest ages are indeed critical, as the issue is not educational preparation or capability, but overcoming historical attitudes about female career choices. Changing a cultural climate takes time, but it is clear that women who work in the field of aviation need to actively serve as mentors and role models for female students, in an effort to counterbalance societies' remaining stereotypical attitudes.

References

Beckman, W. and Barber, P. (2007) Factors Impacting Student Retention in a University Professional Pilot Program. *International Journal of Applied Aviation Studies, 7(2),* 295-311.

Chavanne, B. (2008, August 18). Sea Change. *Aviation Week and Space Technology, 169(7)*, 81-82. Retrieved February 16, 2009, from Academic Search Premier database.

America COMPETES (Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science) Act of 2007, Pub. .L. No. 110-69, (2007).

Ison, D.C. (2008). The Status of Women Faculty in Four-Year Aviation Higher Education Programs. *Journal of Women in Educational Leadership, 6(3),* 201-210. Retrieved February 14, 2009, from Wilson Web database.

Federal Aviation Administration. (2007a) Table 1. Estimated active airmen certificates held: December 31, 1998-2007. Retrieved February 15, 2009 from http://www.faa.gov/data_statistics/aviation_data_statistics/civil_airmen_statistic2007/.

Federal Aviation Administration. (2007b) Table 2. Estimated active women airmen certificates held: December 31, 1998-2007. Retrieved February 15, 2009 from http://www.faa.gov/data_statistics/aviation_data_statistics/civil_airmen_statistic2007/.

Turney, M.A., Bishop, J.C., Karp, M.R., Niemczyk, M., Sitler, R.L., & Green, M.F. (2002). National survey results: Retention of women in collegiate aviation. *Journal of Air Transportation, 7(1),* 69-92.

Walden, S.E., & Foor, C. (2008). "What's to Keep You From Dropping Out?" Student Immigration into and within Engineering. *Journal of Engineering Education 97(2),* 191-205.

COMPARING THE ACCURACY OF PERFORMING DIGITAL AND PAPER CHECKLISTS USING A FEEDBACK INTERVENTION PACKAGE DURING NORMAL WORKLOAD CONDITIONS IN SIMULATED FLIGHT

William G. Rantz, Ph.D.
Western Michigan University-College of Aviation
Kalamazoo, MI
Bryan W. Hilton, B.S.
Western Michigan University-College of Aviation
Kalamazoo, MI

This study examined whether pilots completed airplane digital and paper checklists more accurately when they received post-flight graphic and verbal feedback. Participants were 6 college student pilots with instrument ratings. The task consisted of flying flight patterns using a Frasca 241 Flight Training Device which emulates the Cirrus SR20. An alternating treatment, multiple baseline design across pairs with reversal was used. Visual inspection and statistical analysis of the data suggests that paper checklist accuracy does not differ significantly from digital checklist accuracy during normal workload conditions. The results also suggest that graphic feedback and praise can be used to increase the extent to which pilots use both digital and paper checklists accurately.

Understanding the knowledge and behaviors required to effectively manage risk are an integral component of the professional pilot training curriculum (Western Michigan University: Professional Flight Training Program, 2004). The aviation industry demands that professional pilot graduates understand the inherent risks associated with flight operations and that individuals must continue to practice comprehensive preflight planning, attention to detail, procedural discipline, and run the checklists as printed, Wilson, (2008). Checklists organize tasks into sequences of actions that configure the aircraft and prepare the crew for evolving events. "The major function of the checklist is to ensure the crew will properly configure the plane for flight, and maintain this level of quality throughout the flight, and in every flight" (Degani & Wiener, 1990, p. 7). Checklist devices or methods of presentations are described as paper, laminated paper/card, scroll paper, electromechanical, vocal, and computer-aided/electronic. The most common method of presentation for checklists is the laminated paper/card (Degani & Wiener, 1994; Turner & Huntley, 1991). While this statement may be true for all general aviation aircraft manufactured in the last one hundred years, the rise of lower cost computing hardware and software is rapidly changing how newer aircraft present checklists. (Boorman, 2001a, 2001b).

Within the last two decades electronic or digital checklists have appeared on many regional and major airline flight decks, and some general aviation aircraft. These digital checklists are integrated into the new aircraft panel by the manufacturer with software designed to exclude many paper checklist errors observed in past studies (Arkell, 2006; Boorman, 2001a, 2001b). As avionics prices continue to decline, it is very likely more digital checklists will be installed on smaller general aviation aircraft, thereby expanding the demographics of the pilot users from airline professional to recreational novice. The future challenge is not that pilots understand that the checklist is a presentation method by which flight deck safety is enhanced.

The challenge seems to be recognizing the absence of stimulus control in a varying flight environment which may result in unpredictable checklist use.

An extensive review of the checklist literature has found many interesting areas outside of aviation where checklists are employed (Rantz, 2005). From the accident reports, errors using checklists have and continue to plague the aviation industry in particular. Given the number of aviation studies devoted to checklist use and how tasks are conducted on the flight deck, an extensive search of the aviation checklist literature revealed only one study that has examined whether the traditional paper checklist could be a) used as a dependent variable and b) whether behavioral interventions, could increase the appropriate use of flight checklists (Rantz, Dickinson, Sinclair & Van Houten, In Press, p. 20. The purpose of the present study is to compare and if possible improve the accuracy of both the traditional paper and standard digital checklists.

Method

Experimental Design. An alternating treatment, multiple baseline design with reversal plus an over sixty day probe across pairs of participants was used to compare paper and digital checklists and evaluate the effect of feedback on checklist use. There were four phases of the experiment, baseline, intervention, reversal, and probe. Sessions lasted approximately two hours and 6 participants flew four different flight patterns per session using the Frasca 241, Cirrus SR20 flight training device. Each flight was considered a trial, and checklist performance was scored and graphed separately for each trial. Each flight lasted approximately 20-25 minutes. There were six different flight patterns. The order of exposure to the flight patterns was randomized in blocks of six for each participant. This procedure insured no two patterns were repeated during one session. Paper or digital checklists were randomly assigned for the initial trial at the beginning of each session. For the remaining three sessions, paper or digital checklists were alternately assigned.

The flight checklists. The digital and paper checklists each contained 70 identical checklist items divided into sections that corresponded to each of the eight flight segments. The digital checklist was an integrated function of a multifunctional display (MFD) produced by Avidyne. The MFD model was the Entegra EX5000C used in Cirrus SR20 aircraft. The paper checklist was a spiral bound booklet provided for use in the Cirrus SR20 (Pilot's Checklist Cirrus SR20, 2002), both the digital and paper checklists are used in the colleges' flight training curriculum. The digital checklist display, when used, was in a fixed position ahead and slightly to the right of the pilot's central view. The paper checklist, when used, was positioned on the right leg or lap of the participant and when not used usually remained on the seat beside the participant.

Dependent Variables. The primary dependent variable consisted of the number of paper or digital checklist items completed correctly per flight.

Independent Variable. There were two independent conditions during this study, using paper or digital checklists. The independent variable was the presence or absence of post-flight (a) graphic feedback on the total number of checklist items completed correctly per flight, (b) graphic feedback on the number of items completed correctly, incorrectly, and omitted for each of the eight flight segments per flight, and (c) praise for improvement in the number of checklist items completed correctly.

Inter-observer Agreement (IOA). A second observer watched randomly selected recordings of the flights and scored performance using the checklist observation form. This process was repeated for each participant. This ensured that (a) at least 25% of the sessions were rescored for each participant, and (b) the trials that were rescored were randomly selected. Inter-observer agreement was determined for the total number of checklist items completed correctly. Inter-observer agreement was calculated as follows: number of agreements divided by the number of agreements plus disagreements, multiplied by 100. Inter-observer agreement for correct and incorrect item errors was an average of 95% with a range of 79% to 100%.

Results

Figure 1 displays the total number of paper checklist items completed correctly (open circles) together with the total number of digital checklist items (closed circles) completed correctly for each participant per trial. All participants increased paper and digital checklist performance accuracy over baseline when post-flight graphic feedback was provided and those improvements remained during the withdrawal phase and during a delayed probe. Baseline paper and digital checklist performance varied considerably across participants with participant 1 showing the lowest level of performance in both paper (average 87% error) and digital (average 89% error). Participant 1 had a mean average of 3.37 correct for digital checklist items and 6.11 correct for paper. Both participant 2 and 4 showed the highest level in paper checklist (average 43% error) and both participant scored a mean average of 39.67 correct. However participant 2 averaged 33% error for the highest performance in digital checklist use with a mean average of 44.28 correct. Baseline trends were fairly stable over time for four participants (P1, P3, P4, and P5), with the exception of participant 2, who despite overall high mean average scores, performed one high peak in digital and one high peak in paper performance and showed a steady, overall decline in both paper and digital accuracy from the first trial and participant 6 who showed wide variability between paper (38.92 mean average correct) and digital (37.89 mean average correct) performance. Overall paper checklist baseline performance averaged 62% errors (27.42 mean average) for all participants while digital baseline performance averaged 61% errors (26.57 mean average).

Overall performance in both paper and digital checklist accuracy increased for all participants after the intervention was introduced. There was a dramatic intervention effect using both paper and digital checklists for both individual participants and cumulative across all participants. Two participants (P1 & P3) showed an abrupt level change of over 50% improvement in the first trial, following the introduction of the treatment and then continued an increasing trend. Participant 1 initially increased paper checklist accuracy 71% after the intervention, improving total correct checklist items from 1 item correct out of 70 items to 51 items correct. Participants 3 had the highest initial performance increase across both the digital and paper checklists, increasing level change by 61 % for digital checklist items done correctly and 44% for paper checklist items. Participant 5 experienced an initial increase in level change of 36% improvement for digital and 40% improvement for paper checklist items performed correctly. Two participants showed an increasing level change, for both digital and paper, followed by an increasing trend (P2, P4). Only Participant 6, while initially increasing 13% in paper checklist performance, demonstrated an initial single trial decrease of 1% in digital accuracy followed by an increasing digital trend.

Performance criteria for the reversal phase was established as three consecutive trials in either paper or digital where participant's checklist performance met or exceeded 95% correct on checklist items. All participants reached reversal criteria during paper checklist trials.

Overall across all participants, the average percentage of paper checklist items completed correctly increased from 38% items correct during the baseline phase to 90% items correct during the intervention phase. The average percentage of digital checklist items increased from 39% items completed correctly to 89% items correct during intervention. Improvement continued to near perfect levels for participants during the reversal phase with 100% paper checklist items correct and 99% digital items correct. The average percentage of paper checklist performance declined 3% between a 60 and 90 day delay. The average percentage of digital checklist performance declined 4% during that same time period.

Data contained in Figure 1 were used in the inferential analysis given the model, ($Y_t = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \beta_3 d_3 + \beta_4 d_4 + \beta_5 d_5 + \beta_6 d_6 + \phi_1 \gamma + \varepsilon_t$). The parameters of this model were estimated for each participant using the bootstrap based time-series regression method described in McKnight, McKean, and Huitema (2000). Results were statistically significant for each individual's intervention effect in both paper and digital checklist use. Performance was generally not significant once optimum performance levels were reached during each following phase.

After parameter estimates for each participant were computed they were used as dependent variable scores in the group level analysis. The purpose of this analysis was to provide an overall evaluation of the effects of the interventions for the group of six pilots. The next stage of the group analysis consisted of conventional one-sample *t*-tests to evaluate the hypothesis that each intervention and phase-change parameter value is equal to zero. Once again, results were statistically significant for the overall intervention effect in both paper and digital checklist use. As for each individual's results above, performance was generally not significant once optimum performance levels were reached during each following phase.

The third aspect of the analysis involved computing the difference in performance under the digital and paper conditions at each observation point and testing the difference between the digital and paper means. Once again, the double bootstrap method of McKnight, McKean, and Huitema (2000) was used to estimate the parameters of a time-series model developed to evaluate the hypothesis of zero difference between digital and paper feedback; this is a model that contains only an intercept and an autoregressive parameter. The difference between paper and digital checklist performance was found not to be statistically significant ($t = 1.78$, $p = .08$).

Discussion

This research is a follow up to the study by Rantz, Dickinson, Sinclair & Van Houten (in press) which evaluated the effects of feedback and praise on the use of a simple personal computer aviation training device and a paper checklist. The present study confirmed the findings of the former study, while using a much higher level of simulation. The current study additionally included comparing pilot's performance using both paper and electronic checklists during all phases of the experiment. The results of the present study also suggest using graphic feedback and praise can simultaneously improve checklist reading performance in both traditional paper and modern digital presentation modes. The results also indicated, contrary to common opinion, that the use of a digital checklist did not lead to a reduction in errors compared to the traditional paper checklist in a normal workload environment. This study also suggests a pilot's checklist performance, regardless of presentation method, may be influenced by common underlying rule-based behaviors (learning history), structured feedback, and particular salient environmental prompts.

References

Arkell, D. (2006). From safe to safer: Boeing's electronic checklist marks 10 years of enhanced safety for pilots and travelers. *Boeing Frontiers 4*(11).

Boorman, D. (2001a). Safety benefits of electronic checklists: An analysis of commercial transport accidents. *Proceedings of the 11th International Symposium on Aviation Psychology* (pp. 1-6). Columbus, OH: The Ohio State University.

Boorman, D. (2001b). Today's electronic checklists reduce likelihood of crew errors and help prevent mishaps. *ICAO Journal, 1*, 17-21.

Degani, A., & Wiener, E. L. (1994). *On the design of flight-deck procedures* (NASA Contractor Rep. 177642). Moffett Field, CA: NASA Ames Research Center.

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.

McKnight, S. D., McKean, J. W., & Huitema, B. E. (2000). A Double Bootstrap Method to Analyze an Intervention Time Series Model with Autoregressive Error Terms, Psychological Methods, 5, 87-101.

Pilot's Checklist Cirrus SR20. (2002). P/N 11935-003. Duluth, MN: Cirrus Design Corporation.

Rantz, W. G. (2005). Strategies for Controlling Checklist Reading Behavior: A Literature Review. *Proceedings of the 13th International Symposium on Aviation Psychology* (pp. 497-503). Columbus, OH: The Ohio State University.

Rantz, W. G., Dickinson, A. M., Sinclair, G. A., Van Houten, R. (in press). The effect of feedback on the accuracy of checklist completion during instrument flight training. *Journal of Applied Behavior Analysis.*

Turner, J. W., & Huntley, S. M. (1991). *The use and design of flight crew checklists and manuals*. (Report No. DOT/FAA/AM-917). Washington, DC: Federal Aviation Administration.

Western Michigan University: Professional Flight Training Program. (2004). *Training Course Outline*. Kalamazoo, MI. Author.

Wilson, B. T. (2008). A disturbing trend in professional aviation training. Retrieved from http://flightcog.com/essays/deep.html

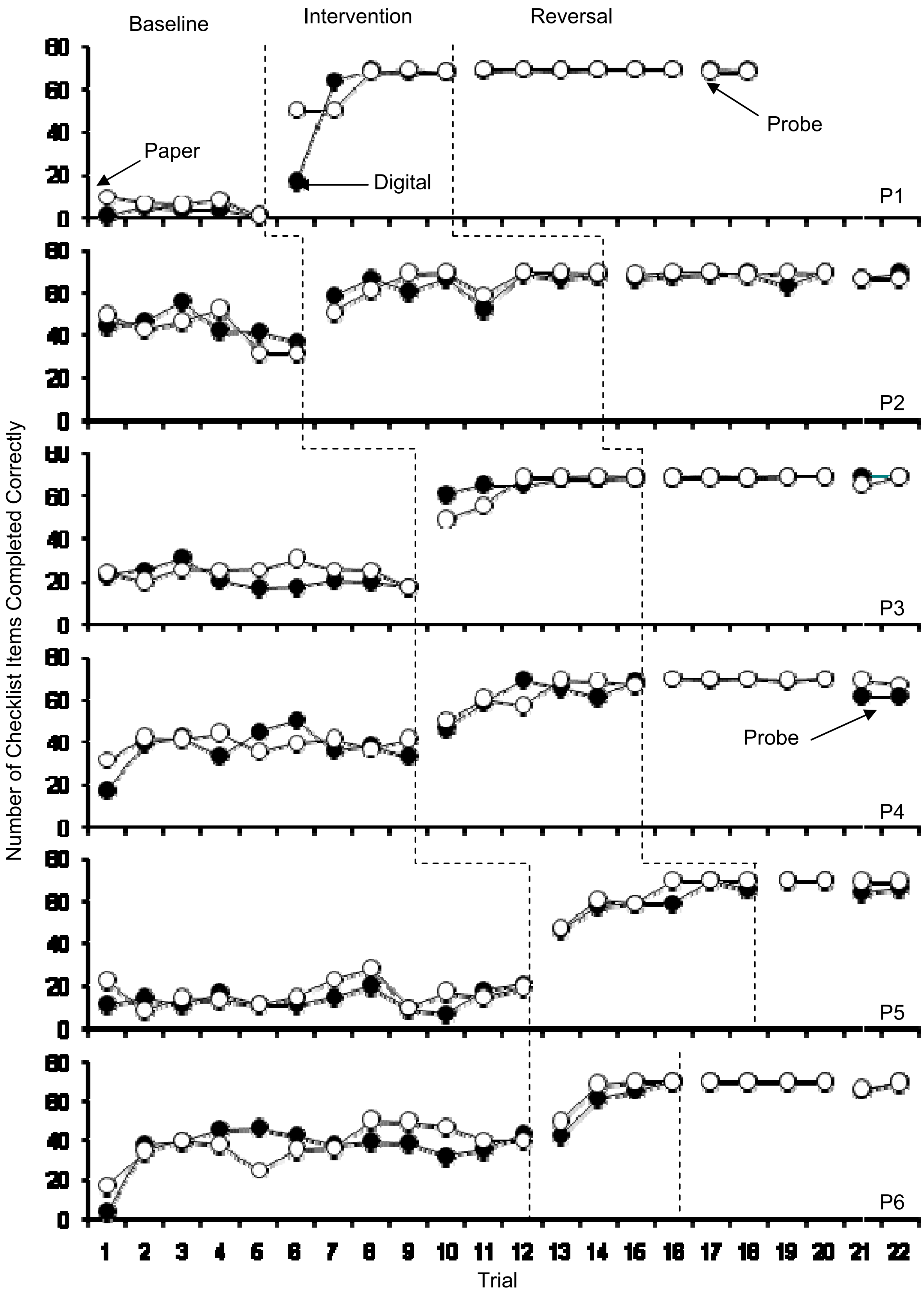


Figure 1. Total Number of Paper and Digital Checklist Items Completed Correctly.

EXAMPLE OF WORK DOMAIN ANALYSIS APPLIED TO TOTAL ENERGY CONTROL SYSTEM

Matthijs Amelink
Thales research & technology NL, DECIS Lab
Delft, The Netherlands
M. M. (René) van Paassen, Max Mulder
Delft University of Technology, Faculty of Aerospace Engineering
Delft, The Netherlands

Automation is often accused of adding to the complexity of a system and unnecessarily increasing operator's workload, and the potential for human error. An approach is needed that guides designers to make the right design choices. Cognitive Systems Engineering (CSE) is a promising approach. However, this field is still young and tangible examples of automation design with an explicit CSE approach do not exist. This paper describes how the design of Total Energy Control System (TECS) that was founded in the late 1970's can be regarded as an example avant la letter. TECS is an automated flight control system designed to solve many of the issues that classical autopilot and auto-throttle systems have. Since TECS has been designed, implemented, and evaluated it could teach valuable lessons on how Work Domain Analysis (WDA) can guide the design of automated systems as the first phase of CSE approach. The application of WDA to TECS is exemplified using the abstraction hierarchy and the abstraction decomposition space.

Finding a design paradigm for automating with the least amount of complexity is our goal. Cognitive Systems Engineering (CSE), Ecological Interface Design (EID), and Cognitive Work Analysis (CWA) are promising design paradigms that guide designers to build better systems for human – machine interaction (Rasmussen et al., 1994, Vicente and Rasmussen, 1992, Vicente, 1999). EID, CWA, CSE have the first step in common, they starts with Work Domain Analysis (WDA) to uncover the constraints and structure of the work domain. This should make visible how to design a system taking into account those constraints.

Above mentioned fields of research are emerging but, unfortunately still few examples exist where WDA has been shown to lead to better systems design. Most examples come from interface design using EID. Dinadis et al. (1999) and Amelink et al. (2005) give examples of ecological interface in the aircraft control domain and Burns et al. (2004) has bundled a number of examples from multiple domain. Examples that apply WDA with the goal to achieve real-world systems design with the least amount of complexity have not been found.

However, this can be illustrated well by the analyzing TECS in retrospect. The main reason why TECS has better performance over classical auto-pilot / auto-throttle systems is because TECS takes the energy management, inherent to flight, explicitly into account. In contrast, classical autopilots are based on representations coming from first principles of small-perturbation flight dynamics, acting on arbitrary states. They are criticized for their complexity and un-human-like behavior under certain conditions. Lambregts (1983a, 1983b, 1996) explains how TECS has better performance and is significantly less complex.

WDA is always about the details in a work domain, therefore a certain depth of knowledge needs to be achieved. First, an introduction to TECS is given so the reader understands the main points of TECS, and the architecture is discussed to show which components the system is made of. Then, the WDA is made using `abstraction' and 'part-whole decomposition', which links the components of TECS the purpose of TECS. We hope that the analysis also facilitates in conveying knowledge about how TECS works and the design rationale behind it.

Introduction to Total Energy Control System

Total Energy Control System is a generalized automatic flight control system that was developed in the late 1970's to early 1980's by Lambregts to overcome a number of issues with conventional autopilots at the time (Lambregts, 1983a, 1983b, 1996). These issues include: unnatural high levels of control activity (especially the auto-throttle), a complex man-machine interface, and functional overlap in control modes causing mode confusion. Lambregts fully recognized the importance of designing with the least amount of complexity added by automation through functional integration. Lambregts' approach starts with an analysis of the fundamental physics of airplane dynamics and designed TECS to act on the energy constraints inherently present in aircraft control. In retrospect this approach coincides with what was later called the *ecological approach*, in this case the *ecology* between automation and the environment. Although TECS was developed before CSE emerged as a research field, the design approach of TECS can be regarded as an ecological approach to automation design avant la lettre. Since TECS has been

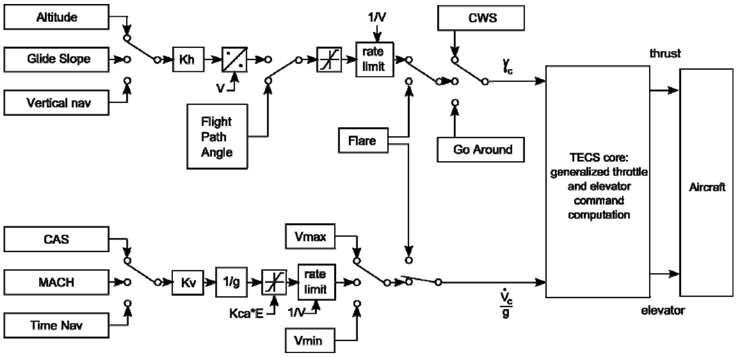


Figure 1: Overview of TECS – control diagram. Adapted from Lambregts(1983a, 1983b).

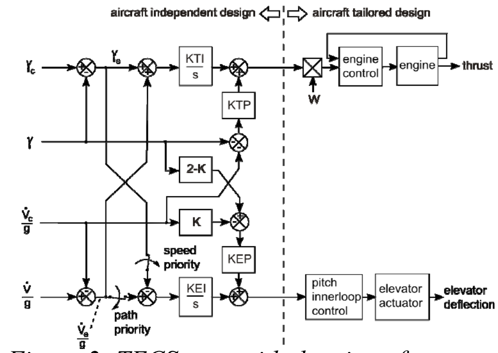


Figure 2: TECS core with the aircraft independent and aircraft tailored parts. Adapted from Lambregts(1983a, 1983b).

designed, implemented, and evaluated it can teach us valuable lessons on how CSE principles should guide the design process of automation. Such a tangible example is not yet available with an explicit CSE approach.

The objectives of the design of TECS were to integrate all longitudinal autopilot and auto-throttle control functions to generate pilot like control, to create a simplified man-machine interface, and to structure the control mode hierarchy to eliminate the overlap in control modes found in conventional autopilots and auto-throttle systems. A methodology was derived for designing a generic elevator and thrust command computation algorithm that provides decoupled flight path and speed maneuver control and is capable of serving all vertical flight path and speed control modes. This overcomes the limitations of separately designed autopilot and auto-throttle systems, and eliminates the need to switch inner-loop controllers with each flight mode. For a complete overview of the performance improvements that TECS offers, the reader is referred to Lambregts (1996, 1983a, 1983b) work.

TECS architecture overview

A conceptual overview of the TECS architecture is given. The complete design is much more complex and cannot be captured in a single diagram or in the scope of this paper. Figures 1 and 2 show the main part of TECS that is considered. Figure 2 shows the content of the 'TECS core' box in Figure 1. The part upstream of the TECS core (Figure 1) is labeled 'mode hierarchy'. The mode hierarchy consists of a number of modes inherited from classical flight guidance and control systems. The modes are organized in a hierarchy to achieve, for example, that the flare mode overrides the V_{min} mode. The commands of the altitude / vertical path modes are transformed to vertical flight path angle commands (γ_c). The commands from the speed modes are transformed into normalized acceleration commands ($\dot{V}_c/g$). These commands are the interface with the TECS core. All modes use this interface and the core processes commands off all modes with the same command computation logic. The aircraft-independent part of the core (Figure 2) computes generic elevator and generic throttle commands based on the energy control logic. The core can compute the commands in three ways, depending on the crossfeed switches positions. In the default setting, the core nulls the path and acceleration errors equally. Either switch can be opened to give priority speed or path commands in the case the required thrust fall outside the engines' thrust range. The computed commands are processed in the 'aircraft tailored' design by the inner-loop engine and inner-loop pitch control to yield the desired engine thrust and elevator deflection. Each of the components and their role in the complete system is discussed below, in the Work Domain Analysis.

Work Domain Analysis

Two relationships between functions are relevant to WDA: abstraction and part-whole decomposition (Rasmussen, 1986, 1994). Abstraction is used to link a function to a function on a higher level of abstraction. Part-whole decomposition is used to split a function into subcomponents or its features on the same level of abstraction. Aggregation is the opposite of decomposition. Rasmussen's(1994) levels of abstraction are adopted. These relations are used to link the functions of components in the control diagram with the functions and goals they achieve. The analysis is based on available information in diagrams, article texts, and correspondences with Lambregts. TECS is delimited by its natural boundary. It includes the aircraft, its dynamics, the control hardware, its functioning, and the TECS control panel that is operated by the pilots. Although 'sensors and feedback signal synthesis' is a critical part of TECS, it can be left out of the scope for this analysis without impacting the principles illustrated. The system spans multiple levels of abstraction and multiple levels of part-whole decomposition.

A first chunking of the system

A 'first chunking' of TECS is made as a first attempt at structuring the knowledge available of TECS, in terms of abstraction and part-whole decomposition. The starting point is the most detailed representation available: the control diagram in Figures 1 and 2 in combination with the literature on TECS. It belongs to the 'generalized function' level of the abstraction hierarchy since it is a conceptual representation independent of physical implementation. The control diagram shows the system in components meaningful to signal processing. At the same time each component has a function that serves the goal of TECS. Abstraction and decomposition relations are used to take the components out of the control theory context and link their functions to the functional goals of TECS. Figure 3 shows the 'first chunking' where the components of Figures 1 & 2 are on the 'generalized function' level and are linked to their abstract function on the 'abstract function' level. In turn, those functions achieve a purpose on the 'functional purpose' level. Part-whole decomposition applies to physical structures as well as conceptual structures. At the top of Figure 3, TECS denotes the entire system as a single concept. It is decomposed into three functional goals: 'safety', 'production', and 'efficiency' according to Van Paassen(1995). Four functions and their abstraction and decomposition relations are further exemplified; they are highlighted with circles in Figure 3 for easy referencing.

(1) The rate-limits in the command signals paths γ_c and $\dot{V}_c / g$ limit the rate of change of these commands ensuring limited commanded maneuvering rates and smooth command generation. The solid arrow shows the abstraction relation between the rate-limits on the 'generalized function' level and the block 'limited maneuvering rates' on the abstract function level. The maneuvering rates are expressed in acceleration normal to the flight path (a_n) and the acceleration along the flight path ($\dot{V}$). In turn, the primary function of limiting the maneuvering rates is to ensure passenger comfort. Passenger comfort is shown to be a part of the 'production goal' of TECS using a hollow arrow meaning: part-whole decomposition. The secondary function of limited maneuvering rates is to protect the airframe loading: envelope protection, which is part of the safety goal of TECS. Passenger comfort requires a lower maneuvering rate limiting than envelope protection hence the order of primary and secondary functions. There is also a decomposition relation between the rate limits and 'equal rate limits' in the 'gain & limiter values' box. The value of the rate limits (and other gain values) is part of multiple functions on the generalized function level. This is visualized by the multiple part-whole decomposition arrows pointing to this block. One of those concepts is 'preserved energy relation', which takes us to the next example.

(2) On the abstract function level the block 'energy based control decoupling' is a main function of TECS. It represents that the energy constraints that work on flight are taken into account in the design of TECS, giving it the basis for its improvements over the classical auto-pilot and auto-throttle design. As visualized by the abstraction relation to the 'functional purpose' level, the 'energy based control decoupling' achieves 'quality of control' and 'efficiency' in terms of fuel economy and engine wear. The implementation of the energy control principles is realized, conceptually, by the structure of the TECS core (Figure 2) and the values of the gains. In the speed and path command signal paths, the energy relationship needs to be preserved in order to achieve control decoupling. This is shown as 'preserved energy relationship between speed and path commands (principle)' which is decomposed into the 'gain values' that instantiate the principle. Note the decomposition taking place from 'energy based control decoupling' to 'default, speed, or path priority configuration'. The latter denotes the three ways the TECS core can compute the elevator and throttle commands (default, path priority, speed priority).

(3) Bandwidth separation is a well known principle from control system engineering to achieve stability and damping in a control systems consisting of nested control loops. This principle is applied to TECS to achieve stability and damping (on the abstract function level). The control frequencies of the different loops (pitch attitude, flight path angle and longitudinal acceleration, speed and altitude) are selected to be a factor of 3.3 to 7.5 apart with the largest gains in the inner-loop. The principle is shown on the generalized function level and is decomposed into the values of the control gains. Note that the gain values are part of satisfying two principles: preserved energy relation and bandwidth separation.

(4) 'Overhead control' is shown on the generalized function level. Available literature does not describe the exact functioning of this block but it is clear from descriptions that there is control logic responsible for (among others) the coordination of mode switches, setting speed/path priority, and detecting speed range violation. Due to lack of information about it, the logic itself cannot be refined at the 'abstract function' level. The switches found in the mode hierarchy and core, are controlled by the overhead logic and are, therefore, part of that sub-system as well. The decomposition arrows originating at the 'overhead control' block show where the TECS control diagram and the overhead control logic are connected. The abstract principles of the overhead control logic remain uncovered by this analysis.

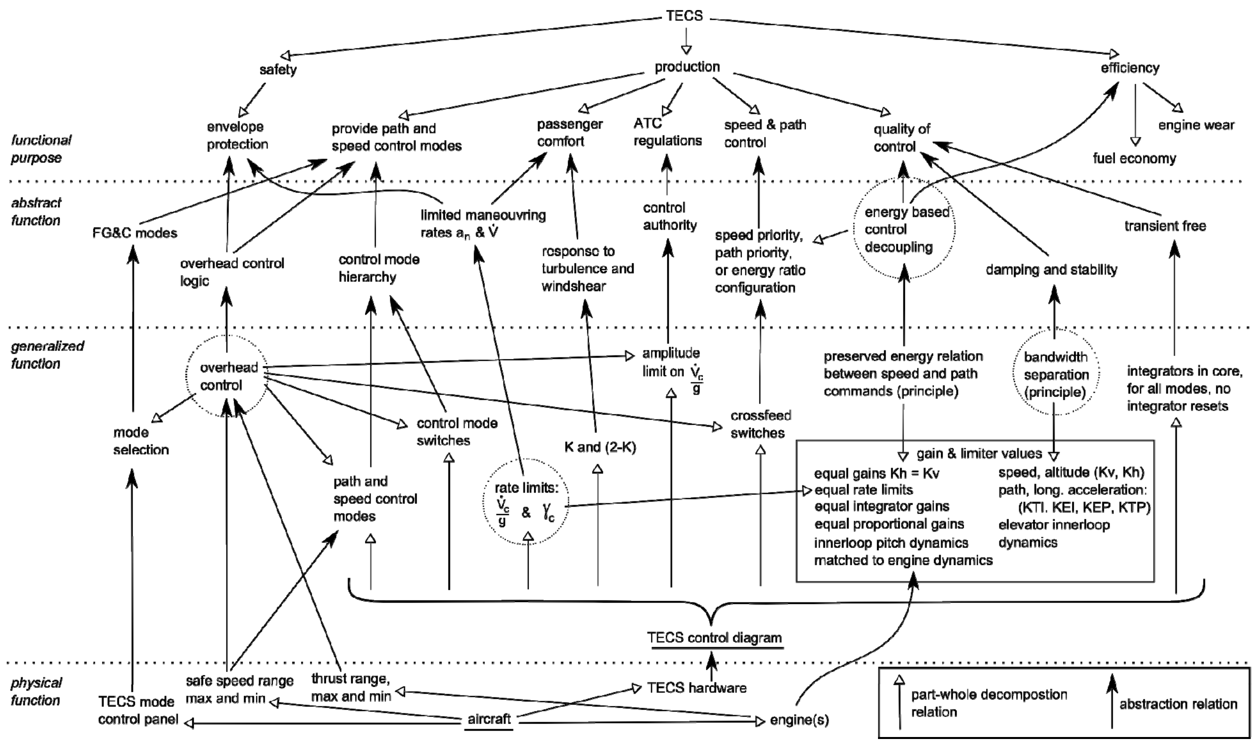


Figure 3: A first chunking of TECS showing functions related through abstraction and decomposition.

On the 'physical function' level the aircraft is shown to visualize that TECS is part of the complete aircraft and to show two non-holonomic (design implied) constraints: the aircraft's speed range and the engine's thrust range. Lack of information on implementation of TECS is reflected on physical function level – hardware on which TECS is implemented is not represented and out of scope for the analysis. Therefore the 'physical function' level is mainly a placeholder for the technologies and allows us to show that aircraft properties are part of the complete TECS system. Note that the relation between aircraft and TECS has not been fully developed in Figure 3.

Aggregation and Abstraction

The first chunking of TECS maps out knowledge of the system with respect to how the system components are designed to achieve the goals of TECS. By making abstractions and aggregations, details are lost but the analysis focuses on the essential abstract principles underlying the functioning of the system and therefore the design choices. The Abstraction Decomposition Space (ADS) is a two-dimensional matrix spanned by levels of abstraction vertically, and levels of part-whole decomposition horizontally. Rasmussen (1994) explains about the ADS: moving up in abstraction and left in level of decomposition does not lead to the same representation as when first moving left in level of decomposition and then up in level of abstraction. Choosing different levels of part-whole decomposition allows the analyst to make different abstractions. This property of the ADS is illustrated here with TECS.

By viewing Figures 1 & 2, it is hard to see the energy relations baked into the design, although the crossfeeds in Figure 2 do give a hint. At this level of part-whole decomposition, the system is viewed in terms of signal processing, gain scheduling and location of the amplitude limits, integrators, etc. A natural abstraction from this representation and level of part-whole decomposition is toward control theoretical analysis covering: control action response, transient response, stability, frequency response and robustness.

In order to make more natural abstractions towards the energy representations (our goal), the level of part-whole decomposition needs to change. Figure 4 shows an aggregation where the components of Figures 1 & 2 have been grouped in such a way that the aircraft independent part of the TECS core becomes the main focus. The arrows in Figure 4 still represent signals but a lot of detail (rate limits, gains, etc.) has been lost including the control theoretical considerations. As a result the main principle is highlighted: control decoupling. The system can be seen in terms of the mode hierarchy that generates the acceleration and path commands, the control decoupling, and the aircraft. These three parts can be visualized on the 'abstract function' level with an analogy to picture the abstract functions. Figure 5 shows the aircraft reservoir analogy (Amelink, 2002, Amelink et al. 2005): the aircraft is shown

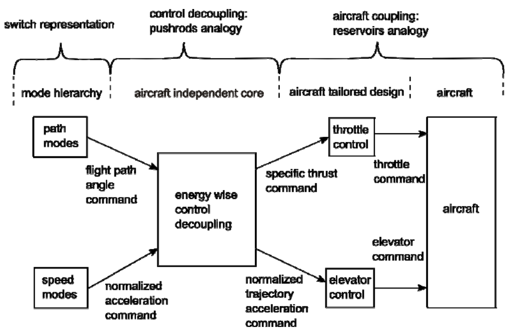


Figure 4: aggregation of TECS into larger functional blocks.

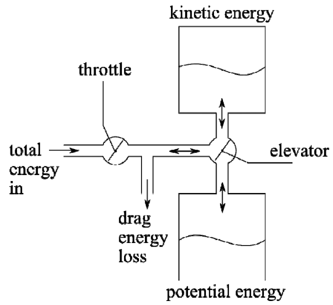


Figure 5: aircraft reservoir analogy.

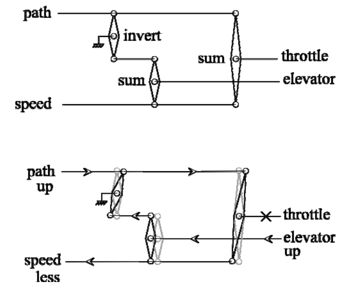


Figure 6:control decoupling pushrods analogy.

as a system storing two kinds of energy in two reservoirs: kinetic energy (speed) and potential energy (height). As shown, the throttle controls the total energy inflow while the elevator controls the distribution of the total energy flow between the two energy reservoirs. The problem of control decoupling is immediately evident: when the (auto-) pilot wants to meet speed and or altitude goals, a coordination of throttle and elevator is needed because neither throttle nor elevator controls speed or vertical path alone. The TECS aircraft-independent core is designed to translate path and speed commands into total energy and energy distributions commands to match the aircraft's energy controls. Figure 6 shows the pushrod analogy that is a simplified mechanical representation of the mathematical relations designed into the aircraft-independent core. One can mentally experiment with the speed and path control inputs to see which throttle and elevator commands are produced. The lower diagram in Figure 6 shows speed and path commands equal and opposite in energy terms, with the result that only the elevator control is needed to exchange speed for height without changes in total energy (throttle). The outputs can be mentally linked to the inputs of the reservoir analogy (Figure 5). The mode hierarchy is best represented by the switching logic already visible at the generalized function level but without the gains to focus on the hierarchical relation between the modes.

From Work Domain Analysis to design

Lambregts (1983a, 1983b) explains that the design of TECS started with the point mass and energy considerations and the idea that the throttle and elevator are the aircraft's energy controls instead of speed and altitude controls. This starting point coincides with the abstract representations in Figures 5 and 6. It is imagined that, from this point a representation similar to Figure 4 emerged, as a first step to implement the abstractions into a conceptual control system at the generalized function level. Figure 4 identifies the energy based control decoupling and draws boundaries between the mode hierarchy, the aircraft independent control decoupling and aircraft tailored design. Figure 3 shows how the system components serve functions to instantiate multiple principles (energy decoupling, bandwidth separation). Finally in Figures 1 and 2 all comes together to form the conceptual control system and explicit signal processing.

Conclusions

In this example work domain analysis we have been able to map a control system onto levels of abstraction and part-whole decomposition. The analysis is however made of an existing system, in retrospect. This does not match the process of designing a new non-existing system and our original goal was to exemplify WDA for designing automation for new systems. The final paragraph of the Work Domain Analysis section describes how the flow of the analysis would be reversed when designing TECS from scratch; starting with the abstract energy representations that need to find their way into the to-be-designed system. It is expected that the analysis for new systems typically start with abstract representations. Abstract representations can be any governing principle that does not have a material presence in the world, like energy flows. Analogies are used to visualize the abstract functions (Figures 5 and 6), to help the reader understand the processes underlying the principle. Mathematics would have been an equally good or a better technical representation but simply stating the law of conservation of energy would not help most people with mental experiments and understand the relations between speed, path, throttle, elevator and energy levels.

The 'first chunking has been experienced as a very helpful exercise to first get things right on the levels of abstraction and later introduce the levels of part-whole decomposition. In such a representation at least two things should be avoided: i) avoid linking everything to everything, and ii) rule out any guess work.

i) In complex systems, especially when optimized, all components interact somehow and arrows can be drawn from many functions to many others. Arrows will start meaning: "somehow relates to …". This is unproductive. The analyst should keep in mind that the essence of functioning needs to be represented by making the right decompositions and abstractions. By making abstractions, details are lost and the essence is highlighted. Similarly, weak couplings disappear and strong ones remain. When all details about the system should be visible, the system should be viewed at the most detailed and least abstract level.

ii) It is tempting to break down a system in those components anyone can see of the top of their head. Unfortunately this does not add to the knowledge of the system. At best it is a start to organize concepts on different levels of abstraction. Above the 'overhead control logic' is exemplified under (4) and it is stated that information is lacking. As a result the analysis gives us little information about the internal workings despite the fact that the overhead logic is represented on three levels of abstraction. The only real information is on the generalized function level where the components it interacts with are linked through part-whole decomposition. If information about a system is missing, some meaningful relations may not be shown. If there is the need to show those relations, the information should be retrieved.

In this particular analysis the reader will note that the links across the levels of abstraction are mostly one-to-one. The explanation for this is that the analysis of TECS mostly spans three levels of abstraction. On the 'generalized function level' TECS is decomposed into those parts that have meaningful abstract functions and dominant one-to-one relation with them. On the 'functional purpose' level these are recombined to the single TECS block.

Perhaps the answer to the question "whether WDA applied to TECS adds value to the design process" can best be given by the reader. If the reader now has some understanding of *how* TECS controls speed and path, *what* the underlying principles are, and *why* they are important, knowledge has been conveyed successfully with this approach. If not, the search continues to find good methods for mapping out the structure of work domains.

Acknowledgements

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

References

Amelink, M. H., Mulder, M., Paassen, M. R. van, & Flach, J. M. (2005). Theoretical foundations for a total energy-based perspective flight path display. *The International Journal of Aviation Psychology*, 15(3), 205–231.

Burns, C. M., & Hajdukiewicz, J. R. (2004). *Ecological interface design*. CRC Press.

Dinadis, N., & Vicente K. J. (1999). Designing functional visualizations for aircraft systems status displays. The *International Journal of Aviation Psychology*, 9(3), 241-269.

Lambregts, A. A. (1983a, October). Integrated system design for flight and propulsion control using total energy priciples. In *AIAA aircraft design, systems and technology meeting.* Fort Worth, TX, USA. (AIAA-83-2561)

Lambregts, A. A. (1983b, August). Vertical flight path and speed control autopilot design using total energy principles. In AIAA guidance and control conference. Gatlinburg, TN. (AIAA-1983-2239)

Lambregts, A. A. (1996). Automatic flight controls concepts and methods. In NVvL annual report.

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

Van Paassen, M.M. (1995). New visualization techniques for industrial process control. In *Proceedings of the 6th IFAC/IFIP/ IFORS/ IEA Symposium on Analysis, Design, and Evaluation of Man-Machine systems* (pp. 457-462). Cambridge, MA: MIT Press.

Vicente, K. J., & Rasmussen, J. (1992, July/August). Ecological interface design: Theoretical foundations. In *IEEE Transactions on Systems, Man and Cybernetics*, 22(4), 589–606.

ON ADAPTIVE AUTIMATION AS A SOLUTION TO THE GLOC CONUNDRUM

Lloyd D. Tripp
Joel S. Warm
Air Force Research Laboratory,
Wright Patterson Air Force Base, OH, USA
Gerald Matthews, Peter Y. Chiu, R. Bruce Bracken
University of Cincinnati
Cincinnati, OH, USA

Gravity-induced loss of consciousness (GLOC) is a major problem facing fighter pilots. In the throes of GLOC, pilots can travel 12 miles without control of their aircraft. The GLOC problem has proven to be difficult to resolve. Solutions involving repeated exposures to GLOC, G-suit pressure manipulation, intense sensory stimulation, and exposure to –G following the GLOC event have been unsuccessful. It is evident that a different approach is needed. One possibility might be an adaptive automation system to warn pilots of the imminence of a GLOC event. A key issue in adaptive automation is the cue to be employed in triggering the onset of automation. Using a centrifuge to simulate gravitational forces together with tracking and math tasks to simulate flight control and navigation, this study assessed the utility of a of cerebral tissue oxygen saturation (rSO_2) measure as a possible triggering mechanism for an adaptive automation recovery system.

Gravity-induced loss of consciousness (GLOC) is a major threat to pilots of modern fighter aircraft. It is brought about by a sudden reduction in cerebral O_2 as a result of increased $+G_z$ force (Tripp, Chelette, Savul, & Widman, 1998). Originally it was thought to consist of 12 sec of complete unconsciousness (absolute incapacitation) coupled with 12 sec of confusion and disorientation (relative incapacitation) However, it is now clear that the course of the GLOC episode is much worse than originally believed. A recent study by Tripp et al. (2006), showed that pilots ceased performing flight tasks approximately 7 sec prior to the onset of the absolute incapacitation phase of the GLOC episode and performance efficiency does not return to baseline values for 55.5 sec following emergence from the relative incapacitation phase (Tripp et al. 2006). Thus, fighter pilots who encounter a GLOC episode can fly approximately 12 miles while not in control of their aircraft. From 1983-1990, GLOC was responsible for the loss of lives of 24 USAF pilots and many other non fatal mishaps (Albery & Van-Patten, 1991).

The GLOC problem is complex and to this point has proven to be difficult to resolve. Measures involving repeated exposure to GLOC, modifications in anti-G suit deflation schedules, intense sensory stimulation deigned to elicit startle responses, and exposure to negative G following the GLOC event have not been successful in attenuating the GLOC problem (Tripp et al, 2006; 2007). In light of these findings, it is evident that a different approach to countering the effects of GLOC is needed.

On a general level, one remedy for reductions in pilot efficiency is *adaptive automation* in which a machine function capable of carrying out duties normally performed by the pilot is activated when the pilot is unable to perform those duties (Parasuraman, Mouloua, Molloy & Hilburn, 1996; Scerbo, 2007; Satchell, 1993; Wickens et al., 1998). Along this line, Scerbo (2007) has argued that in hazardous situations in which pilots are vulnerable and lives are at stake such as GLOC, it is extremely important to have available the capability for an avionics-initiated invocation of automation. Such automation could warn pilots of the imminence of GLOC, thereby alerting them to the need for taking action to decrease the G-force and also assume control of the aircraft if GLOC sets in. As described by Parasuraman, Bahri, Deaton, Morrison, and Barnes (1992), a key issue in adaptive automation is the cue to be used in triggering the onset of automation. This can be achieved through methods based upon critical events, performance measurement, operator monitoring, and physiological assessment. Given the short interval between the onset of high-G and the occurrence of early performance failure and subsequent unconsciousness, it would appear that a physiological cue might be the most effective in the case of a GLOC event.

Although the brain represents only 2% of the human body's weight, it consumes 20% of the body's oxygen requirement (Raichle & Gusnard, 2002). Using near infra-red spectroscopy (NIRS), a non-invasive optical imaging technique for measuring cortical oxygen levels (Gratton & Fabianai, 2007), several studies have shown that noninvasive optical imaging reveals aspects of neuronal activity in the brain (Franceschini & Boas, 2004; Helton et al. 2007; Gratton & Fabiani, 2007; Steinbrink et al., 2000; Tse, Tien, & Penney, 2006). In addition, there are reductions in cortical tissue oxygen saturation during +Gz acceleration (McKinley, Tripp, Bolia, & Roark 2005; Tripp et al., 1998). Accordingly, one goal for the current study was to chart the changes in cerebral oxygen saturation that occur prior to a GLOC episode in order to identify those that could be used operationally to warn pilots of impending loss of consciousness or to provide the trigger for an adaptive automation system that could take over the aircraft until the pilot was able to regain flight control.

Although reductions in cortical tissue oxygen saturation during Gz acceleration have been well documented (McKinley et al., 2005; Tripp et al., 1998), it is critical to note that no data are currently available on the rate of return of tissue oxygen saturation following acceleration offset. Accordingly, a second goal for the present study was to use the NIRS technique to provide the initial examination of the rate of oxygen recovery following Gz offset. Given the prolonged performance recovery time following a GLOC event, one might surmise that the rate of oxygen recovery would be sluggish. That hypothesis was tested in the current study.

Method

Participants Six active duty members of the United States Air Force (three men and three women), participated in the study. They ranged in age from 19 to 34 years, with a mean of 25.5 years. All participants were members of the sustained acceleration stress panel at Wright-Patterson AFB, OH. Participants were required to meet Air Force Flying Class III medical standards prior to their participation.

Facility The study was conducted at the Air Force Research Laboratory's Dynamic Environment Simulator at Wright Patterson AFB.

Acceleration Profiles: A computer control system was utilized to generate a positive Gz acceleration profile. The acceleration profile consisted of a 3G/sec rapid onset to an endpoint of unconsciousness.

In agreement with the flight surgeon, the principal investigator aborted the acceleration profile immediately upon the onset of the GLOC episode. This was followed immediately by a 1.5 sec return to full stop at +1 Gz.

GLOC Criteria: The presence of GLOC was determined using the Whinnery, Burton, Boll and Eddy (1987) criteria that included the following signs: (1) dual eye closure, (2) slumping of the head and upper body, (3) jaw muscle relaxation evidenced by a gaping mouth. All three signs needed to be present in real time surveillance images of the participant in order to determine that the participant had entered GLOC. The principal investigator and the flight surgeon had to be *in total agreement* to make the call. Following the Whinnery et al. (1987) protocol, participants were considered to have regained consciousness when they reopened their eyes. Again, the principal investigator and the flight surgeon had to be in *complete agreement* using real-time observation of the participant.

Performance Tasks A compensatory tracking task used by Tripp et al. (2006) was employed to tap the motor skills required by a pilot to maneuver an aircraft in flight. In addition to the tracking task, participants were required to perform a computation task used previously by Tripp et al., (2006) to tap the higher order cognitive skills needed by fighter pilots to navigate their aircraft. The task involved a series of addition and subtraction problems.

Procedure

Upon arrival at the laboratory, participants were instrumented with a Somentics (Troy, MI) INVOS 4100 Cerebral Oximeter which was used to measure cerebral tissue oxygen saturation (rSO_2) in the right frontal lobe. The self-adhesive oxisensor was affixed to the participant's forehead underneath a flight helmet. Participants wore the standard issue air force flight suit and Gentex helmet with the helmet's visor removed to permit observation of the participant's eyes during GLOC. Oxygen saturation prior to, during, and after the GLOC episodes was measured in terms of percent baseline values. To secure these measures, each experimental session was preceded by a 10-sec resting phase. Following the baseline resting phase, participants remained at rest for another 20-sec prior to the onset of acceleration. Mean oxygen saturation during the initial 10-sec resting phase was the baseline platform from which subsequent oxygen changes in terms of percent baseline were derived. The post-baseline period in which participants remained at rest was necessary to establish the stability of the baseline measure. An unstable baseline platform would render any changes in cerebral oxygen levels associated with GLOC difficult to interpret. Following the O_2 baseline and prior to the acceleration phase, performance baselines for the two tasks were established during a 30 sec testing period in which the gondola was static

Participants were instructed to engage the performance tasks as long as they could before lapsing into unconsciousness, to re-engage the tasks as soon as possible following emergence from the relative incapacitation period, and to continue engagement for five min thereafter. Oxygen saturation was measured continuously from the onset of the pre-acceleration baseline period until the end of the five-min recovery period that succeeded the relative incapacitation phase.

Results

Pre-GLOC Performance The issue of pre-GLOC deterioration in performance was addressed in this study in terms of whether participants *ceased to respond* to either the tracking or the math task prior to the onset of unconsciousness in a GLOC episode. Cessation of response rather than the relative quality of performance was used as the dependent variable because response cessation represents the maximum measure of when pilots are not in control of the aircraft.

One-tailed *t*-tests indicated that the means for the tracking (-3.76 sec) and math tasks (-5.69 sec) were both significantly below zero (or coincident with the onset of GLOC), indicating that in each instance, response cessation significantly preceded the onset of GLOC, $t_{tracking}(5) = 6.71$, $t_{math}(5) = 9.33$, $p_{(Bonferroni\ \text{corrected})} < .05$ in each case. Cessation times for the math and tracking tasks in this study did not differ significantly from each other, $t_{\text{math and tracking}}(5) = 3.46$, $p_{(Bonferroni\ \text{corrected})} > .05$.

Post-GLOC incapacitation performance. The moving window procedures for determining post-GLOC performance recovery times in the tracking and math tasks developed by Tripp et al. (2006) were utilized in this study. These procedures determine the temporal point at which a participant's performance returns to baseline level. A t-test indicated that there was no difference in recovery time between the two tasks. A one-tailed *t*-test indicated that the average recovery time across tasks, **49.45 sec**, differed significantly from zero or immediate recovery from the relative incapacity phase of the GLOC episode,. $t_{\text{one-tail}}(5) = 7.22$, $p_{(Bonferroni\ \text{corrected})} < .05$. However, the average recovery time was not significantly shorter than the 55.50 sec value reported by Tripp and his associates (Tripp et al., 2006) in their initial discovery that participants' performance is degraded for a period of time following GLOC, $t_{\text{two-tail}}(5) = 0.883$, $p_{(Bonferroni\ \text{corrected})} > .05$.

Cerebral Tissue Oxygen Saturation

The mean percent changes from baseline in cerebral oxygen saturation (rSO_2). Data are plotted as a function of successive 2-sec intervals is illustrated in Figure 1. The figure is divided into pre-GLOC, GLOC- incapacitation, and post-incapacitation recovery phases. The acceleration onset landmark reflects the onset of acceleration following the 20-secs of rest that initiated each experimental session. The remaining landmarks reflect the average values across participants for the appearance of math cessation, the beginning and ends of the absolute and relative incapacitation periods, and performance return to baseline.

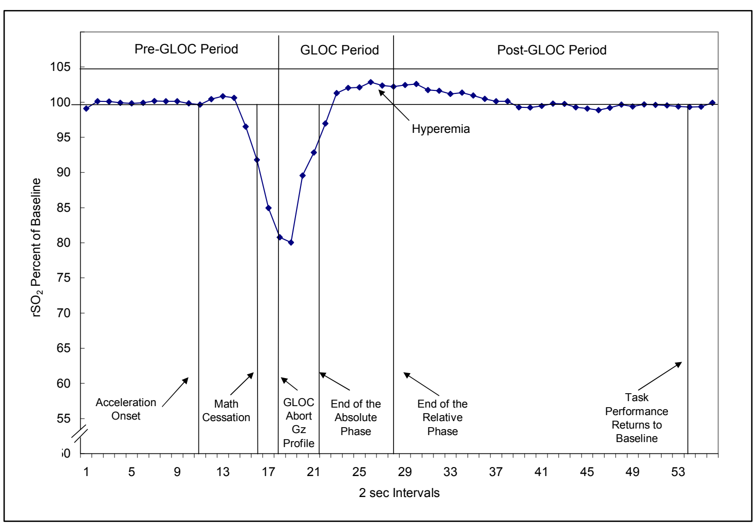


Figure 1. Changes in rSO_2 across time in a composite of the three Gz offset conditions.

A one-way repeated measures ANOVA of the data during the pre-GLOC phase indicated that the rSO_2 scores differed significantly from the onset of the experimental session until the point of GLOC, $F(1.866, 9.330) = 36.715$. $p < .0001$.

Perusal of the figure reveals that rSO_2 remained stable for the 20-sec prior to the onset of acceleration and began a rapid decline at about 28 sec into the acceleration profile until GLOC occurred. Participants ceased performing the math task when the decline in rSO_2 reached approximately 95 percent of baseline and GLOC set in when the decline in rSO_2 reached approximately 80 percent of baseline. A similar ANOVA of the data during the time intervals from the onset of GLOC incapacitation to the point of performance recovery indicated significant differences in the rSO_2 scores across these intervals, F (2.628, 13.140) = 21.108, p <. 0001.

It is clear in the figure that the rSO_2 increased rapidly from the point of deceleration of the centrifuge which occurred at the onset of GLOC, tended to exceed baseline values throughout the relative incapacitation phase, and during the early portions of the recovery phase, and returned to baseline levels at approximately 18 sec into the recovery period, where it remained until performance returned to its baseline level. In all of these ANOVA's, the Box correction was used to compensate for violations of the sphericity assumption (Maxwell & Delaney, 2004).

Discussion

Cerebral Tissue Oxygen and Adaptive Automation

A one goal for this study was to chart the changes in cerebral oxygen saturation that occur prior to GLOC to identify those that could be used by an adaptive automation system (AAS) system in monitoring the pilot in flight in order to warn him/her of impending loss of consciousness and to assume control of the aircraft when the pilot was incapacitated. The results indicate that changes in rSO_2 offer the promise of being useful in this way. The oxygen saturation figure shows that rSO_2 levels tended to remain relatively stable when participants were at rest at +1 Gz but that they declined rapidly from baseline during an acceleration profile. Within that declining function, the figure reveals that participants were unable to process simple mathematical information once rSO_2 fell to a level that was *95 percent of baseline* and that there was an approximately six-sec window before rSO_2 levels fell to *80 percent of baseline* and GLOC set in. These rSO_2 values can be critical landmarks for alerting the pilot that that loss of consciousness is approaching and for triggering the adaptive automation system to assume flight control.

The ability to non-invasively characterize changes in neurophysiology in near real-time in a dynamic flight environment and to use that information to predict a pilot's physiological and cognitive state would be a powerful tool for the high performance aviation community. However, employment of that tool could have serious costs. Consequently, thought is required to determine the scenarios in which it might be utilized advantageously.

Currently, the aircraft cockpit is densely populated with displays that provide the pilot with information regarding variables pertaining to the state of the aircraft such as air-speed, altitude, flight attitude (pitch, roll, and yaw), and hydraulic and electrical system status, but the pilot is virtually blind to pilot-state variables. One remedy might be to provide the pilot with cerebral oxygen information in the form of a three-color light system i.e. green = stable normal rSO2, yellow = compromised cerebral rSO_2, and red = impending GLOC. Such information might serve to reduce the high and often fatal incidence of GLOC that occurs in pilot training (CHI Systems, 2000) by alerting the novice pilot about an impending loss of consciousness so that the pilot could change the energy variable of the aircraft to avoid GLOC, and in cases where GLOC in not avoided, to trigger an auto-recovery system that would return the aircraft to level flight while the pilot recuperates.

A system of this sort might be employed to achieve similar goals in a combat environment. However, there is the question of whether experienced combat pilots would accept such a system because it could be viewed as peripheralizing their roles (Satchell, 1993). Moreover, it may also have negative consequences in combat. The steep climbs and sharp turns that can lead to GLOC in combat are part of the maneuvers often employed by pilots to avoid airborne enemy threats or to gain an offensive advantage over those threats (Shaw, 1985). While taking control from pilots in such situations and returning the aircraft to level flight might avoid the incapacitation induced by GLOC, it could also counter the tactics employed by the pilot to avoid or destroy the enemy and thereby place the pilot's aircraft in harm's way. Hence, it may be more appropriate in the air-combat setting to provide pilots with rSO_2 information that enables them to extend the tactical envelope by allowing them to fly the aircraft to the edge of their physiological capabilities.

A second goal for this investigation was to use the NIRS technique to provide the initial examination of the rate of oxygen recovery following Gz offset and to test the hypothesis, based upon the extended time needed after GLOC incapacitation for performance efficiency to return to pre-GLOC levels, that the rate of oxygen recovery would be sluggish. As can be seen in the oxygen saturation figure, that hypothesis was not confirmed. The rSO_2 level began to increase almost immediately upon termination of the acceleration profile. It rose steeply to a level that exceeded baseline during the relative incapacitation phase and the early portions of the recovery phase, a phenomenon termed *reactive hyperemia,* and settled back to baseline approximately 18 sec prior to the point at which performance efficiency returned to a pre-GLOC level. Two aspects of the time course of return in the level of rSO_2 are noteworthy. The presence of hyperemia is consistent with similar effects observed in medical situations when patients recover from hypoxia. It is due to the dilation of cerebral blood vessels brought about by a build- up of cellular metabolites (Gyton & Hall, 2005). The second key point is that the prolonged period of performance recovery cannot be attributed to delays in the rate of return of rSO_2. Thus, another mechanism must be responsible for the prolonged period of recovery after GLOC. As described by Dirnagl, Iadecola, and Moskowitz (1999), hypoxia causes a critical shortage in brain energy as neurons use glucose and oxygen faster than they are being supplied. At the cellular level, this energy depletion is accompanied by a failure of the Na^+ and K^+ pumps critical for depolarizing neuronal membranes which, in turn, causes conductivity to cease resulting in a significant loss of neural firing. The accumulation of metabolite by-products during ischemic hypoxia delays the recovery of normal neurological function following rSO_2 return. Therefore, it would appear that the need to clear away the metabolic residue of GLOC-induced hypoxia may be responsible for the prolonged performance recovery period even though rSO_2 levels are at or above baseline values. An explanation along these lines reinforces the view that shortening the duration of the GLOC event by fostering a return of blood to the brain would not be a viable alternative to combating the overall GLOC problem, since the performance deficit following GLOC may be more biochemical than hemodynamic in origin.

References

Albery, W.B., & Van Patten, R.E. (1991). Non-invasive sensing systems for acceleration-induced physiologic changes. *IEEE, Engineering in Medicine and Biology, 10,* 52-55.

CHI Systems Inc. (2000, June). *Aircrew experience with G induced loss of consciousness: Problem identification analysis.* (Chi Systems Report No. T0000027.S001). Orlando FL: Author.

Cole, F. (1952). Head lowering in treatment of hypotension. *Journal of the American Medical Association. 27*, 273–274.

Dirnagl, U., Iadecola, C., Moskowitz, M.A. (1999). Pathobiology of ischaemic stroke: an integrated view. *Trends in Neuroscience, 22*, 391-397.

Franceschini, M.A., & Boas, D.A. (2004). Noninvasive measurement of neuronal activity with near-infrared optical imaging. *NeuroImage, 21*, 372-386.

Gauer, O.H. (1950). *The physiological effects of prolonged acceleration,* Chapter VI-B. In German Aviation Medicine Vol. 1. Department of the Air Force, Washington, DC.

Gauer, O.H. (1961). Circulation in man under gravitational stress. In O.H. Gauer & G.D. Zuidema (Eds.), *Gravitational stress in aerospace medicine* (pp. 43-45). Boston: Little Brown.

Gratton, G., & Fabiani, M. (2007). Optical Imaging of Brain Function. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The brain at work* (pp. 65-81). New York: Oxford.

Gyton, A.C., Hall, J. (2005). *Textbook of medical physiology*. (pp. 198). New York: Elsevier Health Sciences.

Helton, W.S., Hollander, T.D., Warm, J.S., Tripp, L.D., Parsons, K., Matthews, G., Dember, W.N., Parasuraman, R., & Hancock, P.A. (2007). The abbreviated vigilance task and cerebral hemodynamics. *Journal of Clinical and Experimental Neuropsychology. 29,* 545-552..

Maxwell, S.E., & Delaney, H.D. (2004). *Designing experiments and analyzing data: A model comparison perceptive* (2nd ed.). Mahwah, NJ: Erlbaum

McKinley, R.A., Tripp, L.D., Bolia, S.D., & Roark, M.R. (2005). Computer Modeling of Acceleration Effects on Cerebral Oxygen Saturation*. Aviation, Space, and Environmental Medicine, 76*, 733-738.

Parasuraman, R., Bahri, T., Deaton, J.E., Morrison, J.G., Barnes, M. (1992). Theory and Design of Adaptive Automation in Aviation Systems. Report No. NAWCADWAR-92033-60

Parasuraman, R., Mouloua, M., Molloy, R., & Hillburn, B. (1996). Monitoring of Automated Systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 37–63). Mahwah, NJ: Erlbaum.

Raichle, M.E., & Gusnard, D.A. (2002). Appraising the brain's energy budget. *Proceedings of the National Academy of Sciences, 99*, 10237-10239.

Satchell, P.M. (1993). Cockpit monitoring and alerting systems. Aldershot, U.K.: Ashgate..

Scerbo, M. (2007). Adaptive automation. In R. Parasuraman & Matthew Rizzo (Eds) *Neuroergonomics: The brain at work* (pp. 239-252). New York: Oxford.
Shaw, R.L. (1985). *Fighter combat tactics and maneuvering.* Annapolis, MD: Naval Institute Press.
Steinbrink, J., Kohl, M., Obrig, H., Curio, G., Syre, F., Thomas, F., Wabnitz,H., Rinneberg, H., & Villringer, A., (2000). Somatosensory evoked fast optical intensity changes detected non-invasively in the adult human head. *Neuroscience Letter, 291*, 105– 108.
Tripp, L.D., Chelette, T., Savul, S., & Widman R.A. (1998). Female exposure to high G: Effects of simulated combat sorties on cerebral and arterial O_2 saturation. *Aviation Space and Environmental Medicine, 69,* 869-874.
Tripp, L.D., Warm, J.S., Matthews, G., Chiu, P.Y., & Bracken, R.B. (October, 2007). +Gz Acceleration loss of consciousness: Use of G-suit pressurization and sensory stimulation to enhance recovery. *Proceedings of the Human Factors and Ergonomics Society, 51, 1296-1300*.
Tripp, L.D., Warm, J.S., Matthews, G., Chiu, P.Y., Werchan, P., & Deaton, J.E. (2006). +Gz loss of consciousness: Time course of performance deficits with repeated experience. *Human Factors, 48,*109-120.
Tse, C.Y., Tien, K.R., & Penney, T.B. (2006). Event–related optical imaging reveals the temporal dynamics of right temporal and frontal cortex activation in pre-attentive change detection. *NeuroImage, 29*, 314-320.
Weil, M.H., & Whigham, H. (1965). Head-down (Trendelenburg) position for treatment of irreversible hemorrhagic shock: Experimental study in rats. *Annals of Surgery, 62*, 905-909.
Wickens, C. D., Mavor, M., Parasuraman, R., & McGee, J. (1998). *The future of air traffic control.* Washington, DC: National Academy of Science
Wiener, E. L., & Nagel, D. C. (1988). *Human factors in aviation*. San Diego, CA: Academic Press.
Whinnery, J.E., Burton, R.R., Boll, P.A., & Eddy, D.R. (1987). Characterization of the resulting incapacitation following unexpected +Gz loss of consciousness. *Aviation, Space, and Environmental Medicine, 58,* 631-636.

ASSESSING NOVEL ADAPTIVE DISPLAYS IMPACT ON PILOT PERFORMANCE

Sean L. Guarino, Karen Harper
Charles River Analytics Inc.
Cambridge, MA 02738

Dahai Liu, Mark Bartosiewicz
Embry-Riddle Aeronautical University
Daytona Beach, FL 32114

Emilie Roth
Roth Cognitive Engineering
Brookline, MA 02445

Dennis Vincenzi
Naval Air Warfare Center Training Systems Division
Orlando, FL 32826-3275

Combat aircraft advances have led to a dramatic increase in the operational tempo facing the Navy pilot, increasing the likelihood for Situation Awareness (SA) failures, biased decision-making and information overload. We designed a system for constructing intelligent adaptive displays to address these issues and, within it, designed and evaluated two interfaces targeted at tactical SA challenges: (1) a *Weapons Employment Zone (WEZ) Display* designed to support awareness of combat geometry; and (2) an *Adaptive Boundary Display (ABD)* designed to warn pilots of impending border infractions that would compromise Rules of Engagement (ROEs). We tested the ability of these displays to improve SA, reduce workload, and improve mission performance in a population of licensed civilian pilots. The WEZ Display significantly improved performance and SA and reduced workload, while the ABD made no significant improvements. We recommend using the WEZ Display to assist novice pilots in understanding and tracking real-time combat geometry.

Introduction

Advances in aircraft performance and operational capabilities have led to a dramatic increase in the operational tempo facing the air combat aviator, reducing the available time to process larger sets of tactically-relevant information and make effective operational decisions based on that information. The technological and information advances of Network Centric Warfare (NCW) have resulted in an explosion in the quantity and complexity of information available to aviators (Shinseki & Caldera, 1999) who must track, monitor, and process information arriving from many disparate information sources and integrate that information into their cognitive decision-making processes (Endsley & Bolstand, 1992). To counter these increasingly complex operational environments, we must develop advanced Human/System Interface (HSI) capabilities that will make optimal use of human operators' cognitive resources.

Modern HSIs do little to address the complexities associated with aircraft operations. While they present the basic information needed for decision-making, they neglect to do so in a manner supporting a proper understanding of the ongoing situation. Most systems do little (if anything) to restrict information flow, regardless of the potential for information overload. Furthermore, most systems neglect to evaluate operator activities, and do nothing to address potential errors caused by decision-making biases. Human decision-making errors that occur when interacting with non-adaptive HSIs can be grouped into three categories: 1) **Situation Assessment Failures**, in which operators fail to detect or correctly interpret the information presented to them, and therefore miss key cues and events central to the formation of an accurate mental model of the situation; 2) **Workload Failures**, in which information overload can lead to information loss and oversight; and 3) **Decision Bias Failures**, in which natural human biases (e.g., a framing of the situation at hand, salience, cues, or heuristics intended to enable rapid decision making) can lead the operator to misinterpreting the tactical situation. To counter these increasingly complex issues, we have developed a middleware system, the *Modular Adaptive Interface Suite* (MAIS), that supports the development of intelligent systems to drive adaptive HSIs, supporting adaptations that optimally use human operators' cognitive resources and promote the formation of accurate mental models of the situation.

In this effort, we used Cognitive Task Analysis (CTA) methods (Mahoney et al., 2008; Bisantz & Roth, 2008; Pfautz & Roth, 2006; Schraagen, Chipman, & Shalin, 2000) including structured interviews, to identify a number of key SA issues arising in air combat aviation. This CTA involved extensive interviews with one former fighter pilot, as well as observations of fighter pilots during training and interviews with current instructors. Several issues were identified, two of which were the focus of the current research: 1) correctly monitoring Rules of Engagement (ROEs) and the location of key geopolitical boundaries to avoid geopolitical incidents during operations near borders; and 2) understanding combat geometry and monitoring the location of the enemy Weapons

Employment Zone (WEZ) and the threat posed to ownship, wingmen and escorted strikers. We designed, developed, and performed a study to analyze display adaptations designed to address these two issues.

First, we developed an *Adaptive Boundary Display* (ABD), designed to address issues aviators have in monitoring geopolitical boundaries while engaging hostile foreign aircraft that attempt to draw them across borders. If provided the opportunity in such situations, the hostile entities will fire missiles. Otherwise, they will attempt to draw military aviators into illegally entering the foreign territory, thus creating a geopolitical incident. Despite ROEs that clearly state the restriction of not crossing the border, these threats can cause the military aviator to lose track of where the border is and inadvertently cross it. To address this issue, we developed an ABD within the Combat Situation Display of the aircraft which dynamically changes the color of the boundary based on the analyzed likelihood that the aviator is going to pass the boundary. The boundary, initially shown in gray, changes to a bright purple and eventually turns red as the need for a reaction from the aviator increases.

Next, we developed a *WEZ Display*, designed to assist aviators in forming an accurate mental model of the combat geometry and understanding the threat posed by enemies. The WEZ defines the missile envelope in front of the aircraft, splitting it into areas in which an enemy is likely to be eliminated and in which an enemy is likely to escape. Realistically, the WEZ is best represented as a bubble in front of the aircraft, as illustrated in Figure 1A. To simplify this representation for aviators, our WEZ display shows a two-dimensional cone representation of the area in front of aircraft (see Figure 1B). The WEZ is characterized in four regions: a *No-Shot* region directly in front of the aircraft, in which the target cannot be safely fired upon; a *No-Escape* region beyond that, in which the target cannot physically escape the missile; a *High Probability of Kill (PK)* region in which a target is unlikely to escape the missile; and a *Low PK* region in which the target's chance to escape is higher, but still not certain. The goal of the military aviator is to achieve a High PK or No Escape shot before providing the opponent with a First Launch Opportunity (FLO) (see Figure 1C). To address this issue, we developed a *WEZ Display* within the Combat Situation Display of the aircraft, showing a constant two-dimensional representation of the known ownship WEZ in front ownship, and a worst-case representation of the enemy WEZ in front of each enemy aircraft icon.

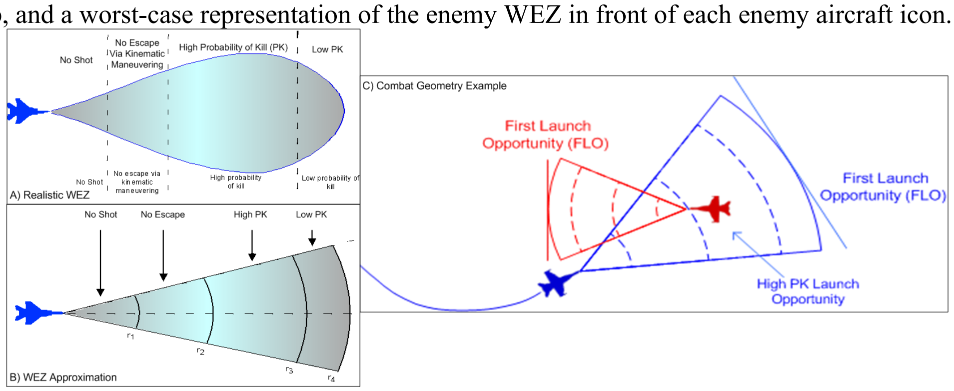


Figure 1: Weapons Employment Zone: A) Realistic WEZ; B) WEZ Approximation; C) Combat Geometry Example

We designed a study to test these adaptive displays as described below. For this study, we made the following hypotheses: 1a) the WEZ Display will improve overall pilot performance by decreasing the enemy FLOs and increasing average shot PK; 1b) the WEZ Display will increase SA; 1c) the WEZ Display will reduce workload; 2a) the ABD will improve overall pilot performance by decreasing instances of border crossings; 2b) the ABD will increase SA; 2c) the ABD will reduce workload; and 3) combining the ABD and WEZ Display will create an improved interaction effect over either display alone.

Methods

Participants: Sixty U. S. citizen participants (54 male, 6 female, average age 22) were recruited from the university community to participate in the study. Participants were all experienced civilian student pilots (59 have private pilot licenses), with an average of 67.3 hours of simulation training, 177.7 VFR hours, 47.3 IFR hours and 57.7 hours of Technological Advanced Aircraft (TAA, i.e., glass cockpit) experience. Most pilots have no air combat experience, although some have related gaming experiences (with an average 5 hours in those types of games).

Experimental Apparatus: In this experiment, we implemented our adaptive displays in a version of OpenEaagles (the Open Extensible Architecture for Analysis and Generation of Linked Simulations, http://openeaagles.org/). Our version of OpenEaagles provides a medium-fidelity air combat simulation environment, including a representation of the Heads Down Display (HDD) of an F-16 with basic instruments and multifunctional displays. We implemented our WEZ Display and Adaptive Border Display as optional components within the OpenEaagles

Tactical Situation Display (TSD). The TSD shows the aviator's ownship at the center of an overhead view and also shows any detected aircraft that are within range of that view. Figure 2A shows a representation of the TSD, and Figure 2B, C, and D shows representations of the Adaptive Border Display, WEZ Display, and combined displays, respectively. Additionally, we enhanced OpenEaagles to record flight path data and a number of key events, such as entering the enemy WEZ, crossing the boundary, or firing a missile at the enemy. In our experimental version of OpenEaagles, each trial immediately ended in one of two cases: 1) when the pilot fired a missile with a PK greater than zero at each opponent, or 2) when the pilot or the escorted striker entered an enemy WEZ.

Figure 2: Tactical Situation Display: A) Basic TSDs; B) TSD w/ ABD; C) TSD w/ WEZ Display; D) TSD w/ ABD and WEZ Display

We implemented an experimental aid component that managed the experiment. Between scenarios, this dialog reported the trial performance to the participant and reminded the participant to fill out between-trial SA and workload measures.

Experimental Tasks and Procedures: Participants first read and signed an informed consent form and filled out a demographic form. Upon completion, they were given a training session on the simulation environment, controls, and concepts of combat geometry and ROEs. Each participant flew an uncontrolled flight scenario to become familiarized with the controls of the simulation environment; this scenario provided no enhanced displays and no border conditions, although it did provide combat situations to familiarize the participant with combat controls. Participants then ran a practice training trial to become familiarized with the format of an experimental trial. This was a three to five minute combat trial that did not count in their final score, in which the participants flew a combat scenario, and performed SA and workload tasks after the scenario. In this trial, participants were also introduced to any additional display features that were available in their condition.

After a break, each participant went through twelve experiment trials. For each trial, the participant flew a fighter aircraft on striker escort missions in a fictional country, called Targetzistan, where his/her job was to protect the striker. Participants were expected to engage in air-to-air combat maneuvers against enemy aircraft coming from Aggressistan. The twelve trials varied in complexity level, following four templates illustrated in Figure 3. Each trial lasted approximately three to seven minutes, and was followed by a SA reconstruction task and a partial NASA-TLX task. After the twelve trials, participants were asked to fill out an overall NASA-TLX survey and an SA questionnaire. It took approximately three hours for one participant to complete the tasks.

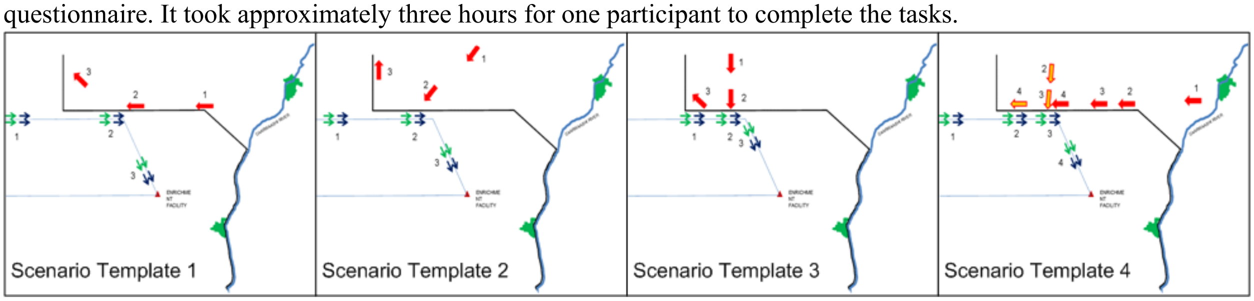


Figure 3: Experimental Scenario Templates

The experiment utilized a 2x2 between-subject design, in which experimental factors were (1) use of the Adaptive Border Display and (2) use of the WEZ Display. This 2x2 design resulted in four display conditions (Adaptive Border Display only, WEZ Display only, both displays, and a control case of no extra displays. Participants were randomly assigned to one of four display conditions (15 participants in each condition).

Data Collection: Three dependent variables were collected, including objective performance, SA, and workload. For each trial, our software automatically measured objective performance based on logged events. This objective performance metric evaluated the participant's success in:

1. **Avoiding breaks in rules of engagement:** -5 points for each instance of breaking the rules of engagement (e.g., each entry into foreign territory).
2. **Minimizing necessary breaks in rules of engagement:** -1 points per every five seconds in foreign territory.
3. **Denying enemy first launch opportunities:** +10 points for denying the enemy first launch opportunity (e.g., avoiding the enemy WEZ and firing on the enemy before the enemy WEZ encompasses the escorted striker).
4. **Achieving an effective launch on enemies:** +0 to +10 points for a launch taken at the enemy, calculated based on the probability of kill (PK) of the shot (0.0 for a 0% PK, +10 for a 100% PK); when there are multiple targets in a trial, this is the mean of the best PK shot taken on each enemy

Note that during training, participants were explicitly informed of the details of the performance metrics, and informed that awards would be based on maximizing performance. The overall performance score was computed by using the four metrics listed above; meanwhile, individual performance metrics were also captured, including Enemy First Strike Opportunities (measuring if the participant or the escorted striker enter the enemy WEZ before the participant makes an effective shot); Shot Probability of Kills (PK) (measuring the probability of shooting down the enemy aircraft) and Boundary Infractions (measuring number of times crossed the border boundary).

SA was measured using an SA reconstruction task designed to assess the level of situation retention after the tasks were completed. In this task, after each trial, participants drew a re-enactment of the scenario, attempting to recall what occurred during the trial. By comparing that data to what really occurred based on our log files, we were able to answer the following five questions:

- Did the participant correctly identify and interpret ownship border crossings?
- Did the participant correctly identify and interpret enemy aircraft border crossings?
- Did the participant correctly identify and interpret enemy first strike opportunities on ownship?
- Did the participant correctly identify and interpret enemy first strike opportunities on escorted aircraft?
- Did the participant correctly identify and interpret own WEZ position?

Based on the answers to these questions, we rated participants from 0-5 to produce an SA score for each trial. Additionally, at the close of the experiment, we assessed overall self-assessment of SA using an SA questionnaire. This questionnaire included fourteen questions using a 0-10 Likert scale to measure participants' perceptions of their own SA. There were 14 questions total in this questionnaire, six were relative to the SA of the geopolitical boundary and eight were related to SA of ownship and enemy WEZ. Score was averaged across the fourteen questions to provide a composite measurement for SA.

Finally, workload was assessed through the NASA-TLX workload rating scale (Hart & Staveland, 1988). The NASA TLX is a subjective workload measure that captures participant's ratings of the mental demand, physical demand, temporal demand, own performance, effort, and frustration level while performing a task. The standard NASA TLX survey is two pages: in the first, participants rate the workload experienced by each of the six specified scales; in the second, they weigh which of the factors they consider more important to measuring workload by performing a pairwise comparison. To shorten the task between trials, we only had participants rate the six scales at the end of each trial. Then, once, at the end of the experiment, they filled out the full NASA TLX for the full experiment.

Results

A between subject Analysis of Variance (ANOVA) was conducted on the three dependent measure. Table 1 summarizes the ANOVA results on all the dependent measures. Based on these results, it is clear that the WEZ display had a significant effect on overall performance score (p=0.008, power of 0.769) and on reducing enemy First Strike Opportunities (p=0.040, 0.541). Figure 4A and 4B illustrate these performance differences for the WEZ Display. Unfortunately, the ABD and interaction between the WEZ and ABD were not significant. Similarly, there was no significant effect from either display on PK of shots taken or on boundary infractions (although, the PK of shots did show a non-significant trend towards improvement when using the WEZ Display (p=0.064, power = 0.459)).

There were two parts of our SA analysis. For individual trials, we analyzed the SA reconstruction task, counting the wrong answers in each case, and finding no significant effects. Additionally, after the experiment, we analyzed an SA questionnaire to access participants' subjective impression of the level of SA provided by their condition. Here, the WEZ had a significant effect on overall SA impression (p=0.013, power = 0.708), and the APD remained non-significant. Figure 4C illustrates these subjective SA differences.

Finally, NASA-TLX measurements were taken to gather subjective impressions of workload during each trial, and in the overall experimental condition. As with SA, the WEZ Display significantly reduced participants' perception of overall workload (p=0.013, power = 0.715), but not of workload during individual trials (although it approached significance, with p=0.051).. Again, the APD remained non-significant. Figure 4D illustrates these subjective SA differences.

Table 1: *ANOVA Analysis Results*

Dependent Measures	**p_{WEZ}**	**Power$_{WEZ}$**	**P_{ABD}**
Overall Performance	0.008*	0.769	0.956
First Strike Opportunities	0.040*	0.541	0.487
Shot PK	0.064	0.459	0.385
Boundary Infractions	0.587	0.084	0.416
Trial SA (Reconstruction Task)	0.238	0.216	0.823
Overall SA (Questionnaire)	0.013*	0.708	0.195
Trial TLX	0.051	0.501	0.644
Overall TLX	0.013*	0.715	0.817

Note. An asterix (*) indicates significant effect.

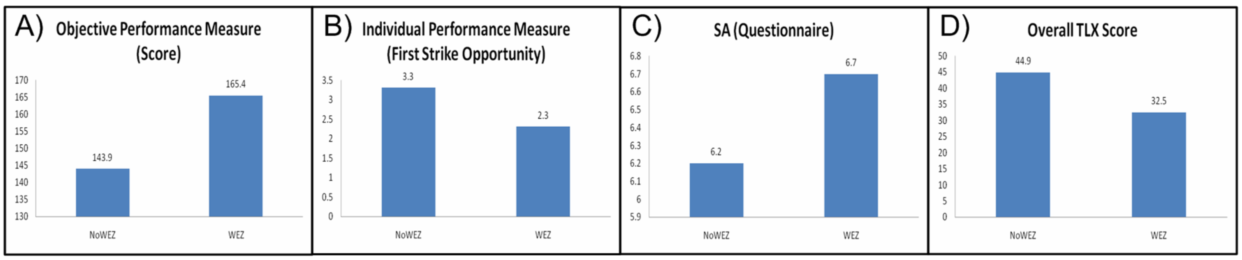


Figure 4: A) Effect of WEZ Display on overall performance, significant at p=0.008; B) Effect of WEZ Display on FSO, significant at p=0.040; C) Effect of WEZ Display on overall SA, significant at p=0.013; D) Effect of WEZ Display on Overall Workload, significant at p=0.013

Discussions and Conclusions

In this study, we investigated two adaptive concepts: the ABD and the WEZ Display. Our first group of hypotheses was that the WEZ Display would enhance performance and SA and reduce workload. Each of these hypotheses was validated by our results. Results showed that the WEZ had a significant effect across most of the dependent measures, including performance measures, overall SA, and overall workload measures. In improving performance, the WEZ Display in particular helped pilots to prevent enemy FLOs. Additionally, it appears that the WEZ Display created a trend of improvement in shot PK (p=0.064). Overall, this evidence indicates that the WEZ Display successfully provided real-time information on the area threatened by enemy aircraft, allowing the pilot to both immediately recognize when his escorted striker was threatened, and maneuver to achieve a high-PK shot without entering the enemy WEZ. With this information aid, aviators can quickly form a more accurate mental model of the combat geometry, enabling them to make a more prompt and accurate response to the situation. Additionally, based on our workload results, the WEZ Display appears to reduce the mental effort required to understand the combat situation and make decisions within the posed combat scenarios.

Our second group of hypotheses was that the ABD would enhance performance and SA and reduce workload. Unfortunately, the ABD had no significant effect on any of these dependent variables. In hindsight, this data is reflective of information we learned in the CTA process, where our SMEs suggested that the issue with monitoring the geopolitical boundary during an engagement occurred primarily because enemies would create a complex “cat and mouse” type of engagement, attempting to draw pilots into geopolitical incidents. In these cases, enemy aircraft would remain just outside of missile range until the pilot was not attending to them and then quickly move in for a strike. In light of this information, we believe that the findings of this study are non-significant because we failed to create this complex type of engagement. In particular, geopolitical incidents only occurred in 30 times across the 720 trials run on all participants in this study. In a follow-up study, we will further investigate the ABD, using more complex scenarios where enemy aircraft attempt to induce the pilots to create an incident by crossing the border, and where enemy WEZ sizes are more threatening.

Our final hypotheses were that the WEZ Display and ABD would combine to enhance both performance and SA and reduce workload over either enhancement alone. Unsurprisingly, the ABD also provided no improvement over the WEZ Display condition. Again, we believe this is because we were not creating scenarios that require the ABD.

Limitations and Future Research

While our results successfully illustrated that pilots can benefit from our WEZ Display, there are a number of limitations on our results. First, as mentioned above, our experiment did not provide the situations required to effectively test the ABD. In a future study, we will test the ABD with more complex scenarios designed to create situations where the ABD would be useful.

Next, our population consisted of a group of civilian pilots, not air combat pilots. By using a population of student pilots at Embry Riddle Aeronautical University, we had experienced pilots, capable of effectively flying scenarios in OpenEaagles with minimum training time. However, while we provided our pilots with initial training in air combat tactics, even Navy novices have more education in recognizing the details of actual combat situations. Clearly, our experiment would be more convincing if we had a military pilot population, and we recommend future research to investigate such a population. Even a small population of retired military pilots showing similar results for WEZ Display would be useful.

A third limitation of our research was the realism of the experimental system. While OpenEaagles provides a solid medium-fidelity air-to-air combat simulation environment, it clearly does not capture all of the complexity of a real aircraft. To begin with, in our experiment we only included a Heads Down Display (HDD) to simplify the task for our non-combat pilots; in a real aircraft, pilots would of course also have an out the window display and a Heads Up Display (HUD), each of which could distract them from our adaptive display. Furthermore, our TSD had no underlying map. Such displays in real aircraft often have a map in the background. This is a significant limitation because it is unclear what effect the map would have on observing both static and adaptive boundary lines. Clearly, performing this experiment in a high-fidelity simulation environment would produce more reliable results.

Nevertheless, based on our results, the WEZ Display appears to be a useful feature for increasing pilot performance in complex air combat situations, helping pilots to understand complex combat geometry problems. We recommend future research investigating the application of these adaptations in more realistic simulation environments and, ultimately, in existing and future military aircraft, and with a population of Navy aviators. We also recommend the development of further adaptations to address other key SA and workload issues plaguing Navy aviators, such as recognition of potential aircraft energy issues, flow issues, and combat timeline issues. Finally, we recommend the application of our intelligent adaptive display technology to development in associated high-tempo domains, such as rotary aircraft piloting or unmanned air vehicle (UAV) piloting.

Acknowledgements

This work was performed under US Navy contract number N61339-07-C-0025 with the Navy Air Warfare Center Training Systems Division (NAWCTSD). The authors thank Dr. Dennis Vincenzi and Mr. Richard Plumlee for their support and direction on this project, and their aid in finding SMEs for our CTA process. Additionally, the authors thank LCDR Michael McLeod (Ret.) for serving as our central SME in this process. Finally, we would like to thank Mr. Randy Levine from ASC/SIMAF for providing support in our use of OpenEaagles for this effort.

References

Bisantz, A. & Roth, E. (2008). Analysis of Cognitive Work. *Review of Human Factors and Ergonomics,* 3.

Endsley, M. R. & Bolstand, C. A. (1992). Human Capabilities and Limitations in Situation Awareness. In *Proceedings of AGARD: Combat Automation for Airborne Weapons Systems: Man/Machine Interface Trends and Technologies*. Neuilly Sur Siene, France.

Hart, S. G. & Staveland, L. E. (1988). The Development of the NASA Task Load Index (TLX): Results of Empirical and Theoretical Research. In P. A. H. Meshkati (Ed.), *Human Mental Workload* (pp. 139-183). North Holland, Amsterdam.

Mahoney, S., Pfautz, J., Roth, E., Fichtl, T., Powell, G., Guarino, S. et al. (2008). Cognitive Systems Engineering for Evolvable Human-in-the-Loop Data Fusion. In *Proceedings of 52nd Annual Meeting of the Human Factors and Ergonomics Society*.

Pfautz, J. & Roth, E. (2006). Using Cognitive Engineering for System Design and Evaluation: A Visualization Aid for Stability and Support Operations. *International Journal of Industrial Engineering,* 36(5), 389-407.

Schraagen, J. M. C., Chipman, S. F., & Shalin, V. L. (2000). Introduction to Cognitive Task Analysis. In J. M. C. Schraagen, S. F. Chipman, & V. L. Shalin (Eds.), *Cognitive Task Analysis* (pp. 3-24). Mahwah, NJ: Lawrence Erlbaum Associates.

Shinseki, G. E. K. & Caldera, L. (1999). *The Army Vision Statement.* US Army.

COMPARING TUNNEL-IN-THE-SKY DISPLAY ON HDD AND HUD FROM TASK OCCUPATION POINT OF VIEW

Kohei Funabiki, Tomoko Iijima and Hiroka Tsuda
Japan Aerospace Exploration Agency
Tokyo, Japan

A series of flight simulations was carried out to investigate the causal factors of attention capture, focusing on a traffic detection task while following a curved trajectory using a Tunnel-in-the-Sky display. The location (head-up or head-down) and size of the display were varied, and traffic detection time and path tracking performance were measured. The results show that the HUD gave the best path tracking at the expense of traffic detection performance, and supports the hypothesis that using a limited viewing volume and high display gain with a Tunnel-in-the-Sky display induces pilots to rely on precise guidance cues instead of the "tunnel" itself, consequently focusing much attention on the control task.

Since the concept of instrument flight was introduced a half-century ago, there have been a number of efforts to develop a display that provides visual cues as intuitive as out-of-window visual cues. A perspective image of a nominal flight trajectory is a typical concept for such a display. In spite of the fact that several such displays have been developed with different names—Channel (Kanal)-, Corridor-, Highway-, Pathway- and Tunnel-in-the-Sky Display—these all share the same basic idea and concept.

The advantages of the Tunnel-in-the-Sky over conventional 2D display formats are reported to be (1) higher tracking performance in manual flight, (2) lower workload, (3) enhanced situation awareness, and (4) greater suitability for curved trajectories. A number of prominent research activities during the past 30 years have led to display design strategies becoming almost established (Grunwald, 1984, Mulder, 1999, and Newman, 2003). Improvements in on-board graphics generation capability and the spread of satellite navigation systems has led to a recent surge in the flight evaluation of such displays, and a few commercial products for general aviation have now appeared on the market. However, although the Tunnel-in-the-Sky has become a common image in near-future advanced cockpits, there are still several issues to be clarified before it can play a dominant role in commercial transport aircraft instrumentation. The most widely recognized of these issues is "Attention" or "Cognitive Capture".

Many studies have investigated attention capture and path tracking performance issues with a variety of display design parameters (Fischer *et al.*, 1980, Wickens *et al.*, 1998, 2003, and Ververs *et al.*, 1998), primarily for flight path guided HUDs (Head-Up Display) but also for Tunnel-in-the-Sky displays. When a HUD is used for precision approach and landing in low visibility conditions, the display presentation should be conformal irrespective of whether conventional flight path cues or a Tunnel-in-the-Sky are depicted. On the other hand, if a Tunnel-in-the-Sky is used in visual (composite) flight, a conformal presentation may be not necessary, and the symbology may be presented on either a HUD or an HDD (Head-Down Display). In such operations, path tracking performance requirements could be relaxed compared with the approach phase, while the ability of the pilot to spot other traffic should be the same as in conventional visual flight. The question is: What is the best combination of display parameters when using a Tunnel-in-the-Sky—conformal or non-conformal, head-up or head-down?

In this paper, we temporarily define "attention capture" to mean "the pilot task is occupied by flight control tasks without notice." This means that if the pilot realizes that his or her attention has become "captured", he or she can recover from the situation by intentionally modifying scanning behavior. In other words, if pilots are sufficiently trained to pay appropriate attention to each item of information, "cognitive capture" should not occur so often. On the other hand, even if a pilot can maintain a good level of optimal scanning behavior, he or she is forced to pay more attention to an instrument if it has poor readability. In this research, we refer to this phenomenon as "attention occupation", distinguishing it from "cognitive capture".

The initial target of this research was to clarify the causal factors of "cognitive capture" associated with Tunnel-in-the-Sky displays. However, in a preliminary experiment, while "cognitive capture" phenomena were not clearly observed in the controlled experimental environment, "attention occupation" was clearly apparent with a degree that seemed to have a close relationship with display configuration. This paper reports the results of a series of flight simulation experiments to investigate these causal factors of “attention occupation” by varying display location and gain.

Assumptions and Hypotheses

Basically, we assume that some display design factors, which differ between the HUD and the HDD, have an affect on how a pilot divides attention between flight control and other tasks. In this research, we select traffic detection as a secondary task. There are several design parameters that might affect both flight control and secondary task performance.

Display Gain and Viewing Volume: It is well known that while a higher display gain increases path tracking performance, it may degrade the stability of the closed-loop system. A study has revealed that the display gain of the conformal HUD is too high and might result in large deviations. Regardless of whether path tracking performance is good or poor, if path error is magnified the pilot’s attention becomes largely occupied by the tracking task, and so attention given to traffic detection is reduced. On the other hand, as the display size is limited, a higher gain results in narrower “viewing volume” of the tunnel image. In general, the viewing volume of a Tunnel-in-the-Sky depicted on an HDD ranges from 60 deg to 80 deg, versus a maximum of 40 deg for a conformal HUD. It is anticipated that a narrower viewing volume reduces position awareness, and may also affect traffic detection.

Location and Focal Point: As HUD symbology is projected at infinity and does not requires the pilot to go “heads-down” to scan instruments, the scanning load for a HUD may be lower than for an HDD. This may led to a HUD giving both increased path tracking and traffic detection performance.

Symbol Overlap: Because HUD symbology is presented superimposed overlapped on the out-of-the-window scene, there is a risk that traffic may be masked by symbols. Particularly in a flight simulation environment, the brightness and color of HUD symbols cannot be well controlled.

Considering these issues, the following hypothesis were set:

1. If display gain is as high as a conformal HUD, the resulting magnification of error will capture pilot attention, and path tracking performance will improve. A reduction in viewing volume may reduce position awareness.
2. If the display focuses pilot attention on the control task, traffic detection performance will decrease.
3. The location and infinity focus of a HUD may reduce scanning load and improve traffic detection or control performance. Consequently, presenting guidance symbols on a HUD but with a reduced display size may enable pilots to pay greater attention to traffic detection while giving a similar level of path tracking performance as an HDD.

Experiment

A set of piloted flight simulations was conducted to investigate the causal factors of attention occupation by comparing HUD and HDD in a task to follow a curved flight path while looking out for traffic.

Simulation Set Up

A research simulator at the JAXA Flight Research Center was utilized. The FOV (field of view) of the out-of-window visual display for a left-seated pilot is –100 to +21 deg horizontal and 35 deg vertical, realized by three SXGA-resolution visual system channels presented by a gapless WAC (Wide Angle collimation) system. HUD symbology was overlaid on a 90% transparent gray-colored “pale” background plane placed in the visual scene as a 3D object to enhance its legibility. The simulated aircraft used the flight dynamics of a Dornier Do.228-202 twin turboprop commuter airplane. All the pilots who participated in the simulation were experienced with this type of aircraft and had actual flight experience with Tunnel-in-the-sky displays (HDD).

Traffic was presented in the visual scene five or six times per flight, one airplane at a time. After entering the scene, traffic aircraft continued flying until either the pilot pressed the microphone push-to-talk (PTT) switch or until 30 seconds had elapsed, before being removed from the scene. A marker was presented at the edge of the cockpit front window as small pink semi-transparent square subtending 2x2 deg from the pilot’s eye point. Markers were presented one at a time, and remained until either the pilot pressed the PTT switch or until 15 seconds had elapsed.

The presented position was varied between nine fixed locations on the upper, lower, left and right edges of the window.

Two types of route were prepared, each consisting of four curved and straight segments with a –4 deg approach path. Pilots were instructed to fly along the displayed trajectory paying sufficient attention to other traffic.

Display Symbology and Geometry

Figure 1 shows an image of the Tunnel-in-the Sky presented by the HDD. Table 1 shows the basic geometry and characteristics of the Tunnel-in-the-Sky display used in the simulation.

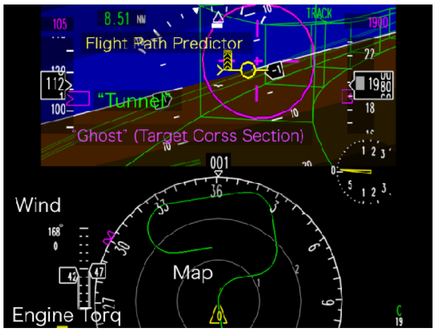


Figure 1. *HDD Symbols*

Table 1. Tunnel-in-the-Sky Characteristics

Parameter	*Experimental Value*
Cross-section Size	100 x 100m
Frame Interface	250m
Tunnel Visual Presentation	Frame (within 0.5NM) Contour (within 4.0NM)
Flight Path Predictor Horizontal Prediction	5 seconds prediction with bank angle
Flight Path Predictor Vertical Prediction	Initial response slaved to pitch attitude

Experiment-1

Display Configuration

A set of flight simulations was conducted to compare four types of display including HDD and HUD, varying display size. A total of six pilots participated, and each pilot flew twice for each type of display. Figure 2 shows the display configuration for the experiment.

HDD-Normal (HDD): A Tunnel-in-the-Sky integrated with a traditional PFD (Primary Flight Display) and horizontal situation display is shown on the instrument panel.

HDD-Large (HDD-L): For comparison with HDD-Normal, the display size is enlarged and the viewing volume narrowed. This results in a display gain three times greater than the HDD-Normal display, and 20% less than the HUD-Conformal case.

HUD-Conformal (HUD-CF): The Tunnel-in-the-Sky is integrated with a traditional “primary mode” HUD format, rather than an “approach” mode format. A conventional PFD and horizontal situation display are shown on the instrument panel.

HUD-Non conformal (HUD-NC): This presents the HDD symbology overlaid on the visual scene, but smaller in size than in the HUD-CF display. The flight path symbol and artificial horizon are not conformal with the visual scene. The symbols are shown in a monochrome, with the same color as in the HUD-Conformal display. Due to the nature of this display, serious clutter occurs during the final approach phase.

HDD	HDD-L	HUD-CF	HUD-NC
Viewing Volume=70	Viewing Volume=36	Viewing Volume=36	Viewing Volume=70
FOV=14	FOV=19	FOV=36	FOV=19

Figure 2. *Display Configurations for HDD and HUD Comparison*

Results

Figure 3 shows the mean RMS values of horizontal tracking error across the display types. HUD-CF and HDD-L show a significant reduction in horizontal error compared to the HDD case (P=0.0010, 0.0047 respectively). Figure 4 shows mean traffic detection times. HUD-CF, HDD-L and HUD-NC show significant increases in traffic detection time compared to the HDD case (P=0.0033, 0.0062, 0.029 respectively). The difference between the HUD-NC and HDD-L cases is also significant (P-0.018). Figure 5 shows mean number of missed markers. A marker was considered as "missed" if the pilot did not push the PTT switch within 20 seconds of its appearance. HUD-CF shows a significant increase in the number of missed markers over the HDD and HUD-NC cases (P=0.032 and P-0.044 respectively).

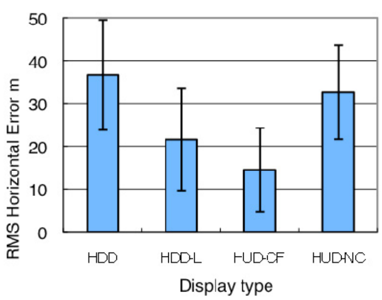


Figure 3. *Horizontal Path Error*

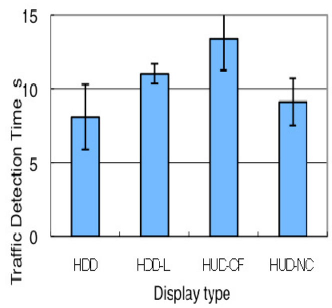


Figure 4. *Traffic Detection Time*

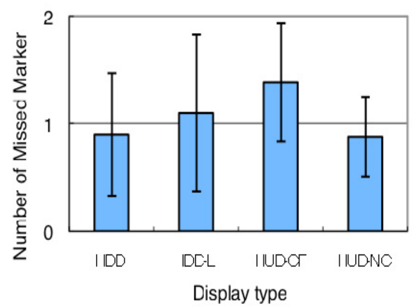


Figure 5. *Missed Marker*

Discussion

The lower path tracking error in the HUD-CF and HDD-L cases shows that a higher display gain enhances path tracking performance. Although most of the subjects complained of an oscillatory tendency with high display gain, they admitted that their tracking performance was better. The higher traffic detection times with these displays shows that this improved tracking performance required greater attention on the display.

Two major guidance and control cues were presented, and control behavior could be to use these in combination; i.e. tracking the target frame ("Ghost") by the Flight Path Predictor, and navigating the ownship within the tunnel by looking at the shape of the tunnel. If precise control is required, the former cue is dominant, and the closed-loop control gain becomes higher. On the other hand, if the path tracking requirement is relaxed, as in the present experiment, the latter cue plays the major role. In this case, a pilot may adopt an "Error Neglecting Control" strategy (Tueunissen & Mulder 1995), resulting in poorer tracking performance.

In the HDD and HUD-NC cases, the subjects seemed to use an error-neglecting strategy. The observed mean horizontal error of around 35m, slightly less than half-tunnel width of 50m, supports this supposition. On the other hand, the limited viewing volume of the HUD and HDD-L displays degrades the position error information that could be perceived from the tunnel, and forces the pilots to abandon the error-neglecting strategy. Consequently, in these cases they might have to rely on the Flight Path Predictor – Ghost cue.

Some subjects complained that a narrower horizontal viewing volume limits the display of future trajectory, especially in curved flight segments. In this particular trial, subjects could not anticipate the descent point well beforehand, and this may have caused them to pay increased attention to the display.

These results can be compared with the similar previous research by Wickens (2003), who found that the higher the display gain, the poorer the tracking performance. Although Wickens's findings appear to be completely opposite to those here, both experiments support the hypothesis that a higher display gain causes scattering of tunnel the symbols over the field of view and prevents the pilot from acquiring position information or guidance cues from the tunnel.

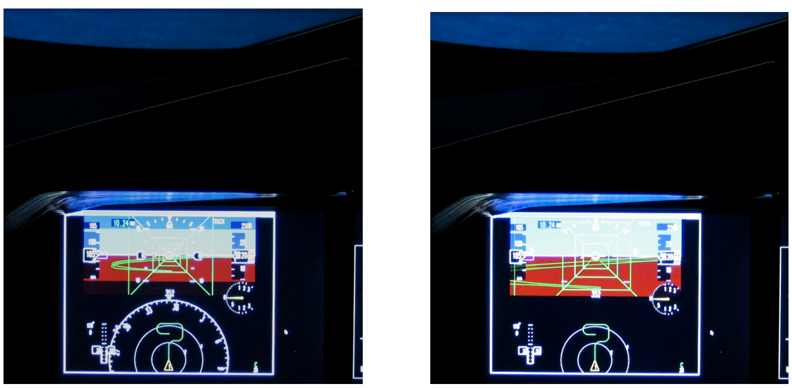
There is another possible explanation for attention occupation considering the effect of the "Ghost" center marker. Wickens's "sliding box" symbol has only a square frame and lacks a center marker, while our corresponding Ghost symbol has a center marker. In the HDD and HUD-NC displays, the size of the Ghost symbol in the FOV is 5deg and 8 deg respectively, versus 20 deg for the conformal HUD. This means that the circular part of the Ghost symbol is outside of the Useful Field of View, degrading the information it conveys. Consequently, the pilot might have had to rely on the Flight Path Predictor – Ghost (center marker) cue, and abandon the error-neglecting control strategy.

As anticipated, all subjects complained that the HUD symbology might mask traffic, or at least that they commented that they sometimes confused HUD symbol elements with traffic. This may have contributed to the longer traffic detection times observed in the HUD-CF case, but considering that HDD-L resulted in a similar traffic detection deficiency, the effect is considered to be small in comparison with the effect of viewing volume. The increased number of missed markers in the HUD-CF case also supports the hypothesis that interaction between HUD symbols and traffic is small, because the markers were presented outside of the HUD symbols.

Experiment-2

Display Configuration

An additional experiment was conducted to examine effect of display size on the trajectory tracking task. In contrast to Experiment-1, neither traffic nor markers were shown, and the subject pilots were instructed to follow the trajectory as precise as possible. The Ghost symbol was removed from the display so that the subjects had to acquire position information from the tunnel, not from guidance symbols. As shown in Figure 6, the HDD and HDD-L displays (FOV= 19deg, same as the HDD) were compared by four pilots.



HDD HDD-L

Figure 6. *Display Configuration*

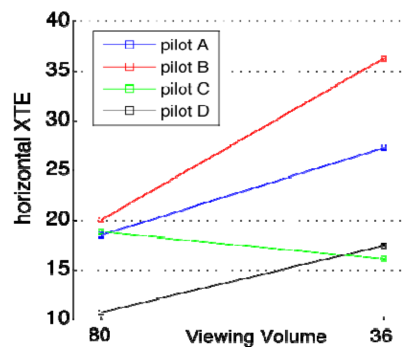


Figure 7. *Horizontal Tracking Error*

Results and Discussion

Figure 7 shows horizontal tracking error. Although it is not significant (P=0.13), three of four pilots had degraded horizontal tracking error for HDD-L, and all the pilots admitted that position awareness was degraded with the HDD-L display, especially in curved flight path sections. This result is consistent with Wickens (2003), and also supports the hypothesis that the center marker of the Ghost symbol plays major role in control strategy selection (error neglecting control or error suppressing control).

Summary and General Discussion

The results support Hypothesis 1 that high gain displays (HDD-L and HUD) give better tracking performance. However, it is considered that the observed high closed-loop gain was the result not of the high display gain but of the narrow viewing volume, which prevented pilots from adopting an error-neglecting control strategy. This view is supported by a comparison with the cases of HDD and HDD-L without the Ghost symbol, which gave opposite results to the with-Ghost cases. The center marker of the Ghost symbol, if it is present, seems to play major role on control strategy selection when display gain is high.

Hypothesis 2, that attention demanded by path tracking will reduce traffic detection performance, is supported by the result that the displays with a narrow viewing volume gave greater traffic detection times, as well as a greater number of missed peripheral markers.

The results do not support Hypothesis 3, namely that presenting guidance symbols on a HUD but with reduced display gain enables the pilot to pay optimized attention to the traffic detection while path tracking performance remains with the same as with an HDD. Although some pilots commented that the scanning load was reduced in the HUD and HUD-NC cases, no objective data were obtained. The HUD-NC display, which presents HDD symbols on the visual scene, showed almost the same tracking and traffic detection performance as the HDD, while the use of only a single color caused much clutter.

Although not detailed in this paper, there are many inconsistencies between the results of the present experiment and previous research. This indicates that attention occupation is influenced by many factors such as the subjects themselves, the detailed design of the display, types of event, and simulation fidelity. Furthermore, it should be noted that a pilot who participated this experiments flew with a HMD with similar tunnel-in-the-sky symbols in an actual aircraft, and commented that spotting traffic through the HMD in an actual flight environment is much easier than in the simulation. Further study is strongly required.

Conclusions

This paper describes a simulation experiment to investigate the causal factors of "task occupation", namely a situation in which "a pilot cannot pay sufficient attention to tasks other than control", in particular dealing with traffic detection while tracking a curved trajectory. As was hypothesized, the use of a conformal HUD resulted in reduced traffic detection performance. This is considered to be caused by the reduced viewing volume of the perspective symbols of a Tunnel-in-the-Sky. A proposed non-conformal HUD showed no superiority over an HDD for detecting traffic.

Acknowledgements

The authors thank all the pilots who participated the experiment. We also acknowledge the contribution made by Bart van Marweijk for building an analysis tool and for pointing out a problem in the former analysis.

References

Fadden, S., Ververs, P. M., & Wickens, C. D., (1998). Costs and benefits of head-up display use: A meta-analytic approach. Proceedings of the Human Factors and Ergonomics Society 42ndAnnual Meeting: Chicago, IL, 16-20.

Fischer, E., Haines, R. F., & Price, T. A., (1980). Cognitive issues in head-up displays (No. 1711). Moffett Field, CA: NASA Ames Research Center.

Funabiki, K., Muraoka, K. and Iijima, T. (2001). Tunnel-in-the-Sky Display for Helicopters, Proc. AIAA Modeling and Simulation technologies conference and Exhibit.

Funabiki, K. and Iijima, T. (2007). Attention Allocation in Tunnel-in-the-Sky on HUD and HDD for Visual Flight, 26th Digital Avionics Systems Conference.

Grunwald, A.J., Robertson, J.B. and Harfield, J.J., (1981). Experimental Evaluation of a Perspective Tunnel Display for Three-Dimensional Helicopter Approaches, AIAA Journal of Guidance and Control, NOV-DEC 1981.

Grunwald, A.J., (1984). Tunnel Display for Four-Dimensional Fixed-Wing Aircraft Approaches, AIAA Journal of Guidance and Control, MAY-JUNE 1984.

Mulder, M., 1999, Cybernetics of Tunnel-in-the-Sky Displays, Delft University Press.

Newman, R. L., (2003). Pathway Displays: A Literature Review, Proc. 22nd Digital Avionics Systems Conference.

Prinzel, L.J. III and Risser, M., 2004, Head-Up Displays and Attention Capture, NASA/TM-2004-21300.

Theunissen, E., and Mulder, M, (1995). Error-neglecting Control with Perspective Flight Path Displays. In 8th International Symposium on Aviation Psychology (pp. 110-115). Columbus, OH: Ohio State University.

Ververs, P. M., & Wickens, C. D., (1998). Head-up displays: effects of clutter, display intensity, and display location on pilot performance. International Journal of Aviation Psychology, 8(4), 377-403.

Wickens, C. D., Alexander, A. L., and Hardy, T. J., (2003). The Primary Flight Display and Its Pathway Guidance, 2003, Workload, Performance, and Situation Awareness, AFHD-03-2, University of Illinois.

Wickens, C. D., Fadden, S., and Merwin, D., (1998). Cognitive Factors in Aviation Display Design, Proc. 17th Digital Avionics Systems Conference.

ECOLOGICAL SYNTHETIC VISION DISPLAY TO SUPPORT PILOT TERRAIN AWARENESS

C. Borst, M. Mulder, M. M. (René) van Paassen
Delft University of Technology, Faculty of Aerospace Engineering
Kluijverweg 1, 2629HS Delft, The Netherlands

A Synthetic Vision Display is generally believed to support pilot terrain awareness. Many studies have shown, however, that perspective views are biased, which can cause pilots to make judgment errors regarding the relative location, height, and ultimately the avoidance of terrain obstacles. Therefore, this system is usually backed by terrain avoidance systems that provide explicit resolutions to circumvent conflicts. They are, however, far from optimal regarding terrain awareness as they fail to present the rationale of the automation. This paper presents an extension to a Synthetic Vision Display that promotes pilot terrain awareness by means of overlays that reveal the functional meaning of the terrain. It is designed to effectively deal with terrain conflict situations while preserving the freedom of maneuvering as much as possible. An experiment showed that the overlays improved pilot situation awareness and decision-making (in unanticipated events) as compared to a command-based interface counterpart.

Since the introduction of the glass cockpit and the technological advances in computing and sensing, the designers of aviation human-machine interfaces can almost freely create the pilot interface that should support situation awareness (SA). Traditional approaches to system and interface design have the tendency to either 1) show as much information as possible on single interfaces in a way that corresponds to a pilot's mental model (Spitzer, 2001), or, 2) to automate and hide the reasoning behind decision-making by showing pilots explicit resolution commands (Pritchett, 2003). From these technology-driven approaches to interface design two systems have emerged in the field of terrain awareness: the Synthetic Vision Display (SVD) the Terrain Awareness Warning System (TAWS).

An SVD shows pilots a perspective view on the surrounding terrain overlaid with primary flight status data. Although it presents data in an intuitive way, perspective views are biased which can cause pilots to make judgment errors regarding the relative location, height, and ultimately the avoidance of terrain obstacles (Wickens, 2002). Therefore, an SVD is usually backed by a TAWS that provides terrain collision warnings and escape maneuver commands. This system, however, is far from optimal regarding pilot terrain awareness as it fails to present the rationale of the automation that could help pilots to understand the nature of the issued alerts and commands (Bisantz & Pritchett, 2003).

Recent studies in SA and interface design claim that the Ecological Interface Design (EID) framework has the potential to support SA and improve decision-making, even in unanticipated situations (Flach, Mulder, & Van Paassen, 2004; Burns, Jamieson, Skraaning, & Kwok, 2007). Previous research in terrain awareness and EID revealed that showing the 'internal' (aircraft performance) and 'external' (terrain) constraints to flight is effective in promoting SA and decision-making (Borst, Suijkerbuijk, Mulder, & Van Paassen, 2006; Borst, Sjer, Mulder, Van Paassen, & Mulder, 2008). Although these designs and the results of pilot-in-the-loop experiments were promising, it was not always clear whether the improved pilot performance and SA could be fully attributed to the ecological interface. The experiment designs compared the ecological interfaces to conventional pilot interfaces, which were not always designed for the same purpose.

This paper describes the design and evaluation of an Ecological Synthetic Vision Display (ESVD) that extends an SVD with functional overlays that show pilots how their maneuvering possibilities are constrained by their own aircraft performance and surrounding terrain. Additionally, an experiment design will be presented that aims to make a fair SA comparison between the ESVD and a viable design alternative.

Enhancing the SVD

A work domain analysis for terrain awareness has been conducted in earlier work (Borst et al., 2006, 2008). The analysis showed that terrain awareness can be achieved by appropriately dealing with the external constraints, imposed by the terrain, and the internal constraints, imposed by the aircraft's climb and turn performance. Analysis showed that in order to effectively promote terrain awareness, an SVD should be enhanced with the following constraints: aircraft maneuvering performance, aircraft energy management, and aircraft-terrain separation.

Requirements

In general, the features on an ecological interface represent the constraints of the work domain. To map the constraints of the work domain into a visual form, EID provides guidelines for an interface design process rather than an interface blueprint. When enhancing an existing interface, however, the design of the visual form is also constrained. The designer is limited to create enhancements that are compatible with the interface “template”. The template of a perspective display enables pilots to perceive relative distances, heights and locations between objects by means of relative angles (with respect to a horizon line), occlusion, and the relative size of objects (Wickens, 2002). To enable pilots to effectively relate the internal aircraft performance constraints to the external terrain constraint on a perspective display, the aircraft performance constraints need to be translated into angular descriptions whenever possible. In the following, the constraints of a Cessna Citation 500 aircraft, of which a non-linear, 6 degree-of-freedom mathematical model was available, will be explored. The content of all plots and figures in this paper are based on that model. Note that for other aircraft the method will be exactly the same.

Exploring the Constraints

Maneuvering In the vertical plane, the aircraft’s optimal climb performance is important for terrain avoidance (Asselin, 1997). The steepest climb angle relative to the air is function of the altitude, weight, roll angle, aircraft configuration, and aerodynamic efficiency. The steepest climb angle relative to the terrain (γ_k^{OC}) can be obtained by adding the influence of wind speed and wind direction to the aerodynamic climb (Asselin, 1997). The turn dynamics of an aircraft, expressed in terms of the ground-referenced turn radius, in coordinated level turns is a function of the airspeed, roll angle, wind speed and wind direction. An important constraint on the turn radius is the maximum allowable vertical load factor n_z, which determines the maximum allowable roll angle. In wind conditions, the ground track of a level turn performed at a constant bank angle will be deformed (Figure 1(b)).

Energy Management The total energy state of an aircraft determines the opportunities for maneuvering. On a perspective display, pilots can perceive the rate of energy exchange by means of the Total Energy Angle (Amelink, Mulder, Paassen, & Flach, 2005). At a constant total energy level, the rate of energy exchange indicates how much potential energy an aircraft is gaining at the cost of kinetic energy and vice versa. Increasing the total energy of an aircraft is done by adding thrust to the system.

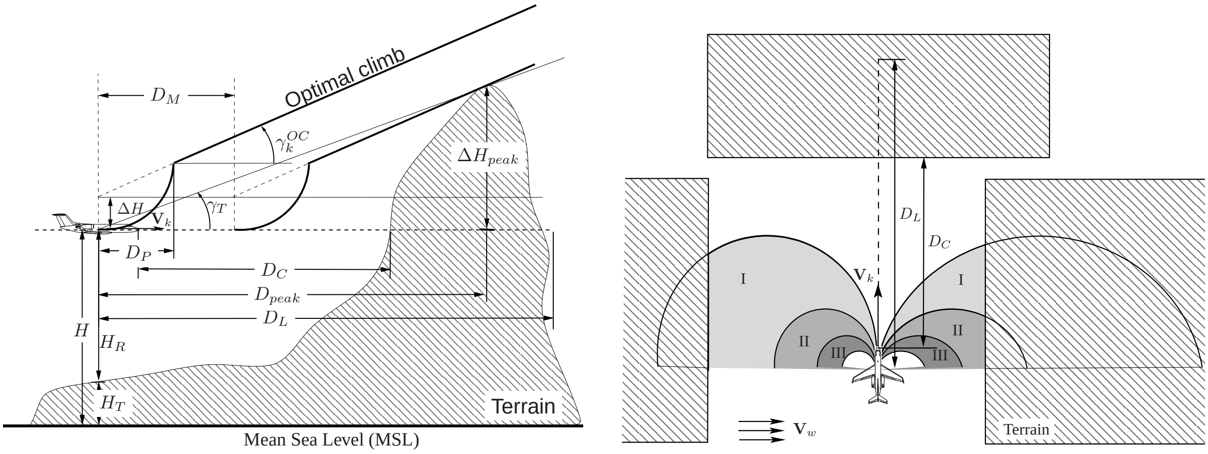
Separation The vertical terrain separation (Figure 1(a)) is expressed by the radio altitude H_R, whereas the forward terrain separation is expressed by the distance-to-collision D_C, which is defined as the distance between the aircraft’s current position and the intersection of the line extending along the current ground-referenced flight direction with a terrain point. For the forward terrain separation, however, the distance-to-maneuver D_M would be more meaningful to the pilot than D_C. Using geometric relations, D_M can be interpreted as the cotangent of the maneuver angle γ_M: $\cot\gamma_M = \cot\gamma_T - \cot\gamma_k^{OC}$, where γ_T is the terrain peak angle. If $\gamma_T < \gamma_k^{OC}$, then $D_M > 0$, meaning that a climb over the terrain would be possible. Assuming a circular pull-up trajectory, D_P is the horizontal distance traveled needed to reduce the current kinematic velocity to the optimal climb velocity during the pull-up.

The distance D_L represents a finite look-ahead distance over which a terrain intersection point can be found. Parameters such as time-to-collision, time-to-maneuver, look-ahead time, etcetera can all be obtained by dividing the above distance parameters with the aircraft ground speed.

The sideward separation constraints are formed by the turn performance of the aircraft and the surrounding terrain. The sideward distance-to-collision is the intersection distance of the aircraft’s predicted curved trajectory. Assuming level turns with a 180-degree heading change at the current airspeed and at a constant roll angle in constant and uniform wind, collision points to the left and to the right can be found using three turn regions (Figure 1(b)): I) all turns with roll angles between 15 and 30 degrees, II) all turns with roll angles between 30 and 45 degrees, and III) all turns with roll angles between 30 and 45 degrees.

Display Mapping

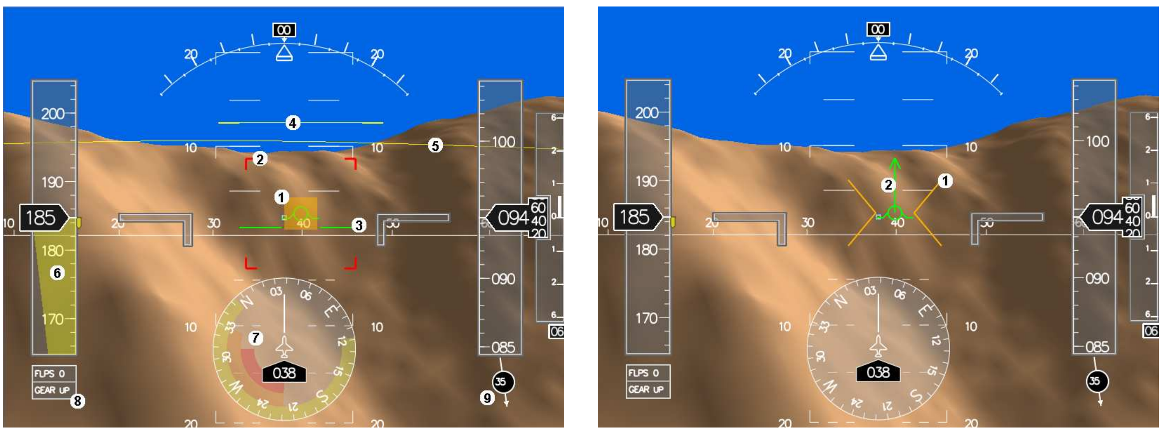
Mapping the above explored constraints resulted in the ESVD as shown in Figure 2(a). For the ESVD, the optimum climb constraint of the aircraft can be projected on the pitch ladder ④. The optimum climb while turning with a 45 degree bank angle is represented by ⑤. All optimum climb angles are computed using an aircraft performance database. By comparing the terrain angle perceived on the ESVD with the steepest climb angle, a pilot would be able



(a) Longitudinal separation (b) Lateral separation

Figure 1: The longitudinal and lateral separation parameters between the aircraft and the terrain.

to see if climbing over the terrain is possible. In Figure 2(a) it can be seen that the aircraft would be able to climb over the mountain. The energy angle is represented by ③. The cue from the energy angle and terrain angle is that the pilot would need to add energy to the system to be able to reach the optimum climb performance. To indicate how much distance and time are left to initiate an escape maneuver, a so-called distance-to-maneuver square was shown (①). As the aircraft approaches the terrain, the inner square will expand to the corner points of the outer square (②). The expansion rate depends on the ground speed at which the aircraft is approaching the terrain. From this a relative time-to-maneuver can be estimated. Furthermore, the inner square changes color from yellow to red, representing “enough” and “little” distance-to-maneuver, respectively. This color-coding corresponds to the caution and warning colors of a TAWS. The yellow area on the speed tape (⑥) indicates that the aircraft has excess speed that can be exchanged into additional altitude. The sideward terrain separations are represented on the compass rose in the form of left and right heading band constraints (⑦). The turn regions described above are used to check terrain intersections, which are color-coded as follows: constraints (or obstacles) in region I are colored yellow, in region II orange, in region III red. In Figure 2(a) it is shown that making any left turn will result in a terrain collision at some point, whereas making a right turn with a bank angle of at least 30 degrees circumvents the collision. Furthermore, pilots could also opt for making a straight climb over the mountain ahead.



(a) The ESVD interface (b) The SVGCAS interface

Figure 2: The interfaces used in the experiment.

Experiment

To evaluate the ESVD, relevant CFIT situations were created and simulated in a pilot-in-the-loop experiment in a fixed-base flight simulator using the Cessna Citation 500 model. The ESVD system was compared to a viable design alternative.

Design Alternative

The F-16 has a so-called auto-GCAS, which takes over the aircraft to prevent a terrain collision. In literature, research has been conducted on using guidance symbology on the HUD such that pilots can manually restore the aircraft to a safe flight condition by following commands (Billingsley & Kuchar, 2001). It was chosen to use these overlays on an SVD, because they have the same safety purpose as the ESVD and they were not designed using EID. This command system will be called SVGCAS (Figure 2(b)). The functionalities of command symbols, however, were tailored to fit the purpose of this experiment and to make a fair comparison between the ESVD and the SVGCAS.

The SVGCAS shows two symbols on the SVD: ① chevrons (><), representing the distance-to-maneuver (and time-to-maneuver), and, ② the ideal evasive maneuver command arrow. The computation of the command is based upon the same look-ahead algorithms and constraints as present in the ESVD.

Subjects

In the experiment a total number of 16 professional glass-cockpit airline pilots participated, with an average age of 29 years and an average experience of 3,000 flight hours. The subjects were instructed to avoid a terrain collision by performing one of the following five escape maneuvers: straight climb up, left climbing turn, right climbing turn, level left turn, or a level right turn.

Independent Variables

The independent variables in the experiment were the display configuration (DISP) and the experiment scenarios (SCENE). DISP (within-subjects) had two levels (ESVD and SVGCAS), SCENE (within-subjects) had 7 levels. The scenarios used to test the effects of the independent variables on the dependent measures are such that in each scenario one escape maneuver is optimal to escape an impending collision. The possible escape maneuvers in the first 5 scenarios were: straight climb up, climbing turn to the left, climbing turn to the right, level left turn, and level right turn. Each of these scenario had two variants (A and B) which featured slightly different terrain, initial aircraft conditions (positions and trim settings), and wind conditions to prevent pilots from recognizing the scenarios. Scenarios 6 (total engine failure) and 7 (flaps retraction failure) were system failure scenarios unanticipated by both the ESVD and the SVGCAS. Hence, pilots could not rely on the information they perceived from the interfaces to avoid a collision and should therefore make a suitable decision based on their knowledge and intuitions.

Dependent Measures

The dependent measures in the experiment were: 1) The decision (escape maneuver choice) and 2) the situation awareness (SA). The decision was rated 0 for non-optimal maneuvers and rated 1 for optimal maneuvers. The SA was measured using a query with simulation freeze which probed the levels of perception (level 1), comprehension (level 2), projection (level 3) and metacognition (self confidence of pilots about their query answers). SA was graded in conjunction with the metacognition as shown in Table 1.

Table 1: *Grade determination of the SA query answers.*

	Query answers	
Metacognition	*Incorrect*	*Correct*
Sure	0	3
Unsure	1	2

Design

In the measurement phase of the experiment, the five anticipated and two unanticipated scenarios (6 and 7) were balanced between two groups of 8 pilots. In the training phase, each group of pilots only flew the two variants of the 5 anticipated scenarios in a different terrain database.

Results

The analysis of the decision and the SA levels was done using repeated-measures Analysis of Variance (ANOVA). The anticipated and unanticipated situations were separately analyzed.

Decision

In anticipated situations, DISP and SCENE had no significant effect on the decision about the escape maneuver. However, from Figure 3 it can be seen that the ESVD resulted in slightly less optimal decisions than SVGCAS. This can be explained due to the fact that the ESVD shows multiple candidate escape solutions, thereby increasing the likelihood that suboptimal escape maneuvers can be chosen. The SVGCAS always showed one escape solution, that is, the optimal escape. In unanticipated situations, however, DISP had a significant influence on the decision ($F(1, 14) = 11.065, p = 0.005$), as well as SCENE ($F(1, 14) = 27.512, p < 0.01$). From Figure 3 it can be seen that in the unanticipated situations pilots made much better decisions about their evasive maneuver when using the ESVD than when using the SVGCAS. Furthermore, flying with the ESVD did not result in any terrain crash, whereas the SVGCAS resulted in 3 crashes.

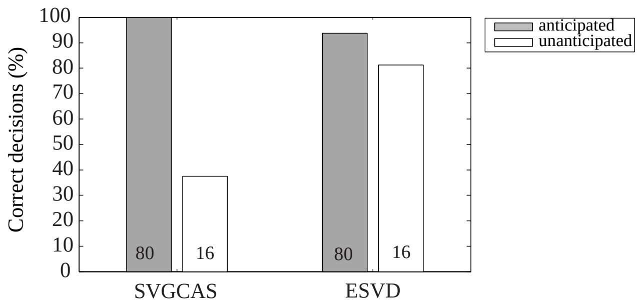


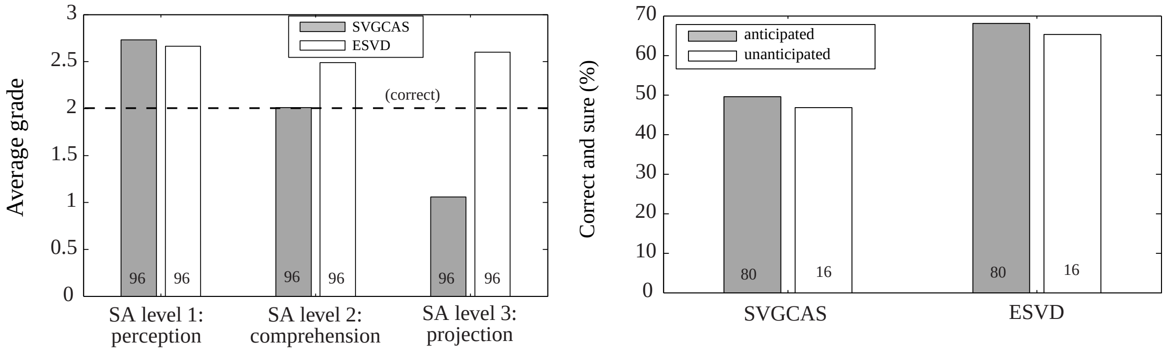
Figure 3: The pilots' decisions in terms of the chosen escape maneuvers. The number of runs are shown below in each bar.

Situation Awareness

In Figure 4(a) the average overall SA grades (anticipated and unanticipated situations combined) are shown, which indicate that pilots could much better comprehend and project the situations when flying with the ESVD than when flying with the SVGCAS. In Figure 4(b) it is shown that pilots were also much more confident about their answers in anticipated as well as unanticipated situations when using the ESVD. In anticipated situations, DISP had a significant effect on SA level 2 ($F(1, 14) = 221.5, p < 0.01$) and SA level 3 ($F(1, 14) = 464.8, p < 0.01$). In unanticipated situations, DISP had also a significant effect on SA level 2 ($F(1, 14) = 115.3, p < 0.01$) and SA level 3 ($F(1, 14) = 478.4, p < 0.01$).

Conclusions

The goal of the ESVD was to improve pilot terrain awareness and decision-making by using EID. The experiment results showed that pilots were more aware of the terrain situations in both anticipated and unanticipated scenarios when using the ESVD as opposed to a command-based display counterpart, the SVGCAS. The ecological approach resulted in an interface that clearly supported the higher levels of cognition and promoted pilot reasoning. The decision-making, however, only significantly improved in the unanticipated events. Despite this result, none of the pilots crashed when using the ESVD.



(a) SA grades

(b) Metacognition

Figure 4: The SA grades and the metacognition, expressed in the percentages of correct answers for which pilots were confident. The number of runs are shown below in each bar.

References

Amelink, M. H. J., Mulder, M., Paassen, M. M. van, & Flach, J. M. (2005, March). Theoretical Foundations for a Total Energy-Based Perspective Flight-Path Display. *The International Journal of Aviation Psychology*, *15*(3), 205-231.

Asselin, M. (1997). *An Introduction to Aircraft Performance* (J. Przemieniecki, Ed.). 1801 Alexander Bell Drive, Reston, VA 20191: American Institute of Aeronautics and Astronautics.

Billingsley, G. O., & Kuchar, J. K. (2001). Head-Up Display Symbology for Ground Collision Avoidance. *The International Journal of Aviation Psychology*, *11*(1), 33-51.

Bisantz, A. M., & Pritchett, A. R. (2003). Measuring the Fit between Human Judgments and Automated Alerting Algorithms: A Study of Collision Detection. *Human Factors*, *45*(2), 266–280.

Borst, C., Sjer, F. A., Mulder, M., Van Paassen, M. M., & Mulder, J. A. (2008). An Ecological Approach to Support Pilot Terrain Awareness After Total Engine Failure. *Journal of Aircraft*, *45*(1), 159-171.

Borst, C., Suijkerbuijk, H. C. H., Mulder, M., & Van Paassen, M. M. (2006). Ecological Interface Design for Terrain Awareness. *The International Journal of Aviation Psychology*, *16*(4), 375-400.

Burns, C. M., Jamieson, G. A., Skraaning, G., & Kwok, N. L. J. (2007). Supporting Situation Awareness Through Ecological Interface Design. In *Proceedings of the 51st annual meeting of the human factors and ergonomics society* (p. 205-209). Human Factors and Ergonomics Society.

Flach, J. M., Mulder, M., & Van Paassen, M. M. (2004). The Concept of the Situation in Psychology. In S. Banbury & S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application* (p. 42-60). Ashgate Publishing, Oxon (UK). (ISBN 0754641988)

Pritchett, A. R. (2003). Human-Computer Interaction in Aerospace. *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, 861–882.

Spitzer, C. (2001). *The Avionics Handbook*. Boca Raton (FL): CRC Press.

Wickens, C. D. (2002, July). *Spatial Awareness Biases* (Report No. AHFD-02-6/NASA-02-4). Moffet Field, CA: NASA.

THE 787 VERTICAL SITUATION DISPLAY HUMAN FACTORS EVALUATION ENHANCEMENTS TO FLIGHT PATH AWARENESS

Jason Hammack
The Boeing Company
Seattle, Washington

Julianne Fox Cummings
Exponent Failure Analysis Associates
Menlo Park, California

Jean Crane
The Boeing Company
Seattle, Washington

A two-phased focused evaluation was run to validate the design implementation of the 787 Vertical Situation Display (VSD), using static and dynamic scenarios in a prototype flight deck. Pilot feedback and pilot questionnaire data were recorded and videotaped for analysis. The validation testing results suggest that overall an effective implementation has been designed. There was one area where issues were found that necessitated requirement changes. The mode transition logic (during takeoff) was redesigned and additional validation testing was conducted to ensure acceptability.

Introduction

One of the basic features of the Boeing 787 airplane is a Vertical Situation Display (VSD) located on the pilots' navigation displays in the flight deck. The VSD is a side view of the airplane's vertical path in space. Together with the lateral view of the airplane's flight path on the Navigation Display (ND), the displays will present pilots with a complete picture of the flight path and related external hazards. The 787 VSD incorporates new display features and is an evolution from the original VSD certified on the Boeing 737.

When a new feature is designed for a pilot display that requires a high level of integration with existing displays, more scrutiny is necessary than that which is required by either a stand-alone display or a display that is similar to one with significant in-service exposure. Within the Boeing flight deck Human Factors design process, these types of new features go through a focused evaluation process to validate the design. Focused evaluations are devised to address concerns, issues, and/or detailed design requirements of new flight deck equipment or functions early in the design/build schedule for an airplane or airplane modification program. This heightened scrutiny has recently been formalized in the certification of new flight deck features by the new European Aviation Safety Agency regulation (CS 25.1302) and its associated Acceptable Means of Compliance (AMC 25.1302) to provide guidance in the certification of new flight deck equipment. The Federal Aviation Administration is expected to adopt similar guidance in the near future.

This paper presents Boeing's 787 VSD design solution and provides two examples of the issues and design details that were assessed during our focused evaluation process to ensure that the pilots could safely, efficiently, and effectively get appropriate information from the VSD.

Background

The primary safety purpose of a VSD is to mitigate Controlled Flight Into Terrain (CFIT) and Approach and Landing Accidents (ALAs), as well as to provide a tool for pilot decision-making related to approach stabilization. The 787 VSD builds upon the Boeing 737 VSD (Jacobsen, Chen, and Wiedemann, 2000; Chen et al., 2000; Houck, Kelly and Wiedemann, 1986). Both the 737 and 787 VSDs provide a mode that depicts the vertical profile along the airplane's current track. To increase pilot awareness of VNAV performance (i.e., the airplane's vertical navigation performance calculated within the Flight Management Computer [FMC]), the 787 VSD has added a mode enabling a graphic representation of the vertical path along the (active) FMC route. This new Path mode gives the pilot a powerful tool to assess the intended FMC vertical profile and its relationship to the approaching terrain, current airplane altitude, selected altitude on the Mode Control Panel (MCP), and decision altitudes.

Without the Path mode VSD, pilots construct a mental picture from FMC VNAV information in conjunction with the graphical terrain depicted on the ND in order to anticipate what the airplane will be commanded to do given the programmed VNAV path. The 787 VSD Path mode will provide a graphical profile depiction of level offs, deceleration/acceleration segments, step climbs, top of climb, top of descent, altitude constraints and vertical targets (e.g., MCP altitude, radio altitude minimums, barometric pressure minimums). This graphical representation of the VNAV path provides an integrated view of all of this information, enabling quick interpretation by the pilot.

VSD Format

The VSD is located on the bottom third of the ND and depicts the following information: airplane current altitude, terrain features, basic navigation data, trend information concerning the airplane's vertical path and speed, as well as some pertinent vertical information, such as Selected Altitude and Minimum Descent Altitude.

There are two modes of the 787 VSD: Track mode and Path mode. The two modes are automatically selected depending upon the state of the airplane automation and the airplane relationship to the FMC route. When the airplane has been selected to fly a lateral path mode of the FMS/autopilot (lateral navigation "LNAV," localizer "LOC," final approach course "FAC," etc.) and is within the set criteria for allowable path deviation, the VSD Path mode is depicted. In all other cases, the VSD Track mode is depicted (e.g., Heading Select "HDG SEL," Heading Hold "HDG HOLD"). In both Track mode and Path mode, a pair of cyan dashed lines (i.e., a swath) is shown on the horizontal view of the Navigation Display that directly corresponds to the information depicted on the VSD. This cyan swath depicts the lateral area over which the VSD is showing terrain and associated waypoint altitude constraints.

In Track mode (see Figure 1), the VSD depicts the information along the airplane's track. The cyan swath aligns with the airplane track and not the FMC route. When the VSD is in Track mode, the LNAV waypoint altitude constraints, underlying terrain, MCP altitude, altitude minimums, and final approach path along the airplanes current track are depicted. In Path mode (see Figure 2), the VSD depicts the information along the FMC route. The cyan swath aligns with the FMC route. When in Path Mode, the terrain at the twelve o'clock position is not within the cyan swath and thus is not depicted on the VSD.

Figure 1. Track mode VSD

Figure 2. Path mode VSD

Project Approach

While validation of the VSD for the 787 occurred throughout an iterative design process, we performed a focused evaluation and validation toward the end of that process. Given that the Track mode VSD has been certified and is being used in service on the 737, the purpose of our evaluation was to validate the Path mode functionality on the VSD, with respect to the following: 1) the new functionality associated with the new Path mode, 2) the integration of the Path mode with the Track mode, and 3) the integration of the new dual-mode VSD into the new 787 flight deck.

We used an iterative process to identify the potential issues associated with the Path mode VSD. We started by bringing together the team that had been involved in the design and oversight of the original VSD, and with the development of the new Path mode, much of which had already been dynamically prototyped. Team member's expertise comprised Engineering, Human Factors and Flight Operations. First, we identified the areas where validation was required (e.g., the new symbology and dynamic behavior associated with the new symbology). Next, we created a matrix that identified the validation objectives and test questions were crafted to address each item requiring validation. Once the test questions were agreed upon, specific test scenarios were proposed for each question. The scenarios were prototyped and presented to the team for review, discussion, and refinement.

A two-phased evaluation was designed to address issues and concerns related to: 1) ND/VSD correlation, 2) Mode transition logic, 3) Holds, 4) Procedure turns, 5) Path depiction during climbs and level offs, 6) Lateral swath depiction, 7) Map range consistency, and 8) Route discontinuities. For the purposes of this paper, we will expound upon the first and second items (i.e. ND/VSD correlation and the takeoff mode transition logic).

In Phase 1, we used static scenarios to solicit early feedback from the pilots. During a static scenario, the pilot viewed static displays in the 787 rapid prototyping device. These scenarios were comparable to what you would expect to see if you took a snapshot of a paused or stopped simulator during a flight. Use of static scenarios was effective in assessing whether the pilots were able to successfully correlate information (e.g., symbology on the VSD to the ND) and whether there was any confusion resulting from the format design. In Phase 2, the scenarios were conducted in the same 787 rapid prototyping device, however during this portion of the testing, the pilots viewed the displays while flying the simulator on autopilot, utilizing the MCP to control the airplane. Questions that could not be addressed by the static scenarios, such as the mode transition logic, were addressed during the dynamic scenarios.

ND/VSD Correlation & Mode Transition Logic – Testable Questions

Given that the selection between Track and Path modes is automatic, it is essential that the pilots maintain an ability to correlate the information on the VSD and ND during all phases of flight. Also it is important that the transitions between Track and Path mode occur at the appropriate time, such that the information depicted is interpreted correctly. Depending upon the mode that the VSD is displaying (i.e. Track or Path), the information (e.g. terrain and waypoints) depicted will often differ. For example, when in Track mode, the VSD displays the information in front of the airplane, along the airplane's current heading. Thus, at the point at which the airplane's route changes heading (i.e., to a direction other than the current airplane heading), the waypoints and terrain along the route are no longer visible from the point where the route alters direction (See Figure 1).

The information displayed while in Path mode is different in that the information along the route programmed into the FMC and executed is depicted on the VSD regardless of whether the route is along the airplane's heading (See Figure 2).

The following are the testable ND/VSD Correlation & Mode Logic questions:

- What do the pilots understand from looking at the VSD and how do they correlate it to the ND map and computer display unit (CDU)? What information was used in making the determination?
- Does the VSD display what was expected?
- Did the point at which the VSD transitioned from Path mode to Track mode align with the pilot's expectations?
- Was there anything that appeared to be conflicting or confusing?

- Are there any potential certification or safety issues associated with the Path mode implementation?
- Were there any items that the pilots would prefer to see depicted differently?
- Is the mode transition logic understandable by the pilots?

The static scenarios focused on presenting the pilots with possible airplane configurations (e.g., LNAV and Path mode, HDG SEL and Track mode, Takeoff Go-Around[TOGA] without LNAV armed and Path mode) followed by an inquiry as to what information on the ND was being displayed on the VSD.

The dynamic scenarios focused on presenting the pilots with possible mode transitions (e.g., airplane in TOGA without LNAV armed during takeoff, Path mode displayed on the VSD until the 400' above ground level (AGL) transition to Track mode) and were followed by an inquiry as to the intuitiveness of the mode transition logic.

Methodology

Participants

In Phase 1, 14 air transport pilots participated in the validation. Eleven out of the 14 pilots had prior experience using the VSD. The pilots had Boeing Designated Engineering Representative (DER) flight test or Boeing Training pilot experience. Some of the pilots also had prior line flying experience with a major airline.

In Phase 2, 12 air transport pilots participated in the validation. Eleven pilots that participated in Phase 1 returned to participate in Phase 2. One additional pilot, who did not participate in Phase 1, participated in Phase 2. Ten of the 12 pilots had prior experience using the VSD in the 737.

Research Environment

Both phases of testing were conducted in the Rapid Prototyping Development System (RaPiDS). RaPiDS is a PC-based 787 dynamic flight deck simulator utilized as a flight deck engineering development tool. RaPiDS provided the needed interactivity and functionality to complete a thorough validation of the VSD.

In Phase 1, nine sets of static display pictures were depicted on their prospective display units: Primary Flight Display (PFD), Navigation Display (ND), Vertical Situation Display (VSD), Engine Indication and Crew Alerting System (EICAS) and Control Display Unit (CDU). Full-scale laminated pictures of the Mode Control Panel (MCP) displaying the appropriately illuminated buttons and settings were presented in the MCP location during each scenario. Additionally, the gear lever and flap lever were set to the correct position for each scenario. Directly after viewing each test scenario, the pilots answered a predefined set of questions and provided feedback.

In Phase 2, the pilots flew six scenarios. The displays, MCP, flaps and gear were operational during these flight scenarios.

Test Conduct

Pilots were brought in twice within a three month period to participate in the two phases of the test. A checklist was utilized to ensure that RaPiDS was configured in the same way for each pilot prior to each scenario. Upon arrival, a pilot experience questionnaire was filled out and a brief written summary explaining the purpose of the testing and what the pilots should expect was provided. Pilots then completed a self-paced Computer-Based Training (CBT) module on both modes of the VSD. After each scenario in the Phase 1 and Phase 2 testing, the pilots provided feedback by completing a questionnaire. Observations were recorded by the test administrator and videotaped (i.e., during completion of the CBT module and during completion of the Phase 1 and Phase 2 test scenarios). Additionally, the majority of pilots provided verbal feedback and commentary throughout the testing, which was also captured.

Phase 1. The pilots viewed nine static scenarios and provided feedback via a questionnaire. The test in its entirety took approximately two hours to complete. In this paper, we will describe four of the nine Phase 1 scenarios.

Phase 2. Phase 2 was primarily focused on validating: 1) the mode transition logic (i.e., during takeoff and during typical transitions between VSD modes) and 2) the display depictions during transitions (i.e., hold and climb depiction). The pilot who had not participated in Phase 1 completed the CBT training upon arrival. In Phase 2, the pilots flew six scenarios and provided feedback via the questionnaire. Test subjects served as the captain (pilot flying). One of the test administrators served as the first officer (pilot not flying). The test in its entirety took approximately 90 minutes to complete. In this paper, we will describe two of the six Phase 2 scenarios.

Results and Conclusion

Phase 1

Scenario 1: Pilots viewed a static display of a VSD in Path mode with LNAV/VNAV engaged. All of the pilots correctly correlated the VSD information to the ND. They all utilized the cyan swath to determine the mode, and nine pilots out of 14 also used the presence of the magenta route on the VSD to assist in correlating the ND and VSD.

Scenario 2: Pilots viewed a static display of a VSD in Track mode with LNAV/VNAV engaged and a cross track error that exceeded RNP. Typically, when LNAV/VNAV is engaged, the VSD would be in Path mode, however when cross track error exceeds RNP, the airplane will not fly the FMC route. Because the VSD depicts the route the airplane will fly, the VSD depicts Track mode in this case. Twelve out of 13 pilots accurately correlated the VSD information to the ND. Twelve out of 14 pilots indicated that they used the cyan swath and five pilots used the absence of the magenta vertical route depicted on the VSD to indicate the VSD was in Track mode.

Scenario 3: Pilots viewed a static display of an airplane in Track mode with HDG SEL engaged. All 14 pilots correctly correlated the VSD information to the ND. Thirteen pilots indicated that they used the cyan swath to make their determination. Ten pilots referenced the HDG SEL roll mode in making their determination and only six pilots indicated that they used the absence of the magenta vertical route depicted on the VSD as a cue to determine that they were in Track mode.

Scenario 4: Pilots viewed a static display of an airplane in Path mode with LNAV/VNAV engaged; the ND range set to 1280 nm and the VSD displaying 2560+ nm. All 14 pilots correctly correlated the VSD information to the ND. Thirteen pilots used the cyan swath to make the correlation. Five pilots used the magenta line on the VSD and three pilots relied on the fact that LNAV was the roll mode in making their determination.

Consistently, throughout Phase 1, the pilots were successfully able to correlate the information on the ND and the VSD. We noted that the cyan swath was the most powerful cue used in correlating the information on the two displays and the magenta vertical route depicted on the VSD was used as a secondary cue.

Phase 2

Scenario 5 and Scenario 6: During each of these two scenarios, the pilots flew the Gypsum 3 departure out of Eagle, Colorado. All pilots flew this scenario with and without the VSD displayed. During Scenario 5, the VSD was depicted and during Scenario 6 it was not depicted. Seven of the pilots flew it first with the VSD (Scenario 5) and five of the pilots flew it first without the VSD depicted (Scenario 6). During both of these scenarios, immediately after takeoff (i.e., at 400’ AGL), the airplane transitioned from TOGA to HDG HOLD and no longer followed the FMC path as was depicted on the ND and VSD prior to and during takeoff up to 400’ AGL. Simultaneously, at 400’ AGL as the airplane was encroaching upon hazardous terrain, the VSD automatically transitioned from Path mode to Track mode. This transition was caused by the absence of LNAV/VNAV being engaged prior to takeoff.

The feedback from all 12 pilots was consistent regardless of the order in which the scenarios were flown. The pilots indicated that they expected the airplane to follow the FMC route after takeoff. They indicated that they were surprised by the 400’ transition and that Path mode was depicted on the VSD, despite the fact that the airplane was not going to fly the path after takeoff. They indicated that the presence of the magenta vertical route on the VSD and the cyan swath on the ND prior to takeoff led them to expect that the airplane would follow the LNAV/VNAV route after takeoff. We also received feedback that when LNAV was not armed at takeoff, the VSD should have been in Track mode and that this implementation posed both a safety and certification risk. The VSD logic was overwhelmingly perceived as conflicting, confusing and/or concerning. Interestingly, one of the pilots that had just finished flying the scenario a second time commented that despite his prior exposure to this scenario, he still assumed that LNAV was armed because of the presence of the cyan swath aligning with the LNAV path on the ND.

We received strong, consistent feedback from the pilot group regarding the VSD implementation and transition logic after flying the Gypsum 3 departure out of Eagle, Colorado. The consensus, regardless of which order the pilots flew the scenarios, was that if the airplane wasn't going to fly the FMC route after takeoff, Path mode should not have been displayed. There was consensus that 400' AGL is not an appropriate time to transition between VSD modes (i.e., from Path to Track). Across the board, pilots agreed that there was a certification and safety risk associated with the current implementation. Hence, a design change was strongly recommended. Additionally, the importance of the cyan swath and the information that it conveyed was strongly emphasized by the pilots during these scenarios.

Based upon the feedback that we received, new logic addressing the certification and safety concerns was developed. In accordance with the direction received, the new logic ensured that the VSD remains in Track mode on the ground when the airplane is not going to fly the FMC route after takeoff. Additionally, there is no longer a trigger point at 400' AGL enabling the VSD to transition between modes. The VSD now consistently depicts the information within the swath along the route that the airplane is configured to fly. A follow-on test with dynamic scenarios was conducted to successfully validate the final design.

Summary

As stated, the purpose of this focused evaluation was to validate that there were not any certification and/or safety issues related to the VSD design implementation and that the VSD Path mode has been implemented in a way that provides the pilot with an improved awareness of the airplane's current position relative to the commanded vertical path. A two-phased effort was effective in conducting the evaluation. Scenarios presented in a static environment were valuable in ascertaining how well information was correlated between the VSD and the ND. Dynamic flight scenarios were instrumental in discovering issues with the mode transition logic. The validation testing (Phase 1, Phase 2, and follow-on testing) results suggest that an effective implementation has been designed. In the area where certification and/or safety concerns existed (i.e., the mode transition logic at takeoff), design changes were incorporated and re-tested.

Acknowledgements

Significant contribution from our colleagues in Flight Operations, Engineering, Human Factors, and Simulation was crucial to our research, development, design, and validation effort for the 787 VSD. The following team members assisted in significantly driving our efforts: Michael Carriker, Sherwin Chen, Alan Jacobsen, Brian Kelly, Keith Johnson, and John Wiedemann.

References

European Aviation Safety Agency. (2007). CS 25.1302 / AMC 25.1302. Installed systems and equipment for use by the flight crew. *Certification Specification for Large Airplanes, Amendment 3*.

Jacobsen, A. R., Chen, S. S., and Wiedemann, J. (2000). Vertical Situation Awareness Display. *Royal Aeronautical Society: Proceedings Situational Awareness on the Flight Deck: The Current and Future Contribution by Systems and Equipment*.

Chen, S. S., Jacobsen, A. R., Hofer, E., Turner, B. L., and Wiedemann, J. (2000). Vertical Profile Display Development. The Boeing Company.

Houck, R., Kelly, B. D., and Wiedemann, J. (1986). An Integrated Display for Vertical Situation Awareness in Commercial Transport Airplanes. Society of Automotive Engineers Aerospace Technology and Exposition. Paper # 861770.

THE MYSTERY OF DISTRIBUTED LEARNING

Gavan Lintern
Cognitive Systems Design
Dayton OH 45439

There are contrasting opinions about the value of distributed learning. Several textbooks on general training issues promote it as an effective training strategy while many researchers who have focused specifically on this topic argue that distributed practice is no more effective than non-distributed practice. It is noteworthy that most who promote distributed learning base their opinion on belief rather than on experimental research while most who argue that it is of no value base their opinions on empirical data restricted primarily to the learning of simple motor skills. Additionally, much of the distributed learning research has employed the experimentally convenient manipulation of distributing learning trials whereas, from a practical perspective, the distribution of sessions would offer a more relevant experimental manipulation. In this paper, I explore the insights that can be gleaned from research that has focused on operationally relevant tasks and in which learning sessions have been distributed.

Anderson (1985, p. 240) observes that distributed practice has profound effects on skill acquisition and Schultz and Schultz (1986, p. 213) claim that it is usually the better training approach. Hopkins, Snyder, Price, Hornick, Mackie, Smillie, and Sugarman (1982) take it for granted that distributed practice will benefit the training of nuclear power station operators. In contrast, Adams (1987, p. 50), Magill (1985, p. 374), and Schmidt (1982, p. 484) argue that distributed practice has substantial effects on performance but minimal effects on learning. My primary goal for this paper is to assess which of these contrasting opinions is the more credible.

Distribution of Trials or Sessions?

Spacing manipulations come in different forms. Sessions that may be in the order of an hour or two long may be spaced by one or more hours (Keller & Estes, 1944) or across days (Hagman & Rose, 1983; Keller & Estes, 1945). Trials that may be in the order of seconds or minutes long can be spaced by seconds or minutes (Magill, 1988; Reynolds & Bilodeau, 1952) or even days (Flexman, Roscoe, Williams, & Williges, 1972). The conflicting opinions about distributed practice that can be found both in scientific and in operational training circles are generally not derived from a consideration of the data from all relevant forms of distributed practice. For example, Adams (1987) considered experiments in which trials have been spaced while Anderson (1985) considered experiments in which sessions were spaced.

The issue of distributing sessions is of interest because the scheduling of training has substantial cost implications especially where operators must return to a central establishment for continued training. In such a case, it will usually be more economical to compress training into as short a period as possible. However, a well-distributed series of sessions may be necessary to establish and to maintain high levels of skill. In this paper I review literature related to distributions of training sessions and its effects on acquisition of action skills; where action skills are to be viewed as those with both a psycho-motor and a cognitive component. A comprehensive analysis of trial-distribution effects on learning of perceptual-motor skills is provided in a review by Lee and Genovese (1988) and in commentaries on that review (Lintern, 1988; Newell, Antoniou, & Carlton, 1988).

In contrast to the supposed benefits of distributed training, it is occasionally argued that compressed training is beneficial. One of the often-mentioned advantages of regularly scheduled flight training, such as that offered by the University of Illinois, is that flight sessions scheduled over alternating days promote faster learning than the irregular distribution of sessions undertaken by many flight students (Shugarts, 1987). In addition, the University's summer semester is occasionally thought to provide a better training opportunity because students fly six days a week instead of three days a week as they do in the Fall and Spring semesters. However, this belief in the advantage of regular and compressed schedules has developed in the absence of any empirical evidence one way or the other.

Opinions relating to the supposed benefits of compressing or distributing trials or sessions are diverse. Furthermore, they are often generated without full consideration of the evidence, and occasionally without consideration of any evidence at all.

Review of Experiments

Three categories of tasks have been used in the research to be reviewed. The first includes relatively well defined technical skills that must be taught over days or weeks in an extended course of instruction. Morse code and typing are the two target skills that have been examined. A second category includes relatively short procedural skills that may be learned within an hour or two. The third includes recreational activities which may take years of intensive practice to reach full proficiency, but which may be learned to a moderate level of competence in several lessons of one or two hours each. It is unfortunate that the target skills are so diverse, but useful data on this issue is difficult to find and, given the strength of the opinions, it seemed worthwhile to explore the implications of any data that might be relevant.

Technical Skills

A set of data used by Anderson (1985) in his discussion of distributed practice is from Morse code research by Keller and Estes (1945). The standard five-week course of Morse Code instruction gave trainees 195 hours of practice, with seven hours of practice on each of five days and four hours of practice on the sixth day of each week. Keller and Estes (1945) compared the standard schedule with an experimental schedule in which 192 hours of practice were distributed over eight weeks, with four hours of practice per day for six days each week. There were no differences between the groups' skill levels at 5 weeks (when Morse code training ended for the massed group), despite the fact that the distributed group had completed only 60% of the training. The distributed-practice group was far more proficient with Morse code than was the massed-practice group at completion of Morse code training (Figure 1).

Two different schedules had been used with those students who had received four hours of practice each working day (Keller & Estes, 1944). One schedule had students learning Morse code in a four-hour block each morning. In the other, the four hours of instruction were distributed throughout the day for five days of the week. Because all trainees were to be released from duty around noon on Saturdays, this group also had a four-hour block of instruction in the morning of that day. Instruction in other communications topics was conducted in the afternoon, Monday through Friday of each week for the students in the blocked condition and during intervening Morse Code sessions for the students in the distributed condition. As described in Keller and Estes (1944), there were no differences in progress or final Morse code performance levels of the two groups (Figure 1).

Three decades later, Baddeley and Longman (1978) manipulated the length and distribution of sessions for instruction of typing. Four groups of subjects practiced typing for one 1-hour session per day, one 2-hour session per day, two 1-hour sessions per day, or two 2-hour sessions per day (i.e., 1, 2, or 4 hours per day). Sessions were scheduled over 5-day working weeks and sessions given on the same day were separated by at least two hours. A test of typing skill administered after 60 hours of training significantly favored fewer hours per day. The number of sessions over which those hours were distributed (a comparison between the use of two 1-hour sessions versus one 2-hour session) had no noticeable effect.

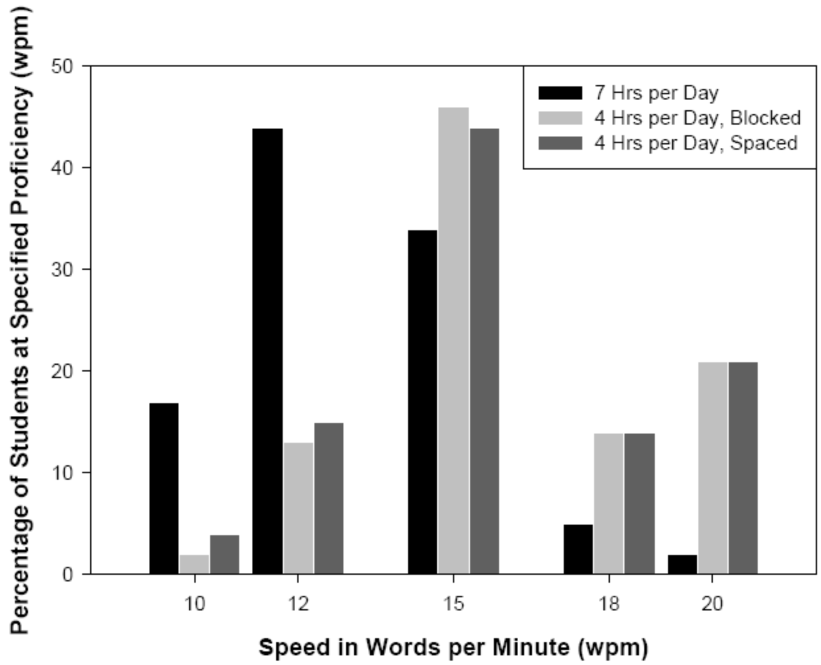


Figure 1. Percentage of Morse Code students who passed various receiving speeds at the end of approximately 200 hours of training (adapted from Keller and Estes, 1944, 1945).

Procedural Skills

Flexman et al. (1972) have provided data that show a learning advantage with compressed trials of a procedural task. A group of student pilots were pre-trained on several flight tasks in a 1-CA-2 SNJ Link ground-based trainer, which simulated the North American T6/SNJ aircraft and had been built from the cockpit of a wrecked T6/SNJ. Following pre-training, these experimental subjects were taught the same tasks to criterion in the T6/SNJ aircraft. A control group of student pilots, not pre-trained in the simulator, were also taught the flight tasks to criterion in the aircraft.

One of the training tasks was the start-up procedure for the T6/SNJ. This task required approximately two minutes to complete and had 16 steps that had to be executed accurately and in correct sequence to start the aircraft safely. Particularly in view of the fact that the simulator had been built from a cockpit of the target aircraft, the simulated starting procedure appeared to provide an excellent representation of the procedure used in the aircraft.

Intensive instruction on this task was not possible in the aircraft because of the battery drain resulting from each start and because of the possibility of overheating the starter motor. Control subjects practiced this procedure once at the beginning of each of their instructional flights, generally separated by one or more days. Experimental subjects first learned the procedure in the ground-based trainer in a session of massed trials that required less than one-hour of instructional time. The experimental subjects required fewer practice trials (ground-based trainer and aircraft trials combined) than did the control subjects (aircraft trials only) to learn the complete procedure and to demonstrate criterion performance in the airplane (three successive trials without error). The faster learning of the experimental group was attributed to the fact that experimental subjects had the bulk of their practice in one massed session of trials rather than distributed one trial at a time over days.

Hagman (1980) examined distributed practice with a procedure in which military trainees were taught an alternator maintenance task either with three massed practice trials on one day or with one practice trial on each of three successive days. This task required approximately 15 to 45 minutes to complete depending on the level of skill of the student. In a single-trial test of alternator maintenance approximately two weeks later, the spaced practice group completed the task in significantly less time and with fewer errors.

Recreational Activities

Young (1954) examined effects of changing the distributions of sessions in badminton and archery classes. There were apparently different students in the badminton and archery classes although there is a possibility that some subjects could have been in both classes. Sixteen badminton and nineteen archery lessons were scheduled over several weeks for two days or four days each week. Some badminton skills were learned more effectively with the more distributed schedule and archery was learned more effectively with the more compressed schedule.

Harmon and Miller (1950) contrasted four schedules of nine lessons for the instruction of billiards. The schedules were one lesson per day (including weekends), three lessons per week, one lesson per week, and nine lessons extended over 55 days in a sequence that became progressively more distributed (i.e., lessons on days 1, 2, 3, 5, 8, 13, 21, 34, and 55). There were no differences, as assessed in the ninth lesson, between groups with the one-lesson-per-day, three-lessons-per-week, and one-lesson-per-week schedules. However, the group with the 55-day progressively-distributed schedule showed better performance in the ninth lesson.

In a follow-up study, Langley (cited in Harmon & Miller) showed that another progressively distributed schedule, in which the nine lessons were distributed over 43 days (lessons on days 1, 2, 3, 8, 15, 22, 29, 36, and 45), was as good as the extended schedule of Harmon and Miller and better than their more compressed schedules. In a second follow-up study, Lawrence (cited in Harmon & Miller) tested the retention of some of the one-lesson-per-day and progressively-distributed-schedule subjects from the Harmon and Miller experiment. Again, an advantage was shown for the progressively-distributed schedule.

Discussion

Some distributions of sessions over days has been shown to assist learning of Morse code (Keller & Estes, 1945), typewriting skills (Baddeley & Longman, 1978), badminton (Young, 1954), billiards (Harmon & Miller, 1950), and alternator maintenance (Hagman, 1980). On the other hand, massing of practice can sometimes offer an advantage (Flexman et al., 1972; Young, 1954) while some variations in distribution of sessions within days and across days do not have any effects (Keller & Estes, 1944; Baddeley & Longman, 1978; Harmon & Miller, 1950). A progressively distributed schedule appears to offer some advantage (Harmon & Miller, 1950) and it should be noted that the schedule used by Flexman et al. (1972) can also be thought of as progressively distributed in that there was early massed practice in the simulator followed by a number of trials in the aircraft spaced by a day or more.

The experiments from which these data have been gathered might be viewed as too diverse to permit any systematic analysis. Nevertheless, there has been little systematic research on this topic and strong opinions about the effects of distributed practice find their way into the published literature. The primary goal for this discussion is to assess whether the empirical work or a rational analysis can lend any support to these opinions.

Cognitive Encoding

Anderson (1985) has attempted to deal with the enhanced learning effect of spaced practice with an appeal to more elaborate cognitive encoding of the representations of skill. Within the framework presented by Anderson, the development of skill follows a path from deductive processing to memory retrieval and pattern recognition. He relates this view to the progressive movement through cognitive, associative, and autonomous stages that are sometimes thought to underlie the acquisition of skill (Fitts & Posner, 1967).

Anderson's appeal to elaboration of cognitive encoding as an explanation of distributed practice effects evolves from a consideration of the spacing effect found in verbal memory experiments where recall is better for those items within a long list of verbal items that are spaced farther apart during learning (Melton, 1970). However, it is a considerable leap to extend an explanation derived from a paradigm in which learning instances within a session are separated by a varying number of other learning instances to one in which training sessions on a complex skill are distributed by hours or days.

Unscheduled Practice Between Sessions

Instead of working on the efficiency of learning during practice, the distributed schedules may promote mental practice (Prather, 1973) during the interval between training sessions, or may even encourage deliberate practice outside of formal classes. In none of the experiments reviewed here was there any reported attempt to control activities between instructional sessions. It is indeed likely that the Morse Code students of Keller and Estes would practice or rehearse Morse code exchanges and routines among themselves after class hours. It is also likely that students enrolled in classes for recreational activities would participate in those activities between classes. Extended intervals between classes would certainly provide expanded opportunities for that extra experience.

If informal practice could be established as the reason for the effectiveness of distributed sessions, it would provide a rationale for developing means of encouraging mental rehearsal and informal practice. One strategy, suggested by the Morse code research, is to establish instructional settings that will encourage students to interact after class hours. Another might exploit the current advances in e-Learning by providing out-of-class opportunities to practice with entertaining simulations of critical tasks.

Nevertheless, appeals to continued learning during intervals between sessions or more efficient practice within sessions as a result of a distributed schedule cannot fully explain the diverse patterns of data observed here. For example, why is a two-day-a-week schedule better than a four-day-a-week schedule for teaching badminton, while there is no difference between the effectiveness of daily, three-per-week, or weekly lessons for teaching billiards. In addition, how can compressed schedules be more effective under some circumstances. The relative success of the progressively distributed schedules of Harmon and Miller (1950) and of Langley (cited in Harmon & Miller) further complicate this issue.

Implications for Instruction

At a more pragmatic level, the data reviewed here are of interest because they show that distributions of practice can effect learning. Although this conclusion may be regarded as little more than folk wisdom by some, others would certainly disagree with it (e.g., Adams, 1987). Those who fail to recognize the effects of distributed practice have paid attention only to data from perceptual-motor experiments in which inter-trial intervals were manipulated and have ignored the data from experiments in which spacing between sessions has been manipulated. On the other hand, it is not clear that those who promote the benefits of distributed practice (e.g., Hopkins et al., 1982) are aware of the data that bear on their view, or of its ambiguity.

Unfortunately, the guidance offered by these data to designers of applied instructional programs is modest. They do support the long-established intuition that distributions of practice can have an important and often facilitating effects. On the other hand, the conditions under which the distribution of sessions has any effect have not yet been fully specified and it remains uncertain whether an enhancement or a decrement is to be expected. Possibly the strongest recommendations to emerge from this review is that intensive early practice followed by less intense sessions will not be detrimental and will often enhance the efficiency of a training program. Further, this review suggests that the amount of instruction provided in any one day or any single topic should be limited. If all potential advantages of session scheduling are to be exploited in the instruction of action skills, a systematic research effort is needed to uncover the underlying factors at work.

References

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. Psychological Bulletin, 101, 41-74.

Anderson, J. R. (1985). Cognitive psychology and its implications (2nd ed.). New York: W. H. Freeman and Co.

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. Ergonomics, 21, 627-635.

Fitts, P., & Posner, M. (1967). Human performance. Belmont, CA: Brooks/Cole.

Flexman, R. E., Roscoe, S. N., Williams, A. C., Jr., & Williges, B. H. (1972). Studies in pilot training: The anatomy of transfer. Aviation Research Monographs, 2(1). Champaign, IL: University of Illinois, Aviation Research Laboratory.

Hagman, J. D. (1980). Effects of training schedule and equipment variety on retention and transfer of maintenance skill" Research Report 1309). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

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

Harmon, J. M., & Miller, A. G. (1950). Time patterns in motor learning. Research Quarterly, 21, 182-187.
Hopkins, C. O., Snyder, H. L., Price, H. E., Hornick, R. J.,

Mackie, R. R., Smillie, R. J., & Sugarman, R. C. (1982). Critical human factors issues in nuclear power regulation and a recommended comprehensive human factors long-range plan. Critical discussion of human factors areas of concern. Human Factors Society.

Keller, F. S., & Estes, K. W. (1944). Distribution of practice in code learning" (OSRD 4330, OEM SR-830, Service Project SC-88). New York: The Psychological Corporation. (ATI 35335)

Keller, F. S., & Estes, K. W. (1945). The relative effectiveness of four and seven hours of daily code practice" (OSRD 4750, OEM SR-830, Service Project SC-88, Report #3). New York: The Psychological Corporation. (ATI 35419)

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. Research Quarterly for Exercise and Sport, 59.

Lintern, G. (1988). Distributed practice: Are there useful insights for application or theory? Research Quarterly for Exercise and Sport, 59, 298-302.

Magill, R. A. (1985). Motor learning: Concepts and applications" 2nd ed.). Dubuque, IA: Wm. C. Brown Publishers.

Magill, R. A. (1988). Activity during the post-knowledge results interval can benefit motor skill learning. In O. G. Meijer K. Roth (Eds.), Complex motor behavior" (pp. 231-246). Amsterdam: Elsevier Science Publishers.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. Journal of Verbal Learning and Verbal Behavior, 9, 596-606.

Newell, K. M., Antoniou, A., & Carlton, L. G. (1988). Massed and distributed pactice effects: Phenomena in search of a theory? Research Quarterly for Exercise and Sport, 59.

Prather, D. (1973). Prompted mental practice as a flight simulator. Journal of Applied Psychology, 57, 353-355.

Reynolds, B., & Bilodeau, I. M. (1952). Acquisition and retention of three psychomotor tests as a function of distribution of practice during acquisition. Journal of Experimental Psychology, 44, 19-26.

Schmidt, R. A. (1982). Motor control and learning. Champaign, IL: Human Kinetics Publishers.

Schultz, D. P., & Schultz, S. E. (1986). Psychology and industry today: An introduction to industrial and organizational psychology" (4th ed.). New York: MacMillan Publishing Company.

Shugarts, D. A. (1987). Combination private-IFR under study. Aviation Safety, VII(5).

Young, O. G. (1954). Rate of learning in relation to spacing of practice periods in archery and badminton. Research Quarterly, 231-243.

STRESS TRAINING EFFICACY IN AN AVIATION CONTEXT

Christopher K. McClernon
The MOVES Institute, Naval Postgraduate School
Monterey, CA

Stress is regularly introduced in training to prepare troops for stressful environments and situations, although there is very little empirical evidence for stress training's effectiveness, implementation and pedagogy. Twenty novice participants were recruited and assigned to either a stress-trained (cold pressor), treatment group or a control group. Stress training was effective at improving the treatment group’s performance during a final criterion session on an aircraft navigation task compared to the control group. In addition, the stress-trained group showed lower criterion heart rate variability, skin conductance, and subjective stress ratings compared to the control group. This research demonstrates stress training as a viable approach for preparing military members for stressful flight environments and combat, in general. Further research addressing the generalization of these results to novel, real-world stressors is proposed.

Stress is often introduced in training so that real-world stress is more familiar and easily mitigated. Stress training relies on transfer-of-training research including Osgood's (1949) similar elements theory, indicating that training should share elements with the transfer task in order for the training to be effective, or as the military often refers to as “train how you fight”. Therefore, to train for a stressful task, a stressful training environment is often utilized in the military. Stress training also relies on Overton's (1964) state-dependent learning where retention and retrieval is dependent on a person's emotional, physiological, and mental states during both training and recall.

A notable component of training pedagogy is Driskell, Johnston, and Salas's (2001) reference to structural and surface features. In an attempt to explain the generalization of stress training to various transfer stressors, they go beyond Thorndike and Woodworth’s (1901) identical elements theory to emphasize the features relevant to both the training and transfer tasks. Structural features refer to “the underlying principles imparted in training,” while surface features refer to “domain-specific characteristics, such as the specific training examples used and the specific attributes of training context” (p. 108). For example, in an aviation context their research suggests that a pilot will benefit from stress training even if the context (surface features) for which he received the stress training (i.e., a low-fidelity flight simulator) is different from the transfer task (i.e., an aircraft). In this case, the stress training itself is thought of as a structural feature, and in their study “it is likely that the stress training resulted in positive transfer to novel settings [and novel stressors] because the underlying principles in training were structurally consistent with the transfer environment” (p. 109).

Stress Training

Stress occurs when “the perceived demands of a situation tax or exceed the perceived resources of the system (individual, group, community) to meet those demands, especially when the system’s well-being is at stake” (Lazarus & Folkman, 1984, p. 8). Early stress researcher Hans Selye (1974) proposed that there are three different responses to environmental stressors. First, from a biological standpoint, an organism can exhibit a catatoxic response to stressors in which a fight response attempts to actively eliminate a stressor. This response can be either impossible or very taxing on an organism. In a combat context, fighting peripheral stressors (e.g., heat, radio noise, ambushing enemy troops, etc.) often distracts from the operational mission. A second response to stressors is the flee response in which an organism attempts to remove itself from a stressor. Again, in a combat context, this response is typically infeasible short of retreat. Finally, and often the most plausible response to stressors as defined by Selye, is a syntoxic response in which an organism learns to co-exist with stress. It is this coping mechanism which allows an organism to exhibit goal-seeking behaviors in the presence of stressors--this behavior is most favored in a combat setting and is the premise for this research endeavor.

Stress can have a profound effect on performance which is demonstrated by the Yerkes-Dodson Law (Yerkes & Dodson, 1908). An optimum amount of cortical arousal corresponds to the highest performance for a given task. Further, as more stress is introduced, resulting in heightened arousal and an appraisal of more demands than resources, performance will deteriorate. The same is true for low levels of stress (and associated lack of arousal) in which boredom and/or complacency can result and performance suffers again. The Yerkes-Dodson Law

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

is also task dependent; simple cognitive tasks, which may insinuate relatively low attention and cognitive requirements, may be subject to higher levels of stress before optimum cortical arousal is achieved, and vice versa. Finally, the Yerkes-Dodson Law is subject to individual differences, where novices may be quickly inundated with stress, while experts may not only successfully moderate the effects of stress but they may *require* higher levels of stress to prevent complacency and optimize performance. Individual differences can also refer to innate personality characteristics and coping strategies employed by different people.

In an attempt to alleviate individual workload and stress many attempts have been made to introduce automation, technologies, intuitive controls/displays, etc. However, these approaches are often very costly, have theoretical limitations, and, in some cases, can actually increase stress. Another approach to alleviate stress in which a dearth of research exists is stress training. However, introducing stress training can often result in two counterproductive results: (1) high levels of anxiety which inhibits both training and post-training performance, and (2) stress which interferes with the acquisition of skills and knowledge for which the training is designed to promote (Friedland & Keinan, 1992). Friedland and Keinan present three stress training approaches in an attempt to (1) effectively impart skills and knowledge on trainees, (2) expose and inoculate trainees to real-world levels of stress, and (3) alleviate the before mentioned shortcomings.

First, a graduated-intensity approach to stress training attempts to slowly "inoculate" trainees with a progressive training protocol while theoretically not impeding task acquisition. While graduated-intensity training can provide trainees with a heightened sense of control and competency, the gradual introduction of stress can breed unrealistically low expectations about transfer task stress levels. Another potential approach to stress training is adapting the training to an individual's needs. By tailoring stress training to each individual, a training system can ensure that both appropriate levels of stress exposure and criterion performance are met, independent of time in training. However, this method relies on the trainee's subjective perception of resources, and may not always indicate proficiency for the task at hand (Friedland & Keinan, 1992). A final training approach that has shown the most promise for alleviating stress while ensuring the acquisition of skills is by following a validated three-phase approach to stress training. Friedland and Keinan defined and tested the interplay between three elementary phases of training: task acquisition (TA), stress exposure (SE), and practice under stress (PUS). They found that introducing stress training in this order resulted in the best performance in a criterion task. These findings indicate that a phased approach to stress training is both effective and efficient. In addition, task acquisition is of foremost importance in any training program; without the proper development of skills, transfer performance will certainly suffer. Lastly, complete compartmentalization of stress exposure or skill acquisition (e.g., exclusion of the PUS phase) is insufficient--both the task and stress exposure must be integrated into a total training plan.

Johnston and Cannon-Bowers (1996) provide further guidance for Stress Exposure Training (SET) which evolved from three main objectives: building skills that promote effective performance under stress, building performance confidence, and enhancing familiarity with the stress environment. Not unlike SET, stress inoculation training (SIT) is a phased training approach in a mostly clinical context, although there are many potential applications of SIT for the training of stressful tasks. As Meichenbaum (1993) points out, the primary notion of SIT is that "bolstering an individual's repertoire of coping responses to milder stressors can serve to defuse maladaptive responses or susceptibility to more severe forms of distress and persuasion" (p. 378). Both SET and SIT consist of three phases, congruent to Friedland and Keinan's three phases.

In a meta-analysis of 37 studies testing the effectiveness of a three phased approach to stress training (i.e., SET or SIT), Johnston and Cannon-Bowers (1996) determined that 67% of studies demonstrated that stress training significantly improved performance. In their meta analysis, Saunders, Driskell, Hall, and Salas (1996) determined that stress training was effective in reducing performance anxiety, reducing state anxiety, and enhancing performance under stress. Despite these findings, no research addresses the applicability of stress training to a stressful flight environment.

Methodology

A study was developed to test the efficacy of a three-phased training approach in an aviation context. Unlike previous research which used a very short-term stressor (e.g., 15 seconds; Friedland & Keinan, 1992; Rosenbaum, 1980), a longer acute stressor (10-15 minutes) was introduced--this is more congruent with aviation and/or combat stressors. In addition to performance measures, human physiological responses and subjective appraisals were used to evaluate the efficacy of stress training in this context.

Participants. Twenty participants (16 males, 4 females) were recruited and ranged in age from 25-39. All participants were required to have 20/20 corrected vision and no known health ailments. In addition, none of the participants had any flying experience.

Equipment. The virtual environment (VE) consisted of 1 desktop computer, 1 visual display, a flight yoke control input, and an audio headset. The only control input for the simulator was a Precision Flight Instruments Cirrus yoke. The rudder (yaw) controls were coupled to the yoke and the throttle was set at a constant setting by the experimenter. This was to limit the amount of training required before testing the applicable research questions.

Trials took place in the X-Plane® v8.0 flight simulator using the included Cessna 172 flight model and instrument panel. The simulator outside view was replaced with instrument meteorological conditions (i.e., visibility less than three nautical mile) and the aircraft display was limited to only three instruments: attitude indicator, altimeter, and directional gyro. Simulator audio output consists of the simulator audio feedback (i.e., engine and wind noise), air traffic control instructions (i.e., *clearances*), air traffic background "chatter", and experimenter/participant voice. The physiological data collection equipment consisted of the Thought Technologies Ltd. ProComp Infinity™ Biofeedback System using both electrocardiogram (EKG) and Skin Conductance (SC) sensors. The BioGraph Infinity 4.0® software was used to collect and analyze the physiological data.

Stressor. The stressor was a cold pressor similar to Friedland and Keinan (1992) and Rosenbaum (1980). The cold pressor method consists of submerging the participant's foot in a bucket of ice water kept at a constant 9°C. This method applied to the hand has proven to effectively introduce stress without harming participants--given the hand dexterity required for this task, the pressor was applied to each participant's left foot. A pilot study determined that this stressor interacted with the primary task and reliably affected physiological responses to stress without undue discomfort to the participant.

Task. Following an administrative portion, participants received training on flying the aircraft from a licensed pilot. They were first taught the functions of the three instruments and the control yoke. They were then taught strategies for flying straight and level, turning, climbing, and descending. Following an instructional video, participants were then given time to practice these procedures. Following training, both control and treatment groups performed a TA session consisting of turns, climbs, and descents to provided clearance headings and altitudes (see Figure 1). A pilot study determined mean times to asymptotic performance during this session which determined the length of this and subsequent sessions.

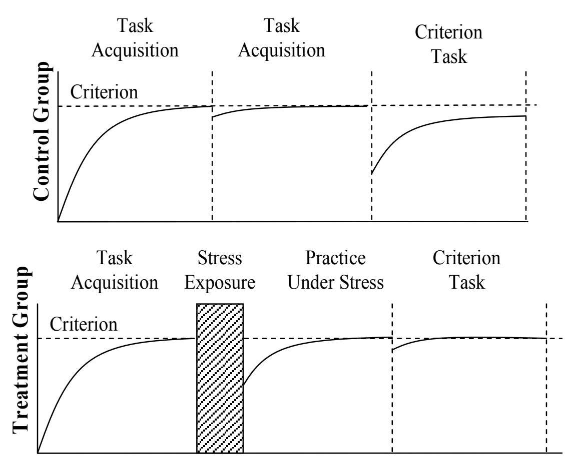


Figure 1. Experiment schedule

Next, the treatment group received stress exposure, separate from the task, where stress mitigation strategies were practiced. The treatment group then performed the flying task while exposed to stress and the control group performed a second TA session to control for the amount of time both groups received in the simulator. Finally, both groups performed a criterion session where the task `was performed during exposure to the stressor. Figure 1 shows hypothetical performance curves for control and treatment groups.

Dependent Variables

Performance. Flight simulator data was first parsed into maneuvering and straight/level portions of flight, and analysis was performed for both phases of flight. For maneuvering portions of flight, roll and pitch were used to calculate the root mean square error (RMSE) from the prescribed criteria (20- and 10- degrees, respectively). These two measures were then combined using a euclidean transformation ($\sqrt{x^2+y^2}$) to formulate a total measure of maneuvering error during each session. Likewise, for straight/level portions of flight, heading and altitude error from the provided clearances were calculated and combined in a similar manner.

Heart rate variability. While heart rate variability (HRV) is a time domain measure of the deviation between heart beat intervals, a power spectral density (PSD) defines the frequency content of this time-based stochastic process by performing a Fourier transformation of the time domain data. The PSD provides the ability to distinguish between different frequency spectra and associated *types* of vagal activity, or autonomic heart rate modulation. Physiological responses to stress often include an increase in LF power (LF; 0.05-0.15 Hz), a decrease in HF power (HF; 0.15-0.5Hz), and an increase in the LF/HF ratio (Berntson & Cacioppo, 2004; Pfieper & Hammill, 1995). Total very low frequency (VLF), LF, and HF power densities were computed as a function of baseline power for each maneuver and the total criterion session. Baseline measures were recorded during the instructional video.

Skin conductance. Another physiological correlate of human emotion is electrodermal response often measured using skin resistance, or "the electrical resistance of the skin to the flow of electromotive current and is measured in ohms" (Grossman, 1967, p. 504). The reciprocal of skin resistance, skin conductance (SC; measured in micro-Siemens, μS), is used to indicate autonomic nervous system activity, and, thus, allows inference regarding emotions. Unlike EKG, SC is a relative measure, meaning only an increase or decrease in individual SC relative to a baseline can indicate a heightened level of emotional arousal. Therefore, all SC measures were calculated as a function of each individual's baseline SC measure.

Subjective stress ratings. Subjective measures are often used as an indicator of human physiology. However, where physiology measures indicate the human body's response to a stress, subjective measurement provides insight about a person's appraisal of a situation given their available resources. Before each TA session, a sample query was conducted to determine each participant's baseline subjective stress level. Participants were asked, "Rate your current stress level on a scale from 1 to 10". Five times during each session, participants were again asked the same stress query. Each query was computed as a function of each participant's baseline.

Data Analysis and Results

The experiment followed a between-subject design and each participant was randomly assigned to either the treatment or control group. An Analysis of Variance (ANOVA) was conducted on all performance, physiology, and subjective stress rating response measures with condition used as a predictor variable. An alpha level of 0.05 was used to identify any significant effect of stress training.

Performance. To account for individual differences, performance was computed as the mean criterion performance divided by each individual's asymptote performance during the final three minutes of the TA session (Cr/TAend). Stress training did significantly improve the performance of the treatment group for maneuvering portions of flight ($F(1,18)=4.43$, $p=0.049$) and total performance ($F(1,18)=4.87$, $p=0.040$; see Figure 2). Variability within each measurement was also collected and analyzed. A method pioneered by Mackie and Miller (1978), called the standard deviation of lane position (SDLP), measures the amount a driver weaves within a lane. For the purposes of this analysis, SDLP was used to determine how much a participant varied regardless of their assigned clearance. This variance measure showed no significant difference for altitude and heading (i.e., straight and level portions of flight; $p>0.05$). However, it did indicate superior treatment group performance when measuring differences in pitch ($F(1,18)=5.98$, $p=0.025$) and roll ($F(1,18)=5.18$, $p=0.035$) for the Cr/TAend measure (see Figure 3). These findings indicate the effectiveness of stress training to improving both precision and accuracy. This finding is not only pivotal for the current research, but it also provides evidence for the use of SDLP in an aviation context.

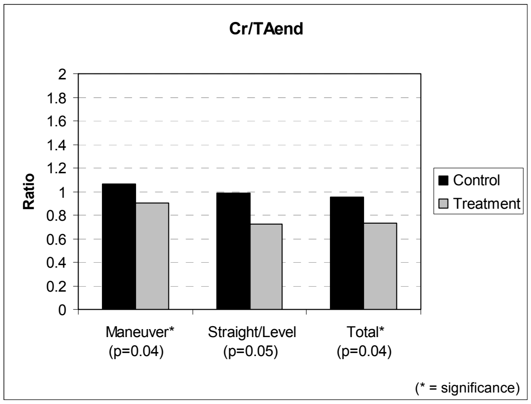


Figure 2. Mean criterion/asymptote RMSE

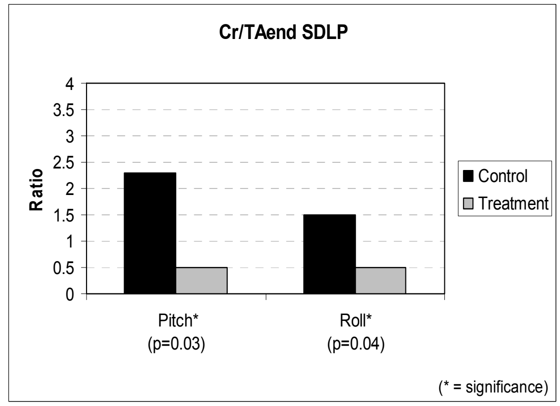


Figure 3. Mean criterion/asymptote SDLP

Heart rate variability. When comparing overall mean session HRV, an ANOVA determined that during the criterion session the control group had a higher mean LF PSD compared to the treatment group ($F(1,18)=7.74$, $p=0.012$). This indicates that the treatment groups' stress training significantly improved the participants' ability to mitigate the stress of the cold pressor in a stressful criterion task.

Skin conductance. After review of the SC data, one participant's data indicated erroneous conductance readings (~90 standard deviations from the group's mean SC). After removing this participant (N=19), ANOVA tests indicated no significant differences for the practice, TA, and TA2/PUS sessions ($p>0.05$), and a significant difference between groups for the final criterion session ($F(1,17)=4.61$, $p=0.047$).

Subjective stress ratings. As expected, analysis of the final criterion session revealed a significant difference between groups for the first ($F(1,17)=13.93$, $p<0.01$), second ($F(1,17)=13.13$, $p<0.01$), third ($F(1,17)=6.49$, $p=0.021$), and fourth ($F(1,17)=4.60$, $p=0.046$) minute in addition to the overall session average ($F(1,17)=8.00$, $p=0.011$).

Conclusions

The results of this study indicate that, similar to Friedland and Keinan (1992), participants who received a three-phased stress training protocol demonstrated superior performance in a stressful criterion task when compared to a control group. Participants were both more accurate and precise when asked to fly to specified headings and altitudes. Yerkes and Dodson's (1908) task dependency was also demonstrated with participants benefiting most during the more difficult maneuvering portions of the criterion session.

Stress training also was beneficial for moderating the physiological responses to stress. Furthermore, when each maneuver and straight/level portion of the criterion session was analyzed separately, the stress-trained group demonstrated lower HRV for six of the eight maneuvers, but only two of the eight straight/level portions of flight ($p<0.05$). This further indicates that stress training may be more effective for decreasing strain only for portions of a task which are inherently stressful to begin with (i.e., maneuvers). In addition, HRV and SC results were most compelling (i.e., more maneuvers found significant) for the early portions of the criterion session. This indicates an increased stress training efficacy for the onset of stress.

There was also a surprising lack of significant difference between the groups' HRV during the second TA session (control group) and the PUS session (treatment group; $p>0.05$). One possible explanation for this insignificance is the effectiveness of the stress exposure training. The stress coping strategies and exposure to the stressor gained during this session may have substantially prepared the participants physiologically for the following PUS session. Another explanation for this lack of finding is the anticipatory response of the control group. During the administrative portion of the experiment, both groups read the Informed Consent Form which stated they would be exposed to cold ice water. As the experiment progressed and the control group was still not exposed to the stressor they may have developed an anticipatory response which was equal in physiological strain to what the treatment group was experiencing under the actual cold pressor stress. Further testing of these hypotheses is needed.

The efficacy of stress training generalization to novel, real-world stressors is still largely unknown. Preliminary empirical evidence indicates the relative importance of training structure features versus domain-specific surface features (Driskell, Johnston, & Salas, 2001; Saunders, Driskell, Hall, & Salas, 1996). Furthermore, these results indicate the promise for training stress exposure in a VE as long as the structural features of the virtual training include the expected stress levels and are congruent with the real-world task(s).

Stress training is an effective means for preparing individuals for stressful flight environments. This training relies on developing the skills to accomplish a task, learning stress coping techniques, and practicing the task under stress. Although this approach is now validated within a laboratory environment, its generalization and transfer to a real-world setting requires exploration. Only when this critical connection is made can appropriate training pedagogy be developed.

References

Berntson, G., & Cacioppo, J. (2004). Heart rate variability: Stress and psychiatric conditions. In M. M. A. J. Camm (Ed.), *Dynamic Electrocardiography* (pp. 56-63). New York: Futura.

Driskell, J. E., Johnston, J. H., & Salas, E. (2001). Does stress training generalize to novel settings? *Human Factors, 43*(1), 99-110.

Friedland, N., & Keinan, G. (1992). Training effective performance in stressful situations: Three approaches and implications for combat training. *Military Psychology, 4*(3), 157-174.

Grossman, S. P. (1967). *A Textbook of Physiological Psychology*. New York, NY: Wiley and Sons, Inc.

Johnston, J. H., & Cannon-Bowers, J. A. (1996). Training for stress exposure. In J. E. Driskell & E. Salas (Eds.), *Stress and Human Performance* (pp. 223-257). Mahwah, NJ: Lawrence Erlbaum Associates.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.

Mackie, R., & Miller, J. (1978). *Effects of hours of service, regularity of schedules, and cargo loading on trucks and bus driver fatigue* (HFR- No. TR-1765-F). Goleta, GA: Human Factors Research, Inc.

Meichenbaum, D. (1993). Stress inoculation training: A 20-year update. In P. M. Lehrer & R. L. Woolfolk (Eds.), *Principles and Practice of Stress Management* (Second ed., pp. 373-406). New York: The Guilford Press.

Osgood, C. E. (1949). The similarity paradox in human learning: A resolution. *Psychology Review, 56*, 132-143.

Overton, D. A. (1964). State-dependent learning produced by depressant and atropine-like drugs. *Psychopharmacology, 10*(1), 6-31.

Pfieper, S. J., & Hammill, S. C. (1995). Heart rate variability: Technique and investigational applications in cardiovascular medicine. *Mayo Clinic Proceedings, 70*, 955-964.

Rosenbaum, M. (1980). Individual differences in self-control behaviors and tolerance of painful stimulation. *Journal of Abnormal Psychology, 89*, 581-590.

Saunders, T., Driskell, J. E., Hall, J., & Salas, E. (1996). *The effects of stress inoculation training on anxiety and performance*. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Selye, H. (1974). *Stress without distress*. New York: New York: American Library.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 8*, 247-261.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18,* 459-482.

Acknowledgements

The work presented in this paper is a portion of the author's PhD dissertation. The author would like to thank his dissertation advisor, Dr. Michael McCauley, and the following Naval Postgraduate School faculty members for their continued support: Drs. William Becker, Anthony Ciavarelli, Rudy Darken, Ronald Fricker, Nita Miller, and Brent Olde. Finally, the author would like to thank the Office of Naval Research and the United States Air Force Academy for their funding and support.

APPLIED THREAT AND ERROR MANAGEMENT: TOWARD CREW-CENTERED SOLUTIONS

Eric E. Geiselman
Aptima, Inc.
Dayton, Ohio

For an operator, a high level of understanding regarding procedures enables appropriate defenses to be built into a robust Threat and Error Management (TEM) framework. Currently, airline flightdeck crewmember training and reference information is concentrated heavily on what and how procedures are performed, but not on why they must be performed a standard way. This missing component of certainty invites misinterpretation of the standards and induces error. I propose a Crew-Centered TEM (CC/TEM) approach designed to arm flight crewmembers with more depth of procedural understanding than that currently afforded. A recent accident where human error was identified as a probable cause is used as an example of how a CC/TEM approach may have prevented the occurrence. CC/TEM solutions have further application within other safety-critical domains, such as medicine and emergency response.

Post investigation statistics indicate 70-80% of air carrier accidents include human error as a causal factor (Shappell and Wiegmann, 2000). Use of the term "human error" is intended to be inclusive, versus the old term "pilot error," in order to recognize that humans other than pilots often contribute to the error chain. The current terminology may be appropriate, but it is also less descriptive and ultimately less meaningful toward the determination of cause. One may ask: "are twenty percent of air carrier accidents truly free of human error?"

One candidate as a "machine cause" accident is the crash of United Flight 232 at Sioux City, Iowa on July 19, 1989 (NTSB, 1989). The three-engine DC-10 was at cruise altitude when the tail-mounted engine experienced an uncontained catastrophic failure. Shrapnel from the engine pierced through the starboard horizontal stabilizer and severed the hydraulic lines where all three of the systems were plumbed together. This resulted in a profound loss of controllability due to the total loss of hydraulic pressure in a triple-redundant system. Through the use of differential thrust via the remaining two engines, the crewmembers were able to approach a runway at the Sioux City airport. Because of the lack of controllability, the aircraft contacted the ground in a wing-low attitude, broke up, and caught fire. Of the 296 passengers and crew aboard, 111 people died.

From the perspective of cause, it is clear that the flightdeck crew were not a factor, but was the accident free of human error? The investigation determined that a maintenance inspection failed to detect a fatigue crack in the engine part found to be responsible for the failure. Even beyond that, the engine cowling was designed to contain this mode of failure. Was the redundancy of the hydraulic system vulnerable due to faulty design? The inclusiveness of the term "human error" extends all the way to the designers' drawing board.

Because aviation is a human endeavor, it can be stated that human error, at some level, is always involved in its failure. Given this, the use of inclusive human error as causal factor in accident should be placed at 100% versus the 70-80% static often used. This does not mean that problem is worse than previously indicated. It means that the determination of human error as the cause of failure serves only as a description of what happened and falls well short of explaining why the failure occurred. In terms of prevention, it is not as important to classify the human error as it is to determine the cause of the error. Human error does not cause crashes; whatever causes human error is what causes crashes. When we fall short of a complete understanding of error cause, we also miss an opportunity to avoid future occurrence by the generation of effective defenses within Threat and Error Management (TEM) applications at the system operator level.

Conventional TEM

Conventional TEM begins with threats that can be expected to be faced by operators (flightdeck crewmembers). A series of defenses are in place to allow the crew to manage threats as they develop. Defenses built into the model include training, procedures, technology, automation, standards, regulations, Crew Resource Management (CRM) practices, etc. The defenses can be thought of as multiple layers of a protective barrier to fend off failure (Reason, 1990). If the barrier is breached, error is generated. The error is now protected by another down-steam barrier with the intent of trapping and correcting the error. If the error is able to propagate along the model, an undesired state is produced. Failure of the crew to detect and correct the undesired state indicates that another defense-set barrier was breached. If left unmitigated by some entity external to the crew (such as air traffic control), then an occurrence is the result. An occurrence can be an accident, incident, regulation violation, or any other negative outcome. In reality, the model is complex, multidimensional, and dynamic. Included are feed-forward and feedback loops which enable errors and undesired states themselves to become threats, which then reenter at the beginning of the dynamic model.

When occurrences are realized, the conventional solution is to patch or add additional layers of defense into the applied TEM model. This is similar to the addition of redundancy within hardware and software systems. But, if those systems are faulty, the addition of redundancy does nothing to prevent individual component failure. Similarly, additional layers of defense within the conventional TEM model do not prevent error generation as much as they are present to reduce error propagation. It is proposed that a higher-level approach is required to help prevent error generation. This is the objective of a Crew-Centered TEM (CC/TEM) approach.

Crew-Centered TEM

It can be generalized that threats come <u>at</u> the crew, and errors come <u>from</u> the crew. By default, this centralizes threat and error management responsibility to the crew. The new concept of Crew-Centered TEM (CC/TEM) is intended to provide crewmembers with high-level resilient defenses (super defenses) against threat and error propagation. This is accomplished though training, manual-based information, and the appropriate modification of standard procedures. The objective of CC/TEM training is to promote standardized conceptual-level understanding of <u>why</u> procedures are required to be performed in a specified manner (procedure explanation). CC/TEM requires that explanation training be added to current training practices, which focus almost exclusively on <u>what</u> and <u>how</u> standard procedures are to be performed (procedure directives). Specific CRM training will reinforce CC/TEM objectives by emphasizing the importance of professional conduct and standards discipline. A CC/TEM manual system will be consistent with and complement training by providing procedure explanation reference material designed to promote crewmember understanding. Within the manuals, procedure explanations should be aligned logically and co-located within the expanded checklist sections. A complete CC/TEM program includes as standards only those procedures that can be explained. In addition to the above, when changes are made to standard procedures, an explanation of why the change is required will accompany the introduction of the change. When the change is implemented, crew training and the manual system will include an updated procedure explanation. Once acquired, this level of standardized conceptual understanding on the part of the crew will produce better judgment and decision making by promoting the execution of standard procedures as they are intended to be performed. The overall objective of CC/TEM is to reduce uncertainty and maintain a stronger defense against all error producing variables. Failure on the part of a crewmember to understand why a standard procedure exists is by itself a significant threat. Uncorrected, the threat may lead to misinterpretation and selective noncompliance errors. The following accident is an example of the type of occurrence which can and does result.

CC/TEM Example

Comair Flight 5191 Accident

Comair (dba: Delta Connection) Flight 5191 was scheduled to depart for Atlanta, Georgia (ATL) from Lexington, Kentucky (LEX) at 6:00am on August 27th, 2006. Having arrived separately the previous evening, the crewmembers (captain, first officer (F/O), and flight attendant) met for the first time approximately one hour prior to scheduled departure time. During the pre-flight setup and checks, a ramp agent delivered the dispatch release documents to the crew indicating that they were on the wrong airplane (a Bombardier CL-600). The crew shutdown the airplane and transferred to the correct plane. Now behind schedule, they continued their pre-flight system checks while the passengers were boarded. The flightdeck conversation was friendly and informal. The captain briefed the F/O and indicated that he was "laid back" and "easy going." The F/O indicated that he "appreciated that." The F/O briefed the captain that he had flown in the night prior and that there were "a bunch of lights out all over the place."

Contrary to company standard procedures, the F/O gave a top-level taxi route briefing to the captain: "right turn alpha taxi to the runway." The F/O referred to runway 24 and was corrected by the captain: "you mean runway 22?" The company Flight Standards Manual (FSM) and Operations Manual (OM) requires the captain to brief the taxi route and refer to the airport diagram while doing so. The tower controller cleared flight 5191to: "taxi to runway 22 cleared for takeoff runway 22." While maneuvering the aircraft, the captain called for Before Takeoff Checklist items to be completed at the F/O's "leisure." While completing the checklist items, the F/O violated sterile cockpit conditions on several occasions. At the end of the runway, the controls of the airplane were transferred to the F/O, who was the pilot flying. The takeoff roll was initiated while the tower controller attended to some administrative paperwork. His back was turned to the active runway.

The runway the aircraft was actually on was not the 7003-foot-long runway 22 for which it was cleared. Instead, the airplane was lined up on for takeoff on runway 26: a 3500-foot-long general aviation runway. The aircraft became airborne but gained little altitude before hitting a small rise and then a fence off the end of runway. The aircraft contacted trees, slid across a field, and burned. Of the 50 people on board, 49 were killed. The F/O survived the crash, but suffered severe injury and permanent disablement.

National Transportation Safety Board (NTSB) Investigation

The following is the probable cause statement from the Flight 5191 NTSB final report. (NTSB, 2007):

> *The National Transportation Safety Board determines that the probable cause of this accident was the flight crewmembers' failure to use available cues and aids to identify the airplane's location on the airport surface during taxi and their failure to cross-check and verify that the airplane was on the correct runway before takeoff. Contributing to the accident were the flight crew's non-pertinent conversation during taxi, which resulted in a loss of positional awareness, and the Federal Aviation Administration's (FAA) failure to require that all runway crossings be authorized only by specific air traffic control (ATC) clearances.*

The probable cause statement addresses what the NTSB believes happened and characterizes it as multiple failures (errors of omission) on the part of the crew to notice cues which should have led them to realize they had lined up on the wrong runway. One other error of commission is the violation of sterile cockpit. There is nothing mentioned within the probable cause statement to indicate what may have caused the errors.

NTSB Findings and Recommendations

The specific findings in the report include 28 factual items. Of those, eight items of crewmember error are indicated. Of the crewmember errors, two specific items of procedural error are included. The procedural errors are cases where the crew failed to comply with standard operating procedure and/or federal aviation regulations. Within the context of conventional TEM, this is indicative of breached layers of defense.

The recommendations include the addition of regulatory changes (e.g. require runway cross-check with other cues) and at least one technology addition change (e.g. develop a moving map display with automatic warning capability). These recommendations represent the addition of redundant defense layers designed to mitigate the effect if the same error chain were to develop in the future. None of these recommendations, applied to the Flight 5191 scenario, would have prevented the aircraft from at least attempting to line up on the wrong runway.

CC/TEM Analysis

To help determine how a CC/TEM approach would apply to the Flight 5191 accident, we must analyze the events from within the scenario from the perspective of the people involved. Additionally, we must investigate events in the absence of information that was not available to the crew at the time of the occurrence. Also, the analysis should proceed without the benefit of knowing the outcome (Dekker, 2006). The objective is to generate questions designed to explore the plausibility of what may have caused the error and why the error was able to propagate through the entire set of TEM defenses. Once this process is complete, the most important step is to build back into the system super defenses which have a reasonable probability of preventing not just the specific chain of events of the current analysis, but also those which are quite different and unanticipated. The following is an example analysis question from Comair Flight 5191:

What if the captain believed there was only one runway at LEX—and that belief was never disproved?

A few of the NTSB findings allude to just such a possibility. Finding number 4: "The captain and the first officer believed that the aircraft was on runway 22 when they taxied onto runway 26 and initiated the takeoff." Finding number 9: "...because they did not cross-check and confirm the aircraft position on the runway before takeoff, they were likely influenced by confirmation bias." (i.e. the crew was only paying attention to those cues which confirmed what they believed to be true, to the exclusion of those cues which should have made them question their beliefs). Finding number 11: "The crew's noncompliance with standard operating procedures (SOPs), including the abbreviated taxi briefing and the non-pertinent conversation most likely created an atmosphere in the cockpit that enabled the crew's errors." According to the NTSB, finding 11enabled the other errors. So, in accordance with the CC/TEM process, we ask: what caused the noncompliance?

Noncompliance Issue

Regarding the CC/TEM analysis questions, the significant noncompliance revolves around the "Captains Standard Taxi Briefing" required action prior to the "Before Starting Engines" section of the expanded checklist. The procedure is defined in and required by the company Flight Standards Manual (FSM) and the associated Operations Manual (OM). Satisfaction of a command (TAKEOFF BRIEF) and response (COMPLETE) line item on the crew checklist is intended to confirm that this briefing was delivered by the captain and received by the F/O. The SOP requires that briefing include at least the following items prior to every flight: anticipated taxi route, runway crossings, hot spots (high threat areas), etc. Furthermore, the procedure requires that the airport diagram (Figure 1) be "out and available" and reviewed during the briefing. There is no information in the manual system that presents an explanation of why the standard procedure exists in its prescribed form.

As the NTSB indicates, the Flight 5191 crew failed to comply with the standard procedure: 1) the captain allowed the first officer to deliver the taxi briefing. 2) The taxi briefing did not include any of the required subject items to the level of detail intended by the procedure. 3) There is no evidence that the airport diagram was referenced by either of the crewmembers. From the airport diagram, it is clear that there is a runway (runway 26) which must be crossed in order to taxi to runway 22 at LEX when departing to the south (Figure 1).

Noncompliance Cause

Reviewing the airport diagram was the captain's first opportunity to confirm that LEX actually has two runways, one of which must be crossed in order to taxi to the end of runway 22. It is plausible that the captain believed that LEX is a single runway airport. From an airline operational perspective, there is only one usable runway at LEX. At single runway airports, there are few critical briefing items which need to be covered. As long as the aircraft proceeds in the right direction toward the correct end of runway, there are few threats. This may have been the reason the captain was satisfied with the F/O's non-standard briefing. The F/O was nearing an upgrade to captain status at the airline and had mentioned that fact to the captain. The captain may have chosen to overlook the non-standard conduct of the briefing to avoid shutting down the amicable rapport he had established among his crew. It is entirely possible that, by his interpretation, the captain believed he was exercising good CRM by allowing this series selective noncompliance to develop and continue uncorrected. He may have considered these infringements to be a low-risk tradeoff compared to creating conflict and coming across as less than "laid back." On the receiving end of this behavior, the F/O is exposed to a willingness on the part of the captain to operate in a less than standard manner. In a more overt indication, the F/O was directed to run and complete standard procedures "at your leisure." This acts only to encourage other downstream "minor" deviations from the standard such as the one-side violation of sterile cockpit procedures on the part of the F/O. These are clear examples of routine violation type unsafe acts according to the Human Factors Analysis and Classification System (HFACS) taxonomy (Shappell and Wiegmann, 2000).

CC/TEM Discussion

Misinterpretation of the intent of SOPs has a negative impact on the accuracy of the information used by the decision maker. The environment for misinterpretation is created and sustained if training and informational resources do not adequately enable an unambiguous conceptual level of understanding with regard to SOPs (as well as policies, best practices, applicable techniques, etc.). This can be considered an undesired state, or at best, a latent threat within the operating model. In the above example, a lack of this level of understanding may have initiated the events leading to the occurrence. Why is the captain required to deliver the taxi briefing and include a review of the airport diagram? It is to assure that the captain gets an accurate “big picture” appreciation for the airport layout. This situation awareness building activity is designed to inform the captain as a decision maker. If asked, line flightdeck crewmembers may not consistently identify the correct objective of the taxi briefing requirement. The following is an example of a likely misinterpretation: the reason that the captain is required to give the briefing is because taxi control is a captain function to be backed up by the F/O’s vigilance during surface movement. There are countless ways this standard procedure may be interpreted, but only one correct explanation. Arriving at the correct interpretation is simple if operators are made aware of what it is in the first place. This awareness should be considered a basic requirement of the operator knowledgebase. This is especially true for those involved in safety-critical and unforgiving endeavors. Those operators must know what, how, and <u>why</u>. In terms of CC/TEM processes, accurate <u>why</u> knowledge is a super defense, and its absence represents a very real threat.

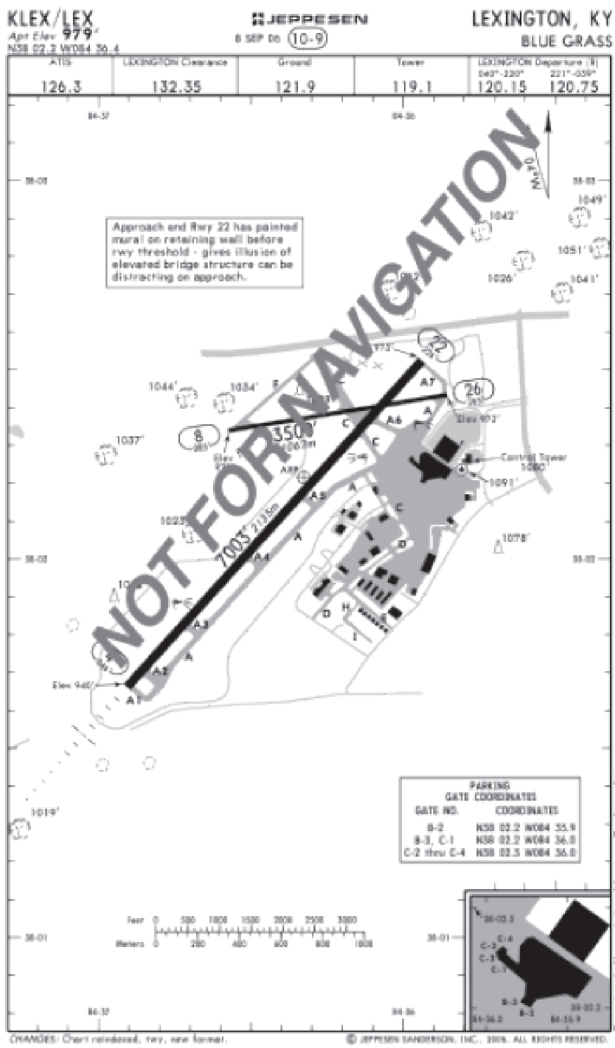


Figure 1. LEX airport diagram chart.

For Comair Flight 5191, the proper conduct of the standard taxi briefing could have resulted in the flight being just another routine completion. The difference may have been as simple as the captain taking the time for a quick glance at the airport diagram. To a large extent, this is all that is required to satisfy the spirit of the standard procedure. Had the airport diagram been reviewed per the SOP, the requirement to cross runway 26 on the way to runway 22 may have been understood before the aircraft pushed off the gate. An objective of a CC/TEM approach is to ensure that the determination of which standards must be followed as proscribed, and which are flexible, is not a matter of crewmember interpretation.

As a new concept, CC/TEM should be considered as an extension of the significant progress made by the application of conventional TEM practices. Going forward, the components of a comprehensive CC/TEM program must be defined, developed, and integrated into existing training and information systems. It is required that the merit of the CC/TEM concept be demonstrated though objective evaluation measurements. Additionally, the applicability of CC/TEM approach should be considered for other domains where the consequence of negative occurrence is unacceptable. This is especially true where team interaction, communication, and coordination are inextricably linked to the successful completion of well-defined procedures (medicine, emergency response, command/control, process control, etc.)

References

Dekker, S. (2006). *The Field Guide to Understanding Human Error.* Burlington, VT: Ashgate Publishing.

National Transportation Safety Board. (2007). *Attempted Takeoff from Wrong Runway Comair Flight 5191 Bombardier CL-600-2B19, N431CA Lexington, Kentucky August 27 2006* (Tech. Report NTSB/AAR-07/05). Washington, D.C.: National Transportation Safety Board.

National Transportation Safety Board. (1990). *United Airlines Flight 232 McDonnell Douglas DC-10-10 Sioux Gateway Airport Sioux City, Iowa July 19, 1989* (Tech. Report NTSB/AAR-90/60). Washington, D.C.: National Transportation Safety Board.

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

Shappell, S.A. and Wiegmann, D.A. (2000). *The Human Factors Analysis and Classification System-HFACS* (Tech. Report DOT/FAA/AM-00/7) Washington, D.C.: Office of Aviation Medicine.

MULTIDIMENSIONAL EVALUATION OF PILOT`S THREAT AND ERROR MANAGEMENT PERFORMANCE DURING COMPLEX FLIGHT MANEUVERS

Ioana Koglbauer and K. Wolfgang Kallus
Department of Psychology, Karl-Franzens University
Graz, Austria

Reinhard Braunstingl
Institute of Mechanics, Technical University
Graz, Austria

Wolfram Boucsein
Institute of Physiological Psychology, Wuppertal University
Wuppertal, Germany

The present research attempts a multidimensional threat and error management performance analysis of pilots flying according to visual flight rules, during the recovery from four unusual aircraft states: extreme pitch, overbanked attitude, full stall and spin. An anticipative training program was developed based on flight mechanical and psychophysiological analysis of an expert's performance. Training took place in a flight simulator and was preceded and followed by check flights with an aerobatic aircraft, a Pitts S-2B, supervised by an expert aerobatic flight instructor. In a between-groups design, a multidimensional assessment was applied, comprising psychophysiological measures of arousal, emotion, subjective, workload and anticipative comparison processes as complementary to technical performance criteria. Our results evidenced benefits of combined theoretical and practical anticipative flight instruction for the threat and error management in complex flight maneuvers.

Incident and accident analyses performed by the AOPA Air Safety Foundation (2003) and by the Boeing Commercial Airplane Group (2004) indicate that loss of aircraft control due to pilot failures in managing complex aircraft states was responsible for a large number of fatalities in general and commercial aviation during the last decades. Causes of complex aircraft states such as unusual attitudes, stalls and spins involve pilot related factors, as well as technical failures and environmental conditions that can not be entirely avoided or controlled. Therefore, it is essential to enhance the pilot's technical and non-technical skills in managing these states within safe psychophysiological boundaries of workload and arousal.

We evaluated the pilot task performance within a model of threat and error management (TEM) described by Helmreich, Klinect and Wilhelm (1999). The model was initially developed to capture specific crew behavior and situational factors during normal flight operations and provide countermeasures at individual and organizational levels. It proposes an observation based evaluation of the TEM process and uses classifiers for pilot's responses to threats and errors such as threat recognition, error avoidance, error detection, additional error and error management, or producing an unsafe condition. According to this model, outcomes of the TEM process are divided into: unsafe action, additional error, or recovery to safe flight.

For analyzing non-technical strategies involved in TEM, we used the theoretical framework of the "anticipation-action-comparison unit" (Kallus, Barbarino & van Damme, 1997) and the concept of the "situation awareness loop" (cf. Kallus & Tropper, 2007). According to the "anticipation-action-comparison unit" concept, goal-relevant future mental pictures are predicted, based on key elements of the situation, previous experience and mental models. Actions are initiated by anticipations of their effects, and comparisons of predicted and actual situational changes close the feedback loop. Situational changes may be pre-classified as match/mismatch evaluations of anticipated and real variations. Anticipatory processes involve different levels of information processing and are manifested on different levels of the central nervous system organization, from complex conscious planning processes to unconscious anticipatory eye-movements (Kallus & Tropper, 2007; Wilson, 2000). Integrated within this model, the concept of situation awareness is referred to as the perception and awareness of key situational elements in a timely manner, the comprehension and determination of their relevance to safety goals and the forecast of their future status (Endsley, 1988). Furthermore, the "situation awareness loop" fosters a controlled sequence of anticipation, perception, comprehension, projection and action in a feedback circuit. Besides situation awareness, anticipation and outcomes of the task management, an operator's workload is another dimension of performance. Workload describes the costs invested by a human operator to achieve a particular level of performance (Hart & Staveland, 1988). The workload or the amount of resources invested in the TEM process is seen as mediator among performance, task difficulty and the operator's skill (Wickens, 2001). Situation awareness, anticipation and workload related arousal can be inferred from physiological measures and subjective ratings (Boucsein 2007; Boucsein & Backs, 2000; Boucsein, Koglbauer, Braunstingl & Kallus; 2009; Kallus & Tropper, 2007; Wilson, 2000 and 2002). In the present research, we attempt to discriminate arousal types using the integrative neurophysiological model of different kinds of arousal provided by Boucsein and Backs (2009), which describes the structural and functional hierarchy within the brain, being involved in general activation, perception, information processing and response preparation. In a previous study we used this four-arousal model in evaluating the adaptive psychophysiological arousal and emotion regulation of an expert pilot during different cognitive, emotional and physical demands of real flight tasks (Boucsein et al., 2009). The present study aimed at the specificity of different types of arousal for anticipatory processes, task performance and post-task echoing involved in the pilot's TEM performance.

Methods

In-flight technical and psychophysiological aspects of the pilot's TEM performance were evaluated by means of an experimental design with two groups: a training and a control group, compared at two times of measurement: before and after simulator training. Statistical analysis of the anticipative simulator training influences was processed using a univariate general linear model. Since there were considerable differences between the training and control group with respects to flight performance and psychophysiological state during the initial flight, and the assumption of homogeneity of regression slopes assumption had not to be rejected, the initial values were entered as a covariate.

Participants and procedure

Twenty-eight male pilots, aged between 21 and 64 years ($M = 38.04$, $SD = 11.45$) volunteered for the study. The participants were randomly assigned to an experimental group ($N = 16$) and a control group ($N = 12$). All pilots held an actual PPL VFR license (private pilot license for flight according to visual flight rules), with their total flight experience ranging from

40 to 430 hours. None of the pilots had aerobatic nor instrument flight rules (IFR) experience. The two groups did not differ significantly in their total flight experience.

The flight task consisted of four flight maneuvers: extreme pitch, overbanked attitude, power-off full stall and left spin with two rotations. Applying the methodological approach for evaluating anticipative processes carried out in our expert case study (Boucsein et al., 2009), each maneuver was split into four phases. An anticipation phase of 15 sec mental preparation preceded each maneuver, followed by the onset, recovery and post-recovery phases. All pilots received theoretical instruction and ground briefing regarding the nature of the maneuvers, their possible eliciting conditions, and avoiding and recovery procedures. Afterwards, each group attended an initial flight, followed by simulator training, simulator test and post-training flight tests. The simulator session of the training group consisted of specific recovery exercises, while pilots of the control group received VFR terrestrial and radio navigation training with similar degree of difficulty. The real flight sessions, supervised by an expert flight instructor (the third author), were performed in a two-place tandem Pitts Special S-2B, a light aircraft certified in the aerobatic category. Training took place in a fixed-base, two-seater generic light aircraft simulator with the following psychological fidelity features: wide screen projection by means of a three-channel visual system to facilitate the simulation of peripheral vision, rudder pressure simulation and an aerodynamic model including stall/spin and unusual attitudes behavior of the aircraft. Generic maneuver representation and response behavior of the simulator were validated at the Institute of Mechanics, Graz University of Technology, with flight maneuver data recorded during a real flight with the Pitts S-2B, using a body fixed coordinate system by means of an inertial platform, which included an aviation-certified laser compass, a MEMS gyro, a GPS sensor and acceleration sensors for each of the three axes (Boucsein et al., 2009). Due to space restriction, only the initial flight and second trial of the final test flight are analyzed in the present paper.

Dependent measures

The TEM performance was evaluated by instructor ratings, ranging from 1 (not acceptable) to 4 (very good). The criterion for a non-acceptable performance was the pilot's failure to respond, manifested as safety-critical omission, wrong prioritization of action or major unsafe acts. Acceptable performance criteria were the presence of major errors that exacerbated the threatening potential of the flight situation or complicated the recovery process, but which were finally mastered by the pilot. Good TEM performance standards included pilot actions that slightly differed in timing and precision from the correct performance, which in turn was rated as very good. The recovery duration was also measured as an additional performance parameter. Self-ratings of the pilot were collected for performance, effort, frustration and task load using the NASA-TLX (Hart & Staveland, 1988). During the entire experiment the pilot's electrodermal activity (EDA) was recorded with the Varioport system (Becker Meditec, Karlsruhe, 2005) as skin conductance from the medial sites of the left foot, adjacent to the plantar area (Boucsein, 1992, Fig. 28). For EDA, skin conductance level (SCL), non-specific skin conductance reactions frequency (NS.SCR freq.) and mean amplitude of skin conductance reactions (SCR amp.) were evaluated using the EDA-Vario software (Version 1.8; Schaefer, 2007). The above mentioned phases of task management were marked with a trigger which was set manually by the flight instructor. The first ten seconds sequence between the 2^{nd} to 11^{th} second of each phase interval was used for psychophysiological evaluation.

Results

TEM Performance and Subjective Ratings

The analysis of instructor ratings yielded that pilots of the training group showed higher overall performance (M = 3.21, SD = .07) than the control group (M = 2.65, SD = .08). The group difference reached significance [F (1, 3.00) = 38.623, p < .05]. The maneuver effects or group by maneuver interactions were not significant. Distinct maneuver analyses indicated a significantly superior performance of the training group in recovering from the extreme pitch attitude as compared to the control group [F (1, 25) = 7.019, p < .05]. Similar results were found for the overbanked attitude [F (1, 25) = 4.304, p < .05] and for the power-off full stall [F (1, 25) = 17.019, p < .001]. The training group showed better performance during the spin recovery (M = 3.05, SD = .17) than the control group (M = 2.59, SD = .20), but the differences were not significant. The recovery duration was slightly reduced from the initial covariate adjusted mean (M = 17.78 sec), not only in the training group (M = 14.22 sec, SD = .44) but also in the control group (M = 14.75 sec, SD = .50). However, the group differences were not significant.

Self-ratings of performance as assessed by the NASA-TLX were significantly better in the training group [F (1, 2.848) = 16.724, p <0.05], while the effort ratings were significantly lower compared to the control group [F (1, 4.899) = 6.756, p < 0.05]. The anticipative simulator training seemed to have a moderate impact on the perceived mental demand, with lower scores in the training group than in the control group, reaching just marginal significance [F (1, 3.714) = 6.546, p = 0.06]. Subjectively experienced physical and temporal demands of the flight task were significantly lower in the training group than in the control group [F (1, 4.066) = 17.604, p < 0.5, and F (1, 4.178) = 63.107, p < 0.001, respectively]. Subjective ratings of frustration did not vary significantly between the groups. Self-ratings of the current physical state and ratings of the psychological state before and after the flight did not show significant differences between the groups.

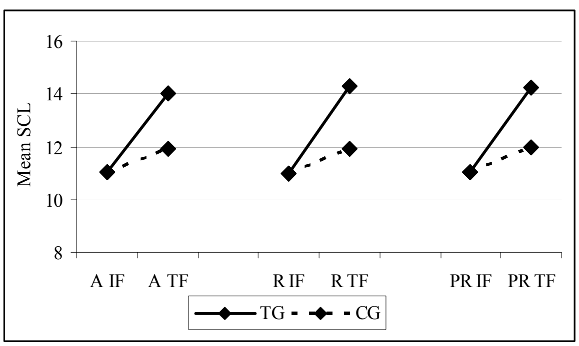


Figure 1. Covariate adjusted means of the Mean SCL (µS) of the training and control group during the initial (IF) and final (TF) test flight. (A= anticipation phase, R= recovery phase, PR= post-recovery phase).

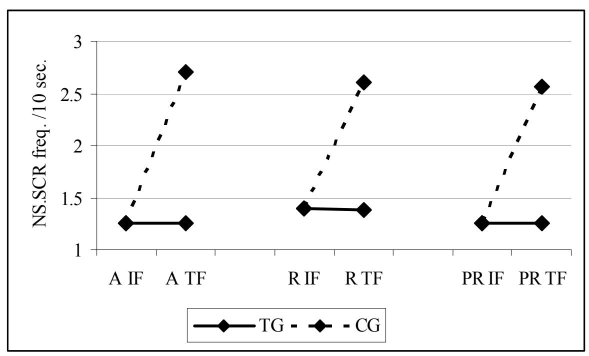


Figure 2. Covariate adjusted means of the NS.SCR freq. of the training and control group during the initial (IF) and final (TF) test flight. (A= anticipation phase, R= recovery phase, PR= post-recovery phase).

Results of electrodermal activity

The simulator training condition seemed to have significantly influenced the mean SCL during the anticipation phase, the training group showing significantly higher values than the control group [F (1, 14.98) = 28.207, p < .001]. As depicted in Figure 1, similar results appeared during the recovery and post-recovery phases [F (1, 14.14) = 23.214, p = .001, and F (1, 22.54) = 24.142, p < .001, respectively]. The mean SCL seemed to also to be influenced by the type of recovery maneuver during the anticipation [F (3, 3.08) = 50.718, p = .005], recovery [F (3, 2.99) = 17.646, p = .05] and post-recovery phase [F (3, 2.98) = 26.984, p = .05]. Group by maneuver interactions were not significant. The mean SCR amp was not significantly influenced by the simulator training condition, nor by maneuver or interactions between the two factors. The NS.SCR freq. (Figure 2) was higher in the control group than in the training group during the anticipation [F (1, 3.04) = 10.740, p < .05], recovery [F (1, 2.99) = 47.010, p < .05] and post-recovery phases [F (1, 3.02) = 30.716, p < .05]. No significant effects were found for the maneuvers and the interaction between experimental groups and maneuvers.

Discussion

Our interest was to determine multidimensional changes of performance under the influence of a specific recovery training in the simulator. The analysis of instructor ratings indicates that the anticipative training performed in a simulator with sufficient psychological fidelity significantly improves the pilot's flight performance in recovering from unusual attitudes, stalls and spins. In general, pilots of the training group improve their TEM performance quality, reaching a level between good and very good, which means that they successfully manage the maneuver threats, only slightly deviating in timing and precision from the correct performance. Pilots of the control group reach a mean performance between the levels of acceptable and good, meaning that their TEM performance generally involves minor and major errors that are, however, finally mastered by the pilots in the given situation. Self-ratings of performance follow the same trend, since pilots in the training group score their own performance significantly higher than pilots in the control group. These qualitative changes in performance are not paralleled by the duration of recovery. Furthermore, the associated workload, in terms of costs of performance, is significantly lower in the training group, since pilots who benefit of the anticipative training report significantly lower effort required by the recoveries than those who could not benefit. Although both groups have mean recovery durations of about 14 sec during the final test flight, subjective evaluations of temporal demand are significantly lower in the training group. Pilots of the training group evaluate the TEM tasks as less physically demanding than pilots of the control group, while the differences in evaluation of mental demand reach just marginal significance.

Autonomic nervous system (ANS) activity as reflected by EDA parameters will be interpreted within the framework of the four-arousal model provided by Boucsein and Backs (2009). The significant increase of mean SCL during all maneuver phases in the training group compared to the control group reflect an increase of general arousal, together with the cortical activation of the motor plans and conditioned behavior patterns permitting timely responses to the anticipated events. In turn, the lower NS.SCR freq. in the training group indicates a significantly decrease of negatively tuned affective responses during all phases and maneuvers. In contrast, the higher NS.SCR freq. in the control group reflects an activation of the affect arousal system, which is responsible for the elicitation of immediate responses such as flight/flight or freezing reactions. These subtle changes in ANS activity are not reflected in the subjective measures of psychological and physical state of the groups. Hence, recording of EDA in flight provides valuable information about the pilot's TEM, which complements performance and subjective ratings.

In conclusion, anticipative flight instruction involving hands-on simulator exercises and recovery procedures split into distinct anticipation-action-comparison units (Kallus et al., 1997) were demonstrated to improve the pilot's TEM performance capability as well as their neurophysiological adaptability to demanding maneuvers like unusual attitudes, full stalls and spins during real flight.

References

AOPA Air safety Foundation (2003). "Stall/spin: entry point for crash and burn?". In *Air Safety Foundation Special Reports*, Maryland.

Boeing Commercial Airplane Group (2004). Statistical summary of commercial jet airplane accidents, Worldwide Operations, 1994-2003. Airplane Safety Engineering, Seattle, Washington.

Boucsein, W. (1992). Electrodermal activity. New York: Plenum Press.

Boucsein, W. (2007). Psychophysiologie im Flug – eine Fallstudie. In P. G. Richter, R. Rau, & S. Mühlpfordt (Eds.), *Arbeit und Gesundheit. Zum aktuellen Stand in einem Forschungs- und Praxisfeld* (pp. 130-142). Lengerich: Pabst Science Publishers.

Boucsein, W., & Backs, R. W. (2000). Engineering psychophysiology as a discipline: historical and theoretical aspects. In: R.W. Backs, & W. Boucsein (Eds.). *Engineering Psychophysiology. Issues and Applications* (pp. 3-30). Mahwah, NJ: Lawrence Erlbaum Associates.

Boucsein, W., & Backs, R. W. (2009). Methods and models for the psychophysiology of emotion, arousal, and personality. In V. G. Duffy (Ed.), *Handbook of Digital Human Modeling* (pp. 35-1 – 35-18). Boca Raton, FL: CRC Press/Taylor & Francis.

Boucsein, W., Koglbauer, I., Braunstingl, R., & Kallus, K. W. (2009) The use of psychophysiological measures during complex flight maneuvers - an expert pilot study. In J. Westerink et al. (Eds.), *Probing Experience II*, Dordrecht: Springer (in preparation).

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society* 32nd Annual Meeting (Vol. 1, pp. 97-101). Santa Monica, CA: Human Factors Society.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.) *Human Mental Workload* (pp. 239-250). Amsterdam: North Holland Press.

Helmreich, R. L., Klinect, J. R., & Wilhelm, J. A. (1999). Models of threat, error, and CRM in flight operations. In *Proceedings of the Tenth International Symposium on Aviation Psychology* (pp. 667-682). Columbus, OH: The Ohio State University.

Kallus, K. W., Barbarino, M., & van Damme, D. (1997). Model of the cognitive aspects of air traffic control (HUM.ET1.STO1.1000-REP-2). Brussels: Eurocontrol.

Kallus, K. W., & Tropper, K. (2007). Anticipatory processes in critical flight situations. *Proceedings of the 14th International Symposium on Aviation Psychology*. Dayton, Ohio, April 23-26, 2007.

Wickens, C. D. (2001). Workload and situation awareness. In P. A. Hanckock, & P. A. Desmond (Eds.), *Stress, Workload, and Fatigue* (pp. 443-450). Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, G. F. (2000). Strategies for psychophysiological assessment of situation awareness. In M. R. Endsley, & D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 175-187). Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *International Journal of Aviation Psychology*, 12 (1), 3-18.

THE USE OF INTRAOCULAR LENSES WITH ADVANCE AVIATION DISPLAYS.

Walter J. Protheroe Jr., MAS
8711 Beau Monde, Houston, TX, 77099-1107
Gerald L. Haynes
19 Brittany Lane, Atoka, TN, 38004.

With an aging aviation population, the use of intraocular lenses (IOLs) has become common place throughout the world for correcting vision acuity, cataracts and eye injuries. The material which comprises an IOL will cause a change in spectral sensitivity seen by the recipient. This is most notable in color variations while viewing Advance Aviation Displays. The results reported here are for one specific composition of IOL given to us by the manufacturer, and are assumed to be or have been in wide use. Cost limitations required that we used color correction filters to simulate colorized IOLs (not actual colorized lenses), but we attempted to use filters with colors simulating the tints of IOLs in production. Additional information has been gathered on contrast variations that should be followed up with additional testing at a later time.

Keywords: intraocular lens, advance aviation displays, colorimeter

Background

Intraocular Lens Implants. Intraocular implants today are used to correct many different medical conditions such as cataracts and physical injuries to the eyes. These implants are now manufactured in both mono-vision (single diopter) and multifocal (multiple diopter) lenses to reduce the use of corrective lenses (glasses) after surgery. The surgery only takes about ten minutes to remove the organic lens and replace it with an IOL.

IOL implants are highly recommended over external corrective lenses today, because of "Fractional Distortion" (Smith & Atchison, 1996, p.108), which is described as either positive (pincushion) or negative (barrel effect) distortion. These aberrations seen at the edges of standard eyeglasses and can cause misreading of dials or displays when glancing at them. Several types of eye-glass frames used today have exacerbated this problem since they allow the complete lower edge and sides of the lens to remain uncovered, which can further distort the image seen by the eye.

However, IOL implants do cause some types of visual distortion. For example, one effect of lens replacement is seeing a "Star pattern" around bring lights observed at night. This star pattern is usually formed by two lines crossing each other (four pointed star) and is variable in luminosity and length, and can exacerbate the problem of identifying distant points of light that are in close proximity to each other. Normally, a person will see a "Halo effect" around bright lights, which does not disrupt the vision as much. A secondary effect of an IOL implant is that a color change can be seen by glancing through the edge of the eye. This secondary effect occurs when light passes through the thin curved edge of the lens.

Another – arguably more significant – type of visual distortion is contrast variations (a darkening of the viewed object; artifacts of sight correction that affect wearers of both corrective lenses and intraocular lenses). This effect is based on the material used to construct the lens and is a direct function of the material's refractive index: as when the index increases, the amount of contrast increases. The refractive index of a Polymethylmethacrylate (PMMA) IOL is 1.49 (Olmos & Roy, 1981), which had been a widely used material for IOLs in the United States and Canada. For comparison, standard reading glasses with polycarbonate lens has a refractive index of 1.586, while a flint glass lens refractive index is 1.700 (Schwartz, 2002).

Changes seen by Intraocular Implant Recipients. Some people who had recent single eye lens replacement notice a change while viewing images on a monitor in that what used to look like a true green now looks blue-green or teal. The reason is that IOL's have specific light absorption characteristics which depend on the material(s) used. As light enters the eye a specific portion of the color spectrum is absorbed by the IOL, causing a shift in observed color.

This noticeable color shift caused by the IOL is a form of limited spectral blindness that is not recognized yet by the Federal Aviation Administration (FAA) (as of May 2007). It has eluded recognition because studies to date have only noted a change in the blue-green spectra, while the aviation medical community only tests for red-green and blue-yellow color blindness (or deficiency). However, this "blindness" is important when looking at a monitor with readouts that rely on subtle changes in color.

The introduction of colorized IOLs has brought new factors to the spectral-absorption issue. These colorized IOLs come in a variety of colors from cyan to yellow to magenta. The choice of what color lens to have implanted depends upon the application and age of the recipient but, obviously, all choices affect the colors perceived by the recipient of the implant. In this manner colorized IOL's are being analogous to slightly-colored sun-glasses which never can be removed.

Legibility of Advance Avionics Displays. At the same time that IOL's are becoming more popular, detecting color variations are becoming more important for cockpit safety. With the introduction of Advance Aviation Displays, flight control information can be displayed in a more realistic setting (graphical realism) by adding color as well as shape. These (primarily) LCD displays, weigh less than previous cathode-ray tubes (CRTs) and associated components, reduce power and space requirements (Helfrick, 1995), reduce manufacturing costs and enhance display characteristics. Moreover, associated software often allows the pilot to select which information is displayed in front of him/her during different times of the flight. This allows the pilot to concentrate on the task at hand and reduces the chances for confusion in the cockpit, assuming that the display is legible.

Many of these advanced avionics displays have the ability to adjust color composition for the user by software selections. However specific colors have become standard in the industry, some of which are problematic for pilots with IOLs. The uses of pure colors such as green or red are easy to set by software, but can be affected by spectral shift for someone who has IOL replacements to the point where a natural green can become teal or red can become brown. When this happens, pilots may overlook important data, especially when a wide variety of information is displayed at one time.

Experimental

Background. The IOLs used this investigation were selected and provided *by the manufacturer*. The claim is that they had represented IOL types manufactured throughout the world and were representative of popular natural lens replacements. The name of the manufacturer has been withheld by prior agreement.

A noticeable difference between the hue and saturation was seen when viewing through the IOL (Garo, 1999). This is caused by the absorption of the lens material at specific wavelengths seen through the IOL (Figure 1), which is a function of its chemical composition. The chemical compositions of the two IOLs types used in this report are Polymethylmethacrylate (PMMA) or poly (methyl 2-methylpropenoate). These are both synthetic polymers of methyl methracrylate, and have a reflectance similar to that of polycarbonate. The haptic [mounting] section of the IOL has a fluorine base, but does not influence the color absorption or excitation pattern of the lens (McCormack & Protheroe Jr., 2008).

Experimentation was done by passing light through an IOL and measuring light received by two different spectrometry systems: (1) a colorimeter (photo-spectrometer) and (2) a wavelength spectrometry system. We modified the colorimeter assembly to "see" light with and without an IOL through a mount assembly. The colorimeter observes a color on the screen and breaks it down into a spectrum. The spectrum is then quantified using a permutation of three light-bands of red, green and blue and the quantity of color observed is displayed on the screen. This instrument has the advantage of being transportable to actual aircraft cockpits.

We use a SPYDER II® colorimeter manufactured by the datacolor Corporation, which is a spectrophotometer used to quantify colors from a luminous source by converting it into a numerical spectrum. It is irrelevant to this spectrometer if the source is a cathode-ray tube (CRT) or a liquid crystal display (LCD), as long as there is sufficient luminescence. This device collects light through a central aperture and then determines its spectrum using a light sensitive integrated circuit package.

The author manufactured an add-on Intraocular Lens (IOL) Colorimeter Mount for aligning the IOLs with the center of the colorimeter (Figure 2). This mount required that the center baffle be removable from the SPYDER II® so that the IOLs could be tested. Additional software from Home-Cinema France (HCFR®) was used to determine the spectral absorption of the IOLs.

A laboratory-sized wavelength spectrometry system measures the spectral wavelength of each color emitted from a screen for user defined acquisition measurements between 350nm and 920nm and downloaded the information to a data file.

The computer and display used for both data acquisition and data reduction were parts of a Hewlett Packard Pavilion dv9000® series laptop computer with a LCD – TFT screen. The power source was from the wall outlet which enabled full backlit capabilities of the monitor throughout the entire test performed within this study. The display was previously calibrated using the datacolor's SPYDER II® Pro version 2.2 software to set a baseline for the examination of the different assemblies tested. Additional tests were done using a MAG® 19" TFT – LCD display to verify TFT response.

Testing Setup. Once the colorimeter was assembled (Figure 2) and calibrated, a color was selected on the screen using the CIE RGB value. The baffle is removed from the colorimeter and replaced by the IOL Colorimeter Mount (see IOL assembly Figure 2). This assembly is placed on the screen to read the color displayed on the monitor. A reading was taken of the color viewed through the IOL to see what spectral shifts occurred. This information is fed back to the computer via the software and displayed in multiple formats.

Results

After insuring that the colorimeter could acquire a complete series of data with different fixtures and configurations, a full series of measurements was done. Different configurations with and without the pre-filter were used to collect data on the changes that occur in the color that is viewed through an IOL. Relative variations in colors caused by IOL's are plotted as Delta-E in Figure 3. [Shifts were most noticeable in the blue and red luminance values.]

"Color-blindness" in even small portions of the spectrum can cause confusion; so we wanted to determine what it would take to alter colors to correct for IOL color loss. Accordingly, data from these tests were graphed to determine if a color correction filter could equalize the color through the IOL and match it with close-to-normal vision. Using the standard configuration set as a baseline (Figure 3), an examination of each graph was done to see if there was a close match within the 90% to 100% luminescence values using the mounted IOL and filter combination. As a result, adding a magenta filter (Charles Beseler Company in Vineland, NJ; part # 8932; value 2.5 – filter factor 1.1) to the mounted IOL adjusted the color response closest to the standard configuration (Figure 3) of all other configurations.

Additionally, using the wavelength spectrometer to view specific colors on the LCD monitor through the IOL, showed a slight increase in specific wavelengths over direct observation. The highest level of contrast differences were up to 27% when testing the color blue (Color value: R,G,B-0,0,255) from 410nm to 560nm. This increase in color quantity and contrast need to be explored further.

Conclusion

With the advent of flying with IOLs in this era of advance avionics displays, we must be assured that confusion in the cockpit does not occur. It will not take much for major changes to occur if there are flight incidents with pilots having IOL implants. Since the colors for the advance avionics displays are software selectable, simple and inexpensive changes can be introduced into the cockpit. Knowing which colors can affect pilots is problematic, but can be determined given the known materials used on the market.

There are more than one hundred manufactures and many chemical variations used in the manufacturing process of IOLs around the world. However, it would not be technically difficult for each manufacturer to test for spectral absorption of each IOL and relay that information into an FAA database. This database will help manufactures to set up software commands to adjust the color displays so that the best contrasting colors are used for all personnel in the cockpit. Another solution would be to approve "Aviation IOLs" that have a specific color absorption pattern; however, current wearers of non-standard IOLs would need to be grandfathered in to keep their flight status. In any case, color displays in cockpits need to be monitored closely to insure that there is no color drift that would be a problem for users of IOLs and other vision correction devices. As we can see, the problem is addressable and the solution can be easy.

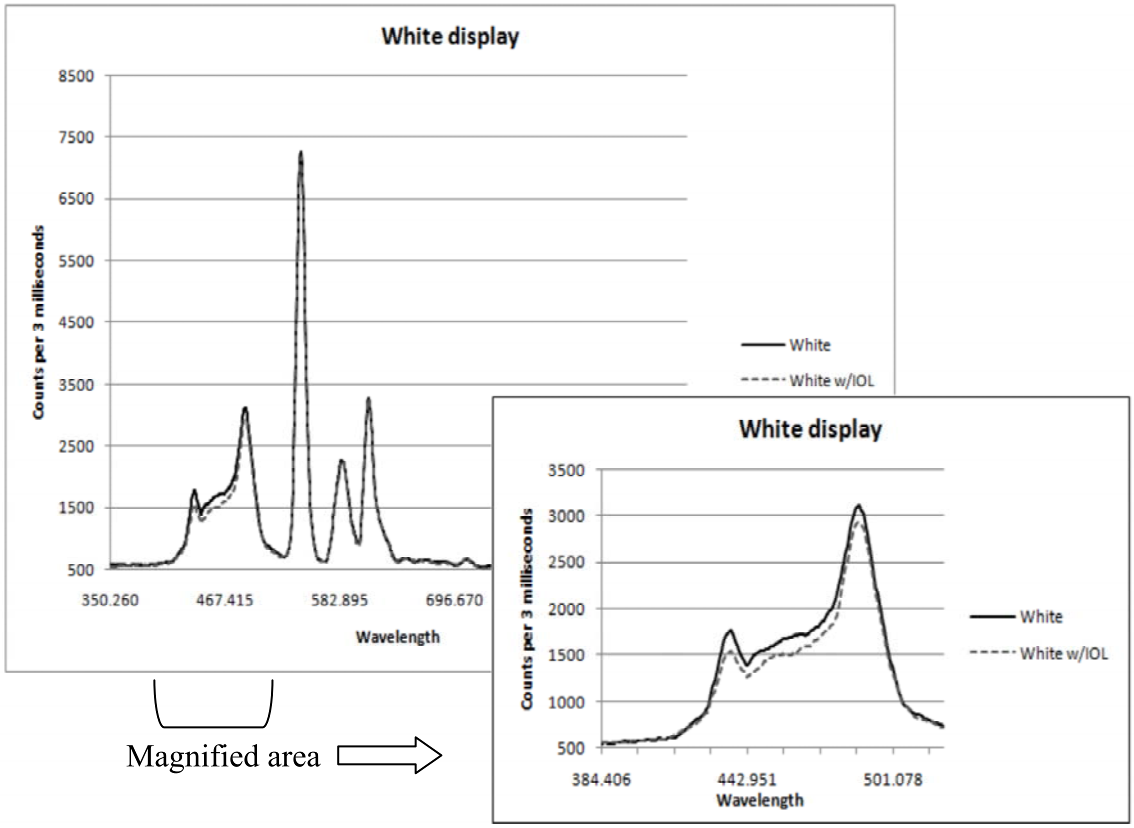


Figure 1. Spectrometer readings from the monitor with a white signal displayed; wavelength spectrometer direct reading (White) and through an IOL (White w/IOL). This shows a variable decrease in the blue to cyan spectrum when viewing the white signal from the LCD monitor through the IOL.

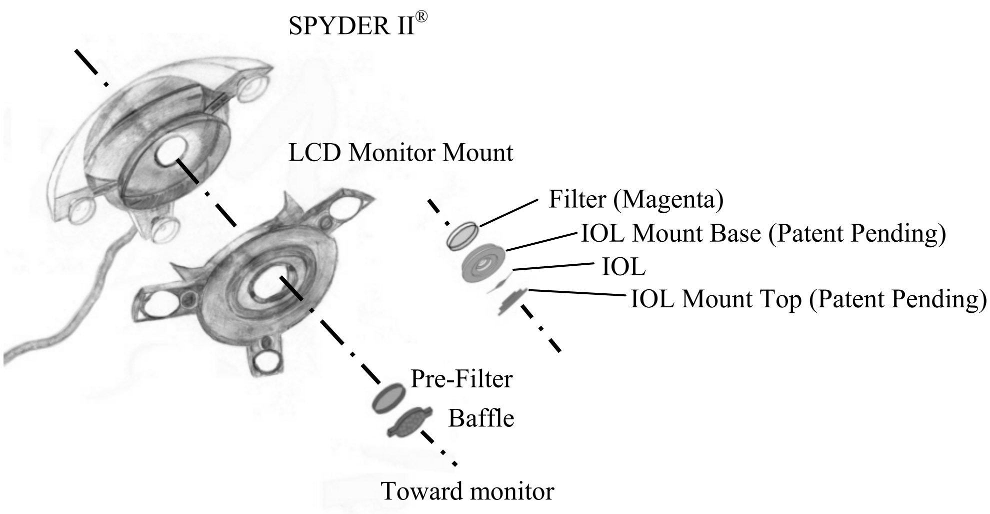


Figure 2. Colorimeter assembly (Sullivan, J. & Protheroe Jr., W. J., 2008). The standard configuration is a combination of the Spyder II®, LCD Monitor Mount, Pre-Filter and Baffle. The test configurations are combinations of the Spyder II®, LCD Monitor Mount, Pre-Filter [with and without], Filter [with and without] (Cyan, Yellow and Magenta filters), IOL Mount Base, IOL [with and without] and the IOL Mount Top.

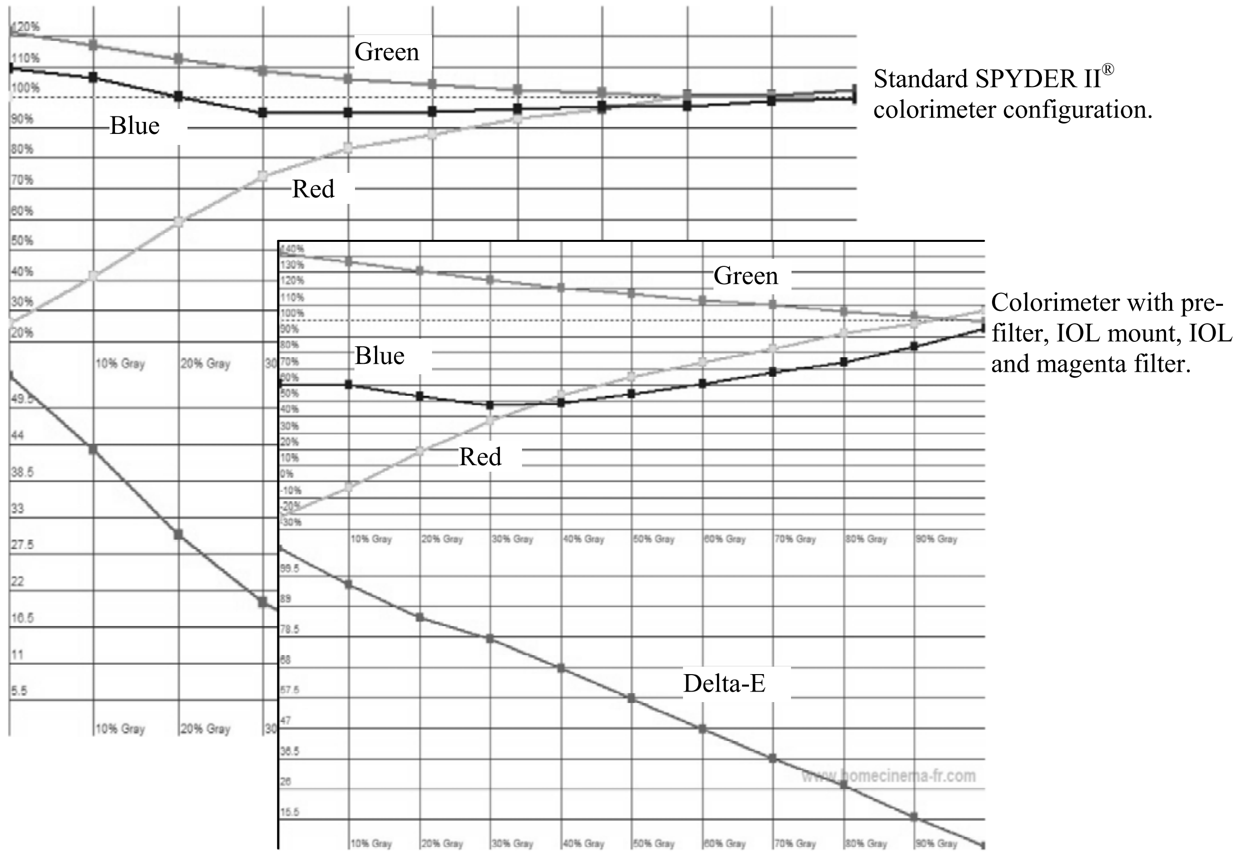


Figure 3. Colorimeter Graphs. RGB and Delta-E plot using HCFR colorimeter software. There is a slight variation in green luminance in the IOL configuration and a much greater variation in the blue and red luminance relative to the standard. None of the other test using the IOL with or without filters showed a tighter grouping seen at the higher percentage of luminance than with the magenta filter. This grouping correlates closer to a more natural vision that is seen in the standard colorimeter configuration.

References

Garo, L. A. B. (1999, May 21). *Lesson 1: introduction to color theory*. Retrieved May 17, 2008, from Color Theory Web site: http://personal.uncc.edu/lagaro/cwg/color/ intro_color.html.

Helfrick, A. D. (1995). *Practical aircraft electronic systems*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 243-244.

McCormack, J. K. & Protheroe Jr., W. J. (2008, June 13). *Eds analysis of lens and haptic sections of iol.* E-mail communication. University of Nevada @ Reno, The Mackay School of Earth Sciences and Engineering.

Olmos, E. & Roy, F. H. (1981). *Intraocular lens*. New York: Praeger Publishers, 61-62.

Schwartz, S. H. (2002). *Geometrical and visual optics, a clinical introduction*. New York, NY: McGraw-Hill, 204.

Smith, G., & Atchison, D. (1997). *The eye and visual optical instruments*. New York: Cambridge University Press, 34, 67, 108.

Acknowledgements

Gerald L. Haynes passed away in 2008 and I would like to take a special note to his family for his contribution to the work presented here; Walter J. Protheroe Jr. (corvos@aol.com). The material used here was based on my Graduate Capstone Project at Embry-Riddle Aeronautical University. I wish to thank Dr. John McCormack (Univ. of Nevada @ Reno) for imaging and EDS analysis. Additionally, I wish to thank *datacolor Corporation* for their supply of SPYDER II Colorimeters used in this research, EOS of Canada for the use of their Cathodoluminescence Spectrometry system and Home-Cinema France for the use of their Microsoft Vista® compatible software.

HOW DIFFERENCES IN SPATIAL ABILITY INFLUENCE INEXPERIENCED USERS IN A VISUAL PERCEPTUAL AVIATION TASK

Michael T. Curtis
University of Central Florida
Orlando, FL

Florian G. Jentsch
University of Central Florida
Orlando, FL

Crystal Maraj
University of Central Florida
Orlando, FL

The approach and landing phase of flight is widely recognized as one of the most difficult phases of flight. More specifically, professionals in aviation training report difficulty in training inexperienced pilots on execution of visual approaches. The current paper focuses on our efforts to develop a perceptual skill trainer using a static image discrimination task. From a perceptual standpoint there are a number of documented visual environmental cues that have been found to impact one's ability to judge distances. These distracting cues can cause individuals to misjudge distance to landing surfaces, and subsequently result in an unstable or unsafe approach. For this study we chose to examine how individual differences in spatial orientation ability predict performance in a visual approach static image discrimination task. As expected, individuals high in spatial orientation ability outperformed those with low spatial orientation ability. More importantly we examine how distracter cues have a differing effect on low and high spatial orientation ability individuals. The results from this study have implications for development of tailored training in aviation training.

The words "clear for visual approach" usually signify favorable conditions for the approach and landing phase of flight. Based on the clear conditions that coincide with visual approach, flight safety would seem to be at its most optimal. Controlled flight into terrain (CFIT), however, remains one of the most prevalent causes of aviation accident (Darby, 2006). CFIT accidents are those in which the pilot unknowingly maneuvers the aircraft into the terrain below (e.g. ground, water, or obstacles). In some cases, this can be attributed to degraded conditions or unexpected events that challenge even the most experienced pilot. Surprisingly, accidents and incidents also occur in clear conditions where visual flight rules (VFR) prevail (Shappell & Weigmann, 2003). Despite technological advancement in training, aircraft instrumentation, and external visual flight aids (e.g., visual approach slope indicator) visual approach still remains one of the most challenging phases of flight. Overall, our research focuses on the investigation of critical skills in visual approach, and the development of supplemental training tools to improve these skills.

Our research efforts are geared toward the investigation of both cognitive and perceptual skills that are involved during the visual approach phase of flight. The current study represents one part of a larger study in which we examine visual perceptual factors that influence distance estimation in a visual approach. Here, we examine how one measure of individual difference might predict performance on a visual aviation task intended to train perceptual performance when conducting a visual approach.

Visual Approach

Visual approach is a flight maneuver that is conducted when pilots have unobstructed visual contact with the landing surface. When cleared by air traffic control for visual approach, pilots rely on their view outside of the cockpit to establish and maintain flight path, while judging ground proximity for final approach and landing flare (Robson, 2001). Distance and altitude regulation are critical to effective maintenance of safe flight conditions. The angle generated by the distance and altitude from the landing surface, or glide slope, is usually recommended at 3° for optimal approach. Anything that deviates too far from the recommended glide slope can result in dangerously steep or shallow approaches. In fact, pilot workload increases as the glide slope angle gets steeper (Boehm-Davis, Casali, Kleiner, Lancaster, Saleem, & Wochinger, 2007). The issues associated with inefficient glide slope maintenance may, as a result, be compounded by additional workload.

Perceptually, the estimation of distance on both the vertical and horizontal plane, known as slant distance, is a key contributor to maintaining a safe flight path. The visual environment at times contains information that may mislead distance judgment. Unfortunately, this can lead to inappropriate flight path alterations, leading to unstable visual approach conditions. Human adaptation to terrestrial viewpoints may have something to do with this. Given that the world is oriented with a bias toward vertical and horizontal orientations (Baddeley & Hancock, 1991), the oblique viewpoint associated with aerial perspective is less familiar. In fact, oblique aerial viewpoints contribute to a number of illusory effects that may cause incorrect judgment of distance in relation to terrain or other obstacles (Leibowitz, 1988). These effects can result from lack of visual information (i.e. black hole effect; Gibb, 2007; Mertens, 1981), variation in runway dimensions (i.e. form ratio, Mertens, 1981; Mertens & Lewis, 1982), or even the relative heading at which an aircraft approaches the runway (Curtis, Schuster, Jentsch, Harper-Sciarini, & Swanson, 2008). Since the geometric principles for approach angle are rigid, the judgment of distance should be relatively straight forward. However, there is a wide variation in visual information that a pilot may experience. Depending on the circumstances, environmental visual cues can act to distract pilot perception. Since visual perception continues to play an important role in aviation, it is important to investigate measures that accurately assess perceptual skill as it relates to the visual approach task. Carefully selected measures of individual difference may provide this accuracy. Ultimately individual difference measures found to predict performance can be a valuable guide in training development.

Individual Differences

Spatial ability is a widely studied individual difference factor that has been investigated in a broad range of domains. Spatial ability consists of a number of widely disputed dimensions (Carroll, 1993; Lohman, 1988; McGee, 1979; Michael, Guilford, Fruchter & Zimmerman, 1957) that are best defined as a representation of an individual's capacity to cope with object relations in space. Spatial ability encompasses skills such as wayfinding, navigation, and object recognition. It has been found to predict task ability and used for selection purposes within the context of numerous domains (Gibbons, Baker, & Skinner, 1986; Ghiselli, 1973; Humphreys & Lubinski,, 1996).

In aviation, spatial ability has been found to predict success in general piloting skill, and for many years in the mid 20th century was used for selection of military pilots (Hegarty & Waller, 2005; Humphreys & Lubinski, 1996). In fact, through a series of spatial measures Dror, Kosslyn and Waag (1993) found that pilots tend to be better than non-pilots in mental rotation and precise distance judgments. The ability to make precise distance judgments is a critical flight skill that has implications especially for execution of visual approaches. Despite the findings by Dror and colleagues (1993), there are still reports of instances where pilots, who should have good distance judgment skill, experience difficulty executing visual approach. So distance judgment skill alone does not guarantee that pilots are impervious to distracting features in the environment.

Many studies of spatial ability of pilots focus on comparing pilot and non-pilot populations. This is informative in distinguishing that experienced pilots have spatial skill advantages to non-pilots, but does little to identify application of spatial skill to specific aviation tasks. Instead of using spatial measures for blanket piloting performance selection criteria (Humphreys & Lubinski, 1996), perhaps spatial measures can serve more to diagnose areas where additional training may benefit.

In visual approach, a pilot's understanding of their location relative to the runway is critical. Unlike the distance judgment task used by Dror and colleagues (1993), visual approach involves accurately judging distance in an environment with a plethora of visual information in the environment. Approach to the same runway from the same distance can have vastly different appearances due to variations in terrain, time of day, and orientation of the aircraft. Measures of spatial orientation, address an individual's ability to recognize how change in viewpoint orientation alters the appearance of the surrounding environment (McGee, 1979). In aviation a spatial orientation measure such as the Guilford-Zimmerman (1948) measure could provide a more meaningful prediction of specific orientation related tasks such as visual approach. Although there is debate over whether spatial orientation and another separately proposed dimension, spatial visualization, measure different constructs (Carroll, 1993); the Guilford-Zimmerman spatial orientation measure more closely resembles an aviation task specifically.

Table 1. ***Manipulated Variables for Visual Approach Discrimination Task***

Manipulated variables	Description
Target variable:	
Slant distance	Combination altitude/distance measure from the focal point of the image (the end of the runway)
Distracter variable:	
Terrain	Quantity and density of environmental features (i.e. buildings, trees, etc.)
Relative approach angle	Heading angle at which the aircraft is facing the runway
Visibility	Meteorological measure of distance at which environmental features can be viewed due to atmospheric conditions
Form ratio	Length width ratio of the runway

Hypothesis

For this study, we sought to investigate how a measure of spatial ability could predict performance on a visual approach task. Using a perceptual discrimination training module geared to improve visual approach skills (Curtis, Schuster, Jentsch, Harper-Sciarini & Swanson 2008), we looked to decipher the predictive nature of a spatial orientation measure. Based on our assertion that the parallels between spatial orientation ability and similar visuo-spatial requirements when flying a visual approach, we hypothesize that individuals with high spatial orientation ability will perform better on the visual approach discrimination task (Hypothesis 1).

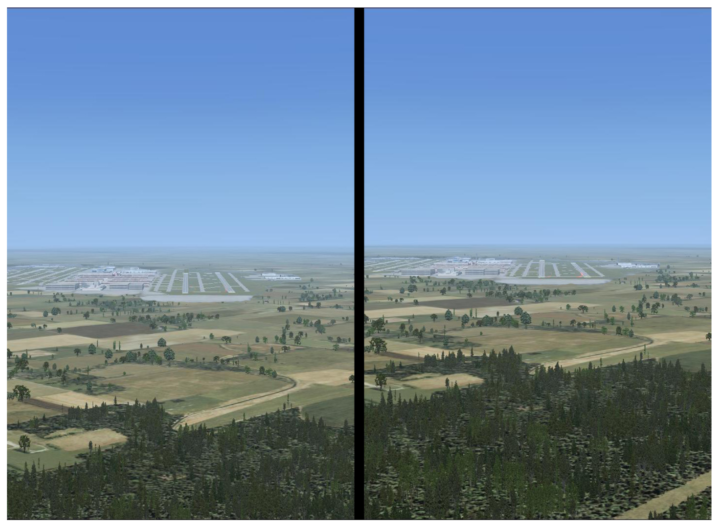


Figure 1. **Sample discrimination task image pair.**

In addition to our primary hypothesis we sought to investigate how those individuals who scored high and low on the spatial orientation measure differed on their responses to the discrimination task. That is, we examined their responses influenced by the distracting environmental cues manipulated in the performance task. Based on our previous hypothesis that individuals scoring high on the spatial orientation measure will score better on the discrimination task, we also hypothesized that individuals in the low spatial orientation group are more susceptible to influence from distracting visual cues (Hypothesis 2).

Method

Participants. For this study participants consisted of 97 undergraduate students recruited from the University of Central Florida. The population was selected due to the assertion that many of the perceptual influences that a non-pilot would experience can be equated to that experienced by a novice pilot trainee.

Measures

Spatial Orientation. For our study we selected the Guilford-Zimmerman (Guilford & Zimmerman, 1948) spatial orientation test. Given the similarity of this test to a pilots experience in the course of an approach and landing task, the Guilford-Zimmerman spatial orientation measure was proposed to predict performance. The test was administered using an electronic format.

Discrimination Task. The performance measure for this study was a discrimination task. The task consists of 270 static image pairs that the participant (Figure 1) was asked to determine if the images are same or different based only on their judgment of slant distance from the end of the runway. In addition to varying the slant distance from the end of the runway, we systematically manipulated a series of distracter variables (Table 1). Accuracy is determined as the number of correct comparison judgments made based on the judgment of slant distance from the end of each runway.

Procedure.

Upon arrival participants were asked to complete an informed consent form, following completion, participants were seated in front of a laptop computer and asked to begin the study. All questionnaires and testing material were presented using MediaLab and Direct RT software. Participants were presented with a timed spatial orientation test. Upon completion of this, the perceptual discrimination task began. A brief tutorial on how to perform the task was provided prior to beginning the session. For the discrimination task, participants were presented with comparisons comprised of two visual approach images presented side by side on the display. Each comparison was presented for a maximum of 10 seconds. Participants were asked to categorize the pair of images as same or different by pressing designated keys on the keyboard. Failure to respond within 10 seconds resulted in an incorrect response. After completion of the discrimination accuracy test, participants were debriefed and dismissed.

Results

All analyses were performed using SPSS 12.0; the alpha level was set at .05, unless otherwise specified. Although data were found to be mildly positively skewed, we decided to forgo transformations to preserve the directional relationship of the variables. Scores from the spatial orientation measure were split at the median to create a high and low spatial orientation score group.

An independent samples *t* test was performed to compare mean discrimination task scores for the high spatial orientation group (M = 160.28, SD = 24.30) and low spatial orientation group (M = 150.82, SD = 18.03). Results indicate a significant difference, t (95) = -2.19, p < .05. This indicates that individuals who scored high on the spatial orientation test did better on the discrimination task than those with lower spatial orientation scores.

Each item on the discrimination task varied on one target variable (glide slope) and four distracter variables (relative approach heading, terrain, visibility and form ratio). An item analysis was performed to investigate whether specific environmental cues were further predictive of performance in either the high spatial orientation group or the low spatial orientation group. We performed a multiple regression to determine if the target variable or any distracter variables predicted performance using backwards removal. For the high spatial ability group there were no variables that significantly predicted performance. In the low spatial orientation score group, form ratio was found to significantly predict performance on the discrimination task, F (1, 268) = 5.377, $p < .05$. Counter to what we expected, this significant effect suggests that participants did worse in the absence of form ratio manipulation (R^2 =.02; Adjusted R^2 = .02). In order to address this conflicting result we further investigated the low spatial orientation group.

Based on findings from Jentsch, Curtis, Schuster and Swanson, (2008) that response predictors differ on same item pairs and different item pairs in the same visual approach aviation discrimination task, we chose to investigate whether a similar pattern exists in the low spatial ability group. We performed two multiple regressions to investigate same image pairs and different image pair responses. Terrain difference was found to significantly predict performance for the low spatial orientation group on same image pairs F (1, 89) = 6.756, $p < .05$ (R^2 =.07; Adjusted R^2 = .06). Furthermore, terrain difference and form ratio were found to significantly predict on different image pairs F (2, 179) = 4.889, $p < .05$ (R^2 =.05; Adjusted R^2 = .04).

Discussion

The purpose of this study was to investigate the predictive capabilities that an individual difference measure has on a visual approach task. Our primary hypothesis was supported in that spatial orientation scores positively predicted performance on the visual approach discrimination task. Given this, it is reasonable to deduce that spatial orientation may provide an accurate prediction of initial individual ability on visual aviation tasks. It is interesting to note that in our second hypothesis there was a lone predictor of performance, form ratio, in the low spatial orientation ability group. At first glance this seems logical, given that form ratio is a known cause of visual misperception in the cockpit (Mertens, 1979; Mertens & Lewis, 1982). However our findings were counter to what one would expect. Individuals were significantly worse at discriminating between items where the form ratio was the same than those where form ratio was manipulated.

Our further investigation on response predictors for same and different discrimination tasks helped to clarify this confusing outcome. Form ratio was found to predict response on different discrimination pairs, but not same discrimination pairs. Participants used form ratio as a criteria for making image pair discriminations. As such, they would correctly respond to different image pairs based on the difference in form ratio instead of the target variable. It is also interesting to note that those in the low spatial orientation group also used terrain for both same and different image pairs. Both this and the form ratio finding suggest that individuals lower on spatial ability may be more prone to distraction from visual features that are known to influence distance estimations. Given this, it is reasonable to suggest that individuals scoring low on spatial orientation should receive additional training geared toward their tendency for distraction by known visual distracters such as terrain and form ratio.

If a number of measures, like the spatial orientation test, can be identified to accurately predict performance for specific aviation based tasks, training could be tailored for each trainee. By adjusting training to individual's strengths and weaknesses, tailored training programs would provide increased efficiency. Individuals who master a skill set will be able to focus on topics in which they are less proficient. Meanwhile, individuals who take more time to grasp topics will be provided additional training to ensure coverage of the topic area.

Most practically, our findings have implications for a very specific skill on a specific perceptual aviation task. Given the wide range of tasks that must be trained to safely operate an aircraft, it is too soon to coronate the tailored approach as the end-all cost saving solution for aviation training. In spite of this, our findings support the notion that individual difference measures can provide prediction of both general skill (i.e., visual discrimination task performance) and more specific performance indicators (i.e., terrain and form ratio variation). Given the promise that this and similar training based research have provided (Curtis, Harper-Sciarini, Jentsch, Schuster & Swanson, 2007), further research and development in tailored training could lead to both cost and efficiency gains in aviation training.

References

Baddeley, R. J., & Hancock, P. J. B. (1991). A statistical analysis of natural images matches psychophysically derived orientation tuning curves. *Proceedings of the Royal Society of London B*, *246*, 219-223.

Boehm-Davis, D.A., Casali, J.G., Kleiner, B.M., Lancaster, J.A., Saleem, J.J. & Wochinger, K. (2007). Pilot performance, strategy, and workload while executing approaches at steep angles and with lower landing minima. *Human Factors,* 49(5), 759-772.

Carroll, J. (1993). *Human cognitive abilities: A survey of factor-analytical studies.* New York: Cambridge University Press.

Curtis, M.T., Harper-Sciarini, M., Jentsch, F.G., Schuster, D. & Swanson, R. (2007). Filling the gaps: An investigation of the knowledge needed for effective human-automation interaction. *Proceedings of the 14th International Symposium on Aviation Psychology*, Dayton, OH: Wright State University.

Curtis, M.T., Schuster, D., Jentsch, F., Harper-Sciarini, M. & Swanson, R. (2008). Perceptual cue effects on distance estimation from different angles of approach. Proceedings for the 52nd Annual Meeting of the Human Factors and Ergonomics Society, New York, NY: HFES.

Darby, R. (2006). Commercial jet hull losses, fatalities rose sharply in 2005. *Aviation Safety World,* 1(2), 51-53.

Dror, I.E., Kosslyn, S.M., & Waag, W.L. (1993). Visual-spatial abilities of pilots. *Journal of Applied Psychology, 78(5),* 763-773.

Gibb, R. W. (2007). *Visual spatial disorientation: Re-visiting the black hole illusion* (Tech Rep No. CI07-0009). Wright Patterson Air Force Base, OH: Department of the Air Force.

Gibbons, R., Baker, R., & Skinner, D. (1986). Field articulation testing: A predictor of technical skills in surgical residents. *Journal of the American Podiatric Medical Association, 73,* 12-16.

Ghiselli, E.E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology, 26,* 461-477.

Guilford, J.P. & Zimmerman, W.S. (1948). The Guilford-Zimmerman aptitude survey. *Journal of Applied Psychology, 32,* 24-34.

Hegarty, M. & Waller, D. (2004). A dissociation between mental rotation and perspective-taking abilitites. *Intelligence, 32,* 175-191.

Humphreys, L.G. & Lubinski, D. (1996). Assessing spatial visualization: An underappreciated ability for many school and work settings. In C.P. Benbow & D. Lubinski (Eds.), *Intellectual talent: Psychometric and social issues,* (pp. 116-146), Baltimore, MD: John Hopkins Press.

Jentsch, F., Curtis, M., Schuster, D. & Swanson, R. (2008). Visual approaches : A continuing training challenge among pilots. Poster presented at *Human Factors and NextGen : The Future of Aviation*, Arlington, TX.

Leibowitz, H.W. (1988). The human senses in flight. In E.L. Weiner & D.C. Nagel (Eds.) *Human Factors in Aviation,* (pp.125-133) Academic Press, San Diego, CA.

Lohman, D.F. (1988). Spatial abilities as traits, processes and knowledge. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 181-248). Hillsdale, NJ: Erlbaum.

McGee, M.G. (1979). Human spatial abilitites: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin, 86,* 889-918.

Mertens, H. W. (1979). *Runway image shape as a cue for judgment of approach angle* (Tech. Rep. FAA-AM 79/25). Alexandria, VA: Federal Aviation Administration, Office of Aviation Medicine. Retrieved May 5, 2008, from http://www.faa.gov/library/reports /medical/oamtechreports/1970s/media/AM79-25.pdf

Mertens, H.W. (1981). Perception of runway image shape and approach angle magnitude by pilots in simulated night landing approaches. *Aviation, Space, and Environmental Medicine,* 52(7), 373-386.

Mertens, H. W., & Lewis, M. F. (1982). Effect of different runway sizes on pilot performance during simulated night landing approaches. *Aviation, Space, and Environmental Medicine, 53,* 463–471.

Michael, W.B., Guilford, J.P., Fruchter, B., & Zimmerman, W.S. (1957). The description of spatial-visualization abilities. *Education and Psychological Measurement, 17,* 185-199.

Robson, D. (2001). *Human Factors for the Professional Pilot.* South Melbourne, Australia: Aviation Theory Centre.

Shappell, S.A. & Wiegmann, D.A. (2003). *A human error analysis of general aviation controlled flight into terrain accidents occurring between 1990-1998* (DOT Report No. DOT/FAA/AM-03/4). Springfield, VA: National Technical Information Service

WHAT CAN A MULTIDIMENSIONAL MEASURE OF STRESS TELL US ABOUT TEAM COLLABORATIVE TOOLS?

Gregory J. Funke[1], Sheldon M. Russell[1], Benjamin A. Knott[2], and Brent T. Miller[2]

[1]Consortium Research Fellows Program, Alexandria, VA, USA
[2]Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA

A substantial body of research literature concerning the effects of collaborative tools on team performance has been generated, but the research has not considered subjective workload and stress associated with tool usage. The current experiment represents an initial, exploratory attempt to characterize the relationship between usage of collaborative tools, mental workload, and the subjective experience of stress. The NASA-TLX and the DSSQ-S were used to assess the workload and stress experienced by participants completing a simulated team command and control task. Task demands and collaborative tool availability were experimentally manipulated. Analysis of the data revealed that participants experienced increases in stress and workload with high task demands which were alleviated by the availability of collaborative tools under certain conditions. The results of this experiment demonstrate the complex relationships between collaborative technologies, workload, and stress.

Collaborative technologies, such as email and instant messaging (IM), are becoming vital tools for military organizations (e.g., Heacox, Moore, Morrison, & Yturralde, 2004). The availability of these tools has dramatically altered the ways in which personnel can communicate and collaborate, allowing organizations to shift from collocated teams to teams that may be geographically and temporally disbursed.

Within the military it has been suggested that collaborative technologies will enable a degree of command decentralization resulting in greater flexibility and adaptivity of forces (e.g., Alberts & Hayes, 2003). However, research indicates that the relationship between collaborative tools and distributed-team performance is complex, in that team task has consistently emerged as an important moderator of the influence collaborative technologies exert on performance (Hertel, Geister, & Konradt, 2005; Martins, Gilson, & Maynard, 2004).

One aspect of the collaborative tool literature that has not been considered yet is their relationship to operator workload and stress. While some research has been conducted examining job stress in fields that rely on collaborative technologies (such as call center workers; Zapf, Isic, Bechtold, & Blau, 2003), research examining the subjective workload and stress associated with tool usage has not yet been initiated, though several researchers have suggested that team affect and mood deserve greater attention from team researchers (Cohen & Bailey, 1997; Mathieu, Maynard, Rapp, & Gilson, 2008).

Modern theories of stress and workload are similar in that they posit that each can be viewed as an interaction between external demands and an individual's cognitive and behavioral responses to those demands (e.g., Gopher & Donchin, 1986; Lazarus, & Folkman, 1984). While workload and stress are considered to be separate theoretical constructs, they may influence performance through similar mechanisms. Attentional resource theories (e.g., Norman, & Bobrow, 1975) suggest that information processing and task performance are dependent on the availability of system resources. Such theories typically propose that system resources exist in a fixed quantity and that resources act as an energizer for information processing. It has also been suggested that subjective workload may represent the proportion of resources required to meet the demands of a task (e.g., Welford, 1978). As task demands increase, more resources are required for task performance and workload increases.

The effects of stress on performance may also be dependent on resource availability either by reducing the amount of resources available for task performance, or because some resources are diverted to processing stressful stimuli (Matthews, Davies, Westerman, & Stammers, 2000). In support of this viewpoint, various stressors, including noise, subjective tiredness, heat, anxiety and prolonged work, have been shown to impair performance most reliably when a task is attentionally demanding (Matthews et al., 2000).

This suggests a possible synergistic relationship between workload, stress, and collaborative technologies. To the extent that collaborative tools reduce operator stress, they may also be expected to reduce operator vulnerability to high workload, and vice versa. The purpose of the current experiment was to explore the influence of several collaborative technologies on subjective workload and stress in a simulated air defense task. Technologies included in this experiment were instant messaging, a virtual whiteboard, and a graphical data display. These technologies were selected because they are consistent with long-term military acquisition goals, and because they conform to anticipated future military capabilities (Sloan, 2008).

Method

Participants

Seventy men and 35 women, drawn from local universities and from a temporary work agency, were fiscally compensated for their participation. Participants were between the ages of 18 and 30 ($M = 21.94$, $SD = 3.16$), and completed the experiment in five-person teams, yielding a total of 21 experimental teams.

Experimental Design

A 3 × 2 × 2 × 2 mixed design was employed in this experiment. Team position was a between-participants factor with three levels (weapons director, strike operator, tanker operator). Within-participants factors included two levels of task demand (low, high), two levels of team communication (standard, enhanced), and two levels of data-display (tabular, graphical). Each team completed 2 trials in each experimental condition, for a total of 16 trials in each experimental session.

Materials

Questionnaires. Operator workload was assessed using the NASA Task Load Index (TLX; Hart & Staveland, 1988), which participants completed immediately following each trial. Subjective stress state was examined using the short version of the Dundee Stress State Questionnaire (DSSQ-S; Matthews, Emo, & Funke, 2005), an experimentally validated measure designed to assess multiple transient state factors associated with stress, arousal, and fatigue. DSSQ-S subscale scores are distributed with a mean of 0 and standard deviation of 1, so that the computed scores for a sample represent deviations from that sample's baseline values in standard deviation units. Participants in this experiment completed the measure immediately before beginning the experiment, and following each two-trial task demand block.

Apparatus. Five-person teams worked together to complete a simulated air defense command and control (C2) task. This task has been used in several previous experiments examining collaborative tool usage in military settings and has been demonstrated to be sensitive to experimental manipulations (e.g., Finomore, Knott, Nelson, Galster, & Bolia, 2007). Participants were randomly assigned to one of three team positions; positions differed in their roles and capabilities. The scenario required two weapons directors (WDs), two strike operators, and one tanker operator. Within the simulation, the WDs' roles were to match friendly fighters with appropriate enemy targets, schedule fighters for refueling and resupply, and communicate their plans with other team members. The role of the strike and tanker operators was to maneuver team assets as instructed, to engage enemy targets, and to provide pertinent information to teammates concerning asset resources.

The asset information available to team members was dictated by the data-display condition of that trial. In the tabular display condition, only strike and tanker operators had access to asset weapon and fuel status, presented in a digital format. WDs, therefore, had to rely on teammates for resource updates.

In the graphical display condition, asset fuel status was displayed in an analog format, and this display was available to all team members. In addition, the graphical display conveyed supplemental information to team members in that its associated asset fuel gauges changed to an amber color when fuel reserves were low, and it featured a black bar which indicated the minimum reserve fuel required to rendezvous with a tanker asset. Examples of both display types are presented in Figure 1.

The number of enemy targets present in each scenario was determined by the task demand condition of that trial. In the low and high demand conditions, 24 or 36 enemy targets, respectively, entered the simulation during the trial. At the conclusion of each trial, participants received a 'team score' based on three performance factors: a) prevention of enemy incursions, b) preservation of team assets, and c) protection of friendly ground forces.

Team communication. Communication between teammates in this experiment was manipulated through the team communication factor. In the standard communication condition, participants could communicate orally using a radio headset. All five team members communicated using the same radio channel to approximate the saturated communications experienced in many 'real world' military environments.

In the enhanced communication condition, participants could communicate using the radio or using two collaborative tools: instant messaging (IM) and a virtual whiteboard. The virtual whiteboard allowed a graphical annotated of participants' tactical displays to be distributed between teammates. This allowed participants to communicate spatial and tactical information (such as routes, enemy locations, etc.) without forcing them to divide their attention across multiple displays (Figure 2).

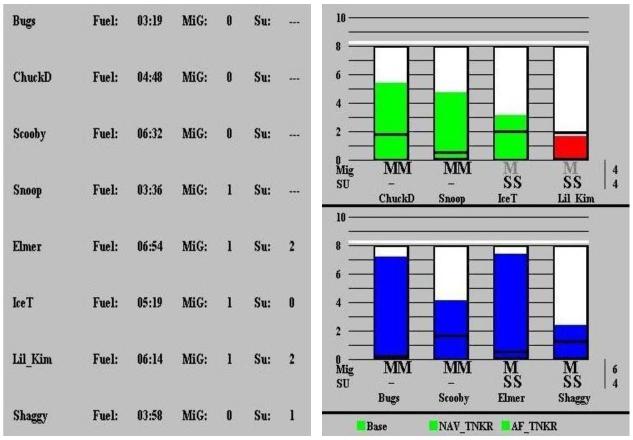


Figure 1. Tabular (left) and graphical (right) data displays. Both displays included information concerning remaining fuel and weapons of team assets.

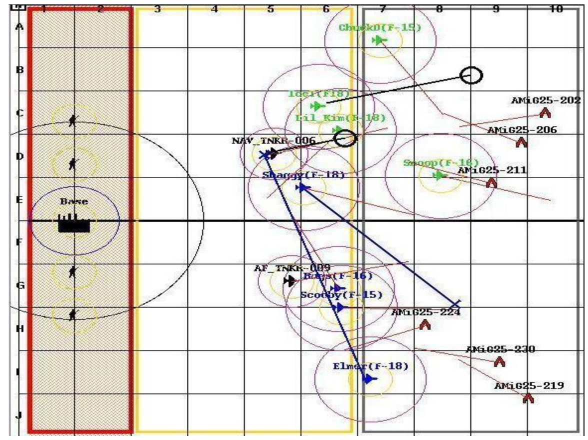


Figure 2. An image from the tactical display. Participant created whiteboard marks (blue and black lines) in the image indicate asset and target route information.

Procedure

The duration of the experiment was approximately 16 hours, conducted across two 8-hour session. The first session was devoted to training and the second to experimental data collection. In addition to 13 practice trials, participants received written and oral instructions detailing the C2 task, the team's goals, the roles and responsibilities of each team position, and the use of the collaborative tools. Participants were instructed on how to complete the DSSQ-S and the NASA-TLX. The experimental schedule of conditions was counterbalanced across teams to control order effects. After completing all experimental trials, participants were asked to complete a post-task debriefing form designed to elicit their impressions of the experimental factors and the C2 simulation. All experimental trials were ten minutes in duration.

Results

A full and detailed accounting of the results of this experiment is beyond the scope of this manuscript. As such, this section is focused chiefly on participants' subjective workload and stress responses to the experimental manipulations.

Team Communication

Following completion of the experimental data collection, audio recordings, instant messenger logs, and DRAW logs of the communications between teammates were compiled and examined. When the tools were available, teams sent, on average, 3.61 IM and 71.44 DRAW messages per trial. As a manipulation check, the mean number messages sent with each collaborative tool were tested against a value of zero using one sample *t*-tests to establish that teams were, in fact, using them. The results of these analyses indicated that participants were communicating at a rate greater than zero using IM, t (20) = 5.94, $p < .05$, and DRAW marks, t (20) = 5.94.

Workload

To test the effects of the experimental manipulations on participants' evaluation of task workload, the mean of each participant's TLX ratings in each condition was calculated. Mean TLX ratings for each experimental condition are presented in Table 1. Mean workload ratings were tested for statistically significant differences between conditions by means of a 3 (team position) × 2 (task demand) × 2 (team communication) × 2 (data-display) mixed-model ANOVA. The results of the analysis indicated statistically significant main effects of task demand, F (1, 102) = 27.91, $p < .05$, and data-display conditions, F (1, 102) = 8.91, $p < .05$, and statistically significant interactions between team position and team communication conditions, F (2, 102) = 5.66, $p < .05$, and between task demand and team communication conditions, F (1, 102) = 4.85, $p < .05$. No other sources of variance in the analysis were significant (all $p > .05$). Overall, participants rated their workload as higher in the high task demand condition compared to the low demand condition, and as higher when using the tabular display compared to the graphical display.

Table 1. *Mean NASA-TLX and DSSQ-S subscale change scores in each experimental condition.*

Team Position	Standard Communication				Enhanced Communication			
	Tabular Display		Graphical Display		Tabular Display		Graphical Display	
	Low	High	Low	High	Low	High	Low	High
	NASA-TLX Workload							
WD	55.14 (2.81)	57.62 (2.92)	51.25 (2.62)	54.65 (2.84)	50.57 (2.71)	54.11 (2.69)	48.08 (2.91)	54.29 (2.59)
Strike	46.76 (3.14)	48.63 (3.18)	44.68 (2.91)	45.94 (3.19)	49.24 (2.80)	52.54 (3.04)	46.18 (2.90)	50.53 (2.80)
Tanker	47.82 (3.41)	48.00 (3.48)	46.17 (3.53)	48.59 (3.34)	45.99 (3.46)	48.85 (3.55)	45.67 (3.89)	48.06 (3.56)
Mean	50.32 (1.84)	52.10 (1.90)	47.61 (1.73)	49.96 (1.86)	49.12 (1.70)	52.43 (1.76)	46.84 (1.80)	51.54 (1.68)
	DSSQ-S Task Engagement							
WD	-.25 (.24)	-.30 (.21)	-.09 (.24)	-.21 (.21)	-.44 (.26)	-.43 (.27)	-.02 (.20)	-.42 (.22)
Strike	.20 (.15)	-.06 (.18)	.17 (.16)	.08 (.17)	.14 (.13)	-.01 (.17)	.20 (.14)	.17 (.13)
Tanker	-.32 (.24)	-.12 (.25)	-.06 (.29)	-.02 (.23)	-.03 (.26)	-.30 (.37)	.11 (.32)	-.08 (.28)
Mean	-.09 (.13)	-.17 (.12)	.02 (.13)	-.05 (.12)	-.13 (.13)	-.23 (.15)	.09 (.11)	-.12 (.12)
	DSSQ-S Distress							
WD	.44 (.18)	.75 (.19)	.12 (.16)	.66 (.21)	.18 (.15)	.57 (.19)	.39 (.20)	.74 (.20)
Strike	.69 (.24)	1.18 (.28)	.35 (.18)	.46 (.19)	.87 (.25)	1.00 (.28)	.47 (.23)	.71 (.22)
Tanker	.19 (.15)	.41 (.25)	-.18 (.15)	-.23 (.17)	-.16 (.18)	-.04 (.26)	.00 (.18)	.69 (.24)
Mean	.49 (.13)	.85 (.14)	.15 (.10)	.40 (.12)	.39 (.13)	.62 (.15)	.34 (.13)	.72 (.13)
	DSSQ-S Worry							
WD	-.46 (.14)	-.41 (.13)	-.48 (.13)	-.61 (.13)	-.61 (.14)	-.47 (.12)	-.67 (.14)	-.54 (.12)
Strike	-.36 (.13)	-.41 (.13)	-.42 (.13)	-.46 (.13)	-.24 (.17)	-.32 (.15)	-.40 (.11)	-.34 (.13)
Tanker	-.25 (.13)	-.15 (.11)	-.27 (.13)	-.50 (.13)	-.54 (.20)	-.49 (.19)	-.41 (.18)	-.48 (.19)
Mean	-.38 (.08)	-.36 (.08)	-.42 (.08)	-.53 (.08)	-.45 (.10)	-.41 (.08)	-.51 (.08)	-.45 (.08)

Note. Values in parentheses are standard errors.

Follow-up post hoc paired sample *t*-tests for the team position × team communication interaction revealed no statistically significant differences between the team positions for either of the communication conditions (all comparisons $p > .05$). In these, and all subsequently reported post hoc analyses, the Dunn-Sidak alpha correction was employed to control Type-I error rates (Kirk, 1995). However, a trend within the data suggested that WDs rated their workload as slightly higher in the enhanced communication condition, and strike operators rated their workload as slightly lower in the same condition (both $p < .10$).

Post hoc paired sample t-tests investigating the task demand × team communication condition interaction indicated statistically significant differences between the low and high demand conditions in each team communication condition, t (104) = -3.60 and -5.63, respectively, $p < .05$. However the mean difference between the low and high task demand conditions was greater in the enhanced communication condition compared to the standard communication condition (i.e., participants' estimates of workload in the enhanced condition were lower in the low demand condition and higher in the high demand condition than those of the standard communication condition).

Stress State

Post-experiment, mean DSSQ-S subscale change scores were computed for each participant in each condition. In addition, a mean post-task subscale score was calculated for each participant as an index of participants' post-experiment state. However, due to a technical error, DSSQ-S data for three teams could not be recovered for analysis. Consequently, all subsequently reported analyses concerning team communications are based on data drawn from the remaining 18 participant teams.

Overall, the mean post-experiment scores indicated that participants' ratings of task engagement were largely unchanged (M = -.08, SD = 1.03), distress increased slightly (M = .50, SD = .95), and worry decreased slightly (M = -.44, SD = .67). Correlations between pre-task and post-experiment DSSQ-S ratings were .59, .60, and .78 for task engagement, distress, and worry, respectively (all $p < .05$), suggesting that participants' mood states were relatively stable from pre- to post-experiment.

Mean DSSQ-S change scores for each subscale are presented in Table 1. Subscale change scores were tested for statistically significant differences between experimental conditions by means of separate 3 (team position) × 2 (task demand) × 2 (team communication) × 2 (data-display) mixed-model ANOVAs.

Task engagement. The results of the task engagement analysis revealed statistically significant main effects for the task demand, F (1, 87) = 7.64, $p < .05$, and data-display factors, F (1, 87) = 5.24, $p < .05$. No other sources of variance in the analysis were significant (all $p > .05$). Participants rated their engagement as lower in the high demand condition compared to the low demand condition, and as lower in the tabular data-display condition compared to the graphical condition. Overall, participants were more engaged when the task was less demanding and when they had access to the graphical data-display.

Distress. For the distress subscale, the results of the ANOVA indicated a statistically significant main effect of task demand, F (1, 87) = 21.03, $p < .05$, and statistically significant interactions between team position and communication conditions, F (2, 87) = 4.75, $p < .05$, team communication and data-display conditions, F (1, 87) = 8.91, $p < .05$, and a three-way interaction between team position, task demand, and data-display condition, F (2, 87) = 3.67, $p < .05$. No other sources of variance in the analysis were statistically significant.

Follow-up post hoc paired sample t-tests for the team position × team communication interaction revealed no statistically significant differences between the team positions for either of the communication conditions (all comparisons $p > .05$). However, a trend within the data suggested that strike operators rated their distress as slightly higher in the enhanced communication condition ($p < .10$).

Post hoc paired sample t-test analyses of the team communication × data-display interaction indicated that participants rated their distress as higher when using the tabular data-display as compared to the graphical display, but only in the standard communication condition. No distress differences were observed between data-display conditions in the enhanced communication condition ($p > .05$).

To further explore the team position × task demand × data-display interaction, separate post hoc 3 (team position) × 2 (task demand) repeated measures ANOVAs were computed for each data-display condition. For the graphical data-display, the results of the analysis indicated that ratings of distress varied by task demand, F (1, 87) = 17.78, $p < .05$. Participants rated their distress as higher in the high task demand condition compared to the low condition when using the graphical display.

The results for the tabular data-display were more complex, in that a statistically significant main effect of task demand, F (1, 87) = 7.09, $p < .05$, and a statistically significant team position × task demand interaction, F (2, 87) = 4.98, $p < .05$, were identified. Subsequent post hoc paired sample t-tests indicated that, in the tabular display condition, WDs rated their distress as significantly higher in the high task demand condition compared to the low demand condition. No such differences were detected for strike and taker operators.

Worry. The results of the analysis for the worry subscale indicated a statistically significant main effect for data-display condition, F (1, 87) = 5.17, $p < .05$, and statistically significant interactions between task demand and team communication conditions, F (1, 87) = 4.34, $p < .05$, and between task demand, team communication, and data-display conditions, F (1, 87) = 4.28, $p < .05$.

To continue examination of the three-way interaction, separate 2 (task demand) × 2 (team communication) repeated measures ANOVAs were computed for each of the data-display conditions. For the tabular display condition, no statistically significant differences between conditions were detected (all $p > .05$). For the graphical display, however, the analysis indicated a statistically significant task demand × team communication interaction, F (1, 89) = 6.34, $p < .05$. Follow-up post hoc paired sample t-tests indicated that in the enhanced communication condition, participants did not rate their worry differentially between task demand conditions. Conversely, in the standard communication condition, participants rated their worry as lower in the high demand condition compared to the low demand condition.

Discussion

The purpose of this experiment was to provide a preliminary attempt to characterize the relationship between collaborative tools and subjective workload and stress. In general, the results of this experiment suggest that collaborative tools and technologies may be both a significant source of, and solution to, operator workload and stress.

Participants' workload ratings were higher in the high demand condition and when using the tabular display. Workload ratings were also influenced by the collaborative tools available to participants, but their effects were moderated by the team position and task demand factors.

Effects of the experimental manipulations on subjective stress response were more nuanced than anticipated. Overall, task engagement and worry decreased, and distress increased from pre- to post-experiment. The observed decrement in task engagement was exacerbated by high task demands and the tabular data-display condition, but was not changed by collaborative tool availability. Distress was further increased by high task demands, but the strength of this effect was dependent on team position, team communication, and data-display conditions. Worry decreased differentially depending on task demand, team communication, and data-display conditions.

Collaborative tools. Access to additional collaborative tools had relatively weak effects on subjective workload and stress in this experiment. Though there were some suggestions of incremental differences in workload and stress experience based on team position and collaborative tool availability, the magnitude of these effects was mostly negligible. This indicates that collaborative tool usage, as implemented in this experiment, does not exert any additional 'costs' in terms of workload or stress (though see below). However, these results do not indicate that organizations should be unconcerned about workload and stress associated with collaborative technologies; collaborative tools may still be a significant source of workload and stress for users for a variety of reasons (e.g., because of poor interface design, inadequate training, laborious implementation, etc.).

Data-display types. Access to the graphical data display decreased subjective estimates of workload and stress compared to the tabular display in this experiment. It is reasonable to assume that while some degree of benefit was provided by the reduction in communication required during a trial (i.e., that relating to WDs and operators exchanging asset weapon and fuel information), some of the observed benefit of the graphical display should also be attributed to its enhanced functionality, which provided WDs with salient cues concerning fuel management. This, in turn, may have allowed WDs to more efficiently allocate team assets to enemy targets, resulting in improved team scores.

The relationship between data display type and team position was also reflected in subjective distress and worry ratings, but this relationship was moderated by task demand and team communication conditions. An interesting aspect of these results is in the complexity of the interactions observed between the experimentally manipulated factors. The results do not 'add up' to a singular representation describing the relationship between subjective stress and the experimental factors. Instead, they illustrate that, under varying circumstances, some team members may be benefitted by the availability of collaborative technologies while others are simultaneously unchanged (or hindered) by exactly the same tools.

This suggests that teams may be better served by *adaptive* collaborative technologies, which may be tailored according to the needs and circumstances of individual team members (Baldwin, 2003). By allowing team members (or an automated decision aid) to flexibly and dynamically alter the functionality of these tools, it may be possible to maximize team performance while minimizing associated negative outcomes such as subjective workload and stress. Determining the nature and behavior of such tools is likely to be a fruitful area of future research.

References

Albert, D.S., Hayes, R.E, 2003. Power to the Edge. CCRP Publication Series, Washington, DC.

Baldwin, C.L. (2003). Neuroergonomics of mental workload: New insights from the convergence of brain and behaviour in ergonomics research. *Theoretical Issues in Ergonomic Science, 4*, 132-141.

Cohen, S.G., & Bailey, D.E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, *23,* 239-290.

Finomore, V.S., Knott, B.A., Nelson, W.T., Galster, S.M., & Bolia, R.S. (2007). The effects of multimodal collaboration technology on subjective workload profiles of tactical air battle management teams. *Proceedings of the International Symposium on Aviation Psychology, 14*, 190-196.

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of human perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 41-1 to 41-49). New York, NY: Wiley.

Heacox, N.J., Moore, R.A., Morrison, J.G., Yturralde, R.F. (2004, June). Real-time online communications: 'Chat' use in navy operations. In *Proceedings of the Command and Control Research and Technology Symposium*, San Diego, CA.

Hertel, G., Geister, S., & Konradt, U. (2005). Managing virtual teams: A review of current empirical research. *Human Resource Management Review, 15*, 69-95.

Kirk, R.E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal and coping*. New York: Springer.

Martins, L.L., Gilson, L.L., & Maynard, M.T. (2004). Virtual teams: What do we know and where do we go from here? *Journal of Management, 30*, 805-835.

Mathieu, J., Maynard, M.T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management, 34,* 410-476.

Matthews, G., Davies, D.R., Westerman, S.J., & Stammers, R.B. (2000). *Human performance: Cognition, stress and individual differences*. London: Psychology Press.

Matthews, G., Emo, A.K., & Funke, G.J. (2005, July). *A short version of the Dundee stress state questionnaire*. Poster presented at the annual meeting of the International Society for the Study of Individual Differences, Adelaide, Australia.

Norman, D.A., & Bobrow, D.G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7,* 44–64.

Sloan, D. (2008, July 28). *Boeing to demo net-centric upgrade on AWACS aircraft at empire challenge*. Retrieved February 27, 2009, from the Boeing Company web site: http://www. boeing.com/news/releases/2008/q3/080708a_nr.html

Welford, A.T. (1978). Mental work-load as a function of demand, capacity, strategy and skill. *Ergonomics, 21,* 151–167.

Zapf, D., Isic, A., Bechtold, M. & Blau, P. (2003). What is typical for call centre jobs? Job characteristics, and service interactions in different call centres. *European Journal of Work and Organizational Psychology*, *12*, 311-340.

HUMAN-ASSISTED LOGISTICS OPTIMIZATION (HALO): SUPPORT FOR TIMELY LOGISTICS DECISION-MAKING

David E. Weldon
JXT Applications, Incorporated

The need for rapid response capabilities and effective joint operations dictates a new approach in which deployment planning is integrated into the mission-planning processes. In the research reported here, we explored new technologies to enable rapid identification of feasible transportation options and enhance shared awareness of commander's intent and inter-command collaboration. Our solution extends state-of-the-art Tabu Search algorithms, producing effective transportation solutions in significantly less time than current models. The human guidance and collaborative components of HALO enhance performance by accommodating dynamic operational requirements. HALO benchmark tests demonstrated superiority to other optimization algorithms on Multi-Vehicle, Pickup and Delivery Problems with Time Windows (MVPDPTW) reported in the literature. When applied to intra-theater MVPDPTW distribution problems, HALO generated feasible, near-optimal solutions acceptable to subject matter experts in less than a minute. We conclude that HALO is a powerful decision tool that is easily integrated into current planning processes with strong user acceptance.

Logistics support, including force deployment and sustainment planning and execution, has traditionally been viewed as a support function to combatant commanders, even though it is one of the key enablers (and limiters) of any military operation. In this model, up-front mission planning and course of action determination is accomplished in a somewhat stovepiped fashion, with relatively little visibility into transportation constraints. These initial requirements are then handed off to Logistics planners to define transportation options that can best support the combatant commander's needs. Through an iterative and cooperative process involving both the supported command (e.g. United States Central Command - USCENTCOM) and the supporting commands (e.g. United States Transportation Command -USTRANSCOM), transportation options are identified, analyzed, and validated. The initial operational plan often needs to be adjusted based on time-phased deployment constraints or shortfalls identified during the transportation option analysis and selection process.

The changing nature of the threat and the associated need for greater mobility, flexibility, efficiency, rapid response capabilities and effective joint operations dictate a new approach. With the recent rapid growth in information and communication technologies, military operations are transforming into a network-centric model that emphasizes shared situation awareness, visibility of a common operating picture and commander's intent, and self-synchronization of distributed forces. This model enables unprecedented levels of collaboration, faster decision-action cycles, and the flexibility to adapt quickly and effectively to changing requirements, priorities and situations. Realization of this model requires that force deployment planning become an integral component of the core mission planning process so that logistics considerations, the opportunities afforded, and the constraints imposed, are known and accounted for in real time during the planning of combat operations.

Thus, our overall goal was to research requirements and design concepts for a Human-guided Tabu Search algorithm that generates an optimal Airlift transportation solution for satisfying operational requirements. To achieve this goal we researched approaches for (1) improving speed of solution convergence, (2) incorporating commander's intent, (3) improving collaboration among operational and logistics planners, and (4) adapting to dynamic preferences and priorities. Based on this research we developed proof-of-concept demonstration software to test hypotheses and verify the efficacy of our approach. We named our demonstration software "Human Assisted Logistics Optimization," or HALO.

Background and Theoretical Approach

Group-Theoretic Tabu Search (GTTS) in Logistics Planning

McKinzie (2004, p. 2) describes the movement of cargo and passengers (PAX) within certain time-window constraints as a highly complex routing and scheduling problem called the Strategic Mobility Mode Selection Problem (SMMSP). The literature characterizes SMMSP as a variant of the Multi-Vehicle Pickup and Delivery

Problem with Time Windows (MVPDPTW), which is a complex generalization of the "Traveling Salesman Problem" (TSP) - a well-known and heavily-studied nondeterministic, polynomial-time, hard (NP-hard) problem (McKinzie, 2004, p.22).

Many types of metaheuristics are applicable to MVPDPTWs as well as other types of logistics and scheduling problems: ant algorithms, Bayesian algorithms, constraint programming, deterministic annealing, genetic algorithms, greedy algorithms, memetic algorithms, multi-objective evolutionary algorithms, simulated annealing, and Tabu Search. Of the numerous deterministic and heuristic search algorithms applied to MVPDPTWs, Tabu Search has proven the most effective (Crino, et al., 2004; Lambert, 2003; McKinzie, 2004). Basic Tabu Search (Glover, 1989, 1990) is a metaheuristic algorithm for solving optimization problems. It is designed to guide other deterministic or heuristic methods so they can escape local minima and prevent oscillations between previously tried solutions; thereby enhancing the likelihood that a global minimum to a "cost function" will be found. A basic Tabu Search algorithm consists of the following:

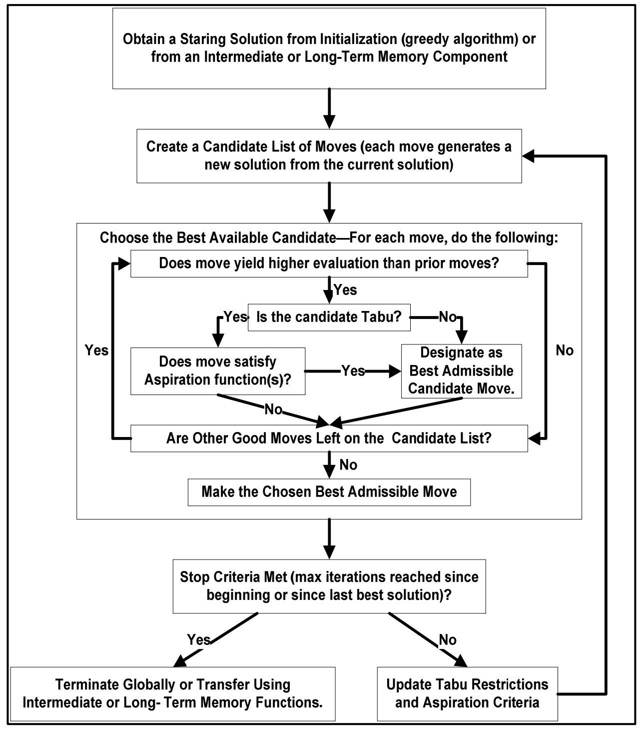


Figure 1. Basic Tabu Search Algorithm.

- A representation of the problem space being searched. In the TSP this would be a matrix representation of the graph consisting of the cities (vertices) and highways or air routes connecting the cities (arcs).

- A short list of previously tried "moves" that are TABU; that is, as long as a move is on this list, it (or its reversal) cannot be tried again. The Tabu list is designed to keep the algorithm from cycling around a local minimum and to encourage breaking out of the local minimum. The length of the list determines how long a move is "Tabu." If there are a number of constraints that apply to the problem, a separate Tabu list may be kept for each constraint.

- Zero or more *aspiration level functions (alfs)*. The purpose of an *alf* is to provide added flexibility to choose good moves by allowing the Tabu status of a move to be overridden (removed early from the list) if the *alf* is satisfied. The form of an *alf* depends heavily on the search problem and includes the cost of the move (however "cost" is defined in the problem).

- Zero or more intermediate and long-term memory functions. These may be added to the basic Tabu Search algorithm to achieve regional intensification of the search or global diversification of the search. By recording and comparing features of a number of "best" solutions reached during a given period of search features common to all, or a majority, of these solutions are used to guide the search by penalizing moves that do not contain these features—resulting in regional intensification. The long-term memory functions serve to diversify the search by deliberately avoiding moves (and solutions) that have common features as defined above.

- The Tabu Search algorithm itself. Figure 1 shows the logical flow of the basic Tabu Search algorithm including all components listed above.

Group Theoretic Tabu Search (GTTS) applies algebraic group theory to the representation of MVPDPTWs. In this approach, vertices and arcs in the MVPDPTW graphic representation are mapped one-to-one onto the finite set A consisting of {1, 2, 3, … , n}, and the symmetric group of n-letters, S_n, is the group of all permutations of set A. This allows the representation of arbitrarily large MVPDPTWs as a 2 x n matrix or array with the first row containing the numbers 1 through n and representing the vehicles and customers[1] in the MVPDPTW. The elements of the second row in its most elegant form contain "cycles" of pickup, transport, delivery, and return—represented by the numbers from row 1 in a short list. (See Crino, et al., 2004, for a detailed and precise description). In this representation a move is a swap of elements within or between cycles, or adding or dropping an element in a cycle. This approach was applied to a large combat theater distribution vehicle routing and scheduling problem (a member of the class of MVPDPTW) with significant success. The best solution was found in just under 63 minutes; however, two near optimal solutions were found after only 11.5 and 24.5 minutes. The use of algebraic group theory in the representation of a given MVPDPTW is a major innovation: elegant in its simplicity yet enormously powerful in its effect, both in solution speed and achievement of near optimal solutions early in the search.

Transportation Optimization as a Joint Cognitive System (JCS)

The role of decision support technology should be to serve the humans who are ultimately responsible for the decision. A JCS is a system in which the human and machine work collaboratively to solve a problem or make a decision (Woods & Hallnagel, 2006). The software component is a cognitive tool that can be wielded by a competent practitioner. This approach exploits the complementary knowledge and "reasoning" processes of the human and software components to obtain better decisions than could be achieved with either alone. In a JCS, the human serves as a manager of knowledge resources that can vary in kind and amount of "intelligence" or power. A JCS is an alternative architecture to the traditional approach of applying computational technology as a stand-alone machine expert that serves as a replacement for perceived human deficiencies; i.e., the "prosthetic" paradigm (Woods & Hallnagel, 2006). JCS architectures avoid many of the problems introduced by the prosthetic design approach (Guerlain, 1999). Problems outside the machine's level of competence no longer lead the human to ineffective solutions. Instead, those aspects of the problem that the machine expert does know about are used effectively to aid in the overall solution. Issues related to trust, complacency, over-reliance, control, and responsibility are decreased.

The JCS approach drove the development of HALO. We inserted the user into the heart of the GTTS algorithm. Users can manually modify candidate solutions, backtrack to previous solutions, modify the tabu list, *alfs*, and any other cost parameters associated with problem elements, and monitor or halt the search algorithm. The User Interface provides an operationally meaningful visualization of the current and other potential search solutions, some intuitive indication of the progress and current attentional focus of the algorithm within the search space, and controls for manipulating and guiding the search algorithm. To be "operationally meaningful" the visualization must represent information and candidate transportation solutions in terms of the operational constituents of the problem set, such as Ports of Debarkation, Ports of Embarkation, waypoints and routes, aircraft assets, cargo, timing profiles, etc. Users must be able to manipulate these objects graphically to obtain detailed information and manage how they are considered within the algorithm. We developed our human guidance component by drawing on recent work in human-guided Tabu Search (e.g. Lesh, et al., 2003; Anderson, et al., 2000) and integrating the JCS approach described above. With respect to the strategic mobility optimization problem and the deployment planning process, the human guided component provided a means of ensuring that commander's intent and practical knowledge of real-world constraints were considered in the optimized transportation solution.

Research Procedures

HALO development process. The following procedural steps were carried out to assess the efficacy of HALO: (1) We acquired existing open source Tabu Search software (OpenTS) and modified and integrated it with human-guidance control functions that would allow users to guide the search by setting and modifying search parameters.[2] (2) We acquired the necessary GTTS objects and methods from the code written by Burks (2006) and integrated them with the OpenTS software. (3) We identified optimization strategies that could be incorporated into the code, implemented them in additional objects and methods, and exposed them to the user interface to put the

[1] Vehicles are the air and ground transports used to move cargo and PAX. Customers are the origins, Ports of Debarkation, Ports of Embarkation, service and destination hubs, and delivery points.

[2] OpenTS can be downloaded from the web site: http://www.coin-or.org/Ots/index.html.

search under human control and guidance. (4) We created a specific Theater Distribution Problem (TDP) scenario to test our hypotheses and refine the JCS architecture. And, finally, (5) we tested HALO using variations of the TDP and compared the results to benchmark solutions for the TDP using Basic Tabu Search and GTTS. We also demonstrated HALO to logistics experts to obtain feedback on the utility and usability of the tool.

TDP scenario description. The selected scenario was a hypothetical, high-intensity, small-scale contingency operation with a highly compartmentalized Area of Operations (AO). There were two stages of operation: deployment and sustainment. The planning goal was to determine the support structure and routing requirements necessary to (1) deploy forces from staging bases in Turkey to Tbilisi, Georgia and Yerevan, Armenia and (2) to sustain combat operations in the AO. We created several variations on the scenario to allow testing and benchmark comparisons. This also allowed us to demonstrate the capabilities of HALO to Subject Matter Experts.

JCS user interface. This interface allows the user to control critical functions in the execution of HALO software while displaying the results of the search in a multi-document display. The GTTS functions under control of the user include (1) starting, stopping, resetting the search, (2) adjusting the impact of thirteen components of the GTTS "cost" function before and during execution of the search, (3) set problem parameters such as the number of planning days, whether vehicles are allowed to refuel enroute, crew work hours, and whether vehicles are allowed to arrive early at a depot, service or destination hub, or a delivery point, and (4) save solutions, reload solutions, and resume solution searches. The user interface display is shown in Figure 2 with the four main windows open for inspection. The four main windows provide the following displays and functions:

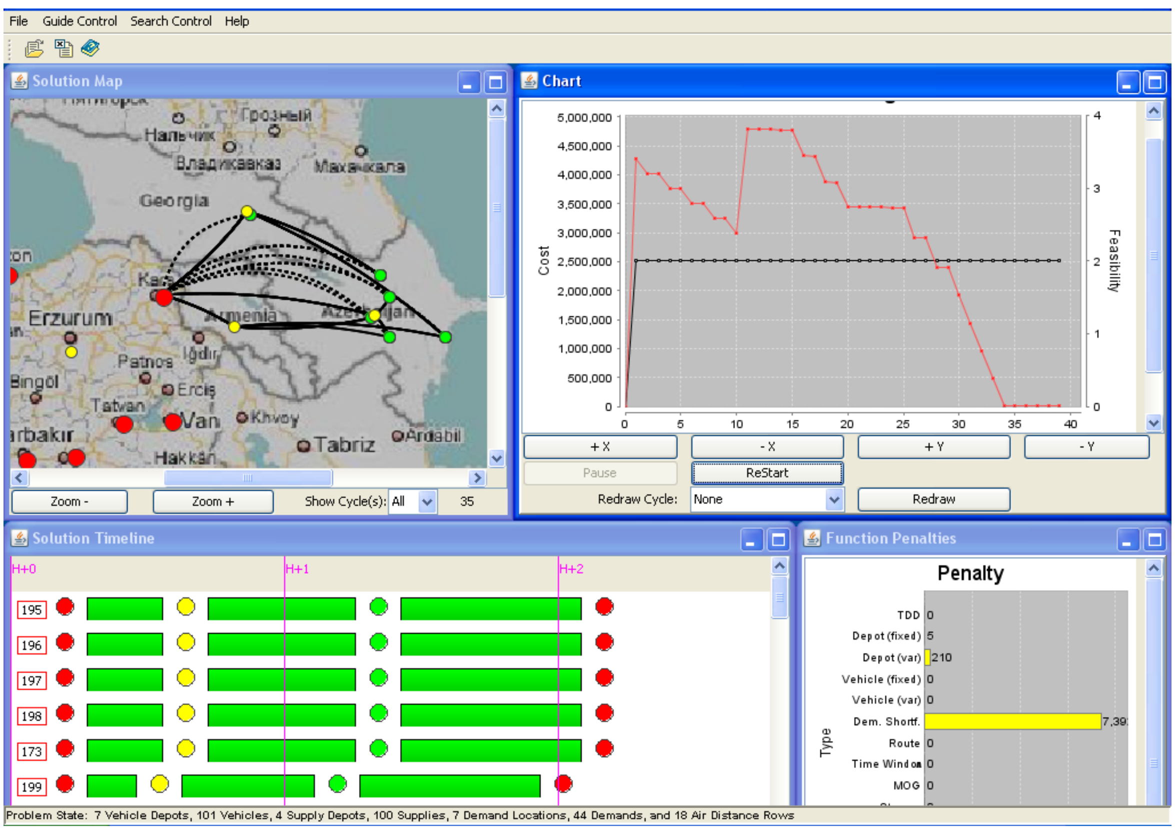


Figure 2. JCS user interface for HALO.

- Map Display Window (upper left quadrant). Displays a map of the Theater of Operations with vehicle depot, supply depot, and demand locations shown by color-coded symbols (red, yellow, and green circles, respectively). As routes are built and removed by the Tabu Server on each iteration, the route changes are displayed in this window.

- Route Timeline Window (lower left quadrant). Displays each vehicle route in the form of a timeline for easy detection of various route violations. This is the main window for examining vehicle and route properties. Users may "mouse-over" a route symbol to see a brief description of the entity represented by the symbol, or "right-click" to get a full description of the selected route and vehicle properties.

- Cost/Feasibility Chart Window (upper right quadrant). Displays changes in the Objective Cost Function and Feasibility of the solution found on each iteration of GTTS. The display is in near-real time. Figure 2 shows the state of the search on the 39th iteration of the search. The search has been paused temporarily to examine in detail the solution found on the 34th iteration. The solution at this iteration is near-feasible (value = 2) and has an Objective Cost of 10,247. The Map Display Window and Route Timeline Window now display their states at the 34th iteration.

- Cost Breakdown Bar Chart Window (lower right quadrant). Displays the individual components of the Objective Cost Function. This bar chart gives immediate visual understanding of the penalty costs that are the cause of the "near-feasible" classification of the solution produced on the 34th iteration. The largest "cost" is Demand Shortfall followed by Time Definite Delivery violations and the Depot cost (these are the largest penalties because the weighted parameters in the cost function have been set to focus on timely delivery of the cargo and PAX. Users may right-click on a bar to see a detailed breakdown of the objects contributing to the objective cost or penalty cost represented.

Figure 3. HALO cost function dialog box.

GTTS cost function control. HALO provides access to the GTTS cost function weights through a dialog box accessible from the "Guidance Control" menu. The HALO default weight settings for the thirteen components of the GTTS cost function are shown in Figure 3 and support a general intent of "minimizing the logistics footprint" in support planning. The first six components are "costs" associated with vehicle depot, supply depot, and vehicle fixed and variable costs (variable costs are associated with vehicle and depot maintenance and ongoing operations). For a military operation requiring tight time windows and no demand shortfalls where the commander's intent is absolute assurance that the warfighter receives supplies when needed (as in the benchmark TDP contingency operation described above), the depot and vehicle cost weights would be minimized and the weights for Time Definite Delivery, Demand Shortfall Penalty, Route Length Violation Penalty, Depot Queue Violation Penalty, and Time Window Violation Penalty would be maximized. If commander's intent is something other than these two scenarios, the thirteen weights would be adjusted to reflect that intent.

Benchmark Results and Conclusions

Tests on HALO were limited to TDPs, which tend to be of shorter duration requiring fewer resources. Nevertheless, we were able to use test data supplied by Burks (2006) as well as several variations on our scenario to obtain both benchmark and scalability results. Tests on these data yielded the following computational-time results: (1) For small TDP scenarios (150-200 nodes), multiple optimum solutions with lowest cost were produced in less than 40 seconds. The initial optimal and feasible solution often appeared in the first 10-20 seconds. And (2) For intermediate TDP scenarios (200-600 nodes), optimum solutions with lowest cost were completed in less than three minutes and low cost, near-optimum solutions were available as early as 45 seconds into the search. This performance exceeded basic Tabu Search (e.g., Tan, et al., 2000) and GTTS without human guidance (Burks, 2006). It also easily surpassed non-Tabu search (genetic) algorithms (e.g., Homberger and Gehring, 2005).

In conclusion, the HALO software architecture represents an optimal approach to collaborative logistics planning and it appears to be fully scalable, although further research is needed to establish firmly its utility in supporting Strategic Airlift Problems and Strategic Mobility Mode Selection Problems. Also, we conclude that human-guidance controls strongly support a JCS architecture for logistics planning. Proper use of these controls can dramatically shorten search time and produce optimal solutions that accurately reflect commander's intent. Finally, we conclude that a JCS contributes significantly to user acceptance and positive regard for the HALO software.

Acknowledgements

This research was supported by a Phase I Small Business Innovation Research grant from the Air Force Research Laboratory under contract number FA8750-07-C-0122. The views, opinions, and/or findings contained herein are those of the author and should not be construed as an official Air Force position, policy, or decision. The author also thanks Scott M. Cone, James T. Moore, and Robert E. Burks for their contributions to this research.

References

Anderson, D., Anderson, E., Lesh, N., Marks, J., Mirtich, B., Ratajczak, D., and K. Ryall., (2002). Human-guided simple search. In *Proc. of AAAI 2000*, pages 209–216.

Burks Jr., R. E. (2006). An adaptive Tabu Search heuristic for the location routing, pickup and delivery problem with time windows with a theater distribution application. *Ph. D. Dissertation, Air Force Institute of Technology, Wright-Patterson Air Force Base, Dayton, Ohio.*

Crino, J. R., Moore, J. T., Barnes, J. W., & Nanry, W. P., (2004). Solving The Military Theater Distribution Vehicle Routing and Scheduling Problem Using Group Theoretic Tabu Search, *Mathematical and Computer Modeling*, **39**(6-8), pp. 599-616.

Glover, F., (1989). Tabu Search—Part I. *ORSA Journal on Computing*, **1**(3), pp. 190-206.

Glover, F., (1990). Tabu Search: A Tutorial. *Interfaces*, **20**(4), pp. 74-94.

Guerlain, S., Smith, P.J., Obradovich, J. Heintz, Rudmann, S., Strohm, P., Smith, J.W., Svirbely, J. and Sachs, L. (1999). Interactive critiquing as a form of decision support: An empirical evaluation. *Human Factors*, **41**, 72-89.

Homberger, J., H. Gehring. 2005. A two-phase hybrid metaheuristic for the vehicle routing problem. *Unpublished Manuscript.*

Lambert, G. R., (2004). A Tabu Search Approach to the Strategic Airlift Problem. PhD Dissertation, University of Texas at Austin, Texas, 214 pages.

Lesh, N., Lopes, L., Marks, J., Mitzenmacher, M., Schafer, G., (2003). Human-Guided Search for Jobshop Scheduling. *Third International NASA Workshop on Planning and Scheduling*, 2003.

Makinzie, K., (2004). A Tabu Search Approach to Strategic Mobility Mode Selection. PhD Dissertation, University of Texas at Austin, Texas.

Tan, K.C., Lee, L.H., and Zhu, K.Q. (2000). Heuristic methods for vehicle routing problem with time windows. *Proc. 6th International Symposium on Artificial Intelligence Math.*, Ft. Lauderdale, FL.

Woods, D., and Hollnagel, E. (2006). Joint Cognitive Systems. CRC Press, Boca Raton, FL.

THE EFFECTS OF SUCCESS RELATED PRESSURE ON INFORMATION PROCESSING STRATEGIES AND PLAN CONTINUATION ERROR

Bourgeon Léonore, Valot Claude, Fauconnet Sophie, Loyau Fanny
French military institute of biomedical research
Brétigny-sur-Orge, France
Navarro Claude
Laboratory of work and cognition – University of Toulouse
Toulouse, France

An experiment was conducted to explore whether plan continuation errors could be explained by two types of perseveration behaviours: Perseveration in a wrong Representation do the Situation (PRS) and Perseveration in a risky Plan of Action (PPA). Effects of success-related pressure and flight phase on pilots performance were also examined. Six scenarios were created where expected or unexpected threats had to be managed. Pilots chose between three plans of actions corresponding to PPA, PRS and Flexibility. Results showed that the two types of perseveration could effectively explain plan continuation errors even though PPA characterised cruise phase and PRS was more frequently chosen when managing threats during the approach phase. An effect of success-related pressure was observed as pilots experiencing high pressure were more flexible than pilots experiencing low pressure.

Plan continuation error or *bias* is an essential component of numerous aeronautical accidents. It occurs when pilots fail to revise an original flight plan despite emerging evidence that suggests it is no longer safe and that a new plan is required (Orasanu, Martin & Davison, 2001). An analysis of accidents reports revealed that nearly two-thirds of decision errors can be classified as plan continuation errors (NTSB, 1994). Moreover, a European safety study showed that between 1991 and 1996, 41.5% of fatal accidents in general aviation were due to perseveration on landing under degraded meteorological conditions while they only represent 4.5% of all accidents (Bureau d'Enquêtes et d'Analyses, 1997). This inability to adapt to changes in the environment which leads to human error can also be related to a general behaviour called *perseveration* and defined in psychology as "the difficulty experienced in switching from one pattern of behaviour or method of working to another" (Coleman, 2001), as opposed to *flexibility*. This being so, perseveration may be observed in a large number of accidents such as a meteorologically changing context or the management of technical failures and may occur at any moment in the flight plan. For example, in military aviation, some accidents occurred when pilots persevered in applying check-lists whose items did obviously not match with the current situation. It is then essential to identify the underlying cognitive processes and factors that impair this decision making process.

Models of aeronautical decision making describe three main processes which are: information perception, elaboration of a mental representation of the situation and selection of a plan of action. Plan continuation error may result from any of these three processes. When an important cue relative to a threat is not perceived or is not interpreted as a threat, the pilot representation of the situation is inaccurate, leading to an inadequate plan of action (Goh & Wiegmann, 2001; Wiegmann & Shappell, 1997). When it is perceived and properly interpreted as a threat, a plan continuation error may still occur if pilots underestimate the risk level associated with the continuation of their action plan and/or they overestimate their capacity to control the situation (Orasanu, Fisher & Davison, 2002; Goh and Wiegmann, 2001). Hence, plan continuation error could be explained by two types of perseveration behaviour: 1) Perseveration in a wrong Representation of the Situation (PRS) and 2) Perseveration in a risky Plan of Action (PPA). Moreover, many reports of accidents happening during the landing phase revealed that pilots made continuation plan errors even though they were aware early on of the deterioration of weather conditions at the destination field. Hence, plan continuation error can occur while changing flight conditions are expected and anticipated. This behaviour may be explained by the PPA type of perseveration where relevant information is perceived and well interpreted but where pilots fail to assess the risk level related to their plan of action. Yet, most studies dealing with plan continuation errors in flight simulation do not manipulate the threat expectancy factor and only refer to unexpected threats. One goal of our study was to verify whether these two types of perseveration could be observed in plan continuation errors and especially by comparing flight situations with expected vs. unexpected threats.

Additionally, while most studies on aeronautical decision making were conducted with commercial aircrews, fewer have been realized with military aircrews (Denihan, 2007; Sicard, Taillemite, Jouve & Blin, 2003). Yet, in this particular domain, flight situations can result in a high degree of complexity due to specific and sometimes hazardous missions (Prince & Salas, 1993). Flying the aircraft may become a secondary task compared to the mission related task (Sicard *et al*., 2003). In this context, organizational pressure may be very high and expressed in the form of pressure to succeed with the mission. A study by Denihan (2007) revealed that naval aviators acted in ways designed to foster their combat mission success over safety. Indeed, interviews of 11 pilots showed that cues related to reducing risk level and considered in the decision making process during non-combat missions were not considered during combat missions. Hence, organizational pressure may increase conflict between mission-related goals and safety-related goals. Yet, in a commercial flight simulation experiment using think-aloud protocols, external pressures represented only 4.2 percent of pilots talk (Orasanu, Fisher & Davison, 2002). Analysis of military pilot decision making could be of interest in determining how organizational pressure can have an impact on plan continuation error. The context of flight in the face of a threat is also an important component of plan continuation errors. An analysis of accident reports showed that plan continuation error is more frequent during approach and landing than during other phases of flight (Orasanu, Martin & Davison, 2001). Still, results from a study conducted in a simulation session where pilots encountered adverse weather did not support this finding (Wiegmann, Goh & O'Hare, 2002). On the contrary, unlike pilots who encountered adverse weather late during the flight, the majority of pilots who faced this event early during the flight decided to continue in accordance with their original flight plan. This result was explained by the authors as the need and the possibility for pilots to verify their assessment of the situation. In our study, we examined the impact of flight phase using various types of threats, such as deteriorating weather conditions, technical failure and external threat.

The purpose of this study was to verify three main hypotheses. First, we hypothesized that plan continuation errors would be explained by two types of perseveration behaviours: Perseveration in a wrong Representation of the Situation (PRS) and Perseveration in a risky Plan of Actions (PPA). On one hand, we expected that PRS would be characterised by wrong diagnosis and PPA by accurate diagnosis. On another hand, we expected that when threats are expected by pilots, plan continuation error should be explained by PPA while when threats are unexpected, plan continuation error should be explained by PRS. Second, we expected that a high organizational pressure would lead to plan continuation error while low organisational pressure would lead to flexibility. Finally, we expected that flight phase would impact decision making processes where the approach phase should lead to more plan continuation error than take-off and cruise phases.

Methods

Participants

Twenty pilots (19 men, 1 woman) from the French Air Force squadron specializing in the transportation of government authorities participated in the study. In flight hours, the participants' total flight experience ranged from 800 to 7,000 hrs and their mean total flying experience was 3442 hrs (*SD* = 1433 hrs). They ranged in age from 28 to 38 years with a mean age of 33 years (*SD* = 3 years). Participation in the study was on a voluntary basis with complete anonymity of the personnel.

Procedure

Participants were first asked to fill out a biographical questionnaire including information regarding their age and their flight experience. They were then given the experiment instructions and started the training session. When they felt comfortable with the use of the interface, they could start the experimental session. The latter was composed of three screens: 1) description of a flight situation (current coordinates of the flight) with contextual information (nature of the mission, flight plan, meteorological conditions, fuel level). Pilots were asked to build a mental representation of the situation and to click on the next stage only when they felt ready. They were informed that from this moment a stopwatch was started; 2) graphic interface representing the cockpit panel. Pilots could click on any instrument or messages they needed to be able to make a decision between three choices of action. Next, they had to complete a confidence level scale from 1 (no confidence in the decision made) to 5 (extremely confident in the decision made); 3) finally, they were asked to write down what elements influenced their decisions and what were the goal(s) they wanted to reach.

Graphic interface

The experiment was conducted with a laptop using the software "E-Prime". This software enables recording of all the actions made by the participants. Hence, analysis of decision making processes was possible with the creation of a specific graphic interface showing the front panel of a A319 cockpit (figure 1). Pilots clicked on a particular instrument to bring up a small information window displaying the information usually provided by this instrument. Additional links were displayed on the side of the panel providing information from Co-pilot, Air Traffic Controller and Cabin Crew. Pilots could open only one window at a time. Participants practiced on a training session until they felt comfortable with the set-up.

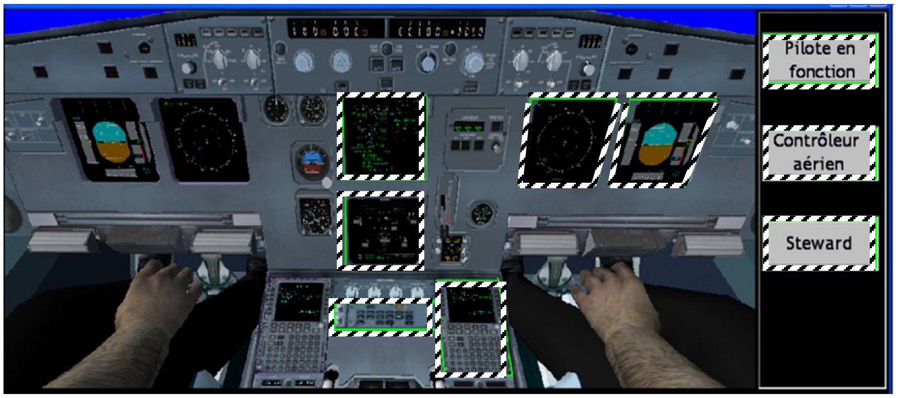


Figure 1. Graphic interface used in the experiment representing an A319 panel. All shaded rectangles could be clicked by participants.

Flight scenarios

Six scenarios were created for the study in collaboration with two pilots who were experts in human factors. They were designed in such a way that each flight situation was ambiguous and where the decision could only be made by the judgment of the pilot with no need for a check-list. Moreover, the threats illustrated by our scenarios had all been involved in incidents or accident databases. The six scenarios reflected three variables employed in this experiment. Threat Expectation (expected threat vs. unexpected threat) and Flight Phase (take-off, cruise, approach) were within-subject variables. Success-Related Pressure (high success-related pressure vs. low success-related pressure) was a between-subjects variable. For the *expected threat* condition, three of the scenarios were conceived such that a potential threat was presented in the first description of the flight situation whereas in the three *unexpected threat* scenarios no potential threat was initially presented. Additionally, each of these conditions occurred during either the take-off, cruise or approach phases of flight. Each participant responded to the six scenarios in random order. Organizational pressure was studied through success-related pressure which was manipulated by the nature of the mission presented at the beginning of the flight situation description. Pilots in the *high success-related pressure* condition had to convey important government authorities whereas pilots in the *low success-related pressure* condition had to convey neutral passengers. Flight plans were the same under both conditions.

Measurement of performance

Three plans of action were presented to participants as a decision choice. They could either divert the flight judging the situation to be much too risky, or they could continue according to the initial flight plan while monitoring flight parameters because of a high risk level, or finally they could continue according to the initial flight plan judging there was no associated risk. These 3 choices corresponded to the perseveration categorizations: Flexibility, PPA and PRS. For ANOVA analyses purpose, these responses were encoded into a numerical variable respectively as 3, 2 and 1, from the most appropriate decision to the least appropriate one. Information processing was analysed through 3 indicators: amount of information accessed, amount of target information accessed related directly to the threat and time spent reading target information as an indicator of the importance of the information for decision making. Finally, participants had to write down all cues that played a role in their decision choice and what goals they wanted to reach. These data were analysed with an *a posteriori* grid coding for building cue categorization and assessing accuracy of the diagnosis.

Results

Decision performance

In order to verify if plan continuation error could be explained by two types of perseveration, we analysed the distribution of the nature of the decision made by participants. Results showed that plan continuation errors were committed on 64 of 120 or 53% of decisions and flexible decisions were taken on 56 of 120 or 47% of decisions. On the 64 plan continuation errors, 48 were PPA or 75% and 16 were PRS or 25%, $p(\chi^2) < .05$. When examining the distribution of the two types of perseveration as a function of Threat Expectancy, results showed that when threats are expected 27 plan continuation errors on of 31 or 87% were explained by PPA and 4 of 31 or 13% were explained by PRS. When threats were unexpected by pilots, 21 of 33 or 64% of plan continuation errors were explained by PPA and 12 of 33 or 36% of plan continuation errors were explained by PRS, $p(\chi^2) < .05$. The distribution between flight phases also showed a significant difference between take-off and cruise phases ($p(\chi^2) = .05$) and between cruise and approach phases ($p(\chi^2) < .05$). Indeed, when threats were managed during the cruise phase, none of the plan continuation errors was explained by a PRS, whereas during the take-off phase 3 of 16 or 19% were explained by PRS and during approach phase 13 of 38 or 34% were explained by PRS.

Effects of Success-Related Pressure, Threat Expectation, Flight Phase and their interactions on decision performance were then analyzed with ANOVAs. Performance was significantly influenced by Threat Expectation ($F(1, 18) = 4.69, p < .05$) and by Flight Phase ($F(2, 36) = 55.7, p < .001$) but not significantly influenced by Success-Related Pressure ($F(1, 18) = 2.53, p > .10$). On one hand, performance was better when threats were expected than when they were unexpected and on the other hand, performance was better during take-off and cruise phases than during approach. The effect of Success-Related Pressure was observed in interaction with Threat Expectation ($F(1, 18) = 9.12, p < .05$): when threats were expected, success-related pressure had no significant impact on performance ($F(1, 18) = 0.72, p > .10$) whereas when threats were unexpected pilots under high success-related pressure performed better that pilots under low success-related pressure ($F(1, 18) = 11.27, p < .05$). Interaction between Success-Related Pressure and Flight Phase was not significant nor was interaction between Threat Expectation and Flight Phase.

Information processing

ANOVAs were conducted for the three independent variables on the amount of information accessed, the amount of target information accessed and the time spent on reading target information. A significant effect of Flight Phase was found on the amount of information accessed ($F(2, 36) = 5.6, p < .05$) where the later threats happened during flight, less information was accessed: around 16 data (±1.5) were accessed during take-off phase, around 12 data (±1.4) were accessed during the cruise phase and around 10 data (±1.3) were accessed during the approach phase. No significant effects were found for Success-Related Pressure, Threat Expectation nor for their interactions. ANOVAs conducted on the amount of target information accessed revealed no significant effect for any of the three variables nor for their interactions. The results of the analyses of time spent reading target information showed only one significant effect of the interaction between Threat Expectation and Flight Phase ($F(2, 36) = 6.35, p < .05$). During take-off and approach phases, time spent reading target information was not significantly different as a function of threat expectancy whereas during the cruise phase, pilots spent more time reading target information when threats were unexpected than when they were expected. Relations between indicators of information processing and the nature of the decisions were analysed using Spearman correlations. No significant effect was found for any information processing indicators since all correlations were close to 0.

Decision cues, diagnosis accuracy and goals to achieve

In order to identify which information was taken into account for decision making, pilots had to give a written account, explaining how they made their decisions and what goals they wanted to achieve. All texts were then classified into: decision making cues, accuracy of diagnosis and goals. Of the 120 decisions made during this experiment, target information was mentioned in 60 decisions (50%), while target information was not mentioned in the 60 others (50%). Distribution among pilots experiencing high success-related pressure and those experiencing low success-related pressure showed that 67% of pilots with high success-related pressure mentioned target information whereas only 33% of pilots with low Success-related pressure mentioned it, $p(\chi^2) < .001$. The link between the number of decisions where target information was mentioned with the nature of the decision taken showed that 71% of flexible decisions were explained with target information whereas only 33% of decisions leading to plan continuation errors were explained with target information, $p(\chi^2) < .001$. On the other hand, no significant difference was found between PPA and PRS, where 37% of PPA decisions were explained with target information for 19% of PRS decisions.

The distribution of diagnosis accuracy with nature of decisions showed a significant difference: 78% of flexible decisions were associated with an accurate diagnosis and for 22% of them no diagnosis was expressed while only 45% of decisions leading to plan continuation errors were associated with accurate diagnosis and 22% of them were associated with a wrong diagnosis, $p(\chi^2) < .001$. Furthermore, 50% of PPA decisions were associated with accurate diagnoses and 17% with wrong diagnoses while only 19% of PRS decisions were associated with accurate diagnoses and 50% of them were associated with wrong diagnoses, $p(\chi^2) < .05$.

Goals to be achieved that were mentioned by participants could be classified into 4 categories: 80% were about maintaining the safety level of the aircraft, 13% were about ensuring that passengers could arrive at their destination, 4% were about maintaining the safety level of passengers and/or aircrew, and 3% were about ensuring effective organization of aircraft repair. The distribution among pilots with a high success-related pressure and low success-related pressure revealed a significant difference ($p(\chi^2)$=.05): 90% of goals mentioned by pilots with low success-related pressure were about aircraft safety levels and 5% concerned the assurance that passengers could arrive at destination while they represented respectively 71 % and 20% for pilots with a high success-related pressure.

Discussion

Our first hypothesis was that plan continuation error could be explained by two types of perseveration behaviour: PRS which describes plan continuation error as committed from a wrong representation of the situation and PPA which describes plan continuation error as committed from an accurate representation of the situation but with a risky choice of action. This hypothesis was verified since both types of perseveration were chosen by participants as the right plan of actions to make when faced with flight threats. Moreover, the link with diagnosis accuracy confirmed that a majority of PPA decisions were characterised by accurate diagnosis while a majority of PRS were characterised by wrong diagnosis. Yet, no difference in the amount of target information mentioned as decision cues was found between PPA and PRS. This may be explained by the fact that several participants did not mention any decision cues at all, even though it is probable that they did take into account some information when making their decision. Additionally, PPA decisions were chosen more frequently to counter flight threats than PRS. Plan continuation errors would then seem to be more frequently due to a difficulty with anticipating the risk associated with a plan of action than a difficulty with assessing the current situation. Yet, our results showed that as a function of threat expectancy, the two types of perseveration are distributed differently. A more important part of plan continuation errors are explained by PRS when threats are unexpected than when threats are expected. This confirms our categorisation of plan continuation errors, since PRS characterizes a wrong representation of the situation and when threats are unexpected it takes more cognitive resources to build an accurate representation of the situation than when threats are expected. Hence, the difficulty encountered by pilots when threats are expected is more about choosing a safe plan of action whereas when threats are unexpected the difficulty is more about finding cognitive resources in order to build a correct representation of the situation.

Our second hypothesis was not verified since pilots with high success-related pressure chose more flexible decisions than plan continuation decisions and inversely for pilots with low success-related pressure. The effect on performance was significant when the threats to be managed were unexpected. Hence, when threats are anticipated pilots may have enough cognitive resources to anticipate flexible solutions whatever the type of pressure. On the contrary, when threats are unexpected pressure has an impact on the decision made by pilots. The presence of important authorities on board seemed to push pilots to privilege safety over mission success. This result is at the opposite of those found by Denihan (2007) where naval pilots on combat mission would rather foster mission success over safety. A bias in the experiment may explain this difference: because pilots had important passengers on board, they could have been tempted to show that this had no influence on their decision. Indeed, results showed that pilots with a low success-related pressure wanted to achieve more safety-related goals than pilots with high success-related pressure who were more concerned about ensuring that their passengers could arrive at destination. Yet, pilots with high success-related pressure mentioned more frequently target information directly linked to threats to be managed than pilots with low success-related pressure, which confirms that pilots with high success-related pressure may have built a better representation of the situation which could explain their better decision performance.

Finally, our results confirmed our third hypothesis that the context of the flight, illustrated here by flight phase also has an impact on plan continuation errors. As expected, performance was better during take-off and cruise phases than when threats happened during the approach phase. In this phase, PRS was most chosen while during the cruise phase plan continuation errors were only explained by PPA. This result suggests that when pilots commit plan continuation errors during approach it could be more due to a difficulty in building an accurate representation of the situation than to a deliberate choice of actions. Indeed, our results showed that the frequency of PRS is in accordance with flight phase workload. The heavy workload of the approach phase could hinder pilots in building an accurate representation of the situation leading to a PRS type of plan continuation error. This result is also supported by the fact that it was during the approach phase that pilots accessed the least amount of information. This result meets those found by Muthard and Wickens (2003) who showed the effect of workload on plan continuation errors in the context of the use of automation.

In conclusion, this study confirmed that plan continuation errors can be explained by two types of perseveration behaviours: Perseveration on a wrong Representation of the Situation (PRS) and Perseveration on a risky Plan of Actions (PPA). This distinction is important to make, since recommendations concerning how to recover from them will focus on different aspects such as specific training in simulator for improving information processing or providing techniques to help pilots to better estimate risks associated with a plan of actions. Finally, success-related pressure illustrated here by the presence on board of important passengers seems to improve decision performance. Yet, further research is needed to complete theses findings and eliminate possible bias.

References

Bureau d'Enquêtes et d'Analyses (1997). Objectif : destination. 1991-1996. Le Bourget : Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile. (http://www.bea.aero/etudes/objectifdestination/objectifdestination.pdf).

Denihan, M. B. (2007). Decision making in naval aviation: contextual factors influencing cue and factor saliency. In *Proceedings of the Human Factors and Ergonomics Society 51th Annual Meeting* (pp. 333-337). Baltimore, MD: Human Factors and Ergonomics Society.

Goh, J., & Wiegmann, D.A. (2001). Visual flight rules flight into instrument meteorological conditions: An empirical investigation of the possible causes. *The International Journal of Aviation Psychology, 11,* 359-379.

Muthard, E. K., & Wickens, C.D. (2003). Factors that mediate flight plan monitoring and errors in plan revision: planning under automated and high workload conditions. In *Proceedings of the 12th International Symposium on Aviation Psychology.* Dayton, OH: The Ohio State University.

National Transportation Safety Board (1994). *A review of flightcrew-involved, major accidents of U.S. air carriers, 1978 through 1990* (NTSB/SS--94-01). Springfield, VA: National Technical Information Service.

Orasanu, J., Fisher, U., & Davison, J. (2002). Risk perception: a critical element of aviation safety. In *Proceedings of the 4th International Congress of the International Federation of Automatic Control*. Barcelona, Spain.

Orasanu, J., Martin, L., & Davison, J. (2001). Cognitive and contextual factors in aviation accidents: Decision errors. In E. Salas & G.A. Klein (Eds.), *Linking expertise and naturalistic decision making* (pp. 209-225). Mahwah, NJ: Lawrence Erlbaum.

Prince, C., & Salas, E. (1993). Training and research for teamwork in the military aircrew. In E.L. Wiener, B.G. Kanki & R.L. Helmreich (Eds.), *Cockpit Resource Management* (pp. 337-366). San Diego, CA: Academic Press.

Sicard, B., Taillemite, J.P., Jouve, E., & Blin, O. (2003). Risk propensity in commercial and military pilots. *Aviation, Space, and Environmental Medicine, 74*, 879-881.

Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment anf flight experience in pilots'decisions to continue visual flight rules flight into adverse weather. *Human Factors, 44*, 189-197.

Wiegmann, D. A., & Shappell, S.A. (1997). Human factors analysis of postaccident data: Applying theoretical taxonomies of human error. *The International Journal of Aviation Psychology, 7,* 67-81.

Acknowledgements

The authors wish to thanks Jean-Claude Ardouin for programming the experiment. This work was supported by DCSSA.

DECISION FACTORS INFLUENCING STIMULANT USE AMONG FIGHTER AIRCREW DURING COMBAT OPERATIONS

Russell K. Gore M.D.
U.S. Air Force Research Laboratory
Dayton, OH
Eric D. Hermes M.D.
Yale University Department of Psychiatry
New Haven, CT

During long combat missions in fighter aircraft, passive in-flight fatigue countermeasures are often not feasible. As a result, stimulant medications (Go Pills) may be used in-flight. The present study attempts to describe the individual decision factors influencing stimulant use during combat operations. *Methods:* 35 deployed F-15E aircrews participated in this study. Prior to the deployment, interviews were conducted to identify factors influencing the in-flight decision to use stimulants. Based on this qualitative information, a novel survey instrument was developed. *Results:* Surveys were completed after 111 sorties. Results were summarized graphically. *Conclusions:* Active and anticipated in-flight fatigue were the most common decision factors across all groups. Leadership influence and perceived repercussions were the least influential. Previous Go Pill experiences and in-flight performance were more influential among sorties using stimulants ($p<0.001$). There were no notible differences in decision factors across deployment experience.

During continuous operations, like those underway in the current combat theater, fatigue represents a significant concern among military aircrews. In a recent survey, 74% of US Air Force aircrews reported flying when drowsy enough to fall asleep (Tan, 2006). Specifically among fighter aircrews, counteracting fatigue is a continuous challenge. During long combat missions, fighter aircrews perform complex physical, cognitive, and emotional tasks without the ability to use passive in-flight fatigue countermeasures. These aircrews, flying in single-piloted tactical aircraft, cannot depend on in-flight napping, activity breaks, or increased cockpit lighting to counteract fatigue (J. A. Caldwell et al., 2009). Often, when passive countermeasures are not feasible, stimulant medications (Go Pills) are used to improve in-flight vigilance.

The use of stimulants is highly regulated and only authorized “after all other fatigue management tools have been exhausted” (Murray, USAF Policy Letter, 2001). Nevertheless, stimulant use in combat is commonplace with 60-65% of fighter and bomber aircrews reporting stimulant use at least once during combat deployments (Emonson & Vanderbeek, 1995; Kenagy, Bird, Webber, & Fischer, 2004). Prior to combat deployments aircrews are required to ground test stimulant medications and attend informal training about stimulant use from the squadron flight surgeon. The authors conducting this study served as fighter squadron flight surgeons and frequently provided this training for combat aircrews. During these training sessions, many aircrews deploying for the first time were noted to ask “when should I take the Go Pill during a combat sortie?”

Many studies have investigated stimulant use in controlled research environments (Bower & Phelan, 2003; J. Caldwell, Caldwell, JL, Darlington, KK, 2003; John A. Caldwell, Caldwell, Smith, & Brown, 2004). However, the decision to use stimulants in these studies is generally controlled as part of the study protocol. In combat, the individual aircrew decision to use stimulants is based on a complex series of in-flight considerations. Military regulations do not specify criteria for in-flight stimulant use and operational fatigue studies addressing this question are few. One study evaluating stimulant use in fighter pilots during the initial phase of

Operation Desert Storm reported that aircrews were instructed to use stimulants "30 min before critical stages of flight if they felt unduly fatigued" (Emonson & Vanderbeek, 1995). Another study of fighter pilot fatigue countermeasures recommends preflight planning of stimulant use in order to avoid a "real-time, fatigue-impaired decision about go-pill use" (Schultz & Miller, 2004). Given the lack of specific guidance and the variability of advice proffered in the operational literature, the present study attempts to describe the aircrew decision to use stimulants in-flight during combat operations.

Methods

This study sought to investigate the complex decision to use or not use stimulants during combat operations. Approval for the project was obtained from the Wright-Site Institutional Review Board (IRB) prior to gathering data.

Study Design

In 2006, an F-15E fighter squadron deployed to a forward operating location in Southwest Asia. During this combat deployment, F15E crews consisting of a pilot and a weapons systems officer (WSO) flew regular combat missions over Iraq and Afghanistan. The use of stimulants during this deployment was authorized in accordance with USAF policy. Aircrews were allowed to consume either five or ten milligrams (mg) of Dexedrine every four hours or 200 mg of Modafinil every eight hours. During the study period, participants were encouraged to complete a post flight survey as frequently as possible after each combat mission. The decision to use or not use in-flight stimulants was assessed using the novel survey instrument described below.

Survey Instrument

Prior to the deployment, detailed interviews were conducted with six experienced F-15E aircrew in order to develop a novel survey instrument. These interviews were conducted to refine our understanding of the factors influencing the individual decision to use in-flight stimulants during combat operations. Based on the qualitative information gathered, we identified 15 primary categories of influence including previous Go Pill experiences, active in-flight fatigue, anticipated in-flight fatigue, preflight fatigue, habit patterns, personal convictions, planned sortie profile, Go Pill availability, crewmate influence, in-flight workload, in-flight performance, perceived repercussions, flight leadership influence, post flight medication effects and command influence. For each category, descriptive statements were developed based on the initial aircrew interviews. Using appropriate descriptive statements, parallel surveys were developed for sorties using stimulants and for sorties not using stimulants. Figure 1 is an example illustrating the questions contained in these parallel surveys. After landing, participants selected the appropriate survey and reported on a visual analog scale the level to which each category influenced their decision to use or not to use stimulants during the sortie.

Data Analysis

Survey responses were compared across stimulant use and aviator combat experience using univariate measures of analysis. Results were summarized graphically based on the mean visual analog scale response. For the purposes of this study, a "sortie" was defined as each individual aircrew survey completed. Although some aircrews completed multiple surveys, we included all 111 completed surveys in our analysis under the assumption of independence.

A.

For the following statements, rate the extent to which each factor influenced your <u>decision to USE the Go-pill</u> during this sortie.

1. **In-flight Active fatigue**–I felt tired/sleepy/sluggish or I was having difficulty staying alert so I decided to take the Go-pill.

DID NOT Influence My decision 0 1 2 3 4 5 6 7 8 9 10 **SIGNIFICANTLY Influenced My decision**

B.

For the following statements, rate the extent to which each factor influenced your <u>decision to NOT USE the Go-pill</u> during this sortie.

1. **In-flight Active fatigue**–I felt rested and alert so I did not need to take the Go-pill.

DID NOT Influence My decision 0 1 2 3 4 5 6 7 8 9 10 **SIGNIFICANTLY Influenced My decision**

Figure 1. Example of survey instrument questions portraying the "In-flight Active Fatigue" category for sorties in which stimulants were used (A) and sorties in which stimulants were not used (B).

Results

Survey Population

The survey population consisted of 35 aircrews, 17 pilots and 18 WSOs, with a mean age of 30 ± 4 yr (range 25 to 41). There were 16 participants with previous combat experience and 19 deploying to combat for the first time. Among the 35 participants, 29 (82.9%) completed a survey after at least one sortie and 18 participants (51.4%) completed surveys after more than one sortie (range 2 to 14 surveys completed). Of the 111 sorties surveyed, the mean sortie duration was 7.6 hrs (range 3.5 to 10.5) and stimulants were used during 39 of the sorties (35.1%).

Survey Results

Figure 2 summarizes the survey results compared across stimulant use displayed in the order of influence for sorties using stimulants. The active and anticipated in-flight fatigue categories were strong decision factors for both groups. Stimulant users report that their decision was more influenced by previous Go Pill experiences ($p<0.001$) and in-flight performance ($p<0.001$). Sorties not using stimulants reported more influence for the preflight fatigue category ($p=0.002$). There were no other notable differences between these groups.

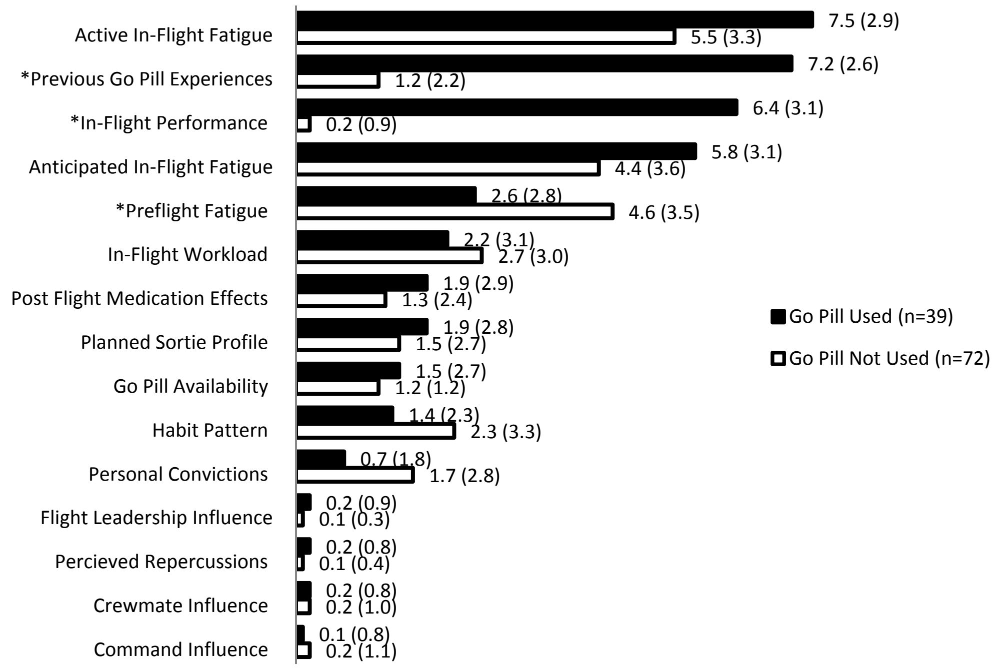


Figure 2. Mean survey responses [avg (SD)] comparing sorties using stimulants and sorties not using stimulants. * Tests of statistical significance for univariate differences between sorties using stimulants and not using stimulants were based on Analysis of Variance. Categories showing significance at $p<0.05$ were Previous Go Pill Experiences, In-Flight Performance and Preflight Fatigue.

Figure 3 summarizes the survey results compared across deployment experience displayed in order of influence for aircrew with deployment experience. Again, active and anticipated in-flight fatigue were the most influential categories for both groups. There were no significant differences between experienced combat aviators and those deployed for the first time. It is notable that the categories reported as the least influential across all groups were command, flight leadership, and crewmate influence as well as perceived repercussions.

Discussion

Among fighter aircrews engaged in combat, the decision to use in-flight stimulants was primarily influenced by preflight and in-flight fatigue as well as in-flight performance decrements. Fighter aircrews were not preplanning stimulant use based mission type and anticipated sortie duration (sortie profile) or personal habit patterns. These results indicate that the decision to use stimulants was in line with the guidance prescribed by Emonson and Vanderbeek, suggesting that aircrews use stimulants if they experience excessive in-flight fatigue (Emonson & Vanderbeek, 1995). Although anticipated fatigue was statistically as strong as active fatigue in this analysis, both of these factors involve an in-flight assessment of fatigue rather than a preflight decision.

The decision to use in-flight stimulants was not influenced by aircrew experience. This result indicates that aircrew strategies for in-flight stimulant use do not change with combat experience. Due to the lack of formal guidance, this suggests that these strategies are either intuitive or communicated informally to new aircrews through observation. Additionally, aircrews appear to be satisfied with these decision priorities so they do not change with more experience. This study did not include a measure of aircrew performance to specifically evaluate the benefits of different stimulant use strategies. Additional studies evaluating performance may reveal strategies, or decision category priorities, that improve combat performance.

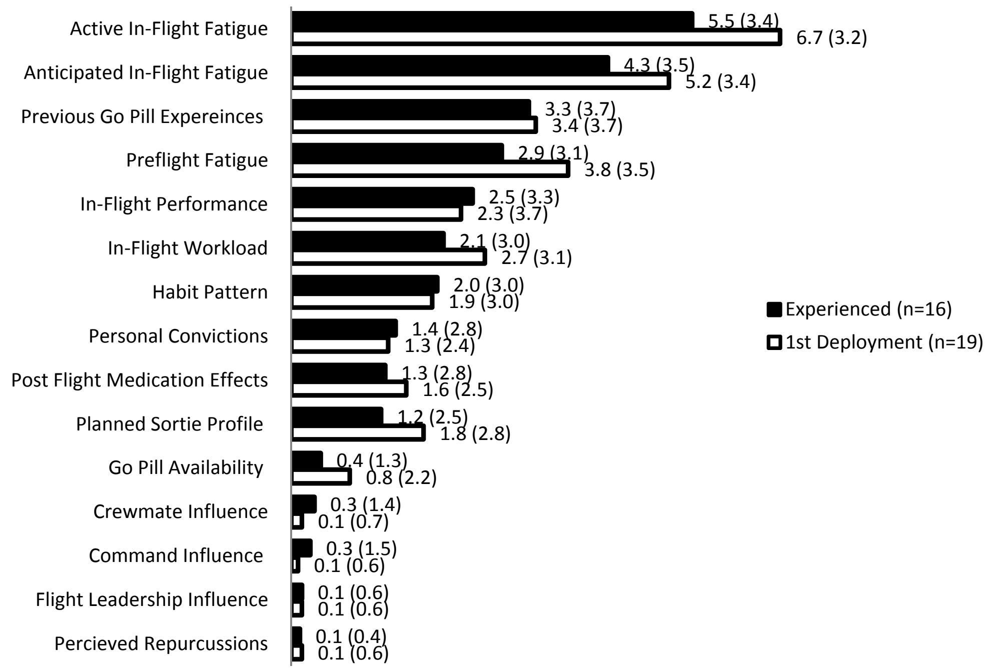


Figure 3. Mean survey responses [avg (SD)] comparing survey responses from aircrews with previous combat experience and aircrews deployed to combat for the first time. Tests of statistical significance for univariate differences were based on Analysis of Variance. No significant differences were found.

Aircrews made the decision to use in-flight stimulants with minimal influence of squadron leadership and minimal concern for post flight repercussions, allowing them to prioritize other decision factors. This finding was consistent across stimulant use and combat experience. Similarly, aircrews in other combat studies have reported minimal "pressure" to use stimulants during long duration missions (Kenagy et al., 2004). These findings contradict media reports suggesting that aviators are occasionally coerced into stimulant use by commanders (Halbfinger, 2003). These findings also contradict the general perception within the fighter community that commanders discourage stimulant use in combat. During this deployment, the lack of leadership influence likely results from local command policies regarding stimulant use. Specifically, stimulant use was approved in advance for the duration of the deployment, there were no command directed limitations, and stimulant medications were readily available.

Acknowledgements

The views expressed in this article are those of the authors and do not necessarily reflect the official views of the U.S. Air Force or the Department of Defense.

References

Bower, E. A., & Phelan, J. R. (2003). Use of amphetamines in the military environment. *Lancet, 362 Suppl*, s18-19.

Caldwell, J. A., Caldwell, J.L., Darlington, KK. (2003). Utility of Dextroamphetamine for Attenuating the Impact of Sleep Deprivation in Pilots *Aviation Space & Environmental Medicine, 74*(11), 1125-1134.

Caldwell, J. A., Caldwell, J. L., Smith, J. K., & Brown, D. L. (2004). Modafinil's effects on simulator performance and mood in pilots during 37 h without sleep. *Aviation Space & Environmental Medicine, 75*(9), 777-784.

Caldwell, J. A., Mallis, M. M., Caldwell, J. L., Paul, M. A., Miller, J. C., & Neri, D. F. (2009). Fatigue countermeasures in aviation. *Aviation Space & Environmental Medicine, 80*(1), 29-59.

Emonson, D. L., & Vanderbeek, R. D. (1995). The use of amphetamines in U.S. Air Force tactical operations during Desert Shield and Storm. *Aviation Space & Environmental Medicine, 66*(3), 260-263.

Halbfinger, D. M. (2003, 15 January). Hearing starts in bombing error that killed 4. *New York Times*.

Kenagy, D. N., Bird, C. T., Webber, C. M., & Fischer, J. R. (2004). Dextroamphetamine use during B-2 combat missions. *Aviation Space & Environmental Medicine, 75*(5), 381-386.

Schultz, D., & Miller, J. (2004). Fatigue and Use of Go/No-Go Pills in Extraordinarily Long Combat Sorties. *Aviation Space & Environmental Medicine, 75*(4), 370-371.

Tan, S. (2006). A Survey of US Air Force Aviation Personnel on Flying, Fatigue and Fatigue Countermeasures. *Aviation Space & Environmental Medicine, 77*(3), 293.

TOWARDS A FOUR-DIMENSIONAL SEPARATION ASSISTANCE COCKPIT DISPLAY

J. Ellerbroek, M. Visser, S. B. J. van Dam, M. Mulder, M. M. van Paassen
Faculty of Aerospace Engineering, Delft University of Technology,
Kluyverweg 1, 2629 HS Delft, The Netherlands.

An initial design of a tactical navigation support tool is proposed, designed to integrate horizontal and vertical separation assistance tools into one display. A novel representation of the separation problem, based on Ecological Interface Design, presents external conflict and performance constraints on an extended, wide-screen Primary Flight Display. Key issues in the current design are discussed, and an experiment is proposed to evaluate the display concept.

In the current airspace environment, congestion problems are expected in the near future, due to rapidly increasing amounts of traffic. Because of the rigid nature of the airspace, which is divided in fixed volumes and route structures, this growth will result in higher workload for air traffic controllers, and reduced efficiency of trajectories. New concepts for Air Traffic Management, such as SESAR, permit a flexible use of airspace, with airborne determination of user preferred trajectories (RTCA, 2002a; SESAR Consortium, 2007). This flexible use is expected to increase airspace capacity, and reduce air traffic controller workload. However, because the separation task is shifted from the air traffic controller to the pilot, it is expected that the pilot needs to be assisted in this task.

Traditional systems, such as Predictive Airborne Separation Assurance Systems (P-ASAS) (Hoekstra, 2001), have been developed to assist pilots in their task of self-separation. Generally such systems support the pilot by presenting a limited set of explicit, ’ready-to-use’, avoidance maneuvers as a solution to a separation conflict. Such automated systems have proven to be effective in terms of conflict resolution and workload reduction, but they limit the pilot in exploring other solutions, and therefore, may prohibit full exploitation of the travel freedom offered by the airspace environment. Also, in a complex traffic environment, non-routine situations may arise, that may not have been foreseen in the automation design. In these exceptional cases, the pilot’s ability to improvise is vital for successful conflict resolution. It is therefore of key importance that automation and instrumentation promote a high level of situation awareness.

At Delft University of Technology, extensive research is being performed on ecological interface design of Airborne Separation *Assistance* Systems: displays designed to visualize the affordances the airspace provides. These displays assist the pilot in their task of self-separation, without relying on resolutions provided by automation. Previously, several concept displays have been developed, for separation assistance in the horizontal plane, as well as the vertical plane (Heylen, van Dam, Mulder, & van Paassen, 2008; van Dam, Mulder, & van Paassen, 2008). Although these displays successfully support pilot decision-making in the task of self-separation, they still map the essentially four-dimensional problem (space and time) onto two displays. This article presents the initial iteration of design of a novel four-dimensional Separation Assistance Interface (referred to as 4D-SAI).

Ecological Approach

Ecological Interface Design (EID) is a design paradigm that originates from the domain of process control. It addresses the cognitive interaction between humans and complex socio-technical systems. Its approach to interface design gives priority to the workers environment (termed ’ecology’), focusing on how the environment poses constraints on the worker (Vicente & Rasmussen, 1992). Rather than taking the worker’s cognitive capabilities as a starting point, EID tries to identify what elements in the environment shape the operator’s behavior: The interface should reveal the possibilities and constraints afforded by the work domain. In other words, EID promises a more systematic approach to unambiguously define ’what is the situation’ the pilot should be ’aware of’ (Flach, Mulder, & van Paassen, 2004). By focusing on the affordances and constraints posed by the work domain, the worker can be supported in actions that go beyond the worker’s anticipated tasks.

EID consists of two steps. The first step consists of determining the goal-relevant properties of the work domain (i.e., what to display), and the second step addresses the actual interface presentation (i.e., how to display). In the first step, a workspace analysis tries to identify functionalities, constraints, and means-end relationships withing the work domain. The main tools for this analysis are the Abstraction Hierarchy (AH), and the Skills, Rules, Knowledge taxonomy (SRK), both developed by Rasmussen (Rasmussen, 1983, 1985). Following the workspace analysis, EID aims to visualize the constraints and means-end relationships in the environment in such a way, as to fully take advantage of the human capacity to directly perceive, and act upon cues from the environment.

Work Domain Analysis

The first step of ecological interface design consists of a workspace analysis, using Rasmussen's Abstraction Hierarchy (Rasmussen, 1985). The abstraction hierarchy is a stratified hierarchical description of the workspace, defined by means-end relationships between the adjacent levels, see Figure 1. Along the vertical axis, the five levels of the AH represent the constraints at decreasing levels of abstraction, starting at the top with the purpose(s) for which the system was designed, all the way down to the spatial topology and appearance of the components that make up the system on the bottom level (Rasmussen, 1985; Bisantz & Vicente, 1994). Along the horizontal axis, components and constraints are arranged from internal elements on the left, to external elements on the right.

At the functional purpose level, the goals of the system are defined, which in the case of flight in general, are flying safely, productively, comfortably and efficiently through unmanaged airspace. Aside from issues such as staying within the flight envelope, safety in aircraft locomotion is assured by maintaining sufficient separation from potentially hazardous objects, such as other aircraft and terrain. In case of the ASAS self-separation application (FAA-Eurocontrol, 2001), this means adhering to the defined separation minima between aircraft. Although more complex in reality, in this paper it is assumed that work is productive, when the distance to the destination is continuously decreasing. For flight in general, comfort poses constraints such as upper limits on maneuver accelerations. The realization of efficiency is much more complicated however, as it depends not only on fuel efficiency, but also on time and position constraints with respect to a flight schedule.

The abstract function level describes the underlying causal relationships that govern the realization of the purpose of the system. In the case of air travel, this level contains the general physical laws that dictate absolute and relative locomotion, and separation (van Paassen, Amelink, Borst, van Dam, & Mulder, 2007).

The general function level describes how the functions at the abstract function level are achieved, independent of the actual implementation of the system. Properties such as weight, lift, thrust and drag, and the maneuvering performance of the aircraft all impose internal constraints on aircraft behavior. External obstructions further constrain aircraft motion, and dictate the (lack of) separation. On the bottom of the abstraction hierarchy, the physical form and functions are described by modeling the internal layout of aircraft components, and external airspace properties such as other traffic, weather, and terrain. The physical function level describes the various components, and their capabilities, and at the physical form level the appearance and location of components, the airspace, and other aircraft are described.

In this paper, the workspace content and boundaries are limited to trajectory planning functions in direct relation with conflict resolution and prevention during cruise flight and in situations with multiple aircraft. Functions related to aircraft control and stability, like staying within the flight envelope and accounting for passenger comfort, are kept out of the analysis. The time interval in which this workspace is analyzed is determined by the applicability of conflict management and is more or less situated between 60 seconds and around 15 minutes. Below 60 seconds, collision avoidance systems like the TCAS II must take over in order to prevent collision (RTCA, 2002b). A 15 minute upper threshold is chosen because the vast majority of conflict resolution and recovery maneuvers take place in less than 15 minutes.

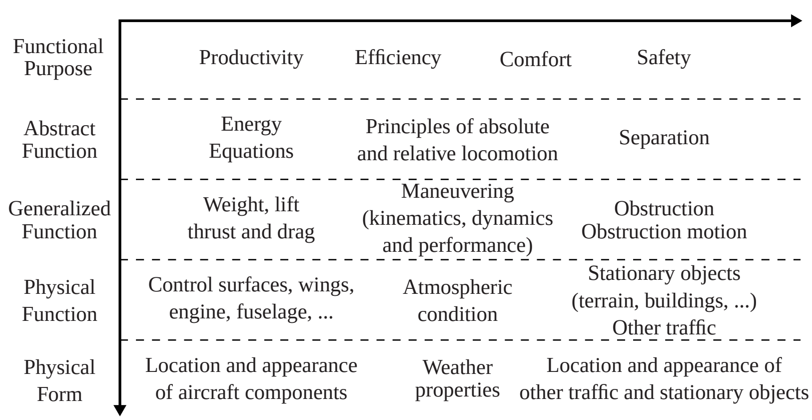


Figure 1: Abstraction Hierarchy for the Separation Assistance Display.

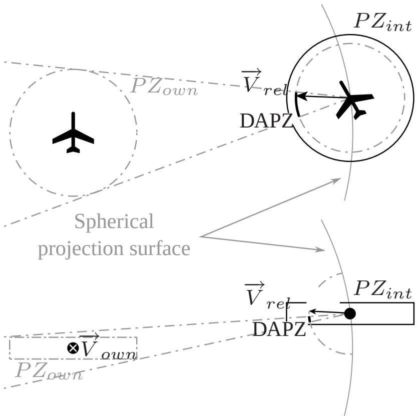


Figure 2: The Danger Area Protected Zone.

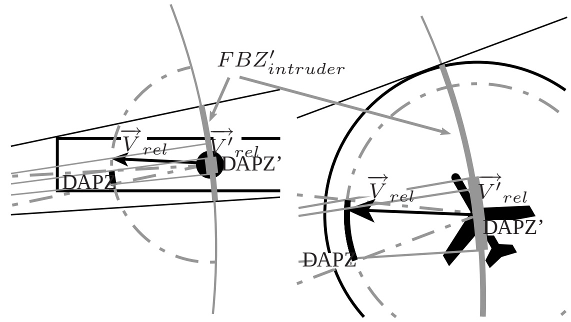


Figure 3: Close-up of the projection geometry for the DAPZ and the FBZ.

Functional Modeling of Aircraft Behavior and Separation

Based on the ecological interface design concept (Vicente & Rasmussen, 1992), the translation of the work domain analysis into an interface design is done through Functional Modeling, which tries to formulate the behavior of a system relevant to achieving its ends. For trajectory planning this implies that the goal relevant affordances must be visualized such, that the pilot's perception of these cues directly triggers desired goal-relevant steering actions.

The visualization of the external constraints in the first concept is based on relative speed, similar to the earlier display designs. There is an important difference, however: While the X-ATP and VSAD display visualizations were based on the ownship velocity relative to the intruder, the present design considers the opposite.

An often heard comment from pilots, in the evaluation of the previous display designs, was that while it featured as a valid and equal option in both displays, velocity changes are rarely used when resolving a conflict. Based on this feedback, the present design uses a cutting plane based on constant velocity to project the 3D situation onto a 2D display. The presentation of conflicts to the pilot is realized by a projection of the separation problem on the surface of an imaginary sphere, with its radius equal to the distance from ownship to intruder (Figure 2). When drawing lines between the borders of the ownship protected zone (PZ_{own}) and the intruder aircraft, a three-dimensional shape is obtained, similar to the Forbidden Beam Zone-concepts developed in the previous designs.

A second sphere is drawn, with origin at the intruder aircraft and with radius equal to the intruder relative speed. The intersection of the three-dimensional FBZ and this sphere is called the "Danger Area Protected Zone" (DAPZ). It represents all velocities with equal magnitude of the intruder relative to ownship that correspond with possible future loss of separation. Both the FBZ as well as the DAPZ can be projected on the imaginary projection sphere introduced above (Figure 3), resulting in a shape that will be referred to as "the puck" (Figure 4). The word "puck" is chosen as the PZ resembles a flat disc, similar to a puck used in icehockey. The curvature of the projection is caused by the circular shape of the puck, and changes as a function of the vertical position of the intruder, relative to the ownship. When the intruder is at the same altitude as the ownship the projection will be rectangular.

Within the puck, the relative speed of the intruder is shown. Clearly, when the tip of this relative velocity vector is located outside the DAPZ, separation is guaranteed. To better indicate the position of the tip of the relative velocity vector, four lines are drawn from the boundaries of the puck towards the velocity vector tip, see Figure 4.

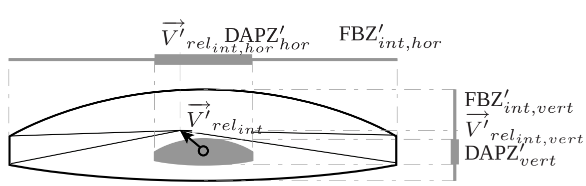
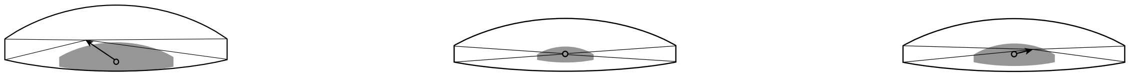


Figure 4: Construction of the 'Puck'.



(a) A grown puck (b) Collision course (c) The puck in a conflict

Figure 5: Some examples of the 'Puck'.

Figure 5 shows what the puck may look like, for three different situations. In Figure 5(a) the DAPZ has grown, indicating that the probability of a loss of separation has become larger. Note that the puck would have grown too in size on what is essentially a three-dimensional perspective projection. From the location of the tip of the velocity vector we can see, however, that no loss of separation will actually occur in this situation, as it is located outside of the DAPZ. We can also see that the intruder aircraft moves upward and to the left, relative to ownship. In Figure 5(b) the relative velocity vector is such that it points directly at ownship, and therefore is located in the center of the DAPZ. This means that in this situation a collision will occur, if no further action is taken. In Figure 5(c) a situation is shown where the relative velocity vector is still inside the DAPZ, indicating a future loss of separation.

The puck shows the relative speed of the intruder, i.e., its relative movement, the urgency of the potential conflict, and the area in which the relative speed vector should not be positioned. It does not show, however, and this is crucial, what the pilot of ownship can *do* to keep the relative velocity outside of the DAPZ. In the current concept, this is one of the main challenges. As the current design is likely to contain perspective elements (using the projection sphere centered around ownship), the “visual angle” design principle, also successfully applied in ecological synthetic vision overlays, was adopted (Borst, Suijkerbuijk, Mulder, & van Paassen, 2006).

Figure 6 shows the problem in 3D perspective. The situation is similar to the previous horizontal/vertical projections, except that now the FBZ is not a two-dimensional wedge, but rather its three-dimensional counterpart. Similar to the transformations applied in the design of the X-ATP display, the constraints on ownship travel can be visualized by a translation of the 3D FBZ with the intruder velocity, resulting in Figure 7. The intersection of a sphere with radius equal to the ownship velocity with the 3D FBZ yields the so-called "Flight-path vector Avoidance Zone" (FAZ). This shape shows the constraints imposed by intruder motion on the ownship flight-path vector, for the current speed of ownship. Future design iterations will investigate how to visualize the effects of changes in ownship velocity. The next step is then to project the FAZ on the perspective projection sphere that is also used for presenting the puck. This is shown in Figure 7. Note that the current derivation of the DAPZ and FAZ assumes instant state changes. It can be shown that this is a safe assumption when a predicted conflict is still in the far future. However, maneuver dynamics will start to play a larger role when conflicts become more imminent: in the case of tactical maneuvers (within 10 minutes of a predicted conflict), unmodeled dynamics will cause significant errors, particularly speed maneuvers (Paielli, 2003; van Dam, Mulder, & van Paassen, 2007). To compensate for such inaccuracies, future iterations of the 4D-SAI will use maneuver dynamics in the presentation of airspace affordances.

Interface design

For the first design prototype of the separation assurance interface, the visual components introduced in the previous section will be presented on a wide-angle Primary Flight Display (PFD), with a heading range of $\pm 180°$, see Figure 8. Clearly, to visualize the separation assistance information regarding all intruder aircraft located within time-vicinity (e.g., 5 minutes), several different options are available. Although the current implementation uses a "omni camera"-like heading presentation on the PFD, alternatives will be considered as well in future designs.

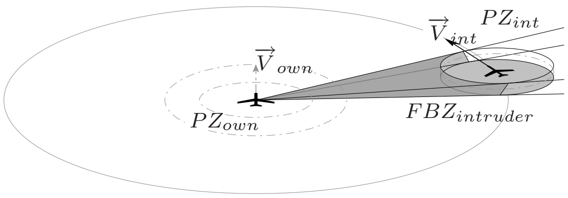


Figure 6: Forbidden beam zone of the intruder.

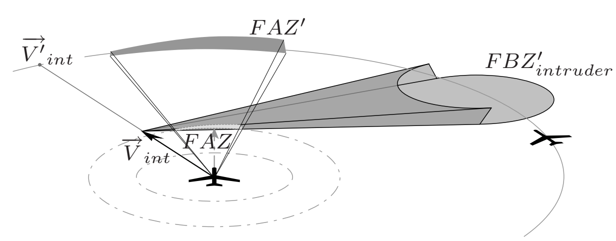


Figure 7: Flight-path vector Avoidance Zone (FAZ)

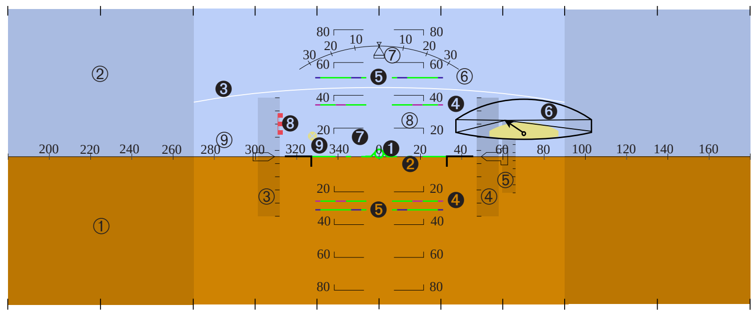


Figure 8: The initial interface design of the 4DSAI showing the example situation

Returning to Figure 8, the numbers indicate the various features of the 4DSAI prototype. Traditional pieces of information are shown using the transparent circles, examples are earth ①, and sky ②, speed ③, altitude ④, bank and slip ⑥/⑦, etcetera. On the horizon the compass headings are shown ⑨. On the speed tape the minimum and maximum velocities can be shown. The altitude and ROC tapes can present similar constraints to the ownship motion. The 4DSAI-related components are shown in the black circles. First, the flight-path vector ❶ shows the current direction of flight. The energy angle is shown as well ❷, i.e., the flight-path the pilot can select to realize a steady climb or descent. The white curved line ❸ shows the maximum flight path angle that can be achieved in a combing turn (max. g-level of 1.4, i.e., maximum bank 45 degrees). The fastest climb (or descent) is shown as the green line with purple stripe ❹, the steepest climb (or descent) is shown as a green line with blue stripe ❺.

Conflicts are shown using the puck ❻; conflicts are only shown when they are predicted to occur within 5 minutes. The small circle in the center of the puck represents the location where the intruder is located. The arrow and its four lines indicate the direction and (projected!) magnitude of the relative velocity of the intruder. When the lines are present the intruder is moving towards ownship, when they are absent the intruder is moving away from ownship. The size of the puck depends on the distance to the ownship (smaller is further away). The DAPZ is the shaded area in the puck and represents the area where the tip of the relative velocity vector should not be located.

The area where the ownship flight-path vector should not be positioned, the FAZ, is shown as well ❼. Note that the FAZ only holds for the current speed. The shading of the FAZ depends on the conflict urgency, from yellow to red. Because the conflict(s) may also be resolved by ownship speed changes, the speeds that are to be avoided are shown as well, on the speed tape ❽. The yellow dot with the cross ❾ gives an indication of the velocity vector of the intruder. Deciding to resolve the conflict by moving the ownship fligh-path vector to this dot will result in a very inefficient resolution, as the ownship will then fly more or less parallel to the intruder (van Dam et al., 2008).

Conclusions

The design of a separation assistance display described in this paper was motivated by the fact that the earlier designs map an essentially four-dimensional problem onto two displays. Using Vicente's Ecological Interface Design paradigm, a first attempt was made with the design of a four-dimensional Separation Assistance Interface. The initial design, presented in this paper, uses a spherical projection of the separation conflict based on a constant velocity. The resulting elements, a flight-path avoidance zone, and a projection of the intruder aircraft Protected Zone, are presented to the pilot on a modified, wide-screen Primary Flight Display. The most important issues in the current design are the method of presenting situations where the conflicting intruder comes from behind the ownship, and the fact that the inside-out presentation of a PFD causes a varying field of view. This means that the separation assistance elements on the display are non-stationary, possibly making interpretation of an impending conflict more difficult. These issues will be adressed in an upcoming evaluation experiment.

Acknowledgements

This work has been co-financed by the European Organisation 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

Bisantz, A. M., & Vicente, K. J. (1994). Making the Abstraction Hierarchy Concrete. *Int. J. Human-Computer Studies*, *40*, 83–117.

Borst, C., Suijkerbuijk, H. C. H., Mulder, M., & van Paassen, M. M. (2006, October). Ecological Interface Design for Terrain Awareness. *International Journal of Aviation Psychology*, *16*(4), 375–400.

FAA-Eurocontrol. (2001). *Principles of Operation for the Use of Airborne Separation Assurance Systems* (Tech. Rep. No. PO-ASAS-V7.1). Federal Aviation Authorities - Eurocontrol.

Flach, J. M., Mulder, M., & van Paassen, M. M. (2004). The Concept of the Situation in Psychology. In S. Banburry & S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application* (pp. 42–60). Ashgate.

Heylen, F. M., van Dam, S. B. J., Mulder, M., & van Paassen, M. M. (2008, August). Design and Evaluation of a Vertical Separation Assistance Display. In *AIAA Guidance, Navigation and Control Conference and Exhibit.*

Hoekstra, J. M. (2001). *Designing for Safety: The Free Flight Air Traffic Management Concept*. Unpublished doctoral dissertation, Delft University of Technology, The Netherlands.

Paielli, R. A. (2003, May). Modeling Maneuver Dynamics in Air Traffic Conflict Resolution. *Journal of Guidance Control and Dynamics*, *26*(3), 407–415.

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.

Rasmussen, J. (1985). The Role of Hierarchical Knowledge Representation in Decision Making and System Management. *IEEE Transactions on Systems, Man, and Cybernetics*, *15*(2), 234–243.

RTCA. (2002a, March). *Airborne Conflict Management: Application Description V2.5* (Tech. Rep. No. RTCA SC-186). Federal Aviation Authorities.

RTCA. (2002b). *Minimal Operational Performance Standards for Traffic Alert and Collision Avoidance System 2 (TCAS2) Airborne Equipement* (Tech. Rep.). Federal Aviation Authorities.

SESAR Consortium. (2007, september). *SESAR Definition Phase D3: The ATM Target Concept* (Tech. Rep. No. DLM-0612-001-02-00). Eurocontrol.

van Dam, S. B. J., Mulder, M., & van Paassen, M. M. (2007, October). Airborne Self-Separation Display with Turn Dynamics and Inturder Intent-Information. In *2007 IEEE International Conference on Systems, Man and Cybernetics.* Montreal, Canada.

van Dam, S. B. J., Mulder, M., & van Paassen, M. M. (2008, November). 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.

van Paassen, M. M., Amelink, M. H. J., Borst, C., van Dam, S. B. J., & Mulder, M. (2007, April). The Chicken, the Egg, the Workspace Analysis, and the Ecological Interface. In *2007 ISAP International Symposium on Aviation Psychology.* Dayton, USA.

Vicente, K. J., & Rasmussen, J. (1992, August). Ecological Interface Design: Theoretical Foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, *22*(4), 589–606.

UTILITY AND RECOGNITION OF LINES AND LINEAR PATTERNS ON ELECTRONIC DISPLAYS DEPICTING AERONAUTICAL CHARTING INFORMATION

Divya C. Chandra
United States Department of Transportation
John A. Volpe National Transportation Systems Center
Cambridge, MA

A study was conducted to explore the utility and recognition of lines and linear patterns on electronic displays depicting aeronautical charting information, such as electronic charts and moving map displays. The goal of this research is to support the development of more standardized and consistent lines and linear patterns for these displays. Data were collected from 273 professional and private pilots. First pilots sorted the names of 65 types of lines and linear patterns in terms of utility of the item. Next they tried to identify nine test linear patterns shown in isolation. Results of the sorting task indicated that the most broadly useful items are controlled and special use airspace. Pilots had difficulty identifying the test patterns, but some patterns were better recognized than others. Results for both tasks varied based on pilot background, such as whether the pilots were qualified for instrument operations or visual operations only.

Current standards and recommendations for electronic aeronautical symbols are documented in the 1997 Aerospace Recommended Practice (ARP) 5289 issued by the Aeronautical Charting Committee within the SAE International Aerospace Behavioral Engineering Technology Committee (SAE G-10). This document contains recommendations for symbols that are primarily shown on charts used during operations under Instrument Flight Rules (IFR), such as instrument approach plates, arrival and departure terminal charts, and enroute charts, although some of the symbols are also found on charts for use under Visual Flight Rules (VFR). Line styles are also recommended in ARP 5289 (e.g., for the missed approach track and airspace boundaries), and there are some general suggestions on using lines of different weights (heavy, medium, and light).

The authors of ARP 5289 (SAE, 1997) expected the recommended symbols to be recognizable by qualified pilots. They also expected that the symbols were simple shapes that could be drawn on the current display technology. Unfortunately, Yeh and Chandra (2005) found that pilots did not recognize some of the recommended symbols. In interviews with manufacturers of electronic moving map displays, it became clear that some of the proposed symbols were difficult to draw on existing displays, so manufacturers were developing their own symbols. Another possible explanation for the lack of standardization in current displays is that manufacturers were not aware of the guidance in ARP 5289 because it was not invoked by a regulatory authority such as the Federal Aviation Administration (FAA). In any case, some of the recommended symbols in ARP 5289 are not in widespread use.

In order to support the development of more standardized symbols, lines, and linear patterns for electronic aeronautical displays, the SAE G-10 Aeronautical Charting Committee is updating ARP 5289; the reissued document will be ARP 5289A. The John A. Volpe National Transportation Systems Center (Volpe Center) is working with this industry committee with funding and support from the FAA. Past research conducted by the Volpe Center in support of recommendations for electronic symbology is documented in various reports and papers (Yeh and Chandra, 2005; Yeh and Chandra, 2006; Chandra and Yeh, 2007; Chandra, Yeh, and Donovan, 2007). The earlier studies focus on pilot recognition and identification of navigation aids, while the later studies also address other symbols (e.g., obstructions and markers) and explore broader issues (e.g., line style conventions and classification of symbols into groups).

The Volpe Center's latest task is to provide objective data upon which to base decisions about what lines and linear patterns should have specific recommendations in SAE ARP 5289A and what those recommendations should be. Therefore the purpose of this study is to understand what lines and linear patterns are most useful to pilots, and to understand which, if any, linear patterns are currently well recognized. A more comprehensive technical report on this study is in Chandra (2009).

Previous studies did not address lines on electronic charts and map displays in detail. Chandra and Yeh (2007) did include a short exploration of line styles, in which pilot knowledge of line style conventions for paper charts and electronic map displays was assessed. The results showed that pilots are fairly knowledgeable about line conventions on paper charts, but that line conventions on electronic displays are not as well known or established. Lines and linear patterns currently in use by several manufacturers are documented in Yeh and Chandra (2008). This study and ARP 5289A address both lines *and* linear patterns, which are similar, but distinct, elements. In ARP 5289A, the term *line* refers to an element typically used to denote a boundary. Lines vary from one another in terms of width (e.g., thick or thin) and/or style (e.g., dotted, dashed, bold). A *linear pattern* may also be used to denote a boundary, but it is represented by a set of repeated patterns or symbols (e.g., several x's along a row).

Method

The study was conducted via paper questionnaire and consisted of two main tasks:

1) Line Sorting. Which lines and linear patterns are most useful?
2) Linear Pattern Recognition. Are there some linear patterns that are well recognized?

The Line Sorting task is designed to address the SAE Aeronautical Charting Committee's goal of identifying which lines and linear patterns should be associated with specific recommendations. The Linear Pattern Recognition task addresses the SAE Aeronautical Charting Committee's goal to understand what current linear patterns are well recognized. Pilots also responded to a set of subjective questions about lines and linear patterns, but those results are not discussed here; see Chandra (2009) for a discussion of the subjective pilot input.

Participant Recruitment and Background Information

Pilots were recruited from United States (US) domestic airlines, international airlines, the Air Force Flight Standards Agency, corporate operators, and private pilot organizations. Some US Government employees from the FAA's Flight Standards Service also participated. International respondents to the questionnaire were based in Australia, Canada, Denmark, England, Germany, Lebanon, Netherlands, New Zealand, and Mexico. Pilots were not compensated for their participation.

The questionnaire was initially sent to 242 instrument-rated pilots. A few months later, it was sent to 355 pilots qualified for visual operations. Pilots were allowed three to four weeks to complete and return the material. Overall, 273 questionnaires were returned with signed informed consent forms, yielding a 46% response rate.

Background information was gathered about pilot ratings and certificates, flight experience, avionics experience, and chart experience. There were 130 pilots in the Instrument Flight Regulations (IFR) Pilot group, which included pilots who reported either Instrument Ratings or Air Transport Pilot ratings. The IFR Pilots included all air transport, corporate, and international operators. The IFR Pilot group also included pilots who conducted military operations, private IFR operations, and even pilots who had experience with operations under Visual Flight Regulations (VFR), but were qualified for IFR operations. The VFR Pilot group had 143 pilots, and included pilots who reported that they held only a private pilot (VFR only) rating. Some of the VFR Pilots had instrument experience but were no longer current in instrument operations.

VFR Pilots reported lower total flight hours; IFR Pilots had a median experience of 9775 flight hours while the VFR Pilots had a median of 377.5 flight hours. The VFR Pilot group also included a higher percentage of pilots 61 and older (34% of VFR Pilots were 61 or older, while just 12% of IFR Pilots were 61 or older). Most pilots reported a typical flight length between one to three hours, with VFR Pilots flying more flights under one hour, and more IFR Pilots with flights longer than three hours. Most VFR Pilots (83%) reported that they only used NACO charts. Some IFR Pilots (39%) used Jeppesen charts exclusively, but many also had extensive experience with other charts, including NACO charts and charts from other sources (e.g. Lido, and charts produced by various governments). Of the pilots flying air transport operations, 80% reported Jeppesen chart experience.

Procedure

There were two sections in the first distribution of the study sent to instrument rated pilots. The first addressed line styles and the second addressed an unrelated research topic; both sections together took approximately 45 to 60 minutes to complete. The second distribution of the study, sent to non-instrument rated pilots, did not include the unrelated task reducing the total experiment time by approximately 15 minutes.

Using the instructions shown below in Figure 1, participants first sorted 65 types of line and linear patterns according to their usefulness. The names of the 65 items were printed on label sheets, one item on each label, in alphabetical order. No images of the items were shown in this task. (For a list of all the items, see Chandra, 2009.) Participants placed the labels for the two most useful categories onto separate sheets of paper, one that was titled

"Items that I find to be very useful in general" and the other that was titled "Items that I recognize and use on occasion." Items that the participant did not commonly use, or did not recognize were left on the label sheets.

(a) *Items that I find to be* ***very useful in general***. These are items that you know well and refer to frequently. They should be easily identifiable. Place these items on the first sheet of paper.
(b) *Items that I* ***recognize and use on occasion***. These are items that you use on occasion, but not as frequently as those you would place on the other sheet of paper. Place these items on the second sheet of paper.
(c) *Items that I* ***do not commonly use, or I do not recognize***. These are items that you seldom use, or you are not sure of their meaning and need more information in order to understand their use. Leave these items on their original label sheet.

Figure 1. Instructions for Line Sorting task

In the Linear Pattern Recognition task, pilots saw nine test linear patterns, which they were asked to identify and indicate their confidence in the response. If they did not know what the linear pattern represented, they were instructed to place a "?" in the response field. A sample question is shown in Figure 2. The linear patterns were selected for this task by subject matter experts on the SAE G-10 Aeronautical Charting Committee. The patterns included two options for the Air Defense Identification Zone (ADIZ) (one that was used by both Jeppesen and Lido, and the other recommended by the ICAO), and one option each for the Air Route Traffic Control Center (ARTCC), Communications, Controlled Airspace, Flight Information Region (FIR), International, Special Use Airspace (SUA), and Time Zone boundaries. In addition, one fake pattern was used as a baseline for comparison.

	Line pattern (or ?): ____________________ 1 2 3 4 5 6 7 **Low Confidence** **Medium Confidence** **High Confidence**

Figure 2. Sample linear pattern question.

Analyses and Results

To understand which items were considered Very Useful overall, responses to the Line Sorting task were tallied within pilot groups (IFR and VFR). A Chi-square test was performed to determine which airspaces and boundaries received a statistically significant number of responses in each response category for each pilot group. The test determined whether the number of responses in the category was statistically different from chance, which would have produced evenly distributed responses (i.e., 1/3 in each of the three response categories).

Results for the Line Sorting task are summarized in Table 1, which lists only the 26 lines and linear patterns that were considered Very Useful by the IFR and/or VFR Pilot groups based on the statistical test. While some items are important to both groups (e.g., controlled airspace) some are understandably Very Useful to only one or the other group (e.g., missed approach procedure tracks for IFR pilots and city patterns for VFR Pilots). For more information about the meaning and use of the individual items, consult Chandra (2009), the Federal Aviation Regulations/Aeronautical Information Manual (FAA, 2007), and/or the FAA Instrument Procedures Handbook (FAA, 2007).

Responses to the Linear Pattern Recognition task varied because of the free-response nature of the task; pilots sometimes used different words to express similar concepts. In order to understand the results, the responses were coded into categories. The categories were constructed with the aid of the SAE G-10 Aeronautical Charting Committee, which reviewed a partial set of data (the first 50 responses) to help the Volpe Center to determine which responses were correct and which were not if there was any question about the response. For example, the Committee determined that "Air Traffic Control *Sector* Boundary" was an incorrect response to the linear pattern that showed an Air Traffic Control *Center* Boundary," because a *Sector* is just one part of the *Center.* In addition, when determining overall accuracy, "Can't Tell" responses (indicated with a question mark in the response) were considered incorrect because the pilot admitted to not recognizing the pattern, whereas Missing responses were excluded from the analysis because the pilot may have left the response blank for other reasons. Final results of the analysis indicate how accurately the symbols were recognized. A similar process for handling responses is described in Chandra and Yeh (2007) in more detail.

Table 1. *Items considered Very Useful by IFR Pilots and VFR Pilots.*

Item	IFR Pilots	VFR Pilots
Air Defense Identification Zones (ADIZ)	x	x
Class B Airspace	x	x
Class C Airspace	x	x
Class D Airspace	x	x
Prohibited Airspace Area (P)	x	x
Restricted Airspace Area (R)	x	x
Enroute ATC Holding Pattern	x	
Missed Approach Procedure Holding Pattern	x	
Missed Approach Procedure Track	x	
Terminal ATC Holding Pattern	x	
Terminal Procedure Flight Track	x	
Terminal Transition or Feeder Route (Arrival, Departure, Approach)	x	
Enroute Airway or ATS Route	x	
Terminal Control Area (TCA/TMA)	x	
Warning Area (W)	x	
Telephone or Power Lines		x
City Pattern		x
Class E Airspace		x
Contours		x
Lake or Pond		x
Military Operations Area (MOA)		x
Railroad (single or multiple track)		x
River or Stream		x
Road (single or multi-lane)		x
Shoreline		x
Temporary Flight Restriction Area (TFR)		x

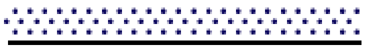
Results for the Linear Pattern Recognition task are summarized in Table 2, which shows the percentages of Can't Tell and Missing responses, as well as responses accuracies for the IFR and VFR Pilot groups for each of the nine test patterns. Preliminary testing of the linear patterns had indicated that the identification task would be difficult without context (e.g., cues about the linear pattern's shape, size, and relative location on the display), and the final data confirmed this expectation. Notice that the fake pattern was actually the most difficult pattern for the participants to identify as expected. The IFR and VFR Pilot groups differed in their response accuracies to six of the nine linear patterns; statistical significances for the differences are shown in the rightmost column of Table 2.

Discussion

Results of the Line Sorting task identified lines and linear patterns that were very useful to IFR and VFR Pilots. Chandra (2009) provides a more detailed breakdown of these results based on the pilot background. These results may be used by the SAE G-10 Aeronautical Charting Committee to determine which lines and linear patterns should be assigned specific recommendations in ARP 5289A. For example, recommendations may be most useful for the items considered Very Useful by *both* IFR and VFR Pilots listed in Table 1, such as the different Airspace Classes, and Prohibited/Restricted Areas. Regulatory authorities can then use either the full results of this study, or the ARP5289A document to determine if the information needs of the pilots are met by a given display.

The full results of the Line Sorting task presented in Chandra (2009) can also be used by manufacturers to determine what lines and linear patterns would be useful to pilots given a particular type of flight operation. For example, Chandra (2009) provides a breakdown of results for flight operation type. Manufacturers of displays for private pilots, as an example, may want to review Chandra (2009) to understand the results across all their potential customers including Private VFR, Private IFR, and Private Business operators.

Table 2. *Summary of results for the Linear Pattern Recognition Task.*

Test Item	Linear Pattern	Can't Tell Responses Overall	Missing Responses Overall	IFR Pilot Accuracy (N=130)	VFR Pilot Accuracy (N=143)	IFR vs. VFR Pilots Statistical Significance
ADIZ Option 1 (ICAO)		36%	22%	22%	49%	*F*(1, 210) = 18.7, *p* < 0.001
ADIZ Option 2 (Jeppesen/Lido)		68%	11%	28%	4%	F(1, 241) = 30.1, p < 0.001
ARTCC		58%	11%	20%	16%	No significant difference
Communications		60%	10%	34%	0%	F(1, 243) = 67.3, p < 0.001
Controlled Airspace		24%	26%	37%	49%	No significant difference
Fake Pattern		70%	13%	n/a	n/a	n/a
FIR		44%	26%	47%	7%	F(1, 199) = 47.4, p < 0.001
International Boundary		32%	25%	35%	50%	F(1, 204) = 5.16, p < 0.05
SUA Boundary		22%	11%	54%	48%	No significant difference
Time Zone		37%	26%	45%	14%	F(1, 200) = 26.2, p < 0.001

Results of the Linear Patterns Recognition task may be used in identifying whether some linear patterns are currently well recognized, and should be recommended for use as is. Recognizing linear patterns in isolation was a difficult task and overall recognition rates were relatively low, particularly in comparison to the recognition rates obtained for identifying specific symbols such as the navigation aid symbols and other general symbols that were evaluated in Chandra and Yeh (2007). The most recognizable linear pattern was the Special Use Airspace Boundary, which obtained a 51% recognition rate overall, whereas navigation aid symbols were typically recognized by pilots 80% of the time or better. Although the recognition rates for linear patterns were relatively low overall, some patterns were better recognized than others and these results may be used by the SAE G-10 Aeronautical Charting Committee to determine which linear patterns should be included in ARP5289A. Even if the linear pattern is not recognized by a majority of pilots, reusing an existing symbol will aid pilots who are familiar with it, and it may reduce future potential conflicts with that symbol.

Summary and Conclusions

This report provides an overview of a study conducted to explore the utility and recognition of lines and linear patterns on electronic displays depicting aeronautical charting information, such as electronic charts and moving map displays. Further details about this study are reported in Chandra (2009).

The results of this study provide valuable information for the development of an industry recommendations document that will help manufacturers and regulatory authorities assess whether the information needs of the pilots are met by various electronic displays of aeronautical charting information. In order to maximize the applicability of the results, data were collected from pilots who fly all types of operations, from around the world. Items that were useful to different pilot groups were identified based on pilot qualifications, types of flight operations, and typical flight length. Recognition of a test set of nine linear patterns was difficult, but some patterns were more recognizable than others.

Results of this study will be considered in the development of an updated industry recommendations document, specifically, the SAE International Aerospace Recommended Practices (ARP) document on Electronic Aeronautical Symbols (ARP 5289A). The FAA, other civil aviation authorities, or ICAO the may choose to adopt this industry document by reference at a later date. Note that this research applies to any electronic display that shows the lines and linear patterns tested in this study, regardless of the intended function of the display, so its applicability may be far reaching.

References

Chandra, D. C and M. Yeh (2007). Pilot Identification of Symbols and an Exploration of Symbol Design Issues for Electronic Displays of Aeronautical Charting Information (Report Nos. DOT/FAA/AR-07/37, DOT-VNTSC-FAA-07-07). Cambridge, MA, US DOT Volpe National Transportation Systems Center.

Chandra, D.C., Yeh, M. and C.M. Donovan (2007). Pilot Identification of Proposed Electronic Symbols for Displays for Aeronautical Charting Information. *In Proceedings of the 51st Annual Meeting of the Human Factors and Ergonomics Society.* Santa Monica, CA: Human Factors and Ergonomics Society. pp. 65-69.

Chandra, D.C. (2009) Utility and Recognition of Lines and Linear Patterns on Electronic Displays Depicting Aeronautical Charting Information (Report Nos. DOT/FAA/AR-09/01, DOT-VNTSC-FAA-09-03). Cambridge, MA, US DOT Volpe National Transportation Systems Center.

Federal Aviation Administration (2007a). Instrument procedures handbook (FAA-H-8261-1A). Washington, DC: Author.

Federal Aviation Administration (2007b). Federal Aviation Regulations/Aeronautical information manual (ASA-07-FR-AM-8K). Newcastle, WA: Aviation Supplies & Academics.

Society of Automotive Engineers (SAE) (1997). Electronic Aeronautical Symbols, ARP 5289. Warrendale, PA, SAE.

Yeh, M. and D. C. Chandra (2005). Designing and Evaluating Symbols for Electronic Displays of Navigation Information: Symbol Stereotypes and Symbol-Feature Rules. (Report Nos. DOT/FAA/AR-05/48, DOT-VNTSC-FAA-05-16) Cambridge, MA, US DOT Volpe National Transportation Systems Center.

Yeh, M. and D.C. Chandra (2006). Pilot stereotypes for navigation symbols on electronic displays. In Reuzeau, F., Corker, K. & Boy, G. (Eds.) Proceedings of the International Conference on Human-Computer Interaction in Aeronautics. (pp. 66-73). Toulouse, France: Cépaduès-Éditions.

Yeh, M. and D.C. Chandra, (2008) Survey of Symbology for Aeronautical Charts and Electronic Displays: Navigation Aids, Airports, Lines, and Linear Patterns. (Report Nos. DOT/FAA/AR-07/66, DOT-VNTSC-FAA-08-01) Cambridge, MA, US DOT Volpe National Transportation Systems Center.

Acknowledgements

This report was prepared by the Behavioral Safety Research and Development Division at the Volpe National Transportation Systems Center. It was completed with funding from the FAA Human Factors Research and Engineering Group (AJP-61) in support of the Aircraft Certification Service Avionics Branch (AIR-130) and the Technical Programs and Continued Airworthiness Branch (AIR-120). I would like to thank the FAA program manager, Tom McCloy, as well as the FAA technical sponsor Colleen Donovan for their assistance with this project.

I would also like to thank the members of the SAE International G-10 Aeronautical Charting Committee for providing technical guidance and insight, and thanks to the International Federation of Air Line Pilots Associations (IFALPA), the National Business Aircraft Association (NBAA), the Airline Pilots and Owners Association (AOPA), and the Air Force Flight Standards Agency for assistance in recruiting participants. Particular thanks go to the many pilots who donated their time for the study. Finally, I would like to thank the Volpe Center staff who helped with the study: Catherine Guthy, Michelle Yeh, Michael Zuschlag, Raquel Rodriguez, and Rachel Selgrade.

The views expressed herein are those of the author and do not necessarily reflect the views of the Volpe National Transportation Systems Center, the Research and Innovative Technology Administration, or the United States Department of Transportation.

TRAINING TO REDUCE AVIATION MAINTENANCE ERROR: ASSESSING MAINTENANCE RESOURCE MANAGEMENT PROGRAMS IN COMMERCIAL AVIATION

Erin E. Block
National Center for Aviation Safety Research, Saint Louis University
St. Louis, MO

Edward J. Sabin
National Center for Aviation Safety Research, Saint Louis University
St. Louis, MO

Manoj S. Patankar
National Center for Aviation Safety Research, Saint Louis University
St. Louis, MO

Maintenance resource management (MRM) training is intended to integrate existing technical skills of maintenance employees with interpersonal skills/human factors knowledge to improve communication effectiveness & maintenance safety. The FAA suggests successful MRM training not only teaches error avoidance, but also the adoption of attitudes that support a safety culture. This coincides with FAA encouragement to incorporate systems theory into MRM training to put human factors issues in larger organizational context. Many programs currently use the MRM/TOQ survey to assess the impact of MRM training & its effectiveness in changing safety-related attitudes.

Previous research has identified key factors in MRM/TOQ items & argued that the instrument has good reliability/validity; however, interrelationships among factors have not been closely examined, nor have there been systematic attempts to understand how these four critical areas of human factors training fit into aviation safety frameworks. With growing relevance of systems theory to aviation, MRM training assessments should be based in systems framework & MRM/TOQ results analyzed therein. The paper reviews MRM training in a commercial aviation organization from a systems perspective in order to improve training assessment & confirm reliability/validity of MRM/TOQ. Findings indicate revision of MRM/TOQ is necessary to accurately assess training; also present evidence to support using systems framework to evaluate MRM training. This work is part of ongoing programs at the National Center for Aviation Safety Research.

Development of MRM Training

In the aviation industry, successful organizational performance is often considered with respect to safety - the avoidance of accidents and incidents and the promotion of behaviors/organizational norms considered safe. While many organizations are concerned with employee safety, aviation is a "high-consequence" industry; that is, an industry in which the consequences of poor safety performance are far more significant and generally more public than in other industries (e.g., commercial aircraft disasters, maintenance damage to multi-million dollar equipment, stakeholder fatalities). Aviation organizations are thus significantly invested in identifying and appropriately measuring factors that may affect safety performance. Intra- and interpersonal human factors as significant sources of error have received increased scrutiny in recent years for their potential impact on aviation safety (Patankar and Taylor, 2008).

MRM training developed from existing CRM programs finding success during the 1980s (Taylor & Patankar, 2001). According to Taylor & Patankar, the first reported CRM program geared toward aviation maintenance workers began in November 1989; this and other programs eventually became known as MRM programs after the term "Maintenance Resource Management" was coined in 1992 (Taylor & Christensen, 1998). According to the FAA, MRM is a "process for improving communication, effectiveness and safety in aircraft maintenance operations" (2000, p. 6), and was developed to address "teamwork deficiencies within the aviation maintenance environment" (p. 6). A review of four generations of MRM programs by Taylor and Patankar (2001) demonstrates a pattern of changing interest in the focus on organizational system variables and longitudinal stability of post-training attitude and behavior changes. Fourth generation MRM programs are interested in gauging participant attitudes not only on directly safety-related issues (such as whether participants feel their work impacts passenger safety), but also on organizational context issues such as leadership and coworker interaction. This represents the incorporation of applied psychological principles into awareness training for safety-critical attitudes and behaviors, in an industry in which the consequences of failure are quite literally disastrous. Research that assists the aviation industry in appropriately integrating these principles is essential to their continued adoption and efficacy.

MRM Training Assessment

Based on the development and goals of CRM training programs, the Cockpit Management Attitudes Questionnaire (CMAQ) was developed to assess flight crew attitudes regarding human factors issues (Helmreich, Foushee, Benson, & Russini, 1986). Just as MRM evolved from initial efforts involving flight crews, so too did evaluation methodologies for MRM training evolve from those initially created to assess flight crew changes following CRM. Taggart (1990) was among the first to adapt CRM evaluation methods for the maintenance environment, revising the CMAQ for use with aviation maintenance employees. This modification was called the Crew Resource Management/Technical Operations Questionnaire (CRM/TOQ); later renamed the Maintenance Resource Management/Technical Operations Questionnaire (MRM/TOQ).

Research on the CMAQ (Gregorich, Helmreich, & Wilhelm, 1990; Sherman, 1992) confirmed four constructs based on the survey items given to flight crews: communication/coordination, shared command responsibility, recognition of stress effects, and avoidance of interpersonal conflict; though the conflict avoidance factor is inconsistent in CRM data and discarded in subsequent analyses (Gregorich et al.). Aviation industry accident/incident reviews suggest that behaviors related to these four constructs underlie many of the human factors errors that have occurred. Additionally, these four constructs incorporate aspects of a systems approach to promoting safety, attempting to gauge alignment among organizational factors such as leadership with interpersonal factors such as communication and conflict avoidance. The CMAQ, and later, the MRM/TOQ were both designed to measure changes in these four attitude constructs prior to and immediately following resource management training.

Gregorich et al. (1990) ran confirmatory factor analyses across three different samples to confirm the factor structure of the CMAQ with flight crew employees. While an exploratory factor analysis for maintenance employees has been conducted by Taylor (2000b), this analysis found different results than those presented by Gregorich et al. (such as strong evidence for the conflict avoidance factor), increasing the level of ambiguity regarding the four factors that result from evaluations of resource management training. In addition to these findings, Taylor was not focused on details regarding the nature of the pre- and post-training attitude relationship, meaning the analyses done by Gregorich et al. with flight data have yet to be done using data from maintenance employees. Given the existing discrepancies between the general factor structures found for each sample, as well as knowledge of the prevalence of difference types of error for each sample, it makes sense to compare both flight and maintenance sample data to identify relevant similarities or differences between the two populations.

While these four attitude constructs have been assessed with maintenance employee samples, the relationships among the constructs have not yet been identified. The present study attempts to fill this gap by analyzing maintenance data from the MRM/TOQ to replicate the work of Gregorich et al. (1990). If their results regarding flight crew attitude differences following training can be recreated with data from maintenance employees, this may increase the generalizability of the Gregorich et al. findings and suggest key leverage points when implementing human factors initiatives in aviation (Block, 2008). Additionally, confirmatory factor analyses with a large sample should clarify the relationship among the MRM/TOQ items for each factor and the presence of the conflict avoidance factor for maintenance employees, as well as support its continued investigation with that population, even though published data regarding flight crews (e.g., Gregorich et al.) may not support this factor as a part of resource management or human factors training. The present study thus also hopes to provide support for the continued use of the MRM/TOQ to measure maintenance employee attitudes, as the MRM/TOQ is based on a tool initially created to measure flight crew attitudes. Consistent results for both flight crews and maintenance employees would add additional support for the MRM/TOQ.

Participants & Design

Data were collected from 1458 aviation maintenance employees from a major U.S. airline who had participated in MRM training within the last 14 months. The mean age of respondents was 49.74 years (SD = 8.03), mean years in maintenance at the target organization was 17.55 (SD = 7.07), and a majority of the participants were male (96.6%).

The MRM/TOQ is a 17-item questionnaire developed to measure the attitudes and intentions of participants in airline maintenance communication and safety training workshops (Taylor, 2000b). As mentioned previously, the

MRM/TOQ has been adapted for maintenance employees from the Cockpit Management Attitudes Questionnaire CMAQ; the attitude areas as measured by the CMAQ have been validated as predictors of outcome factors (expert performance ratings), suggesting attitudes as measured by the CMAQ (or, in this case, the MRM/TOQ) may be used to indicate performance outcomes (Helmreich & Foushee, 1993).

Participants were asked to complete the MRM/TOQ immediately before the class started, and again at the end of the day (immediately following training). This survey gathered information on the four attitude areas identified in the work of Taylor (2000a, 2000b) for MRM training, as well as information on individual demographics. The post-training questionnaire also collected responses on three general outcome items: 1) this training has the potential to increase aviation safety and crew effectiveness; 2) this training will be useful for others; and 3) this training is going to change behavior on the job.

Factor Comparisons
Confirmatory factor analyses (CFA) using LISREL were conducted for both the pre- and post-training data to ascertain if the four-factor structure (communication/coordination, relational supervision, recognition of stress effects, and conflict avoidance) described by Taylor (2000b) for maintenance employees using the MRM/TOQ is consistent with the present data sample. Results of the post-training factor analysis are presented in Table 1. For pre-MRM training data Taylor's four-factor structure failed to converge when all items were included. Eliminating the weakest item ("A truly professional team member can leave personal problems behind when working") allowed a four-factor structure to converge, but still indicated poor item loadings for the two remaining items related to "recognition of stress effects". Complete elimination of the stress effects factor did not change the fit indices of the three remaining factors: root mean square error of approximation (RMSEA) = 0.046, Tucker-Lewis index = 0.98, and comparative fit index = 0.98.

For post-training data the initial four-factor solution converged, but again indicated poor item factor loadings (lambda values) for "recognition of stress effects"; the CFA procedure was re-run with a three-factor structure, and results of both analyses may be seen in Table 1. Eliminating the stress effects factor contributed to a slight improvement in the RMSEA (from 0.061 to 0.056); other fit indices were unchanged by the removal of this factor: Tucker-Lewis index = 0.98, and comparative fit index = 0.98. Although Gregorich et al. (1990) identified conflict avoidance as a fourth factor, and found instability in this factor (made up of two items: "maintenance personnel should avoid disagreeing with one another" and "it is important to avoid negative comments about the procedures and techniques of other team members"), the present study found the three-factor structure more consistent in both pre-training and post-training data. The data indicate the factor that should be eliminated is not conflict avoidance, but recognition of stress effects. In his sample of aviation maintenance employees, Taylor (2000b) similarly found conflict avoidance as a stable third factor, although he suggests a four-factor structure is present and appropriate for data collected with the MRM/TOQ.

Pre-Post Training Comparisons
To test whether the same pattern of significant pre- and post-training attitude changes would be found in the present data as was found by Gregorich et al. (1990), a repeated-measures ANOVA procedure was performed on each pair of pre- and post-training composite scores, including site location and type of maintenance job held by participants as potential interacting variables. This is similar to the cross-organization and cross-job title analyses conducted by Gregorich et al. Analysis showed a significant main effect of training for two of the four factors: communication/coordination ($F_{1,1249} = 4.48$, $p < 0.05$) and conflict avoidance ($F_{1,1249} = 17.42$, $p < 0.05$). Analysis of the remaining two factors showed marginally significant change following training: $F_{1,1249} = 3.12$, $p = 0.07$ for relational supervision; $F_{1,1249} = 3.75$, $p = 0.053$ for recognition of stress effects. These results, however, may be due more to the large sample size contributing to the liberality of the F-test rather than to the presence of meaningful post-training differences (e.g., the net change in the composite score on relational supervision from pre- to post-training is 0.54, slightly more than one-half of one scale point). None of the interactions for any of the four factors were significant.

According to Gregorich et al. (1990), if resource management training was successful there should be diminished response variation following training; and if training enhanced pre-existing attitudes, response variation is likely to have increased following training. To determine whether such a variance reduction had occurred in the present sample, a t-test for the difference between correlated variances (e.g., testing for heterogeneity of variance) (Gregorich et al., 1990; Ferguson, 1971; Howell, 2002) was computed for each pair of pre- and post-training factor scores. Contrary to results obtained by Gregorich et al., only the relational supervision and conflict avoidance factors showed a significant change in variability following training, and both of those factors actually demonstrated *increased* variability following MRM training, rather than the anticipated decrease in mean variability. Prior to MRM training, the mean variability (calculated as the squared deviation from the mean; Howell, 2002) for the relational supervision factor was $s^2 = 18.55$; following training the mean variability was $s^2 = 20.00$ ($t_{1457} = -2.595$, $p < 0.01$). Prior to MRM training the mean variability for conflict avoidance was $s^2 = 4.98$; following training the mean variability for this factor was $s^2 = 5.89$ ($t_{1457} = -5.378$, $p < 0.001$).

Discussion

Attempts to confirm the four-factor structure identified by Taylor (2000b) and provide additional support for the use of the MRM/TOQ in assessing MRM training produced mixed results. While the conflict avoidance factor (characterized by the items "maintenance personnel should avoid disagreeing with one another" and "it is important to avoid negative comments about the procedures and techniques of other team members") has been found to be inconsistent and under-identified in samples of flight crews (Gregorich et al., 1990), both Taylor (2000b) and the present study found consistent responses for this factor among maintenance employees. The reasons for this distinction between the two broad categories of aviation employees are unclear. Perhaps maintenance employees perceive interpersonal disagreements and procedural disagreements as highly similar types of conflict, whereas flight crews make a distinction between the two – this may cause the items to diverge and may contribute to inconsistent results for flight crews on this factor. Additionally, the format of conflict training for maintenance employees may differ in approach from that used with flight crews, leading to discrepancies in interpretation of this factor. Future research comparing flight crews and maintenance employees on human factors issues may wish to address this in greater depth.

The factor "recognition of stress effects" proved to be problematic in the present study. Although Taylor (2000b) found this factor to be consistent among maintenance employees in the air carriers he studied, the present work (which may not have included the same carriers) found little support for continued inclusion of this factor in its present state. Confirmatory factor analysis indicated a poor degree of item fit for the three items intended to measure stress recognition. Results would suggest that stress recognition as a factor be dropped from subsequent investigations using the MRM/TOQ. For the aviation industry, however, stress recognition is considered an important aspect of human factors knowledge that should be conveyed to employees—meaning it should *not* be discarded from future MRM research and training. The observed data on this factor strongly suggest that modifications are warranted for either 1) items used in the MRM/TOQ to measure understanding and agreement with stress recognition human factors training; or 2) the method of presentation of stress recognition information in the MRM courses. The data on this factor indicate a conceptual disconnect between the teaching of stress recognition as it is currently provided and the understanding of stress recognition as it is captured by the current MRM/TOQ. If the industry wishes to continue emphasizing the importance of stress recognition as a component of MRM training, the measurement methods for this factor and the training content must be in alignment.

The present study sought, in part, to identify whether maintenance employees respond to resource management training in a manner similar to that reported by Gregorich et al. (1990) for flight crews. Inter-item reliability for the items used in the MRM/TOQ factors was comparable to that reported by Gregorich et al. and Taylor (2000b), except for the factor "recognition of stress effects" (Cronbach's $\alpha = 0.29$ in both pre- and post-training data). Of interest is the question of what constitutes "good" reliability. Both Gregorich et al. and Taylor report "acceptable" or "good" reliability for the items in each factor, with Cronbach's alpha values ranging from 0.47 – 0.67 (Gregorich, et al.) and from 0.51 – 0.77 (Taylor). Generally, α levels above 0.70 are desired to indicate "good" reliability (as an $\alpha = 0.70$ would indicate slightly less than half of the variance in the item is attributable to measurement error). The reasons for the instability in this factor are unclear. Perhaps maintenance employees do not clearly perceive the relationship between the effects of stress and safety outcomes, perhaps the items are being misinterpreted, or maybe the material is not being clearly explained in the course of MRM training.

Gregorich et al. (1990) argued that successful resource management would be indicated not only by post-training attitude change, but also by diminished response variation following resource management training; otherwise, if training bolstered existing attitudes, response variability should increase post-training. While Gregorich et al. found that, as hypothesized, variability decreased following training, the present study found the opposite: for two of the four factors (relational supervision and conflict avoidance) variability significantly increased following MRM training. According to Gregorich et al., these findings would suggest that MRM training has not been completely successful in changing important safety-related attitudes for maintenance employees. An alternative view may be that, for the majority of maintenance employees, MRM training is changing attitudes in the desired direction (hence the significant and near-significant post-training attitude changes); however, this is causing the gap between employees who exhibit desired change and employees who "boomerang" following training to widen, which would appear as greater variability in post-training responses.

Table 1: Post-Training Data Confirmatory Factor Analysis

	Four Factor Structure		**Three Factor Structure**	
	λ	**R^2**	**λ**	**R^2**
Communication/Coordination				
Employees should make the effort to foster open, honest, and sincere communication.	0.89	0.79	0.88	0.78
We should always provide both written and verbal turnover to the oncoming shift.	0.86	0.73	0.86	0.74
My work impacts passenger satisfaction/safety.	0.88	0.77	0.87	0.76
A debriefing and critique of procedures and decisions after a significant task is completed is an important part of developing and maintaining effective crew coordination.	0.82	0.67	0.82	0.68
Having the trust and confidence of my coworkers is important.	0.76	0.58	0.77	0.59
Start of shift maintenance crew meetings are important for safety and for effective crew management.	0.72	0.51	0.73	0.53
My coworkers value consistency between words and actions.	0.63	0.39	0.62	0.39
Relational Supervision				
My supervisor can be trusted.	0.84	0.70	0.83	0.70
My supervisor protects confidential or sensitive information.	0.79	0.63	0.79	0.63
Mechanics' ideas are carried up the line.	0.67	0.45	0.67	0.45
My suggestions about safety would be acted on if I expressed them to my lead or supervisor.	0.74	0.55	0.74	0.55
I know the proper channels to route questions regarding safety practices.	0.55	0.30	0.55	0.30
Conflict Avoidance				
It is important to avoid negative comments about the procedures and techniques of other team members.	0.75	0.56	0.75	0.56
Maintenance personnel should avoid disagreeing with one another.	0.71	0.50	0.71	0.50
Recognition of Stress Effects				
Even when fatigued, I perform effectively during critical phases of work.	0.05	0.00		
Personal problems can adversely affect my performance.	-0.19	0.06		
A truly professional team member can leave personal problems behind when working.	-0.25	0.03		

With the exception of the overall factor structure and the analyses of response variability, the present study confirmed that many of the characteristics found in samples of flight crews who have participated in CRM training might also be found among maintenance employees who have participated in MRM training. This suggests that the underlying principles of both CRM and MRM are contributing to attitude change with regard to human factors issues. The differences in the consistency of the factor structure between flight crews and maintenance employees may suggest that there are differences in the material emphasized for flight crews and the material emphasized for maintenance employees (such that conflict avoidance is given more time/weight in MRM courses), that flight crews and maintenance employees have different mental concepts for "conflict avoidance", or that conflict avoidance is a more relevant concept for maintenance employees. These differences in the order and consistency of the factor structure for these items suggest that there are different characteristics of the flight crew and maintenance populations, and that these differences should be taken into account when modeling attitude change, developing resource management training, or making comparisons on outcomes between the two groups. Future studies that capture the underlying concepts behind employees' evaluations of both conflict avoidance and recognition of stress effects would help to clarify these discrepancies, and better explain the relationship between each of the four factors, general job area (flight crew or maintenance) and safety.

Acknowledgements
The authors are grateful to the Federal Aviation Administration for support of this research, conducted under grant #2003-G-013, Maintenance Aviation Safety Action Programs, Principal Investigator Dr. Manoj Patankar. The authors also wish to thank Dr. Matt Grawitch for his editorial guidance and insightful comments in the development of this research. This paper is based on a doctoral dissertation by E. Block.

References
Block, Erin E. (2008). Maintenance Resource Management in aviation: A systems approach to measuring training. Ph.D. dissertation, Saint Louis University, United States -- Missouri. Retrieved March 16, 2009, from Dissertations & Theses @ Saint Louis University database. (Publication No. AAT 3324137).

Federal Aviation Administration. (2000). *Maintenance resource management training.* (AC No. 120-72).

Ferguson, G. A. (1971). *Statistical analysis in psychological education.* New York: McGraw-Hill.

Gregorich, S. E., Helmreich, R. L., & Wilhelm, J. A. (1990). The structure of cockpit management attitudes. *Journal of Applied Psychology, 75,* 682-690.

Helmreich, R. L. & Foushee, H. C. (1993). Why crew resource management? Empirical and theoretical bases of human factors training in aviation. In E.L. Wiener, B.G. Kanki, & R.L. Helmreich (eds), *Cockpit Resource Management.* San Diego, CA: Elsevier Press, 3-45.

Helmreich, R. L., Foushee, H. C., Benson, R., & Russini, W. (1986). Cockpit resource management: Exploring the attitude-performance linkage. *Aviation, Space, and Environmental Medicine, 57,* 1198-1200.

Howell, D. C. (2002). *Statistical methods for psychology*. (5th ed.). Belmont, CA: Duxbury Press.

Latorella, K. A. & Prabhu, P. V. (2000). A review of human error in aviation maintenance and inspection. *International Journal of Industrial Ergonomics, 26,* 133-161.

Patankar, M. S. & Taylor, J. C. (2008). MRM training, evaluation, and safety management. *The International Journal of Aviation Psychology, 18,* 61-71.

Sherman, P. J. (1992). New directions of CRM training. *Proceedings of the Human Factors Society Thirty-sixth Annual Meeting.* Santa Monica, CA: HFS.

Taylor, J. C. (2000a). The evolution and effectiveness of maintenance resource management (MRM). *International Journal of Industrial Ergonomics, 26,* 201-215.

Taylor, J. C. (2000b). Reliability and validity of the maintenance resources management/technical operations questionnaire. *International Journal of Industrial Ergonomics, 26*, 217-230.

Taylor, J. C. & Christensen, T. D. (1998). *Airline Maintenance Resource Management: Improving Communication.* Warrendale, PA: Society of Automotive Engineers.

Taylor, J. C. & Patankar, M. S. (2001). Four generations of maintenance resource management programs in the United States: An analysis of the past, present and future. *The Journal of Air Transportation World Wide, 6,* 3-32.

ADAPTATION OF CRM TRAINING FOR THE RAILWAY INDUSTRY OPERATIONAL SAFETY BENEFITS

Matthew T.S. Tsang
Kowloon Canton Railway Cooperation (KCRC)
Hong Kong, China

Hans-Juergen Hoermann
German Aerospace Center (DLR)
Hamburg, Germany

In aviation, Crew Resource Management (CRM) was developed to address safety issues derived from accident and incident investigations. As CRM has proven its effectiveness by improving teamwork, communication and staff responses to operational hazards, there have been many attempts to expand this concept into other high-risk sectors such as medical, nuclear, or military. Although some work was also conducted to modify CRM for the railway industry, no such experiences yet existed in China or Hong Kong. Having observed the effectiveness of CRM and Line Oriented Training (LOT) in aviation, this paper documents the introduction and initial evaluation of CRM and LOT in Hong Kong in the West Rail (WR) division of the Kowloon-Canton Railway Corporation (KCRC). Results of an initial evaluation study with 120 operative crewmembers provide empirical support for the chosen approach.

Recent world headline disasters such as the September 11 attacks, 2005 London transport bombings, and 2003 Daegu subway arson attacks sent explicit warnings to all nations that mass-transit systems can and do become luscious targets of terrorist attacks with catastrophic consequences. Whilst it may be impossible to ever eliminate all forms of threats, one thing remains clear— the effectiveness of crew response to emergency situations can have a huge influence on its outcome. Since railway and aviation industries share many similar characteristics and vulnerabilities, transferring training strategies for flight crews into the railway industry may be advantageous in developing better training and safety programs for train drivers and traffic controllers. A CRM oriented program for operative train crews could improve performance in managing threats and errors, reducing the consequences of eventual emergencies in the railway sector (Morgan, Olson, Kyte, Roop, 2007; Dedale, 2006).

Human Factors Issues Encountered in the Railway Industry

Railway technology has gone a long way to identify which trains are on the track and controlling their progress and whereabouts by keeping safe but efficient train to train separations, and managing communications between trains and the operations control centre (Wilson, Norris, Clarke & Mills, 2005). However, train operations communications studies showed that the design, implementation and operation of train operations communication systems generate a host of new human factors problems. Although the implementation of automatic train protection (ATP) was able to significantly reduce the number of fatal accidents due to signal passed at danger (SPAD), driver experience erosion due to control automation was detrimental to safety (Wright, Turner, Antonelli & Bendig, 2004). The reliance on control automation during normal operations may mean that when manual control is demanded the crews and drivers task performance have decreased efficiency. After a long period of smooth normal train operations with increased levels of automation, traffic controllers or train drivers could become less efficient in manual control under degraded train operations due to lack of exposure. This was demonstrated by brief survey of railway stakeholders in the UK (Wilson & Norris, 2005), in which 80% of the experts agreed that experience erosion is a potential safety issue.

Lessons learned in railway industries consistently demonstrate failures in organizational and human performance factors as being causal or contributory to accidents, prompting the RSSB (2006-2007) to list major needs and recommendations in safety, many of which can be satisfied by the more interactive and effective styles provided by CRM and LOT training:

- Introduction of structured refresher trainings for train crew;
- Improved co-ordination between train crews, station controllers and traffic controllers for manpower back-up;
- Emphasis on interactive training in Standard Operating Procedures (SOP's);
- Optimisation of training costs by using e-learning, computer-based training (CBT), integrated training facilities (ITF);
- Assess effectiveness of using a decision support system (DSS) to support train crews in coping with emergencies;
- Develop self-learning portals for train crews;
- More efficiency in training time, type-rating tests for train drivers, station controllers and traffic controller train crews due to job rotation.

Superseding the conventional training courses, CRM and LOT type of training in the railway industry could provide a more interactive and more effective environment for learning.

Methods

Drivers, station and traffic controllers from the West Rail (WR) division of the Kowloon-Canton Railway Corporation (KCRC) took part in a three days CRM training program. The program used lectures, video aided training facilities (VAT), and integrated training facilities (ITF) to expose staff to CRM and other safety related concepts in a series of workshops, experiential learning, role plays, video exercises, case studies, discussion groups, team building exercises and social and leisure activities (Tsang, 2007). ITF is a centralized training facility that links the main control system simulator, the train control and signaling system simulator and the cab simulator in a networked, integrated team-training environment. It was used in LOT programs to simulate emergency and abnormal operation scenarios to test and train the ability of teams to efficiently handle such situations. Feedback was provided by computer, video and voice records, as well as CRM/LOT trainer debriefings. An attitude survey was administered to provide pre- and post-training scores, to track changes in safety related attitudes attributable to the training. In addition, a non-CRM trained control group was available, which provided performance data during an emergency drill as compared to the CRM trained group (Tsang, 2007).

Personal Attributes of the Trainees

The trainee population consists of 120 participants: 80 train drivers, which is 80% of West Rail working population (KCRC, 2005); 20 station controllers (60% of WR population), 20 traffic controllers (60% of WR population). The male to female ratio is 1:0.15, which is representative of the overall employee population. Trainees were divided into four groups of 30 members based upon age, education background and years of experience in the railway industry. Participants who satisfied any of the 3 factors listed in table 1 were accepted into the respective group, prioritizing from group A to D. Therefore, if a participant fitted multiple categories they were accepted into the higher ranked group.

Table 1. *Criteria for sorting participants into groups.*

Group qualification criteria	Age (years) *or*	Highest Education Qualification *or*	Railway Experience (yrs.)
A	>45	Degree or higher	>15
B	40-45	Diploma/certificate	10-15
C	35-40	Post Secondary	5-10
D	<35	Secondary	<5

CRM Training Program

Five designated CRM-trainers first underwent a CRM class instructor course before they facilitated 4 sessions of three-day CRM programs for the operating train crews. The CRM program used for the KCRC covered the following methods and activities: (see Tsang, 2007 for further details).

- Presentations
- Experiential learning
- Role play
- Video exercises
- Case studies
- Discussion Group
- Team building exercises
- Social and leisure activities.

The training was further supported by automated facilities including integrated training facilities (ITF), computer based training (CBT) and a decision support system (DSS).

Results

Evaluation of Changes in Safety Attitudes

Before and after CRM training, train crews were asked to fill out an attitude survey on safety climate consisting of 20 questions focusing on safety awareness and safety practices in the organization to gain indications on whether there have been changes in safety attitudes. T-Tests for paired samples were used to compare each item pre and post for all 120 subjects and multivariate Analysis of Variance (MANOVA) with one between subject factor (Group) and one within subject factor (pre and post CRM) was conducted. The subjects showed generally stronger team working attitudes after training. There is a strong within subject effect for the attitude items ($F = 11.28$, $p < .001$). All changes were in the desired safety related direction. There was only a small between subject (group) effect ($F = 1.42$, $p < .05$) and no interaction of both. The attitude questionnaire showed positive changes of safety attitudes of the trained subjects after training. After the CRM training crews displayed an increase in safety oriented attitudes probably because of the new concept of CRM was stimulating new ideas in their way of thinking. This indicated that CRM was able to introduce the safety critical concepts and ways of thinking that crews should adopt in their day to day roles.

Performance Observations from Drills

Another test of training effectiveness was conducted during an emergency drill (fire inside tunnel drill) at East Rail Beacon Hill Tunnel (2.75 km in length) and West Rail Tai Lam Tunnel (5.5 km in length) in conjunction with the Hong Kong police, fire and hospital services by comparison of a CRM-untrained control group and a CRM-trained group. An observer panel scored behaviours of the two groups. It was reported to display better incident handling capabilities, stronger teamwork and communication skills throughout the exercise. They were also able to offer a larger number of possible explanations for the simulated 'incident', with an increase in the number of explanations classified as situation awareness, decision-making, communication, and supervision. Overall, the train crew who attended CRM training worked more cohesively as a team, maintained stronger situation awareness and adopted "readback message" as double confirmation before execution of safety critical commands as compared to the CRM-untrained control group. The use of accident scenarios to evaluate CRM training effects had been used similarly before for air traffic controllers (Andersen & Bove, 2000).

Post Training Feedback

Feedback was obtained with a specific course feedback questionnaire regarding the relevance, interest, standards of teaching, exercises, videos and course exercises from all 120 train crews who participated in the course. Feedback was generally positive, and participants showed the desire for a longer course to help consolidate concepts. Overall, the CRM courses were well-accepted by the course participants and management. Based on these post course survey results, increased management attention had emerged at corporate level to provide more teambuilding workshops, more consultancy studies on human factors aspects such as fatigue awareness and safety awareness training for all levels of KCRC staff in 2007.

Workplace Performance

Tangible evidences of success were observed from train services achievements in 2005 and 2006 (KCRC, 2006, 2007a). Two methods were adopted to measure the workplace performance post CRM training to check information retention and if skills learnt were applied on the job. The first method involved a six months workplace performance summary report feedback from the General Manager of West Rail Operations Department (KCRC 2006, 2007a, 2007b). The items in this survey included:

- Immediate results: Computer records from ITF interactive responses
- Performance results: Ride checks and workplace performance audits
- Overall performance results of West Rail (accident rate)
- Complaint and Commendation letters from the customers
- CRM and computer-based training (CBT) computer records;
- Train "Black box" data and ride checks by train crew inspectors to monitor train drivers performance.

A Real Challenge

On 14 February 2007, a train fire occurred inside West Rail Tai Lam tunnel. Due to effective teamwork in incident handling, 800 passengers evacuated from the incident train to the nearest station from the tunnel within 20 minutes with no injuries reported. Train services had recovered 221 minutes after the outbreak. Most of the train crew in this incident were participants of the prior CRM training. The train crew team received commendations from the CEO of the Hong Kong Special Administrative Region government.

Post CRM Training Observations.

Some points of interests have also been noted in this research that may hold relevance for further studies. Firstly, train crews with more working experience/ qualifications were noted to have better knowledge of concepts and work related issues, and presented themselves as more emotionally stable in incident handling (Tsang, 2007). Therefore, it is probably of benefit to the team to assign them as team leaders such that they may remain calm in the face of danger and issue strategic orders for the rest of the team. Observations and reports also revealed that, as a whole, female train crews have better written and verbal communication skills, being able to describe situations more vividly to ensure others have a clear understanding, making them ideal candidates at the Incident Control Point (ICP) to observe and report to internal and external parties of any incidents. Train crew with a better education background and qualifications (see table 1) displayed better cognitive skills to deal with the human machine interface. This may be due to their prior familiarity with computerized equipment and procedures. This means that they are able to learn and remember equipment/machinery operation procedures more quickly and be at ease in such posts. The video aided training (VAT) portal, as part of CRM training platform is now becoming a knowledge portal in other KCRC railway lines such that they are able to draw upon the experiences and knowledge from West Rail. The VAT can now be accessed through KCRC intranet, serving as an effective tool to refresh and aid crew in regularly refreshing learnt concepts after the training.

Discussion and Conclusions

The KCRC CRM programme was designed to introduce models of effective CRM behaviours. Such positive CRM models allowed the crews to engage in role-playing exercises during which such behaviours were practiced, and provided feedback to the crews with respect to their performance. As a result of the training and team building exercises, there is now increased synergy between crew members and most can work better and more effectively as a team. Radio communication quality had increased amongst West Rail members compared to other CRM-untrained KCRC divisions. Trainees' concepts of team coordination have been reinforced as reflected in the post training questionnaire and also on-the-job performance involving communication and coordination between train drivers, station controllers and traffic, better incident handling and the ability to reduce operational impacts due to equipment failure.

Therefore, this development and implementation of the CRM program in KCRC represents a significant example of team performance research being translated into practice in an industry other than aviation as long as the training materials can be customized for the required domains, on the basis of requisite psychological research. Apart from the training materials, the effectiveness of CRM training also relies on suitable training facilities, and the existing CBT, VAT, DSS and ITF facilities were all found to be valuable to achieving training objectives. Equally important, integrated training based on CRM and LOT will help to narrow if not bridge the gap between the performance of experienced and inexperienced, as well as highly educated versus fairly educated employees.

From the organizational point of view, the implementation of CRM has put the focus on expecting better performance. There is an explicit tendency of working towards a higher level of safety and risk awareness in KCRC work culture, as well a higher standard of occupational safety and health at the workplace. The adaptation of CRM concepts signify a new era in the KCRC training philosophy, and it is expected that more scenario-based training under the platform of CRM will soon be evolving and emerging. Positive results and commendations of teams trained by CRM resulted in increased commitment from senior management to such applications with the commitment of increasing the safety level and confidence in service reliability to the stakeholders.

This trial of CRM revealed that set on top of traditional training, the effectiveness of CRM has been strengthened and the coverage has been widened giving the crew a more in-depth realization to various incidents and the efficiency of team work as a whole. Although the adoption of CRM training in KCRC is still in its infancy, it is too early at this stage to draw a definitive conclusion of how successful it will be in the long run. However, with results demonstrated so far across a variety of measured and real life performance indicators, the training had help to satisfy the safety centric needs demanded in modern railways and hopefully improving human performance when emergency situations do arise.

References

Andersen, V. & Bove, T. (2000). A feasibility study of the use of incidents and accidents reports to evaluation effects of team resource management in air traffic control, *Safety Science, 35*, 87-94.

Dedale Asia Pacific (2006). *National rail resource management project: Review of best practice implementation issues and task analysis*. Interim Report. Dedale: Albert/VIC, March 2006.

KCRC (2006). *KCRC West Rail Performance Report 2005*. KCRC: Hong Kong.

KCRC (2007a). *KCRC West Rail Performance Report 2006*. KCRC: Hong Kong.

KCRC (2007b) *KCRC Annual Report 2006* (pp 23-30). KCRC: Hong Kong.

Morgan, C.A., Olson, L.E., Kyte, T.B. & Roop S.S. (2007). *Railroad crew resource management (CRM): Pilot rail CRM training development and implementation*. Final report. US DOT, Federal Railway Administration. DOT/FRA/ORD-07/03.

RSSB (2006). *Understanding human factors – a guide for the railway industry.* CD ROM – Rail Safety and Standards Board: UK.

RSSB (2007). *Annual safety performance report*. Rail Safety and Standards Board: UK.

Tsang, M.T.S. (2007). *Rail human factor training: Adaptation of Crew Resource Training in KCRC to enhance modern railway safety*. Master thesis. Massey University: Auckland/NZ.

Wilson, J. & Norris, B. (2005). Rail human factor: Past, present and future, *Applied Ergonomics, 36,* 649-660.

Wilson, J., Norris, B., Clarke, T. & Mills, A. (2005): *Rail Human Factors Supporting the Integrated Railway*, Ashgate, USA, 3-19, 41-50, 309, 319-331, 93-405.

Wright, M. Turner, D., Antonelli, A. & Bendig, M. (2004). The impact of experience erosion on safety *Final Report for RSSB* 20-T153-GSBI-55-TRT, RSSB UK.

THE MANAGEMENT OF VERBAL COMMUNICATIONS IN COMPLEX AERONAUTICAL SYSTEMS

Fanny Loyau
Marie Rivenez
Sophie Quiblier
French military institute of biomedical research - IRBA
Brétigny Sur Orge - France

This study addresses the question of human factors in verbal communications in AWACS (Airborne Warning and Control System). It aims to improve selection, training and learning by developing a tool focused on attentional division and multiple communications management. Radio communications management seems to be highly dependent on crew expertise. Indeed, many instructors say that several years of experience on AWACS are required in order to develop communications management skills and to improve their intelligibility. On the basis of our preliminary observations, we propose three hypotheses which would account for this expertise effect, focused on automatisms, volumes management of radio networks and attentional division. In the first hypothesis we assume that experts would benefit from more attentional resources for communications management compared to beginners because experts automate technical tasks. A laboratory experiment was designed for testing this hypothesis. Our goal is to improve technical training for operators in communications management.

Aeronautical military systems are increasingly complex. They involve many participants, fighter pilots, transport pilots, and air traffic controllers. They have a common tool, verbal communication, which is crucial to mission success and to maintaining a high level of safety. But verbal communication is still a human tool. It requires good transmission and good understanding. Thus, there is an element of uncertainty, which can lead to misunderstandings, impairment of the mission, incidents or accidents.

A Pre-Study: Context and Hypothesis

A preliminary study revealed difficulties associated with radio communications in AWACS (Airborne Warning and Control System), integrated into the aircraft E-3F Boeing 707 (Picture 1). These planes belong to the French Air Force and are part of the EDCA (« Escadron de Détection et Contrôle Aéroporté ») on the Avord base. Their mission is to monitor and control air (and sometimes sea) operations through detection and data transmission. The crew of an AWACS consists of 18 persons, 4 members for the flight crew and 14 for the mission crew. This mission crew performs several functions: control, surveillance and technology.

Picture 1. An E-3F Boeing 707 AWACS

All crew members express the same difficulty: understanding verbal communications, because of the large number of personnel on board associated with external partners. This generates many verbal interactions, much audio information in addition to non-verbal sounds such as alarms, and noise from the environment: engine noise and aerodynamic noise. This can lead to misunderstanding, loss of information, particularly for novices. According to instructors, expertise plays an important role in the ability to manage voice communications through the development of skills, including technical tasks routinization.

Routinization of Technical Tasks

Characteristics of the AWACS Sound Environment

The sound environment in AWACS is composed of simultaneous sounds and messages. This leads to the phenomena of masking, energetic and informational, which causes degradation of speech intelligibility. Energetic masking results from spectral and temporal superimposition of simultaneous speech signals (Moore & Glasberg, 1987; Brungart, 2005). When this occurs it is impossible to understand one of the signals. This type of masking cannot be reduced by technological means. However, our goal is not to act on the interface, but on the operator and practices he uses to remove ambiguities. We do not therefore look at this type of masking. Informational masking occurs when relevant and interfering messages are audible, but auditors are not able to understand the relevant message while ignoring interfering messages (Brungart & Simpson, 2002).

Operators' Adaptation

In the context of AWACS, there are several ways of minimizing this informational masking, including routinization: Operators reduce the attentional load required for technical tasks by automating these tasks in order to provide more resources for communications management. The principle of routinization is a saving of cognitive resources through the development of automatic skills operation. More resources are then available to perform another task in parallel. This is a standard component of expertise (Amalberti, 1996). Indeed, AWACS instructors say that with experience they have more attentional resources to manage verbal tasks because technical tasks, with routinization, do not require much attention. In contrast, novices are more likely to favor one of these tasks at the expense of another.

This can be explained by the fact that the existing training tool does not assess the level of routinization achieved by staff and manages very little verbal communication.

A Double Task Experiment

We would like to develop a tool that allows novices both to carry out routinization of technical tasks, and to further improve their management of verbal communications. This tool should be an indicator of routinization level and contribute to this practice, with the hypothesis that a double learning task is more efficient than a single task. A laboratory experiment was set up, consisting of a routinization task and an intelligibility task, which could simulate technical tasks and operator communications in real situation.

The Technical Task

The technical task should be a slow routinization task, i.e. a task that becomes increasingly automatic as learning proceeds. As part of AWACS, it corresponds to the routinization of tasks by experts but not by novices at the end of their practical training. It is the only type of routinization (as opposed to rapid routinization and non routinization) for which training will enable novices to be operational on early missions. The technical task can be simulated by using a protocol which has been tested several times and has a good indicator of the degree of routinization (Amato, 2005; Bourgy, 2007).

This task was a simulation of piloting at Charles de Gaulle Airport. The operator had to learn the procedure without making mistakes: he had a plan, which we asked him to ignore as soon as possible. The route consisted of several views of the airport. In each case, there were several possible directions, only one was the right direction. By simply clicking in the desired direction he moved forward. Each test resulted in a score measured in route time and in error percentage. This task was considered as a routinization task when the error percentage was virtually zero and when the route completion time was fast and stable.

The Verbal Task

The verbal task in this experiment should interfere with a non-routinization task and interfere little with a routinization task. It was built on the premise that the intelligibility of a speech message can be measured using the Coordinate Response Measure (CRM), developed by Moore (1981) and tested by Brungart (2001). This tool enables evaluation the degradation of intelligibility due to informational masking.

The CRM is a verbal instruction asking the participants to choose a number of a certain color on a screen. The message is in the form "Ready [call sign], go to [number][color] now"(e.g. "Charlie Ready, go to green four now »). It is presented at the same time as one or several other messages of the same form but with different call sign, number and color. Listeners recognize the relevant message through the target call sign (in our example, "Charlie"). The ability to follow the relevant message is determined by measuring the percentage of attempts in which subjects select the number specified in the relevant message.

In our experiment, this methodology was used in part: verbal stimuli were 16 phrases in English, from the combination of a call sign ("Charlie"), 4 colors ("red", "blue", "green" "white") and 4 numbers ("one", "two", "three", "four") pronounced by a woman. Each sentence takes approximately 1500 milliseconds. Participants give their answers in a box with 4 colored buttons (red, blue, green, white) and 4 white buttons showing numbers (1, 2, 3, 4). Each test is used to obtain a score of intelligibility, as measured by percentage of correct answers (correct identification of color and number).

Procedure

The experiment is conducted in two phases: (1) - a learning phase and (2) - a test phase.
(1) - The first phase (single task phase) is to learn the route and, separately, to learn the verbal task. A group of participants should make 54 route tests in order to achieve routinization of the task (routinization group). Another group of participants should make only 9 route tests in order to just familiarize themselves with the task (familiar group). Both groups were subjected to 15 sets of 100 sentences (1500 MRC phrases) to familiarize themselves with the use of the answer box.
(2) - The second phase (double task phase) is the completion of the route simultaneously with the verbal task. All participants perform 63 tests in the double task.

Participants

Twenty-four naive participants were involved in the experiment, 14 women and 10 men, for a mean age of 32 years. Half of them were placed in the routinization group and the other half in the familiar group. These subjects had no hearing trouble, and their laterality was respected: they used the mouse with their dominant hand.

This protocol is used to determine the best method of learning (associate the verbal task from becoming familiar with the technical task or after routinization of the technical task). The best way of learning is obtained by comparing the number of tests necessary to obtain equivalent performances between routinization and familiar groups.

Materials

Participants were in a soundproof room, facing 4 screens on a desk. These screens were connected to a computer running Windows. To respond, the subjects had a mouse and an answer box with standard key spacing. The experimenter had a computer with Matlab, to load the sounds for the intelligibility task, and a TDT, a processor used to start each sound and to collect responses from the answer box.

First results

Before discussing the results themselves, it was necessary to examine certain pre-conditions.

First, group homogeneity was verified using the single task. There was an equivalency of route time (for the first nine) (Figure 1). It's in line with what we expected in that we assumed that the groups had similar performance to the technical task because they are based on the same level of experience, which is zero.

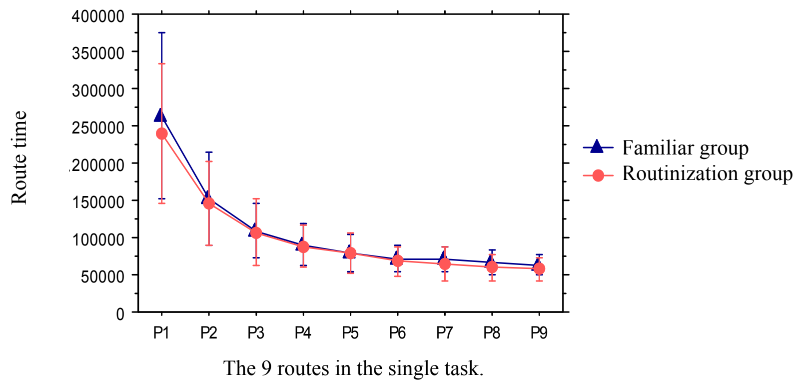


Figure 1. Route times for the nine first routes.

For performance of the MRC task (Figure 2), there is a difference between the groups: the familiar group has the lowest performance at the MRC in single task. This performance difference may simply be due to chance. It may also be that the technical task, carried out before the single task, is longer for the routinization group, which could lead to a difference of concentration or motivation.

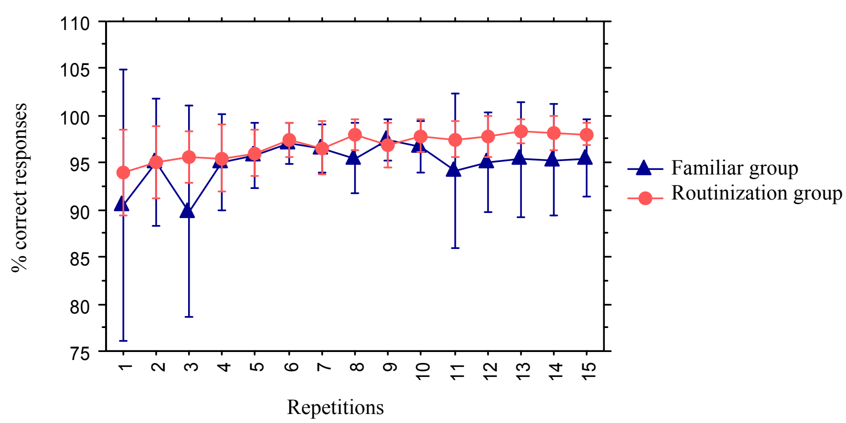


Figure 2. Performance of the MRC task.

Second, in the double task, performance of the MRC task was lower for the familiar group (Figure 3). In order to compare the participants in terms of attentional resources allocation to the verbal task, it is necessary they have the same performance. For this, only the rates of correct responses above 95% were retained. There is a very significant effect of repetition for the two groups: more they make routes, more they are progressing. There is also a significant effect of group: the routinization group is significantly better than the familiar group.

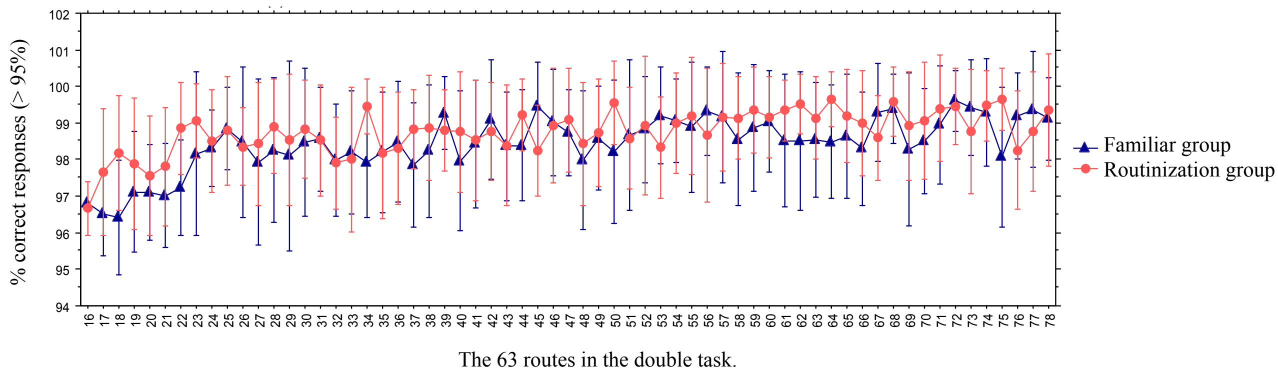


Figure 3. Performance of the MRC task in the double task

Conclusions

Prerequisites have been verified; the homogeneity of the groups for the technical task was confirmed. In contrast, performance for intelligibility on the double task could not be verified. Indeed, the familiar group focused less on the intelligibility task and therefore spent more time on the path times.

The next step of the study will consist of the statistical analysis of results.

On the one hand, the cost of the MRC task will be determined. Our hypothesis is that, during the transition from single task to double task, there would be a cost of the MRC task for the familiar group, and it would be low or even zero for the routinization group, ie the familiar group would take more time than the route time for the single task.

On the other hand, the contribution of learning on the single or the double task will be established. Our hypothesis is that learning on the double task is more effective than on the single task. The question is to determine how many repetitions it took the familiar group to achieve the performance of the routinization group, if they achieve this performance.

Perspectives

The tool used in our experiment could be an indicator of the level of routinization, if it interferes more with a non routinization task than a routinization task. This tool could be improved in order not to interfere at all with a routinization task.

In the longer term, the applied aim of the project is the creation of an operational tool for the training of the AWACS mission crew, which implies respect for the ecology of the technical task. Two possibilities are envisaged: either the addition of the verbal task to the existing simulator, or the creation of a simulator dedicated to training on the double task. A feasibility study will be necessary.

We shall then consider two others aspects of expertise in relation to the management of verbal communication in AWACS: strategies for the management of sound levels, and division and focalization abilities of attention.

Acknowledgments

The authors would like to thank Véronique Chastres for her helpful for the statistical analysis.

References

Amalberti, R (1996). La conduite des systèmes à risques. Paris: PUF, Le Travail Humain.

Amato, J.-N., Ardouin, J.-C. & Amalberti, R. (2005). Le contrôle cognitif des activités routinières en situation dynamique : cadre théorique et montage expérimental. Rapport de recherche. Institut de Médecine Aérospatiale du Service de Santé des Armées. Paris, France.

Bourgy, M. (2007): Le contrôle cognitif des activités routinières en situation dynamique : étude expérimentale des régulations implicites liées à l'humeur du sujet. Rapport de recherche. Institut de Médecine Aérospatiale du Service de Santé des Armées. Paris, France.

Brungart, D. S. (2001). Evaluation of speech intelligibility with the coordinate response measure. *Journal of the Acoustics Society of America*, 109, pp. 2276-2279.

Brungart, D. S. (2005). Informational and energetic masking effects in multitalker speech perception. In P. Divenyi (Ed), Speech separation by human and machine, Kluwer Academic Press: USA, pp. 261-267.

Brungart, D. S. & Simpson, B. D. (2002). Within-ear and across-ear interference in a cocktail-party listening task. *Journal of the Acoustics Society of America*, 112, pp. 2985-2995.

Moore, T. J. (1981). Voice communication jamming research. *In AGARD Conference proceedings 331*: Aural Communication in Aviation, Neuilly sur Seine, France.

Moore, B. C. J. & Glasberg, B. R. (1987). Factors affecting thresholds for sinusoidal signals in narrow-band maskers with fluctuating envelopes. *Journal of the Acoustics Society of America*, 82, pp. 69-79.

USING MULTIPLE IMPERFECT DIAGNOSTIC AUTOMATION

David Keller
Stephen Rice
New Mexico State University
Las Cruces, New Mexico

Due to the difficulty of performing multiple tasks, operators in complex environments are often aided by automation. Because automation is not always perfect operators must decide how much to trust in and depend on the automation aids. Theoretically operators can adjust their level of trust using either a component-specific or a system-wide trust strategy. This study tests these two theories. 36 participants monitored two gauges, each with an automated aid at different reliability levels, while engaged in a pursuit tracking task that simulated an unmanned aerial vehicle (UAV) mission flight. The data suggest that participants do not evaluate the reliability of each gauge independently (i.e. component-specific trust), but instead combine their experience from each automation aid and derive one overall perceived reliability value for the entire system consistent with the system-wide trust hypothesis.

Operators when engaged in complex environments are frequently responsible for multiple tasks (Dixon & Wickens, 2006). Often the number or complexity of the tasks can not be completed safely or efficiently by the operator alone (Sheridan, 1987). Therefore, automation has become prevalent to aid the operators in completing these tasks.

Automation can be classified into four different stages that perform synthesis, diagnosis, response selection and response execution functions (Parasuraman, Sheridan & Wickens, 2000). This study focuses on diagnostic automation. One goal of diagnostic automation is to alert operators of relevant information (e.g. Wogalter & Laughery, 2006) to ensure that the important information is quickly processed (Wickens & McCarley, 2008). The addition of a diagnostic aid allows operators to perform difficult multiple tasks without the need to constantly switch attention between automated and non-automated tasks.

Unfortunately, in many cases automation is not perfectly reliable. Therefore it is up to the operator to decide how much to trust in and depend on the automation aids. Thus, it is important to understand how trust works, and what factors moderate how operators determine how much they will use an automated aid. Much is known in regards to trust in diagnostic automation for single automated aids (e.g. Dixon & Wickens, 2006; Dixon, Wickens, & McCarley, 2007; Lee & Moray, 1994; Parasuraman, Sheridan, & Wickens, 2000; Rice, in press; Rice, Clayton & McCarley, in press; Rice & McCarley, 2008; Rice, Clayton, Wells & Keller, 2008; Rice, Hughes, McCarley & Keller, 2008; Rice, Trafimow, Clayton & Hunt, in press). However, less is known about trust in multiple automated aids. The current study focuses on these issues.

Some researchers have used the term *focus of trust* when discussing trust in multiple agents. The focus of trust can be described by the level of detail that the trust is centered on, which varies from trust in a specific agent to general trust (Couch & Jones, 1997). Because so much research has been done using only one automated aid, researchers and system designers may believe that operators are able develop trust in individual aids which we call component-specific trust. If this were the case then operators would have the ability to evaluate the reliability of an individual aid independently of the performance of the other aids. Therefore we would hypothesize that an operator using component-specific trust, using two aids of different reliability levels, will use each aid differently in accordance to how reliable the aids really are.

However, given the complexity of using multiple automation aids it is also possible that operators may actually rely on a more general focus of trust at the system-wide level. At the system-wide level of trust, operators would determine the reliability of each aid based on the performance of the entire system. Operators using system-wide trust would combine experiences from each automation aid and derive one overall perceived reliability value for the entire system. If this were the case, operators would respond to each individual alert according to the overall perceived system reliability. We would hypothesize then, that when using two aids each with different reliability levels, the operator will use both aids the same as if they both had the same reliability level.

The current study tested these two hypotheses against each other by having participants monitor two gauges, each with its own diagnostic aid. The reliability of each aid varied from perfectly reliable (100%) to moderately reliable (85%) to fairly unreliable (70%). Participants were also responsible for a concurrent pursuit tracking task that simulated a typical UAV flight task while increasing overall workload.

Method

Participants

36 participants (17 females) from New Mexico State University participated in the experiment. The mean age was 20.1 (SD = 2.89) with a range from 18-29.

Apparatus and Stimuli

The experimental simulation ran on a Dell Vostro 200 computer, with a 21” Dell monitor using 1600 x 1200 resolution. The experimental display (see Figure 1) consisted of two areas to allow for both the tracking and monitoring tasks. The tracking task was performed using the top portion of the display. The tracking task consisted of participants controlling a crosshair image to “track” a computer controlled aircraft image on the display. Participants controlled the crosshair image using an Attack3 Logitech joystick. The program was designed so that participants had to exert constant feedback on the joystick. If participants did not, the crosshair image would drift toward the outer edges of the display. The computer controlled aircraft moved around on the screen in a random pattern. The simulation program determined the random movements of the aircraft image by randomly selecting a new movement direction (up, down, right, left, up-right, up-left, down-right, down-left) roughly every second. The tracking task was set against a backdrop consisting of clouds and blue sky (see Figure 1).

Figure 1. Screenshot of experimental display.

The monitoring task consisted of participants monitoring two gauges for system failures. The two gauges were located at the bottom center of the display. Each gauge consisted of 10 black numbers equally spaced around the inside of a circle. The numbers were displayed sequentially (clockwise) around the circle, starting with the number 0 at the top center of the circle and ending with the number 9.

The values of the gauges were represented using two needles (one black and one red). The black needle represented units of 1000. The red needle denoted units of 100. Thus, the values of the two gauges depicted in Figure 1 are approximately 3,690 for the gauge on the left and 1,290 for the gauge on the right. The movement of the black needle was driven by the sum of four sine waves ranging in bandwidth from 0.04 to 0.43 Hz. The movement of the red needle was dictated by the movements of the black needle. As the black needle oscillated back and forth in its random patterns, the red needle followed in a linear fashion just as the minute hand moves in accordance to the hour hand on an analog clock.

Above each gauge were two boxes, outlined in black, with numerical values. The number in the left box indicated the ideal value for a safe system. The number in the right box indicated the range of safety for the system. Thus, for the example shown in Figure 1 the gauge on the right indicated a safe system as long as the needles stayed within 3800 ± 400 (3400-4200). The gauge indicated a system failure (SF) if the needles went out of this range. If the participant believed that a SF had occurred, the participant was expected to press the appropriate button on the joystick as quickly as possible. When a SF occurred, the needles stayed out of the acceptable safe value range until the trial ended.

All participants were aided on the monitoring task by an automated aid for each gauge the participants had to monitor. The automated aids sounded an auditory alert (i.e., a synthesized human voice pronouncing the word "alert one" for left gauge and "alert two" for the right gauge) when they detected a SF. The automated aids were 100%, 85%, or 70% reliable. The automation aids, expressed in the framework of signal detection theory, provided hits, false alarms (FA), or correct rejections (CR). Each aid, regardless of reliability, made 20 hits during the 100 trial experiment. The perfectly reliable aid made 80 CRs, and 0 FAs. The 85% reliable aid made 65 CRs, and 15 FAs. The aid that was 70% reliable made 50 CRs and 30 FAs.

Trials

There were 100 experimental trials that each lasted 30 s. At the beginning of each trial, the target safe value changed to a new random value between 1,000 and 9,000, rounded to the nearest 100. The target safe range also changed to a new random value between 100 and 900, rounded to the nearest 100. Also at the beginning of each trial, the SF gauge itself reset to the target safe value and then immediately began oscillating. SF and non-SF trials were randomly ordered. SFs and automation FAs always occurred within a temporal window beginning 5 s and ending 12 s from the start of the 30 s trial interval, thus giving the participant at least 18 s to detect the failure and respond. Only one gauge at a time indicated a system failure and there was never more than one SF or automation alert per trial. Trials lasted the entire 30 s, regardless of whether or not a SF occurred or was detected. During each trial, participants were allowed to make only one SF response and were not allowed to change a response. Once a 30 s trial ended, participants were no longer able to respond to that particular trial. At the end of each trial, participants were informed if their response was correct (green border flash) or incorrect (red border flash).

Design

The experiment used a 2 x 3 factorial mixed design with Automation Reliability as the within-participant variable and Condition as the between-participant variable. This design resulted in three experimental conditions: (a) 100/100 – consists of both automation aids being 100% reliable; (b) 85/100 - consists of one imperfect aid at 85% reliable and one perfect aid; (c) 70/100 – consists of one imperfect aid at 70% reliable and one perfect aid. The reliability level of the alerts was counterbalanced between the left and the right gauges.

Procedure

Participants were first asked to read and fill out a consent form. Participants then received extensive verbal instructions, followed by a 20-trial practice session. Participants were told that both tasks (tracking and monitoring) were of equal importance. After the practice trials, participants then were told about the automation aids. Each participant was told that the automation might or might not be perfectly reliable and that the imperfect automation would err by producing false alarms (no misses). Participants were also told that the aids worked independently of each other and that the reliability of each aid might differ from each other. They were not told exactly how reliable the aids were. Participant completed a total of 100 experimental trials, after which they were debriefed.

Results

The following dependent measures were analyzed: SF sensitivity (*d'*), response time (RT), and tracking error (TE). Overall ANOVAs were performed on the data, followed by post hoc comparisons when necessary. The independent variables included Reliability and Condition. Reliability consisted of each participant viewing one perfectly reliable aid and one imperfect aid (except in the condition where both aids were perfect) at the same time. The Condition variable consisted of the imperfect aid (from the automation reliability variable) having a reliability level of 100%, 85% or 70%.

Sensitivity

SF sensitivity was assessed using the signal detection measure *d'*. Perfect scores (e.g., zero operator misses or FAs) were adjusted by assuming ½ a miss or FA. These data are presented in Figure 2. A two-way ANOVA using Condition (70/100, 85/100, and 100/100) and Reliability (whether or not the automation was perfect or imperfect) revealed a main effect of Condition, $F(2, 33) = 12.41$, $p < .0001$, $\eta^2 = .43$, no main effect of Reliability $F(1, 33) = 0.22$, $p = 0.644$, $\eta^2 = .01$, and no interaction between Condition and Reliability, $F(2, 33) = 0.56$, $p = 0.579$ $\eta^2 = .03$. These results are consistent with the idea that participant trust in the perfectly reliable gauge was reduced as a function of it being linked with an imperfectly reliable gauge (i.e. system-wide trust).

Post Hoc comparisons revealed that participants' sensitivity was significantly lower when using perfect aids in the 85/100 condition, $t(22) = 2.85$, $p = .009$, $\eta^2 = .27$, and the 70/100 condition), $t(22) = 3.79$, $p = .001$, $\eta^2 = .40$, when compared to the corresponding aid in the 100/100 condition (see Figure 2). When linked with an imperfectly reliable aid, the perfectly reliable aid suffered compared to an identical perfectly reliable aid that was linked with another perfectly reliable aid. This effect was both statistically and practically significant regardless of whether the imperfect aid was 85% reliable or 70% reliable.

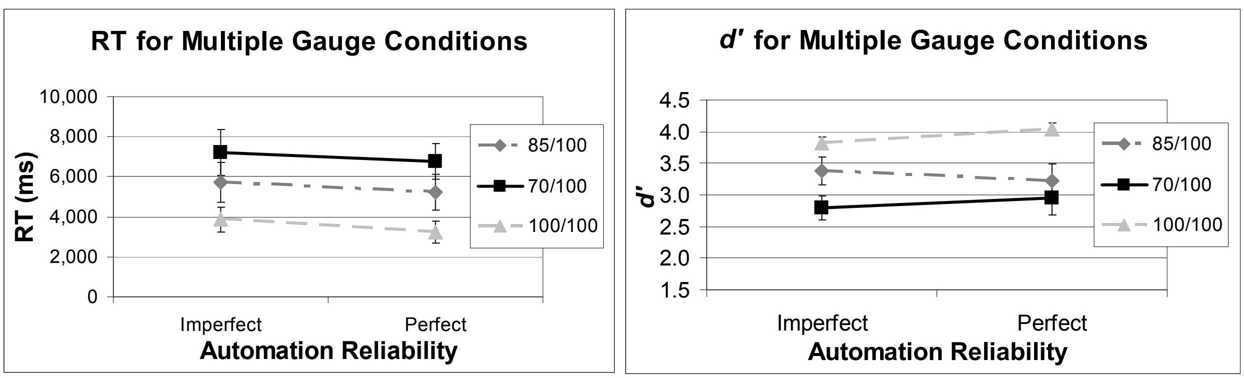


Figure 2. RT and *d'* as a function of Condition and Reliability. SE bars are included.

RT

RT was measured from the time that a system failure occurred to the time the participant responded by pressing the appropriate button on the joystick. A two-way ANOVA using Condition (70/100, 85/100, and 100/100) and Reliability (whether or not the automation was perfect or imperfect) revealed a significant main effect of Condition, $F(2, 33) = 4.15$, $p = .025$, $\eta^2 = .20$. There was also a significant main effect of Reliability, $F(1, 33) = 4.41$, $p = .043$, $\eta^2 = .12$, which was not entirely unexpected since one would expect that the perfectly reliable aid would facilitate quicker RTs than would the imperfectly reliable aid. However, because this finding did not address any of our theoretical points, we did not pursue further analyses on this effect. Lastly, there was no interaction between Condition and Reliability, $F(2, 33) = .04$, $p = .963$, $\eta^2 = .002$, (see Figure 2) a finding consistent with the data from *d'*, which indicated that when linked with an imperfectly reliable aid, the perfectly reliable aid suffered compared to an identical perfectly reliable aid that was linked with another perfectly reliable aid.

In addition, post hoc comparisons revealed that participants' RT was significantly slower when responding to SFs with perfect aids in the 85/100 condition, $t(22) = 1.85$, $p = .038$, $\eta^2 = .14$, and the 70/100 condition, $t(22) = 3.35$, $p = .003$, $\eta^2 = .34$, when compared to the corresponding aid in the 100/100 condition (see Figure 2). Again, performance suffered for perfectly reliable aids when paired with an imperfect aid compared to performance when both aids are perfect. This effect applies when the imperfect aid was 85% reliable or 70% reliable.

Overall Tracking

Tracking error was calculated using the Root Mean Square (RMS) error, which was defined as the distance between the positions of the controlled UAV image and the chase plane image. A one-way ANOVA with condition as a between-participants factor revealed no effects of condition, $F(2, 33) = .48$, $p = .6204$, $\eta^2 = .003$. These data alleviate any concerns about a tracking/monitoring performance tradeoff. Only the monitoring task was affected by the experimental manipulations, while the tracking task remained stable, presumably because operators were "protecting" the tracking task, as they should in aviation.

Discussion

Theoretical Implications

The current study tested two competing theories of trust when using multiple aids. The component-specific theory suggested that operators would determine the reliability of each automated aid separately and therefore use them differently. System-wide trust predicted the opposite, which is that operators would determine the reliability of each aid according to the performance of the overall system and not according to the performance of each individual aid. Therefore operators using system-wide trust would treat each aid the same as if they both had the same reliability.

The results show that system-wide trust best predicts how operators, in the context of this study, use systems with multiple automation aids. Overall operator performance on the monitoring task (*d'* and RT) declined, when using an aid that was 100% reliable along side an imperfect aid (see Figure 2). In addition, performance was no different for both aids, when operators were using one aid that was perfect at the same time as another aid that was imperfect. This effect was evident for both *d'* and RT, indicating that there was no speed-accuracy tradeoff. Likewise, as the tracking error did not differ significantly between conditions, we can conclude that there was no tradeoff between the tracking task and the monitoring task, indicating that operators possibly protected the tracking task at the expense of the monitoring task.

Theoretical Limitations. The theoretical limitations of this study are fairly straightforward and suggest future research. First, it is premature to generalize these findings to other paradigms because the nature of the current task is highly specialized. Second, the current study only employed FAs, and not automation misses. Third, because both SF gauges were identical and close together, it could be the case that the physical characteristic of the gauges generated a stronger trust merging behavior than one would expect from dissimilar SF gauges.

Practical Implications

From a practical perspective designers and operators of systems with multiple aids should be aware of how operators may spread their trust across the different automated components, especially when each aid differs in reliability. One would hope that operators can and do treat each aid differently according to the true reliability of each aid. However, as this study demonstrated, it is possible for the performance of one automated aid to significantly affect how an operator will treat other aids in the system. One obvious problem of system-wide trust is that if operators determine their use of an aid based on the overall system performance, instead of on the reliability of the individual aid, operators may lose trust and therefore disuse highly reliable automation aids, as was shown in this experiment. The disuse of the reliable aid could entail unnecessary monitoring and/or inappropriate overruling of the perfectly good automated aid (Muir & Moray, 1996). The disuse of automation will not only hinder performance on the task aided by the highly reliable aid but will take valuable attentional resources of the operator away from other concurrent tasks causing a decrease in performance there as well.

Conclusion

This study demonstrated operators may very well use system-wide trust when using a system with multiple automated aids. The results showed that performance (RT and *d'*) suffered when using a perfectly reliable aid when the operator was also exposed to an imperfect aid. Due to system-wide trust the imperfect aid penalized trust in the perfectly reliable aid. Operators and system designers should be aware of the possibility of system-wide trust in their systems and carefully consider the consequences.

Acknowledgments

The authors wish to thank Nathaniel Bean for his help in collecting data and Amjad Nusayr, Raja Kothembaka, and Shane Thomas for their help in programming the simulator. This research was funded by an Air Force Grant (Index #111915). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors.

References

Couch, L., & Jones, W. (1997). Measuring levels of trust. *Journal of Research in Personality, 31,* 319-336.

Dixon, S., & Wickens, C. (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., Wickens, C., & McCarley, (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors, 49*(4), 564-572.

Lee, J. & Moray, N. (1994). Trust, Self-Confidence, and operator's adaptation to automation. *International Journal of Human-Computer Studies, 40*(1), 153-184.

Muir, B., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation, *Ergonomics, 39*(3), 429-460.

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. (in press). Examining single and multiple-process theories of trust in automation. *Journal of General Psychology*.

Rice, S., Clayton, K. & McCarley, J. (in press). The Effects of Automation Bias on Operator Compliance and Reliance. *Human Factors Issues in Combat Identification*.

Rice, S., Clayton, K., Wells, A. & Keller, D. (2008). Manipulating Trust Behaviors in a Combat Identification Task. *Human Factors in Combat Identification.*

Rice, S., Hughes, J., McCarley, J. & Keller, D. (2008). Automation Dependency and Performance Gains Under Time Pressure. *Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: Human Factors and Ergonomics Society.

Rice, S. & McCarley, J. (2008). The Effects of Automation Bias and Saliency on Operator Trust. *International Congress of Psychology*.

Rice, S., Trafimow, D., Clayton, K. & Hunt, G. (in press). Impact of the contrast effect on trust ratings and behavior with automated systems. *Cognitive Technology Journal*.

Sheridan, T.B. (1987). *Handbook of Human Factors.* New York: Wiley.

Wickens & McCarley. (2008). *Applied Attention Theory.* Boca Raton, FL: CRC Press, Taylor & Francis Group.

Wogalter, M. S., & Laughery, K. R. (2006). Warnings and hazard communications. In G. Salvendy (Eds), Handbook of Human Factors and Ergonomics (pp. 889-911). New York: Wiley.

AUTOMATION DEPENDENCY UNDER TIME PRESSURE

Stephen Rice, David Keller, Gayle Hunt, David Trafimow
New Mexico State University
Las Cruces, New Mexico

Previous research has identified many factors that affect human dependence on automated systems. Some of these factors include automation reliability, types of errors, and training. This study introduces a new factor, time pressure, which is directly related to operator dependence on automated aids. Participants were asked to perform a simulated UAV target-detection task with the aid of diagnostic automation. Two factors were manipulated in this study: reliability of the automation and time pressure. The results indicate that participants faced with time pressure were more likely to depend on the automation than participants who had more time to evaluate the recommendations. The subsequent increase in dependence due to time pressure was beneficial to overall performance when the automation was highly reliable. In conditions with low reliability, overall human-automation performance suffered due to time pressure. The results imply a potential technique for eradicating the problem of under-dependence on highly reliable automated systems.

The term automation can be described as a mechanism which serves to substitute or enhance human performance. Recent research indicates that while a human-automation team often outperforms a human alone, it rarely measures up to the sole performance of the automation (e.g. Dixon, Wickens, & McCarley, 2007; Dixon & Wickens, 2006; Rice, in press; Rice, Clayton, Wells & Keller, in press; Rice & Hunt, 2009; Rice & Keller, 2009; Rice, Keller, Hunt & Trafimow, 2009; Rice & McCarley, 2008; Rice, Trafimow, Clayton & Hunt, in press). In short, this finding reveals that automation, when left to its own devices, is often more accurate than a human-automation team where the individual is allowed to override the automation. There has been a growing concern in regards to human under-dependence on automation. Specifically, it may be harmful when humans second-guess automation, because they too often disagree with the automation when it is correct, and too often agree with the automation when it is incorrect.

Based on this information, the most intuitive option may be to take the human out of the equation altogether; however, there could be several serious consequences of this action. For example, there are real-world episodes of automation failure (e.g. the auto-pilot in an aircraft fails), and in this type of situation, it is essential that a human operator is present and able to regain control of the system. Although current research indicates that human-automation teams are typically less accurate overall than automation alone, this does not eliminate the possibility that with practice and training, the human-automation team may eventually become *more* accurate than automation alone. Removing the human from this equation would eradicate the possibility of obtaining performance levels exceeding the abilities of the automation. Thus, it is important to investigate any possibility of advancing human-automation performance, rather than doing away with the human factor altogether.

Automation can be divided into four different stages, modeled after human cognitive processing (Parasuraman, Sheridan, & Wickens, 2000). These stages are information acquisition, diagnosis, response selection, and response execution. For this study, our focus will fall primarily with the second stage, or diagnostic automation, which is a common function found in settings such as Unmanned Aerial Vehicles (UAVs). An example of diagnostic automation would include warning alarms and target-detection.

There are several different ways in which humans can interact with diagnostic automation (Parasuraman & Riley, 1997). One of these ways is disuse, which indicates a neglect of the automation. This often occurs when the automation consistently performs poorly (e.g. high false alarm rates), causing the operator to frequently ignore the suggestions made by the aid. This type of interaction may be beneficial in the situation where the human operator is more accurate than a consistently poor and unreliable automation. On the other hand, disuse can be highly dangerous when a human operator's performance is inferior to the automation, leading to sub-optimal results. In this situation it is imperative to correct this disuse behavior.

In reference to an automated aid, dependence is a behavior that is typically mediated by a subjective assessment of trust (a mental state). Trust in automation can be defined as an attitude indicating one's level of confidence that the aid will be successful in helping reach one's goals (Lee & See, 2004).

There are many factors that may cause a shift in trust, which in turn causes a shift in dependence. One major factor is the accuracy of the automated aid. A study by Parasuraman, Molloy & Singh (1993) revealed that when the automation is exceptionally reliable, a human operator may depend too heavily on it and may fail to detect its infrequent errors. Conversely, Dixon and Wickens (2006) found that when the automation is scarcely reliable, a human operator may refrain from any dependence on the aid, to the extent that they may ignore even its correct predictions.

A second factor affecting the level of trust in automation is the type of error the automation makes; that is, if the automation misses a target or makes a false alarm. According to Meyer (2001; 2004), trust and dependence on automation are affected by which of these two errors the automation is more prone to make. Specifically, when the automation is prone to false alarms, lower operator compliance (response when the automation reports an event) is typically the result. When the automation is prone to misses, lower operator reliance (response when the automation does not report an event) is typically the result. A study conducted by Dixon, Wickens and McCarley (2005) found that false alarms affected reliance as well as compliance, but agreed that the two errors affected dependence in different ways. Two additional studies took these findings one step further, demonstrating that the different error types affected different cognitive processes, which lead to different behaviors regarding dependence (Rice, in press).

It is important to note that trust is not the only factor that may influence operator dependence on automation. For example, if an operator is required to complete multiple tasks simultaneously, she may have no choice but to depend more heavily on the automation in order to keep focus on other tasks. In this case, without a change in trust, dependence has still increased.

Another situation where dependence may be affected without a change in trust is when it is of utmost importance to pay attention to warning signals, regardless of how accurate the automation may be. For example, the Traffic Collision Avoidance System (TCAS) is highly prone to false alarms, but in order to avoid the risk of an airplane collision, operators must respond to each and every warning alarm, even knowing that the TCAS has a high false alarm rate.

Figure 1 demonstrates how various factors may contribute to dependence on automated aids. This model is not inclusive of all factors that may affect dependence. It is simply meant to show how some factors may be mediated by trust (e.g. opacity of automation, prior information, and automation reliability) and others may not.

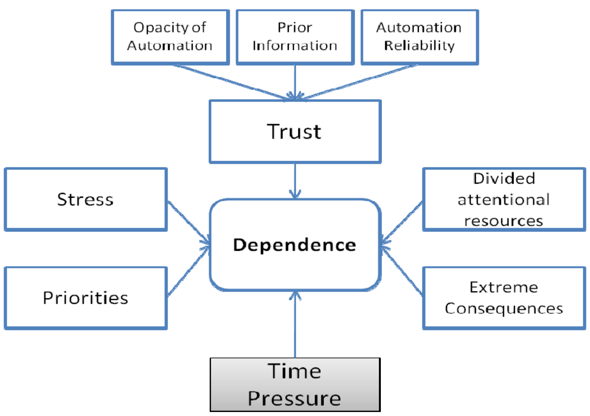


Figure 1. A Model of Automation Dependence.

The purpose of this current study is to introduce another possible factor, time pressure, which may directly influence dependence on automation, while bypassing trust. We believe that when time pressure is introduced (in the absence of outside stressors or multiple tasks), operators will exhibit different behaviors than if there was an abundance of time to make a decision. We predict that those experiencing time pressure will be more dependent on the automation than those with no time pressure. Because the reliability of the automation remains the same across time conditions, there is little reason to suspect that there will be any change in trust, so any difference in behavior should be the direct result of the time manipulation.

Such an outcome may be both beneficial and harmful, depending on the situation at hand. With our manipulations, we plan to demonstrate how automation dependence may greatly influence human-automation performance both for better and for worse.

For this study, two hypotheses are proposed. First, participants given less time to make a decision will depend more highly on the automation than participants given more time to make the same decision. Second, as a result of the increased dependence, participants under a time constraint will perform with greater overall accuracy than their counterparts when the automation is highly reliable, and will have lower overall accuracy when the automation is less reliable.

Methods

260 participants (143 female, 117 male) from New Mexico State University participated in this experiment in exchange for course credit. Participant ages ranged from 17-38, with a mean age of 20.1 (*SD* = 2.63). All participants were screened for normal or corrected-to-normal vision.

The experiment was presented to each participant using E-Prime 1.1 on a Dell computer with a 20" monitor, using 1024 x 768 resolution. E-Prime recorded accuracy and agreement rates for each trial. Images with no target consisted of 50 aerial photographs of Baghdad. Images with a target were created by placing a small tank image onto the 50 aerial photographs using Photoshop CS3. Altogether, there was a total of 100 images to be presented as stimuli—50 with the target present and 50 with the target absent.

An automated aid was used with four different reliability ratings (100%, 95%, 80%, and 65%). Reliability ratings were randomly assigned between subjects and all errors were false alarms. As a control, a condition with no automated assistance was included.

After signing an informed consent, participants were seated 21" from the experiment display with a chinrest controlling head position. Specific instructions were presented on the computer screen and any additional questions were answered by the experimenter. Instructions explained that participants were going to be presented with 100 aerial images of Baghdad and their task was to decide if an enemy tank was present in each image. If a tank was detected, they should respond by pressing the "J" key. If no tank was detected, they should respond by pressing the "F" key. All participants were asked to respond as accurately as possible within their time limits. Participants were informed about the automated aid, given the exact reliability rating, and told that the automation would only err by false alarm. Finally, participants were told that they would have either 2 or 8 seconds to view each image.

Participants were instructed to press any key when they were ready to begin. Each trial began with an automated recommendation stating either "The automation has detected a tank!" or "The automation has determined that there is no tank present!" After 1000 ms, the image was presented. Each image was presented for either 2 seconds (speeded condition) or 8 seconds (unspeeded condition), after which participants made their decision by pressing either the "J" or the "F" key and were then presented with feedback for each trial. Images were presented in random order and each participant viewed all 100 images only once.

Immediately following the completion of the computer portion of the experiment, participants were asked to complete a questionnaire (Dixon & Wickens, 2006) regarding the reliability of the automation and their own trust in the aid.

Including the baseline condition, there were five levels of automation reliability. In addition, there were two time manipulations, resulting in a total of 10 conditions. All subjects were randomly assigned to only one condition in this between-subjects design.

Results

The analyses that follow are separated into two parts: accuracy performance and dependency effects. The percentage of correct trials to total trials was used to measure accuracy. These data can be found in Figure 2. The measure *d'* was not used in this analysis because a very high number of perfect scores were found in the 100% and 95% conditions. However, it should be noted that the effects of *d'* were almost identical to those of accuracy.

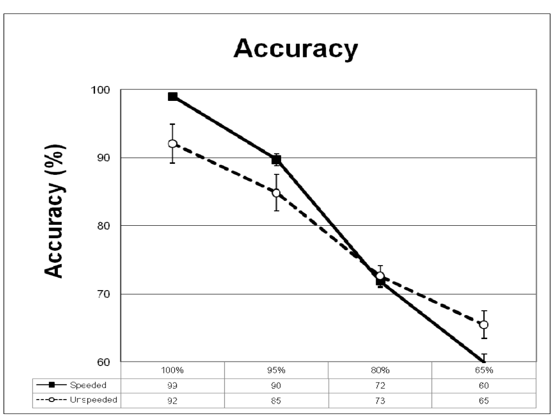


Figure 2. Accuracy Data as a Function of Automation Reliability and Time.SE bars are included.

The first analysis was performed on the Baseline conditions in order to confirm that the given task was challenging enough to warrant the aid of an automated system. It was also important to determine that the time manipulation would cause a tradeoff between speed and accuracy. Indeed, when participants experienced time pressure, their performance suffered tremendously, $t(50) = 4.43$, $p < .001$, $d = 1.25$. In fact, baseline performance in the speeded condition was barely above random performance.

A two-way between-participants ANOVA on the 8 automation conditions, using Reliability and Time as factors, revealed a main effect of Reliability, $F(4, 250) = 161.77$, $p < .001$, $\eta_p^2 = .72$, and no main effect of Time, $F(1, 250) = 1.15$, $p > .10$, $\eta_p^2 = .005$. However, there was an interaction between Reliability and Time, $F(4, 250) = 9.06$, $p < .001$, $\eta_p^2 = .13$, indicating that time pressure was beneficial to general performance at certain reliability levels, but was harmful at other levels, as seen in Figure 3.

Planned comparisons revealed that the 100s (speeded) condition produced higher accuracy than the 100u (unspeeded) condition, $t(50) = 2.43$, $p < .01$, $d = .69$, and the 95s condition produced higher accuracy than the 95u condition, $t(50) = 1.73$, $p < .05$, $d = .49$; however, the 65s condition produced lower accuracy than the 65u condition, $t(50) = 2.32$, $p = .01$, $d = .66$. There was no significant difference between the 80s and 80u conditions, $t(50) = .43$, $p > .10$, $d = .12$. The 95s conditions produced higher accuracy than the 80s condition, $t(50) = 13.77$, $p < .001$, $d = 3.89$, which in turn produced higher accuracy than the 65s condition, $t(50) = 7.65$, $p < .001$, $d = 2.16$.

The rate of participant agreement with the automation was used as a measure of operator dependence, as it is assumed that high dependence on automation indicates high rates of agreement with the automation (Dixon & Wickens, 2006; Dixon, Wickens & McCarley, 2007). It must be noted that high agreement rates to automation hits or correct rejections could possibly be due to high performance both by the automation and the operator independently. In the situation where the automation errs only by false alarm and the operator agrees with this failure, it is only reasonable to assume that this is due to high rates of dependence. For this reason, agreement rates for hits, false alarms, and correct rejections are all analyzed separately.

Planned comparisons revealed that the 100s condition generated higher agreement rates than the 100u condition, $t(50) = 2.51$, $p < .01$, $d = .71$, the 95s condition generated higher agreement rates than the 95u condition, , $t(50) = 1.67$, $p = .05$, $d = .43$. There were no differences between the 80s and 80u conditions, $t(50) = 1.08$, $p > .10$, $d = .31$, or between the 65s and 65u conditions, $t(50) < 1.0$. These data provide some support for the notion that participants under a time constraint tend to have higher agreement rates with the automation.

Planned comparisons revealed that the 100s condition generated higher agreement rates than the 100u condition, $t(50) = 2.33$, $p = .01$, $d = .66$, the 95s condition generated marginally higher agreement rates than the 95u condition, , $t(50) = 1.50$, $p = .07$, $d = .42$. There were no differences between the 80s and 80u conditions, $t(50) < 1.0$, or between the 65s and 65u conditions, $t(50) < 1.0$. These data are consistent with the automation hits data.

Planned comparisons revealed that the 95s condition generated higher agreement rates than the 95u condition, $t(50) = 3.04$, $p < .01$, $d = .86$, the 80s condition generated higher agreement rates than the 80u condition, $t(50) = 2.51$, p < .01, $d = .71$, and the 65s condition generated higher agreement rates than the 65u condition, $t(50) = 1.84$, p < .05, $d = .52$. These data confirm participants typically have higher rates of agreement with the automation when they are under a time constraint even when the automation is incorrect.

Date collected from the trust questionnaires were not surprising, since participants were told exactly how reliable the automation would be prior to beginning their task. Reliability ratings after the experiment did not differ significantly as a function of time manipulation (all *p*s > .10). Furthermore, participants' ratings of general trust in the automation also did not differ significantly as a function of time manipulation (all *p*s > .10). These data provide evidence that levels of trust in the automation were not significantly affected in this study and that the time manipulation did, in fact, affect operator dependence directly.

Discussion

The results of this study have both theoretical and practical implications. As was discussed in the introduction, many factors influence trust, which may have an effect on an operator's level of dependence on automated aids (Dixon & Wickens, 2006; Dixon, Wickens & McCarley, 2007; Parasuraman & Riley, 1997; Rice & McCarley, 2008; Rice, Clayton & McCarley, 2008; Rice, Clayton, Wells & Keller, 2008). This model (see Figure 1) implies that trust (a cognitive process) is a mediator between external factors and dependence on automation (a behavioral response).

There are also many external factors that may affect operator dependence directly, which may also be seen in Figure 1. The current study has introduced a new factor, time pressure, which affects dependence directly and is not mediated in any way by trust.

The results of this study clearly show that when participants were faced with a situation involving time pressure, they were more compliant with the automation than those who had more time to make a decision. While this general finding regarding hits and correct rejections may also be explained by superior performance by both the automation and the operator independently, compliance during trials where the automation produced a false alarm were also higher in the time pressure condition. This finding can only be explained by a higher level of dependence in speeded conditions, regardless of automation reliability.

As suspected, this increased dependence had both positive and negative effects. Positively, when the automation was highly reliable, the overall performance of the human-automation team was improved as a function of the time pressure. The reason for this improvement was clearly due to the added dependence on the automation, which is typically more accurate than the human alone (as discussed in the introduction). The simple addition of a time factor was able to produce significant improvements in human-automation performance.

Negatively, the increased dependence was not specific to the highly reliable conditions. Dependence also increased in conditions with very unreliable automations, leading to poor overall human-automation performance. In these conditions, participants with more time were able to override the automation's suggestions with some degree of confidence and ultimately performed better.

In an effort to ensure that time pressure was not mediated by trust, three deliberate steps were taken. First, participants were clearly told the precise reliability rating of the automation and exactly what type of errors to expect. Second, participants received feedback after each trial in order to allow them to gauge for themselves how accurate the automation really was. Finally, participants were asked to fill out a survey upon completion of the experiment, which ultimately indicated no difference in trust between the speeded and unspeeded trials. Based on this information, it is logical to assume that time pressure, in fact, is not mediated by trust.

There are at least two major practical implications warranted by the results of this study. First, it is essential that designers of automated aids carefully study the environment that their devices will be used in. In an environment with a great deal of time pressure, they must consider that the operator will likely depend highly on the automation, regardless of how reliable it is. If the aid is not highly reliable, a dangerous situation could occur. The opposite is also true. If there is no time pressure, an operator may have ample time to second-guess the automation even if it is highly accurate, in which case a dangerous situation could also occur. Designers must take this into consideration and adjust the environments accordingly as to avoid catastrophic situations.

The second practical application is in the training of operators. When an automation is known to be highly reliable, it may be beneficial to train operators using a time pressure situation in order to increase their dependence on the automation. Should operators learn to depend highly on the automation without constantly questioning its recommendations, the human-automation performance will likely improve. Despite the findings in this study, however, future research should be done in regards to this training technique to discover its long-term effects, especially after the time pressure has been removed. Further, this training method is unlikely to be beneficial in situations where the automated aid is no more reliable than the unaided human. When using low-reliability automated systems, a different training technique should be used to teach operators how to comply most effectively with the automation.

Conclusion

This experiment has shown that the external factor of time pressure causes a higher level dependence on automated aids without affecting operator trust in the automation. It must be highly stressed that this increased dependence is only beneficial when the automation is highly reliable. In this study, performance suffered greatly when the automation had a lower level of reliability. Though previous studies have identified other factors that may increase dependence on automation (e.g. increasing trust), these other factors are very difficult to manipulate in an applied setting. This study has introduced a simple factor that can easily be applied to a situation where operator dependence on automation is necessary and beneficial. To those in aviation and other related fields, it is of utmost importance to be able to detect dangers from an aerial viewpoint very quickly and accurately. Thus, it is essential to further our knowledge and understanding of how operators may triumph over their instinct to second-guess these highly reliable automated systems.

Acknowledgments

The authors wish to thank Amy Wells and Jackie Chavez for their help in collecting data. This research was funded by an Air Force grant (Index #111915). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author.

References

Dixon, S. & Wickens, C. (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., & Chang, D. (2005). Mission control of multiple unmanned aerial vehicles: A workload analysis. *Human Factors, 47*(3), 479-487.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse the misses? *Human Factors. 49*(4), 564-572.

Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, and abuse. *Human Factors, 39*(2), 230-253.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics, 30*(3), 286-297.

Rice, S. (in press). Examining single and multiple-process theories of trust in automation. *Journal of General Psychology*.

Rice, S., Clayton, K. & McCarley, J. (in press). The Effects of Automation Bias on Operator Compliance and Reliance. *Human Factors Issues in Combat Identification*.

Rice, S., Clayton, K., Wells, A. & Keller, D. (2008). Manipulating Trust Behaviors in a Combat Identification Task. *Proceedings of the 2008 Human Factors in Combat Identification Workshop.*

Rice, S. & Keller, D. (2009). Determining reliability of multiple automation aids using system-wide trust: A comparison of four hypotheses. *Proceedings of the 5th International Conference on Technology, Knowledge and Society*.

Rice, S., Keller, D., Hunt. G. & Trafimow, D. (2009). The effects of time pressure on automation dependency. *Proceedings of the 15th Annual International Symposium of Aviation Psychology*.

Rice, S. & McCarley, J. (2008). The Effects of Automation Bias and Saliency on Operator Trust. *International Congress of Psychology*.

Rice, S. & Hunt, G. (2009). Proving a multiple-process model of trust in automation. *Proceedings of the 15th Annual International Symposium of Aviation Psychology*.

Wickens, C. D. & Hollands, J. G. (2000). Engineering Psychology and Human Performance, 3rd Edition. Upper Saddle River, NJ: Prentice Hall.

PREPARING FOR THE FUTURE OF COLLABORATIVE AIR TRAFFIC MANAGEMENT

Philip J. Smith and Charles Billings
The Ohio State University
Columbus, Ohio

One of the key areas identified in the NextGen Concept of Operations focuses on Collaborative Air Traffic Management (CATM). Within the NextGen Concept of Operations, this collaboration is expected to continue to occur in a highly distributed environment where operational staff are distributed across many locations and organizations, including flight operations centers, pilots, the Air Traffic Control Systems Command Center, Enroute Centers, TRACONs and Airport ATC Towers (or their future functional equivalents). Significant changes can be expected, however, due to the introduction of new decision support technologies which will enable different approaches to traffic management. These changes in CATM require careful consideration of human factors issues at many different levels, ranging from strategies for distributing roles and responsibilities among different people and technologies, to the design of sensing, communication and decision support technologies and the associated human-computer interfaces, to the training of new procedures and the tools that support them. In this context, this paper focuses on two issues:

- What operational concepts for CATM are likely to be developed over the next 10-15 years?
- What human factors issues need to be considered in the design and implementation of these new operational concepts?

Current projections indicate that, in order to support the needs and desires of the U.S. economy and air travelers over the next 10-20 years, the National Aviation System (NAS) needs to support two to three times current air traffic levels. In order to meet these projected needs, major technological advances have been proposed. These technologies focus on a variety of areas, ranging from the application of new sensors to provide data about aircraft movement on the airport surface, to the installation and use of more advanced Communication, Navigation and Surveillance (CNS) capabilities to support trajectory based operations, to the introduction of advanced decision support tools and automation to support strategic decision making and tactical operations.

One of the key areas that is critical to safely meeting this need for significant enhancements of the NAS concerns Collaborative Air Traffic Management (CATM). The FAA describes this priority area as:

> "strategic and tactical interactions with the operators to mitigate situations when the desired use of capacity can not be accommodated. CATM solution set includes the flow programs as well as collaboration on procedures that will establish balance by shifting demand to alternate resources (e.g. routings, altitudes, times)," (FAA, 2007).

It is clear that the success of CATM in the future system will be heavily dependent upon the ability of NAS Service Provider operational staff (traffic managers and controllers in their future roles) and airspace users to collaborate effectively to make use of advanced technologies within new operational paradigms.

Furthermore, this collaboration and coordination must take place in a highly distributed environment where operational staff are spread across many locations around the country, including Flight Operations Centers (FOCs), airline ramp control towers, the Air Traffic Control Systems Command Center (ATCSCC), ARTCCs, TRACONs, and Airport ATC Towers (ATCTs) (or their future functional equivalents), along with the pilots in the aircraft. It is also clear that, in many cases, these individuals will not always be free to interact in real time, requiring different forms of asynchronous communication and coordination. Thus, although these roles will evolve and change over the next 2 decades, it is clear that the NAS will remain a highly distributed, human-centered work system.

CATM is likely to be of central importance in order to enable the airspace users to operate safely and efficiently, both areas of increasing concern as the volume of traffic and fuel prices increase, making it more challenging for airlines, air taxi operators, freight carriers and general aviation operators to maintain viable, cost-effective operations. The operational concepts that have been proposed for CATM for the next 10 years and beyond therefore require careful consideration of human factors issues at many different levels, ranging from:

- Strategies for distributing roles and responsibilities among different people and

technologies (automation)
- The design of procedures that can be effectively carried out
- The selection of sensing, communication and decision support technologies and associated functionalities that should be used
- The design of the human-computer interface
- Training of these procedures and the new tools that support them.

There are also important human factors issues that need to be considered in terms of the integration of human factors considerations into the design process itself.

Because of the significant changes in CATM that are being proposed, including new operational concepts for the functioning of the NAS, very different forms of collaboration and decision making, and technologies that are quite different from those used today, it is critical to that the necessary foundational human factors research be completed in a timely fashion. This basic research must be identified and then completed to ensure effective human-systems integration in the NAS. Thus, the focus of this paper is to identify the human factors engineering issues of greatest concern to the evolution of CATM in the NAS over the next 10-15 years so that necessary research can be conducted to fill existing gaps in these areas.

Methods

In addition to reviewing the relevant literature, structured interviews were conducted with operational staff at the FAA (traffic managers and controllers), with operational staff for flight operators (ATC coordinators, dispatchers and ramp controllers), with aviation human factors experts, with CATM and Traffic Flow Management (TFM) technology developers and with FAA program managers with responsibilities relevant to CATM/TFM. This included:

- Ramp Control - COA at EWR, DAL at JFK, UPS SDF, FedEx at MEM
- Airport ATC Towers - DTW
- ARTCCs – ZOB, ZNY
- FOCs (Flight Operations Centers) – AAL, UAL, SWA, DAL, NWA, NetJets, Jet Blue
- ATCSCC
- NY Port Authority
- FAA William J. Hughes Technical Center (human factors expert)
- MITRE (human factors experts and CATM/TFM technology developers)
- VOLPE (CATM/TFM technology developers)
- MIT Lincoln Labs (CATM/TFM technology developers)
- CDM and NextGen program managers
- Lockheed Martin (technology developers)
- Metron Aviation (technology developers)
- Sensis (technology developers)

Results

Below, we summarize findings regarding likely directions for the future evolution of CATM. We also identify key human factors issues that need to be considered as part of this evolution.

CATM Research and Development Focus Areas

The FAA Service Roadmap for CATM (FAA, 2008) highlighted seven broad categories as the focus for future research and development relevant to CATM:

- Continuous flight day operations ("Performance analysis, where throughput is constrained, is the basis for strategic operations planning. Continuous (real-time) constraints are provided to Air Navigation Service Provider (ANSP) traffic management decision-support tools and National Airspace System (NAS) users. Evaluation of NAS performance is both a real-time activity feedback tool and a post-event analysis process. Flight day evaluation metrics are complementary and consistent with collateral sets of metrics for airspace, airport, and flight operations.")

- Full collaborative decision making ("Timely, effective, and informed decision-making based on shared situational awareness is achieved through advanced communication and information sharing systems. Decision-makers request information when needed, publish information as appropriate, and use subscription services to automatically receive desired information through the net-centric infrastructure service. Decisions are made with an awareness of system-wide implications, including an increased level of decision-making by the flight crew and flight operations centers.")
- Traffic management initiatives with flight specific trajectories ("Individual flight-specific trajectory changes resulting from Traffic Management Initiatives (TMIs) will be disseminated to the appropriate Air Navigation Service Provider (ANSP) automation for tactical approval and execution. This capability will increase the agility of the NAS to adjust and respond to dynamically changing conditions such as bad weather, congestion, and system outages.")
- Management of SUAs ("Airspace for special use assignments, schedules, coordination, and status changes are conducted automation-to-automation. Changes to status of airspace for special use are readily available for operators and Air Navigation Service Providers (ANSP). Status changes are transmitted to the flight deck via voice or data communications. Flight trajectory planning is managed dynamically based on real-time use of airspace. ")
- Trajectory flight data management ("Trajectory Flight Data Management will improve the operational efficiency and increase the use of available capacity by providing for improved flight data coordination between facilities. This will enable access to airports by readily facilitating reroutes. Additionally, it will support more flexible use of controller/capacity assets by managing data based on volumes of interest that can be redefined to meet change to airspace/routings. Trajectory Flight Data Management will also provide continuous monitoring of the status of all flights – quickly alerting the system to unexpected termination of a flight and rapid identification of last known position.")
- Provide full flight plan constraint evaluation with feedback ("Timely and accurate NAS information allows users to plan and fly routings that meet their objectives. Constraint information that impacts proposed flight routes is incorporated into Air Navigation Service Provider (ANSP) automation, and is available to users for their pre-departure flight planning. Examples of constraint information include infrastructure outages, and significant congestion events. special use airspace status, significant meteorological information (SIGMET).")
- On-demand NAS information ("National Airspace System (NAS) and aeronautical information will be available to users on demand. NAS and aeronautical information is consistent across applications and locations, and available to authorized subscribers and equipped aircraft. Proprietary and security sensitive information is not shared with unauthorized agencies/individuals.")

More detailed insights are provided by ongoing concept exploration and development projects. Abstractly, some of the major areas that these CATM projects focus on include:

- Strategic planning and coordination to inform tactical decision making, taking into account priorities and constraints of both the NAS Service Provider and the flight operators (e.g., SEVEN, which will allow NAS customers to submit prioritized lists of alternative routing options for their flights; Integrated Collaborative Routing and the use of Flow Constrained Areas, which allow the flight operators to try to resolve the problems associated with a constraint before the NAS Service Provider needs to intervene)
- Integration of information about weather and traffic uncertainty into CATM/TFM decision making (e.g., RAPT, which predicts the likelihood that departure fixes will be open or closed for the next 1-2 hours; weather ensembles to characterize uncertainty in the weather; MITRE's Probabilistic Air Traffic Management system; Use of NCAR's NCWF to characterize uncertainty in the weather)
- The use of digital communication to support collaboration among FAA facilities, FOCs and pilots (e.g., TFDM, an integrated system for airport Tower Flight Data Management and communication with flight operators; DFM or the Departure Flow Management tool, a system to support

electronic communication and coordination between airport towers and ARTCCs in order to assign release times for aircraft affected by MIT (Miles-in-Trail restrictions)

- The use of automation to allocate arrival and departure slots to flights to support strategic or tactical metering (GDPs, or Ground Delay Programs to meter arrivals at an airport; AFPs, or Airspace Flow Programs to meter flights through a certain region of airspace; adaptive AFPs, or strategies for tactically adjusting AFPs to respond to the evolution of a weather pattern; DFM; TMA, or Traffic Management Advisor, which tactically meters flights to runways at arrival airports; Control by Required Time of Arrival at a fix)
- Integrated airport surface and terminal area airspace management (e.g., new procedures and decision support tools to help Airport Tower, TRACON, ARTCC and ATCSCC traffic managers better manage departures and arrivals on the airport surface in collaboration with dispatchers, ramp supervisors and pilots; dynamic departure rerouting using Route Segment Coded Departure Routes to reduce delays in dynamically rerouting departures)
- Airspace redesign and trajectory-based operations (e.g., the FAA Big Airspace concept; dynamic airspace reconfiguration; advanced RNAV arrival and departure routes; application of concepts based on performance based services)
- Integrated application of complementary traffic flow management strategies (supported by tools for predicting the integrated impact from applying multiple TFM strategies simultaneously; e.g., integration of TMA with GDPs; integration of RAPT, DFM, TFDM and SEVEN; integrated use of GDPs and AFPs).

Note that the point of identifying these ongoing efforts is twofold. First, the aviation human factors community needs to provide guidance regarding the viability of the associated operational concepts from a human performance perspective. Second, the human factors community needs involvement in the design details. In some cases, existing human factors expertise is sufficient to provide the necessary guidance. However, in other cases, new human factors research is needed.

Relevant Human Factors Research Issues

The primary goal of these interviews was to elicit insights regarding the major human factors challenges associated with CATM, and to identify critical tasks where they need to be addressed over the next 10-15 years. A small but illustrative sample of the resultant input is provided below, followed by a more abstract summary of human factors dimensions that the interviews highlighted as critical considerations when defining future operational concepts for CATM, and for designing the supporting procedures and tools.

Collaboration, Distributed Work and Communication

"Electronic negotiation of routes is becoming an increasingly important issue. We need a way to ensure that the traffic managers and dispatchers are collaborating with the same picture. Then everyone needs to be able to propose solutions, look at predicted results generated by a model, and collaborate to arrive at a collaborative assessment."

"With reroutes, different players are best equipped to deal with different parts of the problem. How do we let everyone provide their inputs and then pull all of that into a coordinated whole?"

"Today, the customers have constraints that the service provider doesn't know. How do we get that information into the mix when considering reroutes?"

"There's a tendency to push traffic management responsibilities up higher in the system to provide the bigger picture. That doesn't necessarily make the system more efficient, as knowledge of certain details may be critical to determining the best solution. How do you make sure decision making is pushed to the right level or person, matching the tasks with the people who have the needed expertise, time, motivation and access to data?"

"Some form of chat is very useful because it's persistent. You can go back and look at what was exchanged if you're not free at that exact moment."

Dealing with Uncertainty

"We currently present uncertain numbers as if they are deterministic. The users have to fall back on rules of thumb to interpret them. We need tools that really do consider uncertainty and inform the users so they can make decisions."

"There are different ways to estimate the uncertainty of a weather forecast. Which approach is better? How should the uncertainty be communicated?"

"We can't burden traffic managers and dispatchers with probability distributions. It works better to tell the user the tradeoffs associated with different options depending on how the situation develops."

"The hard part is making sure your strategic plan leaves enough options open to handle the uncertainty. In this system, uncertainty changes over time. You need options that you can use when uncertainty is reduced."

"Helping with risk management is a big issue. What should we display and how should we display it? How do we provide different levels of detail to different people."

Information Access and Function Integration
"You've got to be careful with SWIM. We want certain people, such as pilots, to focus their attention on what's really important for their tasks."

"SWIM could change the patterns of interaction and information access. That could be good or bad. We need to understand these impacts."

"Providing appropriate situation awareness is hard. You have different players with different objectives. What do you display to them?"

"We get lots of input when we're building a new tool, which pushes us to build in options and subtleties that allow flexibility in dealing with specific situations. The end result is often that there are so many options that it's difficult to do the basic functions that are used 99% of the time."

"Information overload is a big problem now that only promises to get worse. How do we present the most relevant and most important information? Should we give them more glass or let them access different information and tools on the same glass? What are the best ways to help the user navigate and display the information he needs for different tasks? How do we make sure critical information isn't overlooked?"

"Right now we have a bunch of stand alone tools for traffic managers and dispatchers. If we want to develop automation suites for these different people, what is needed in terms of functional integration and display integration?"

Mental Workload and Attention
"You've got to deal with reality. If the workload is too low, you tend to have loss of situation awareness. And if it's too high, you also lose situation awareness. For example, the dispatchers stop listening to telecons and hot lines when they're too busy. That information is lost. On the other hand, it may be more work to stop and type in that information so others can see it. What's the right balance?"

Training
"The need for better training is a constant refrain we hear out in the field. This applies in general as well as to new collaborative decision making tools and procedures."

More abstractly, this full set of interviews emphasized the following issues:

- Considering peak workload demands on performance and staffing requirements
- Developing and sustaining or supporting expertise (feedback, experience and training requirements)
- Integrating tools: Information display and navigation requirements
    - Incorporating intelligent alerts

- Using auditory vs. visual alerts
  - Designing to support distributed work
  - Tailoring interfaces to specific audiences
  - Ensuring appropriate situation awareness
- Dealing with uncertainties
  - Using alternative technologies to collect and process data to develop forecasts
  - Using alternative technologies to help identify and evaluate different strategies dealing with uncertainty (cognitive compatibility)
  - Providing effective information access and display
  - Using distributed work solutions to reduce cognitive complexity
  - Avoiding overconstrained solutions
  - Supporting strategies and tactics that enable effective adaptation
- Incorporating effective safety nets
  - Responding to different contributing causes
    - Brittle technologies
    - Human error
    - Unanticipated scenarios
  - Considering different classes of solutions
    - Technological
    - Human-centered
    - System-level solutions
- Designing effective roles and responsibilities for automation and people
  - Human as monitor
  - Learning and maintaining skills
  - Automation as backup for automation
  - Human adaptation
- Supporting communication
  - Information overload
  - Digital vs. voice
- Integrating of HF in the design process
- Designing organizations, work teams, individual job functions and physical facilities
- Selecting the best form of coordination and collaboration
  - Management by directive
  - Management by permission
  - Management by exception
  - Real-time interactive collaboration
  - Asynchronous coordination through constraint propagation.

Finally, while such a list of important human factors considerations serves to direct attention to important issues, it does not give justice to one of the major overriding human factors concerns:

How do the design decisions based on consideration of these individual issues interact to influence performance in the actual work environment?

Acknowledgements

We would like to thank Dino Piccione and the Human Factors Research and Engineering Group for their support of this effort.

References

FAA, (2007). Improve collaborative air traffic management http://www.faa.gov/about/office_org/headquarters_offices/ato/publications/oep/version1/solutionsets/catm/

FAA, (2008). http://nas-architecture.faa.gov/nas/view_exec/home.cfm

AIRPORT DEPARTURE FLOW MANAGEMENT (DFM): FINDINGS FROM FIELD TRIAL TESTING

Amy Spencer
Cognitive Systems Engineering, Columbus, Ohio
Ted Carniol and Jason Pepper
Metron Aviation, Inc., Dulles, Virginia
Philip J. Smith
Ohio State University, Columbus, Ohio

In this paper, we discuss the field testing of a Departure Flow Management (DFM) capability that has been developed by the FAA to reduce manual airport Call For Release (CFR) coordination requirements and workload, while increasing airport departure throughput and reducing delays. This field test consisted of shadow and operational phases and utilized both qualitative and quantitative methods. This study took place February and March 2008 at the Los Angeles (ZLA) Air Route Traffic Control Center (ARTCC) and Burbank (BUR), Las Vegas (LAS), Los Angeles (LAX), Ontario (ONT), and San Diego (SAN) airports. This test provided insights into how this tool changes roles and responsibilities, and how specific design features and functionality influenced the performance of the human operators. Human factors design improvements are discussed, along with the broader implications of the results of this case study for the introduction of new tools and automation into a distributed work environment.

In today's National Aviation System (NAS), flights from different airports within an ARTCC often compete for slots at a departure fix, in an overhead stream, or at a destination airport. This requires coordination in terms of the sequencing and timing of departures in order to efficiently utilize shared resources. Today, airports accomplish this mainly through a manual and time consuming approval request (APREQ) process. Note that the ZLA facility uses the term Call for Release (CFR) rather than APREQ.

The CFR process involves a phone call from a controller in the airport Tower to the overseeing ARTCC in order to request a departure release time for any flight included in a traffic management initiative (TMI) such as a miles-in-trail (MIT) restriction. The Tower controller provides the earliest time that the flight in question can depart. The ARTCC traffic management coordinator (TMC) fielding the phone call uses the tools at their disposal, including the radar display, to determine whether the departure time being requested by the Tower is feasible given the TMI and the current situation. This decision making process includes consideration of a variety of factors including local and downstream airspace and arrival airport restrictions. In today's environment, this manual CFR process is very time-consuming for the ARTCC TMC, requires significant collaboration, and does not produce optimal efficiency.

The DFM Capability

DFM automates the calculation, communication, and assignment of departure release times from multiple airports over shared NAS resources and into overhead traffic flows via improved display, decision support and digital communication capabilities. DFM introduces significant enhancements in the ARTCC and Air Traffic Control Tower (ATCT) environments including digital communications via both ARTCC and ATCT displays. Further, DFM pushes the decision making to the Towers by providing them all the information required to assign release times without the need to communicate with the ARTCC. These enhancements significantly reduce the hundreds of daily phone calls currently required to coordinate airport departure management.

Today, ATCT users manage CFR procedures with little information regarding the availability of slots in the overhead flows of traffic. They communicate with an ARTCC TMC who has this information in order to coordinate the release of certain departures - often a significant number of departures on any given day (400-500 at ZLA and around 900 at ZOB).

The DFM capability represents a significant change in the distribution of airport departure release time decision-making and workload. The DFM ATCT display automatically populates with all flights requiring CFR, identifies available departure times, and displays these departure slot availabilities to the ATCT user responsible for obtaining release times. The ATCT user can then request (in the case of Manual approval mode) and/or assign (in the case of Automatic approval mode) departure times at their facility via the DFM interface. The DFM ARTCC display in turn displays all departure release time requests to the ARTCC traffic manager who is responsible for either approving release time assignment (in the case of Manual approval mode) or simply monitoring assignments (in the case of Automatic approval mode).

Below, we focus on functionality and human factor issues related to the DFM interface, operational environment and user roles and responsibilities. Recommendations and findings regarding specific functional requirements and display capabilities are based primarily on insights gained through our observational studies and feedback from the participants. These findings are organized into four basic categories:

- Supporting situational awareness
- Decreasing ARTCC and ATCT communication workload, response time and head-down time (as it relates to departure release time approval and management)
- Increasing usability of the DFM interface
- Resolving Automatic approval mode issues

Supporting Situational Awareness

The shift in responsibility introduced by DFM must be supported through effective situational awareness for ATCT and ARTCC users, as well as shared situational awareness across these two groups. Specifically, the ARTCC traffic manager must be supported in maintaining an adequate mental model of air traffic in order to know when to intervene or change a release time, particularly in the case of automatic release time assignments. In addition, the ATCT traffic manager (or supervisor) must be supported in selecting appropriate and effective departure release times. Note that one of the findings of the ZLA field test is that because DFM does not include information regarding arriving traffic or airport surface constraints the ATCT user will likely require information outside of that currently provided within the DFM display. This is particularly important with single runway operations, such as SAN ATCT, where departure release time availability is subject to arriving traffic.

In order to support situational awareness, DFM includes the use of both visual display vocabularies and audible alerts. There are several events that require ARTCC and/or ATCT user notification that should be supported by this functionality including:

- ATCT requesting a Manual approval release time from the ARTCC
- ATCT requesting a Manual approval release time from the ARTCC within 5 minutes of requested departure time
- ARTCC change to requested and/or assigned release time to ATCT (including the removal of a release time request or assignment)
- ARTCC approval of release time assignment to ATCT
- Flight delayed by more than 15 minutes due to TMI
- Earlier slot open for a delayed flight

Note that careful consideration must be given to determine the types of events best indicated through audible indications. The best design limits the number of different versions of audible alerts, to indicate to the DFM user that something important has happened, and to then rely on the visual display of information for specification. In addition, it is unlikely that a final design would rely solely on audible alerts to indicate all of these various events.

Decreasing ARTCC and ATCT Communication Workload, Response Time and Head-down Time

Phone calls to perform CFR procedures often dominate the time and attention of ARTCC and ATCT personnel, hence the desire for automation. One critical design feature of DFM is the ability to effectively inform the user whenever an action or acknowledgment is required. Interface design methods must focus on limiting the amount of time that it takes DFM users to notice and react to events and must limit the head-down time required to interact with the system.

The majority of ZLA participants noted that they may not notice DFM events without some audible cue to draw them to the display. Implementing audible alerts, as discussed above, allows the ATCT user to step away from the DFM screen while waiting for a response from the ARTCC and decreases the length of time spent looking at the screen waiting for a response thus mitigating head-down issues. In addition, using DFM to reduce the amount of time that the ATCT spends on the phone with the ARTCC has the side benefit of increasing the amount of time the ATCT user can stay on frequency communicating with flight crews.

Increasing Usability of the DFM Interface

Both ARTCC and ATCT DFM displays use data tag color coding to indicate a variety of flight states including: Automatic Approval flight (cyan), Manual Approval flight (yellow), Flights from airports without DFM (purple), Manual Approval flight pending approval (inverse yellow), Manual Approval flight pending approval and within 5 minutes of release time request (inverse orange), flight with a release time assignment (green), flight 2 or more minutes past its departure release time (red) and en route flights (grey). In addition, whenever a Manual approval request is made or when a release time request or assignment is changed, the data tag will include both a release time acceptance button (represented by a checkbox) and a release time rejection/undo button (represented by a looped arrow). In addition, data tags contain ACID, originating airport, requested release time, assigned release time and aircraft type.

The DFM ARTCC Display (Figure 1) consists primarily of Flow Timelines. One of the more significant interface enhancements made between the ZOB and ZLA field tests was changing the timelines from representing Traffic Management Initiatives (TMI) to representing a specific flow. ZLA participants remarked favorably on this approach.

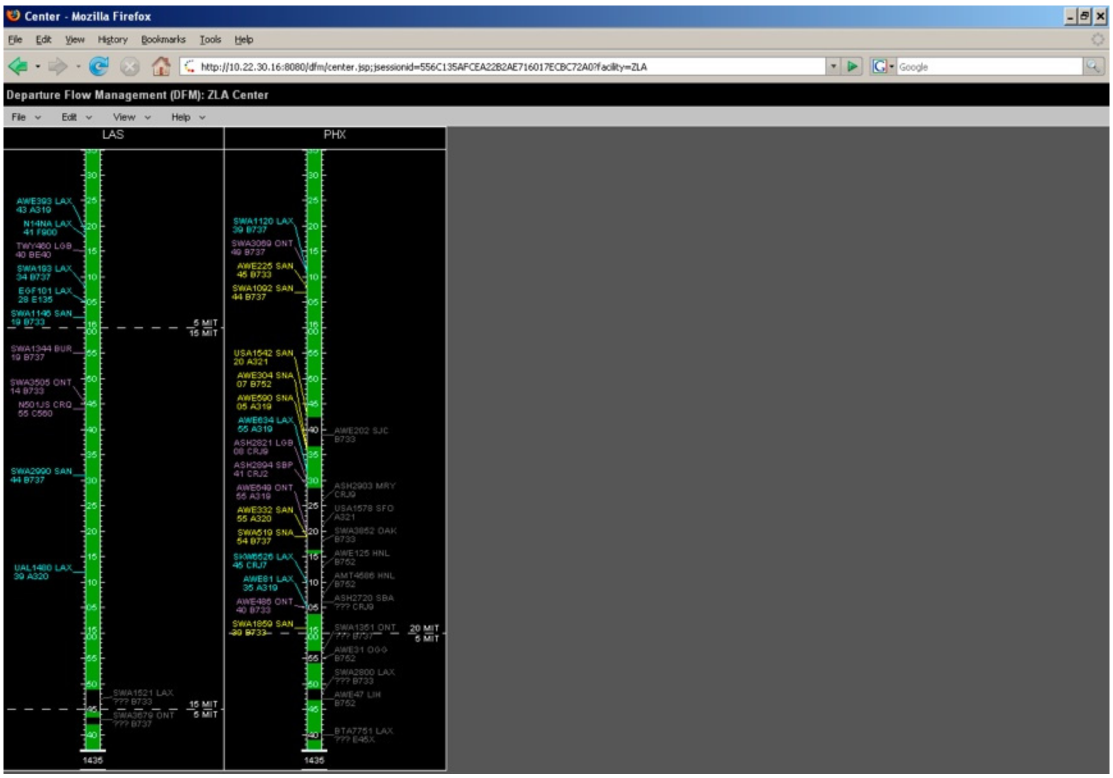


Figure 1. ARTCC DFM Display

Specifically, each Flow Timeline is double-sided and contains those flights that are expected to depart via the designated flow. The left side of the ARTCC timeline contains flights without a departure release time while the right side contains those flights that have requested a release time (pending approval), flights with a release time and en route flights. The timeline itself is color coded to represent available slots within the flow (green), unavailable slots (black) or to indicate that there is no TMI requiring CFR (blue).

In terms of managing the CFR process, the ARTCC traffic managers main interaction with the display is approving or rejecting release time requests by either clicking the approve or reject buttons provided within the flight data tag or changing the release time request or assignment by dragging the flight to a different release time within the timeline.

The DFM ATCT display (Figure 2) consists of a single one-sided timeline that automatically updates the display of slot availability depending on which flight (or, more specifically, which flow(s) associated with the flight) is selected. To the left of the timeline display is a list of all flights subject to CFR without a release time assignment (the Need Release

Times table); to the right of the timeline is a list of all pre-departure flights that have requested a release time (pending approval) and flights with a release time (the Have Release Times table). The ATCT user (traffic manager, supervisor or controller) requests a departure release time within this display by dragging flights from the Need Release Times table to the desired departure release time in the timeline.

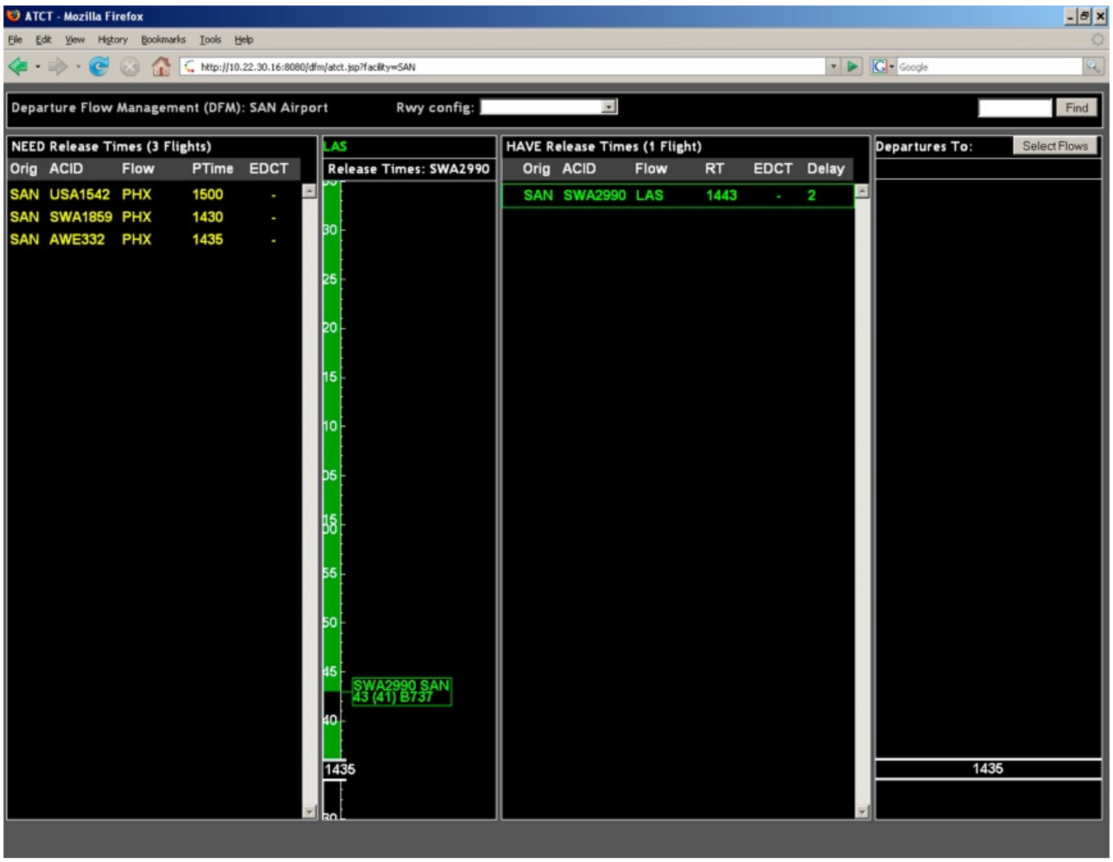


Figure 2. ATCT DFM Display

Human factors and functional recommendations derived from the field test include:

- Increase display font size to support the user's ability to glance at the display from a distance and discern that an action is necessary and to minimize head-down time, particularly in the ATCT environment
- Provide a "snap to" functionality to promote better accuracy when users drag flights to the timeline to make release time requests and assignments
- Provide display configurability in terms of font size, data tag elements, timeline duration (including the ability to see a history), displaying 2-digit vs. 4-digit times, flight filtering, and other features.

- Provide a "swap release time" capability to allow the exchange of release times between two flights
- Consider the integration of data from other tools such as ETMS and TMA

Resolving Automatic Approval Mode Issues

As described above, DFM supports two different kinds of departure release time modes: Automatic and Manual approval. In the case of an Automatic release time request, the ARTCC traffic manager simply monitors departure and en route demand to ensure that no flight receives reportable delay. In the case of a Manual request the ARTCC traffic manager must explicitly approve the request. In both cases, the ARTCC traffic manager has the ability to override or change any release time assignment or request at any time and the DFM uses a variety of visual and audible aids to maintain situational awareness.

In terms of Automatic approval mode, the assumption is that DFM can indicate available gaps because it has a sufficiently complete model of the situation. This then allows the ATCT user to select effective release time assignments. In many cases, it is likely that DFM will have a sufficiently complete model of the situation in order to identify available gaps. However, when there is an exception, features such as audible and visual alert functionality and the inclusion of meta-knowledge will support ARTCC traffic manager decision-making. These types of capabilities support the ARTCC traffic manager's ability to manage by exception, rather than having to monitor every flight. Such meta-knowledge would support the identification of cases where DFM may not know enough to assign an effective release time.

Conclusions

The ZLA field test validated the overall DFM capability concept, and provided insights for enhancements related to functionality, interface and human factors issues. ARTCC and ATCT users showed overwhelming acceptance of the concept and eagerness to see the capability put to operational use. In particular, users commented that they believed the capability supported greater situational awareness, operational flexibility and planning and created more time for managing other tasks and responsibilities. Kurt Rammelsburg, LAX STMC stated, "After the Field Trial, DFM was rated for functionality, usefulness and effectiveness. No one gave it a rating less than 80-100% positive rating in any area. Unheard of for a first field system trial."

Acknowledgements

This ZLA Field Test built upon fall 2007 testing conducted at the Cleveland (ZOB) ARTCC and Cleveland (CLE), Detroit (DTW) and Pittsburgh (PIT) airports. This work was performed for the Federal Aviation Administration's (FAA) Collaborative Air Traffic Management Technologies (CATMT) program.

References

[1] Carniol, T., et. al., (2007) Field Trial Report for Departure Flow Management (DFM) Testing at the Cleveland Air Route Traffic Control Center (ZOB).

[2] Carniol, T., et. al., (2008) Field Trial Final Report for Departure Flow Management (DFM) Testing at the Los Angeles Air Route Traffic Control Center (ZLA).

HIGH-FIDELITY SIMULATION TO COMPARE THE TOWER OPERATIONS DIGITAL DATA SYSTEM TO FLIGHT PROGRESS STRIPS

Todd R. Truitt, Ph.D.
Federal Aviation Administration
Atlantic City International Airport, NJ, USA

The current experiment used a high-fidelity, human-in-the-loop simulation to compare the Tower Operations Digital Data System (TODDS) to paper Flight Progress Strips (FPSs) during zero-visibility Airport Traffic Control Tower operations. Sixteen current controllers participated in groups of two. Each group received touchscreen and TODDS training before completing eight practice and eight test scenarios. The participants worked at both the ground and local control positions under four experimental conditions. The participants used one of four systems – the Integrated TODDS, FPSs with Airport Surface Detection Equipment – Model X, Perceptual-Spatial TODDS, or FPSs only – to control airport traffic. The participants had a Standard Terminal Automation Radar System display in all conditions, but did not have an out-the-window view. Dependent measures included the number and duration of airport operations, the number and duration of communications, TODDS usability, and participant opinion. We found advantages for surface surveillance and TODDS, and Integrated TODDS provided additional benefits.

Airports are central to implementing the Next Generation Air Transportation System (NextGen), and the Federal Aviation Administration (FAA) must change the way in which airports operate to fully realize its benefits. Two key capabilities discussed in the NextGen concept of operations (Joint Planning and Development Office, 2007) are Equivalent Visual Operations (EVO) and Network-Enabled Information Access. A subcomponent of EVO, the Staffed NextGen Tower (SNT) concept, proposes to reduce the cost of physical Airport Traffic Control Tower (ATCT) infrastructure with the ability to manage airport traffic from a remote location. The development of Electronic Flight Data (EFD) will take advantage of network-enabled information access, which allows stakeholders to access and share all air traffic information related to the National Airspace System. The implementation of EFD may also alleviate some of the human performance constraints inherent in the current paper Flight Progress Strips (FPSs). For example, EFD can reduce the controllers' need to search for information presented in visually separate locations and provide the opportunity to integrate flight data with other often-used information sources, such as surface surveillance and weather information.

To address the role of EFD, Engineering Research Psychologists from the FAA Human Factors Team – Atlantic City designed two prototype Electronic Flight Data Interfaces (EFDIs) for use in ATCTs (see Truitt, 2006a, 2006b). The Integrated EFDI combined EFD with a surface surveillance capability. The Perceptual-Spatial EFDI provided a way for controllers to spatially organize EFD using a surface map of an airport without surface surveillance. We have recently refined the concepts to create the Tower Operations Digital Data System (TODDS), as described by Truitt (2008). To design the Integrated TODDS (I-TODDS) and the Perceptual-Spatial TODDS (PS-TODDS), we used a process based on "The Bridge" methodology (Dayton, McFarland, & Kramer, 1998) that relies in part on usability testing throughout the development process. By examining task flows and paper prototypes, we were able to ensure that the resulting interfaces would function as expected, and we could address numerous problems before the software development began. We continued the usability testing during software development. Once the initial prototypes were functional, we conducted formal usability testing (Truitt & Muldoon, 2007).

We refined the newest version of TODDS to address the results of the usability test and to expand the scope of the interfaces beyond flight data management. In addition to making the most difficult features easier to use, we added the ability to issue digital taxi-out clearances, perform taxi conformance monitoring, indicate closed runway and taxiway segments, and access integrated weather information, including advisories for wake turbulence separation. We also designed a touchscreen training protocol to better familiarize users with the interface hardware. We conducted the current experiment to evaluate TODDS against comparable conditions using FPSs.

Method

We conducted the experiment in the Research, Development, and Human Factors Laboratory at the FAA William J. Hughes Technical Center. The experiment placed current ATCT controllers in a high-fidelity simulation to compare TODDS to FPSs under zero-visibility conditions (i.e., no out-the-window view). The experiment used a

2 (run number – first vs. second) X 2 (flight data type – TODDS vs. FPS) X 2 (surface surveillance – yes vs. no) repeated measures design.

Participants

We recruited 16 current ATCT controllers from busier (level 10 and above) facilities and received volunteers from Phoenix, Las Vegas, Miami, Philadelphia, and Salt Lake City ATCTs. The participants had a mean age of 42.4 years and had actively worked in an ATCT for an average of 17.8 years.

Apparatus

We used three 21.3″ 1,600 x 1,200 pixel touchscreen displays: Two contained the TODDS ground and local control positions, and one contained the Airport Surface Detection Equipment-Model – X (ASDE-X) (no touchscreen capabilities). A fourth display presented Standard Terminal Automation Replacement System (STARS) data on a 20″ Tower Display Workstation. A fifth display showed a screen of the Information Display System (IDS), including the current Automatic Terminal Information Service (ATIS) code, wind direction, speed, gust, and runway visual range. We constructed two FPS bays that fit over the touchscreens for use in the appropriate experimental conditions.

Procedure

The participants arrived and worked in groups of two. Before the experiment, they signed an Informed Consent Statement, completed a Biographical Questionnaire, and received a briefing on the simulated airport operations. The participants then completed the touchscreen and TODDS training protocol.

The touchscreen training protocol consisted of three specific tasks (select a single button, select two buttons in sequence, and drag a button to a target area), with 10 different button sizes across multiple trials. The button sizes ranged from 1.5 X 1.5 in. (3.8 X 3.8 cm) to 0.4 X 0.4 in. (1.1 X 1.1 cm). The buttons and target zone in the drag task appeared at random locations on the touchscreen. After completing all three tasks to criteria for a button size, the participants repeated the tasks with the next button size. The entire touchscreen training protocol lasted about 2 hr. The participants then received training on TODDS using a structured protocol. Half of the groups received training on I-TODDS first, whereas the other half received training on P-S TODDS first. The TODDS training lasted about 2 hrs.

After training, the participants completed eight practice and eight test scenarios. In the I-TODDS condition, the participants had an ASDE-X display that was integrated with EFD, weather information, digital taxi clearances, and taxi conformance monitoring on a single display. In the FPS + ASDE-X condition, the participants had FPSs, ASDE-X, and IDS. In the P-S TODDS condition, participants used EFD integrated with weather information and digital taxi clearance, but ASDE-X was unavailable. In the FPS-only condition, the participants used FPSs and the IDS, but ASDE-X was unavailable. The participants had a STARS display in all conditions. The participants worked two consecutive scenarios in each condition, alternating between the ground and local control positions. The participants controlled the airport traffic and maintained flight data for each aircraft. They did not have an out-the-window view, but they could assess aircraft position from pilot reports, STARS, and surface surveillance (if available). We counterbalanced the order of conditions. The participants completed the Post-Scenario Questionnaire (PSQ) at the end of each test scenario and completed the Post-Experiment Questionnaire after completing all scenarios.

Subject matter experts developed one 40-min airport traffic scenario based on Boston Logan International Airport using runways 27, 33L, and 33R as the active runways. The scenario included 49 departures and 31 arrivals, with arrivals and departures on all three runways. There were a variety of aircraft types, including civil and commercial aircraft, all of which were capable of data communications. We then created 16 different “versions” of the base scenario by changing only the aircraft call signs to reduce the potential effects of traffic demand and aircraft type while preventing the participants from recognizing that each scenario was identical. We presented each version of the scenario in the same order for all participants, although the participants experienced them in a different combination of experimental conditions. The scenario began with aircraft already on the airport surface and in the air. Five simulation pilots communicated with the participants and entered commands to move aircraft through the scenarios.

Results

We analyzed the data using the appropriate repeated measures analysis of variance (ANOVA) for each dataset and Tukey's Honestly Significant Difference (HSD) post hoc test, as needed. All statistically significant results used the criteria of $p \leq .05$.

Number of operations. The mean number of arrivals did not differ between conditions. However, the participants were able to depart approximately 50% more aircraft when surface surveillance was available, $F(1, 7) = 114.94$. There was no difference in the number of missed approaches (see Table 1).

Table 1. *Mean (SD) Number of Airport Operations by Type and Condition.*

	Arrivals	Departures	Missed Approaches
I-TODDS	29.2 (0.9)	33.8 (6.8)	0.9 (1.1)
FPS + ASDE-X	29.3 (0.9)	31.8 (6.6)	0.8 (1.1)
P-S TODDS	29.2 (1.3)	20.5 (7.0)	0.8 (1.3)
FPS	29.3 (0.8)	20.3 (3.8)	0.7 (0.8)

Ramp waiting time. We recorded the time that each departure aircraft reached a ramp spot and the time of each aircraft's first taxi movement. Using these times, we calculated a mean ramp waiting time. There was a main effect of surface surveillance presence, $F(1, 7) = 54.77$, and flight data type, $F(1, 7) = 17.35$ (see Figure 1). When surface surveillance was unavailable, aircraft waited on the ramp about 80 s longer than when surface surveillance was available. The ramp waiting time was about 37 s longer when the participants used TODDS instead of FPSs. Once a departure aircraft reached a ramp spot in the TODDS conditions, the pilot requested a taxi clearance via data communications. The ground controller then issued a digital taxi clearance via TODDS. It took as long as 30 s for a digital taxi clearance to reach an aircraft and for the pilot to accept the clearance. The ground controller then instructed the aircraft via voice communications to begin taxiing. The process of sending, receiving, and acknowledging digital taxi clearances took a significant amount of time due to data transmission.

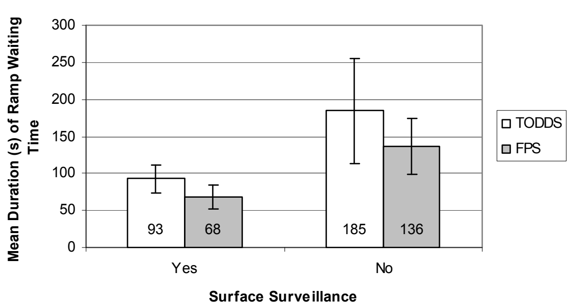


Figure 1. Mean duration (s) of ramp waiting time by surface surveillance presence and flight data type.

Taxi operations. For taxi-out operations, we recorded the duration from when an aircraft made its first taxi movement until departure (i.e., wheels up), and we found a significant interaction of surface surveillance presence and flight data type, $F(1, 7) = 6.68$. A planned comparison confirmed that aircraft took significantly less time to taxi out (106 s on average) when the participants used I-TODDS compared to FPS + ASDE-X, $F(1, 7) = 6.35$. For taxi-in operations, we recorded the duration from when an aircraft landed until it reached an arrival ramp spot, and we found a significant main effect of surface surveillance presence, $F(1, 7) = 44.52$. Taxi-in operations were over 1 min shorter when the participants had surface surveillance. A planned comparison showed that taxi-in operations were significantly shorter (35 s on average) in the I-TODDS condition than in the FPS + ASDE-X condition, $F(1, 7) = 10.79$ (see Figure 2). Neither taxi-out nor taxi-in times were significantly different when surface surveillance was unavailable.

Departure delays. We counted a departure delay when the time between an aircraft's first taxi movement and departure exceeded 20 min. There was a significant main effect of surface surveillance presence on the number of departure delays, $F(1, 7) = 14.74$. There were 2.5 fewer delays when surface surveillance was present. Departure delays were significantly shorter by 202 s when surface surveillance was present, $F(1, 7) = 25.91$. There were 1.2

fewer delays and those delays were 43 s shorter in the I-TODDS condition compared to the FPS + ASDE-X condition, but these differences were not statistically significant (see Figure 3).

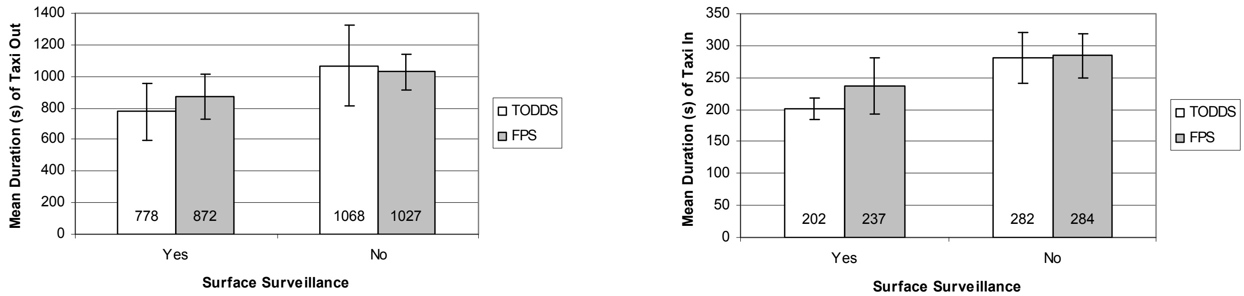


Figure 2. Mean duration(s) of taxi-out (left) and taxi-in (right) operations by surface surveillance presence and flight data type.

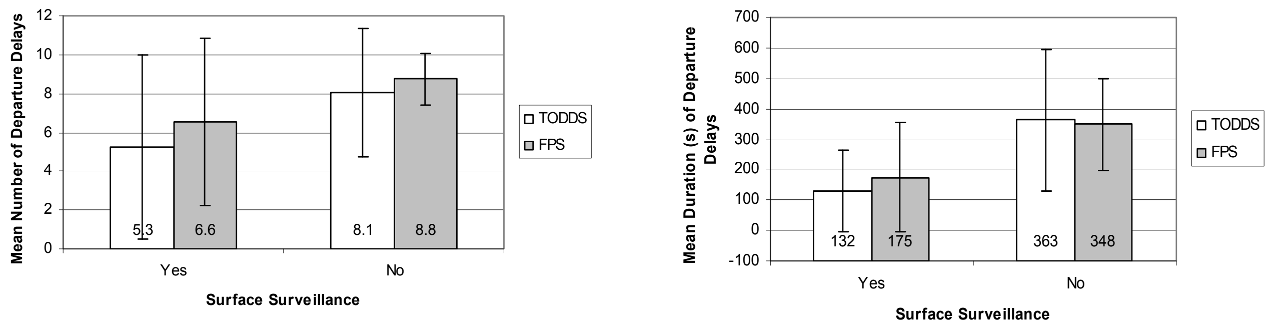


Figure 3. Mean number (left) and duration (right) of departure delays by surface surveillance presence and flight data type.

Radio transmissions. The participants made two fewer transmissions per minute when surface surveillance was present, $F(1, 7) = 38.96$. The participants also made two fewer transmissions per minute when using TODDS compared to using FPSs, $F(1, 7) = 17.93$. When using TODDS, the participants' transmissions from the ground controller position were significantly shorter, $F(1, 7) = 79.02$. When using FPSs, the participants had to give a full taxi clearance for departures (e.g., "United one niner, taxi to runway two seven via alpha and echo, hold short runway three three left."). In contrast, once the pilot had acknowledged a digital taxi clearance issued via TODDS, the participants only had to tell the pilot to "resume taxi" to start an aircraft's ground movement (see Figure 4).

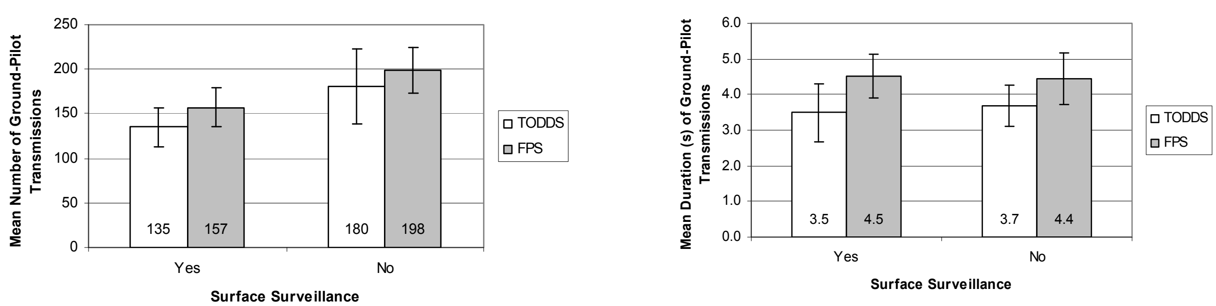


Figure 4. Mean number (left) and duration (s) (right) of ground controller to pilot transmissions by surface surveillance presence and flight data type.

Usability - Integrated TODDS. We calculated an error rate (ER) percentage for each TODDS action type by dividing the number of successful actions (S) by the sum of successful actions (S) and failed actions (F), and then multiplying the result by 100, so that ER% = S/(S + F) X 100.

There were 29 distinct actions that the participants could perform with I-TODDS. Of these actions, they performed 18 of them at least once on average. There was substantial variability between the participants in how often they performed each action. With the exception of Flight Data Element (FDE) selects at the local control position, the error rate for the most commonly performed actions decreased compared to the initial usability study (Truitt & Muldoon, 2007). The overall error rates (calculated over all actions, regardless of frequency) decreased from 11% to 4% at the ground control position and from 13% to 4% at the local control position (see Table 2).

Table 2. *Mean (SD) Number of Touchscreen Actions, Error Rates, and Percentage Change in Error Rates from the Initial Usability Study for the Ground and Local Control Positions with Integrated TODDS.*

	Ground			Local		
Touchscreen Action	Mean (*SD*) Number of Actions	Mean Error Rate (%)	% Change	Mean (*SD*) Number of Actions	Mean Error Rate (%)	% Change
Data Block Select	158.8 (73.19)	1	-3	108.3 (63.29)	3	-4
FDE Select	35.1 (45.35)	5	-8	18.4 (21.61)	10	+4
Reposition	45.2 (3.31)	4	-4	62.3 (49.96)	5	-9
List Transfer	36.4 (2.19)	4	-2	NA	NA	NA
Position Transfer	39.1 (4.11)	2	-4	29.4 (2.28)	4	-10
External Transfer	28.6 (1.93)	4	-2	31.9 (6.23)	3	-11
ATIS Update Ack.	2.4 (5.27)	25	-28	1.6 (5.21)	0	-35
D-Taxi	40.4 (3.67)	6	NA	NA	NA	NA
Total Actions	389.2 (99.44)	4	-7	253.2 (101.64)	4	-9

Usability – Perceptual-Spatial TODDS. There were 27 distinct actions that the participants could perform on PS-TODDS. Of these actions, they performed 11 of them at least once on average. There was substantial variability between the participants in how often they performed each action. With the exception of FDE repositions, the error rates for the most commonly performed actions decreased compared to the initial usability study (Truitt & Muldoon, 2007). The overall error rate decreased from 7% to 4% at the ground control position and decreased from 9% to 4% at the local control position. The participants performed the Taxi-into-Position-and-Hold (TIPH) clearance at the local control position with a lower error rate because we locked the TIPH buttons in place so that the participants could not move (i.e., drag) them when selected (see Table 3).

We attribute the overall reduction in error rates primarily to the touchscreen training protocol and to a slight increase in familiarity with the TODDS interfaces prior to data collection. The dramatic reduction in the error rate for FDE ATIS update acknowledgments was due to a redesign of the touch sensitive area for this particular element. Compared to the initial usability study, the participants made more touchscreen actions, with fewer errors, in the current experiment.

Table 3. *Mean (SD) Number of Touchscreen Actions, Error Rates, and Percentage Change in Error Rates from the Initial Usability Study for the Ground and Local Control Positions with Perceptual-Spatial TODDS.*

	Ground			Local		
Touchscreen Action	Mean (*SD*) Number of Actions	Mean Error Rate (%)	% Change	Mean (*SD*) Number of Actions	Mean Error Rate (%)	% Change
FDE Select	278.4 (67.99)	2	-3	209.6 (54.05)	4	0
FDE Reposition	154.5 (49.07)	9	+1	93.6 (41.97)	7	+2
FDE Resequence	0.1 (0.34)	0	-18	1.1 (1.48)	0	-5
Position Transfer	23.3 (5.42)	3	-2	28.9 (1.65)	2	-3
External Transfer	27.1 (2.28)	1	-4	20.8 (4.93)	1	-4
FDE Recall	1.2 (1.42)	13	NA	0.0 (0.00)	0	NA
TIPH	NA	NA	NA	34.8 (9.63)	1	-15
Depart. Clearance	NA	NA	NA	20.8 (5.71)	3	-7
Highlight	0.3 (0.87)	0	NA	2.8 (7.47)	0	NA
ATIS Update Ack.	8.4 (10.85)	3	-11	5.9 (7.20)	2	-34
D-Taxi	29.9 (5.66)	3	NA	NA	NA	NA
Total Actions	524.9 (119.78)	4	-3	418.6 (103.84)	4	-5

Post-scenario questionnaire. Overall, the presence of surface surveillance had the largest effect on the PSQ ratings. With surface surveillance, the participants reported that they needed less effort to maintain flight data and issue taxi clearances; they were better able to detect aircraft on a runway; they were more aware of projected aircraft positions; they had a greater awareness of potential runway incursions; they were more aware of the overall traffic situation; and they had lower workload due to controller-pilot communications. Also, when working at the ground control position with surface surveillance, the participants reported that they were better able to find flight information. At the local control position they were better able to find weather information and had a better awareness of the current location of aircraft when surface surveillance was available.

When the participants used TODDS, they thought that it was easier to issue taxi clearances from both the ground and local control positions. When they worked at the local control position with TODDS, they reported a greater awareness for potential runway incursions. When working at the ground control position with TODDS, the participants reported lower workload due to controller-pilot communications. The participants rated their awareness for current aircraft position as being low at the ground control position in the FPS-only condition, but rated it even lower in the PS-TODDS condition. The participants rated their awareness of current aircraft locations equally high when surface surveillance was available, regardless of the flight data type.

Post-experimental questionnaire. The participants reported that the elements of the readout area, weather information, and FDEs of TODDS were readable. They also gave high ratings for the readability of data blocks in the I-TODDS. The participants rated the effort to use the touchscreen in I-TODDS as moderate, whereas the PS-TODDS took a little more effort, perhaps because they had to move each FDE multiple times. The participants thought that the I-TODDS would have a moderately positive effect on their ability to control airport traffic, but PS-TODDS would have only a slightly positive effect.

Conclusion

The presence of surface surveillance significantly improved airport efficiency by increasing the controllers' awareness of the traffic situation and the number of departures; and by reducing ramp waiting time, number and duration of departure delays, number of ground controller-to-pilot transmissions, and controller effort. TODDS increased ramp waiting time, but decreased the number and duration of ground controller-to-pilot transmissions. I-TODDS decreased the duration of taxi-out and taxi-in operations. I-TODDS also provided an operational increase in the number of departures and a reduction in the number and duration of departure delays, but these differences were not statistically significant. The overall error rate for TODDS usage was 4% – a reduction from the initial design concept. The participants found TODDS useful and thought it would have a positive effect on ATCT operations, especially when integrated with surface surveillance, as with I-TODDS. However, they had some reservations about PS-TODDS because it required more effort and could mislead the ground controller regarding aircraft position. Based on the results of this experiment, I-TODDS may be able to support SNT operations as an alternative to an out-the-window view. Future experiments should compare TODDS with and without an out-the-window view.

References

Dayton, T., McFarland, A., & Kramer, J. (1998). Bridging user needs to object oriented GUI prototype via task object design. In L. E. Wood (Ed.), *User Interface Design*. CRC Press.

Joint Planning and Development Office. (2007). *Concept of operations for the next generation air transportation system* (Version 2.0). Retrieved January 8, 2009 from http://www.jpdo.gov/library/NextGen_v2.0.pdf.

Truitt, T. R. (2006a). *Electronic flight data in airport traffic control towers: Literature review* (DOT/FAA/CT-05/13). Atlantic City International Airport, NJ: FAA William J. Hughes Technical Center.

Truitt, T. R. (2006b). *Concept development and design description of electronic flight data interfaces for airport traffic control towers* (DOT/FAA/CT-06/17). Atlantic City International Airport, NJ: FAA William J. Hughes Technical Center.

Truitt, T. R. (2008). *Concept development and design description of the tower operations digital data system* (DOT/FAA/TC-TN08/09). Atlantic City International Airport, NJ: FAA William J. Hughes Technical Center.

Truitt, T. R., & Muldoon, R. V. (2007). *New electronic flight data interface designs for airport traffic control towers: Initial usability test* (DOT/FAA/TC-07/16). Atlantic City International Airport, NJ: FAA William J. Hughes Technical Center.

Identifying Psychosocial Factors Associated with Work-Related Musculoskeletal Disorders in Flight Attendants in a Taiwanese Commercial Airline

Kai-Hui Lee
Cabin Crew Division, China Airlines
Li-Hua Kao, Ph.D.
College of Business, Chung Yuan Christian University, Taiwan

Psychosocial factors have the potential for causing psychological or physical harm, perceived psychological demands, job stress, or work-related musculoskeletal disorders (WMSDs). The purpose of the current study was to study the relationship between psychosocial factors including *job demands, job control, managerial and colleague support, relationships at work, role conflict,* and *organizational change,* and psychosocial factors, health and the well-being, and WMSDs among flight attendants. A survey, mainly based on the "HSE Indicator Tool" developed by Health and Safety Executive and the Copenhagen Psychosocial Questionnaire (COPSOQ), was distributed in 2009 to flight attendants in a Taiwanese commercial airline by convenience sampling. A total of 145 flight attendants returned the survey. Data showed that 115 (79.3%) of flight attendants received musculoskeletal injuries while at work. It was found that job demand is significantly related to work-related musculoskeletal disorders.
Keywords: Flight attendant; Work-related musculoskeletal disorders; Psychosocial factors; Job demand; Stress

Background

There is substantial evidence that musculoskeletal disorders have been the largest, single work-related illness and injury problem in the United States for the last decade. The USA Bureau of Labor Statistics (BLS) reports that there were 333,760 musculoskeletal disorders involving reported days away from work in 2007. Musculoskeletal disorders (MSDs) are not all related to work activities. Nevertheless, MSDs are among the most severe injuries facing American workers (BLS, 2009). Ergonomic injuries and illnesses that affect the connective tissues of the body such as muscles, ligament, nerves, tendons, joints, cartilage, spine, or spinal disks can be described by the term "musculoskeletal disorders" (OSHA, 2002; BLS, 2009). Workers who are exposed to force, pressure, repetition, awkward postures, overextension of limbs etc. in the workplace over a period of time can suffer a variety of disorders and other conditions. These conditions, collectively referred to as musculoskeletal disorders, are thought to result, amongst other causes, from certain psychosocial factors such as job dissatisfaction, monotonous work, and limited job control (OSHA, 2009).

In the aviation industry, the International Air Transport Association (IATA) reported that a higher occurrence of occupational injury and illness has been associated with flight attendants compared to other aviation employees. (IATA, 2006, cited in Kao, Stewart, Lee, 2009). A study conduced by the Environment, Health and Safety Group at AirBC (a former Air Canada airline) concluded that about 58 percent of the injuries to AirBC flight attendants were musculoskeletal injuries, involved the back, neck or shoulders, and were due to ergonomic risk factors (FSF, 2002). Kelleher and McGilloway (2005) found that flight attendants working for an Irish airline experienced high levels of work-related stress because of psychosocial factors and personal characteristics. Another study indicated that the majority of flight attendants working on long-haul international commercial airlines who are exposed to ergonomic stressors are likely to experience work-related musculoskeletal symptoms (Lee, Wilbur, Conrad, Mokadam, 2006).

MacDonald, Deddens, Grajewski, Whelan, Hurrell (2003) and Siu , Phillips, Leung (2004) concluded job stressors such as high mental or job demand, imbalance between job demands and outside obligations, low supervisor support, and emotional load have a positive relationship with the job dissatisfaction and work-related accidents and injuries (MacDonald, et al., 2003; Cooper, Sutherland, 1987, Holecom et al., 1993, Hoffmann, Stetzer, 1996, Murray, Fitzpatrick, O'Connell, 1997, cited in Siu et al., 2004). Psychosocial factors have the potential for causing psychological or physical harm, perceived psychological demands, or job stress (Sauter and Swanson, 1996, cited in Kumar and Kumar, 2008). The Health and Safety Executive (HSE) defines stress as *"The adverse reaction people have to excessive pressure or other types of demand placed on them"*. The National Institute for Occupational Safety and Health (NIOSH) has defined job stress generally as "*The harmful physical and emotional responses that occur when the requirements of the job do not match the capabilities, resources, or needs of the worker. Job stress can lead to poor health and even injury*" by (NIOSH, 1999). A study by Kumar and Kumar (2008) reported that the psychosocial factors can include, amongst other terms, job dissatisfaction, monotony of work, lack of social support, overexertion due to excessive work rates, workload, time pressure, and no control over work/rest patterns.

Over the last few decades, psychosocial factors and their relationship to occupational injury have been studied by various researchers or national and international organizations. In 1979, Karasek proposed a demand-control model which represented the relationship between job characteristics, i.e., job demand and job control (Karasek, 1979, cited in Saastamoinen, Laaksonen, Leino-Arjas, Lahelma, 2008). Recently, the Karasek model has been modified. In 2004, MacDonald argued that stress response is multidimensional and can directly increase both error rate (with possible increases in accident risk) and the risk of WMSDs (MacDonald, 2004). There is more and more concern about the influence of psychological factors along with personal characteristics on WMSDs, but so far little is known about the relationship between psychosocial factors and WMSDs among flight attendants in Taiwan.

Growing evidence shows that there is a significant relationship between psychosocial factors and WMSDs (Cohen, Gjessing, Fine, Bernard, and McGlothlin, 1997, Davis and Heaney, 2000, Hopkins, 1999, Linton, 1999, Ryan, Bampton, 1988, cited in Kumar and Kumar, 2008). For example, Davis and Heaney (2000), and Cohen et al. (1997) noted that job dissatisfaction, monotonous work, limited job control, and lack of social support result in musculoskeletal disorders (Kumar and Kumar, 2008). Another similar study conducted by researchers Devereux and Buckle, found that both work-related stress and WMSDs have similar causes, and they concluded that stress is a significant predictor of physical pain (Devereux, Buckle, 2000, cited in MacDonald, 2004).

This study uses as its base MacDonald's ergonomics model (MacDonald, 2004) and an extensive literature review. Key constructs within MacDonald's model are psychosocial factors, job demands, personal factors, mental workload, stress and fatigue, and health and well-being. MacDonald proposed that job demands among the most frequently cited stressors can result in WMSDs. In the present paper, the focus is primarily on the relationship between these factors and work-related musculoskeletal disorders. Specifically, the purpose of this study was:

- To identify what kind of psychosocial factors (i.e. *job demands, job control, managerial support, colleague support, relationships at work, role conflict,* and *organizational change*) along with personal characteristics are associated with well-being (i.e. *job satisfaction*, *general health*, *mental health*, and *vitality*) or WMSDs among flight attendants
- To define for airline management risk factors to consider in the prevention of occupational injury of flight attendants

Methods

Procedure

In the current study, a paper-based questionnaire based on the "HSE Management Standards Indicator Tool" developed by HSE for measuring psychosocial factors and the Copenhagen Psychosocial Questionnaire (COPSOQ) was developed. The questionnaire looks at the job demands, job control, managerial support, colleague support, relationships, role, change, job satisfaction, general health, mental health, vitality, and WMSDs among flight attendants to gather data about psychosocial factors that may influence WMSDs. Data were obtained from 145 flight attendants employed at a selected major Taiwanese commercial airline by convenience sampling during a three-week period from January 20, 2009. One of the authors, who is a flight attendant, distributed and collected surveys, along with a plain language statement and a consent form, face-to-face, from her colleagues. Participants were reimbursed NT$ 100 (US$3) for their participation. Data were manually keyed in and stored in Microsoft SPSS 14. Negatively worded items were recoded before averaging so that higher scores on all items reflected a positive response. One-way ANOVA, multiple regressions, and logistic regression were performed.

Measurements

Work-related musculoskeletal symptoms (pain, ache, or discomfort) that subjects experienced during the last 12 months were measured by four questions derived from the Nordic Musculoskeletal Questionnaire (NMQ) and the National Institute for Occupational Safety and Health (NIOSH) Symptom Survey (Lee et al., 2006). Respondents were asked to indicate if they had experienced aches, pain, or discomfort in nine body regions (neck, shoulders, upper back, lower back, wrists, elbows, hips, knees, and ankles) that they considered work related at any time during the prior 12 months. If the respondents answered "yes", then they were instructed to continue with the questions about symptom severity (frequency, duration, and intensity) derived from the NIOSH survey. Psychosocial factors, used as part of the study questionnaire, were measured by a Chinese language version of the HSE Management Standards

Indicator Tool. The Tool comprises seven subscales, i.e., *job demands, job control, managerial support, colleague support, relationships at work, role conflict,* and *organizational change*. Thirty-five questions measured the above-mentioned subscales. A five-point Likert scale, range from 5: *always*, to 1: *never* (or 1, 2, 3, 4, or 5 for the items negatively formulated), was used. Subscales of the health and well-being conditions were derived from the COPSOQ. These were translated into Chinese by a professional translator. Fourteen items were used measuring *job satisfaction* (4 items), *general health* (1 items), *mental health* (5 items), and *vitality* (4 items). They were measured by a five-point Likert scaling where 5 is *all of the time*; and 1 is *none of the time*. The survey also collected data on various personal variables. These variables included *gender, age, tenure of employment* (years of employment as a flight attendant), *job position, marital status, children, number of block hours* (air travel time) that they typically worked per month, and *the frequency of international routes* that they flew per month.

Participants

All participants in this study are flight attendants whose base city is Taipei. They work for the cabin crew division of a major Taiwanese airline operating international flights to twenty-six countries and seventy-three destinations using forty-seven wide-bodied jets. The cabin crew comprises flight attendants, assistant pursers who are the leading flight attendants in economy class, and the purser, a line manager or supervisor who is the person in charge of the entire flight. Block time was estimated from questionnaire information describing total block hours a month. Based on the findings, there is no flight attendant who works alone on board as a solo flight attendant. Of the 159 questionnaires distributed, 145 (91.19%) were returned and validated for analyses.

Results of Data Analysis

Characteristics of the study population

The majority of the participants were females (80%). Half (49.7%) of the respondents were young women aged from 25 to 34 years and 82.8% had worked as a flight attendant for more than 5 years. Of the participants, only 9% were pursers who had management responsibilities. Fifty-one percent were married, and 35.9% had children. About three quarters (73.8%) of the flight attendants reported that their block hours were 75 to 85 hours per month and half (46.2%) worked primarily on eastbound flight route twice a month. An interesting finding is that the majority (79.3%) of the flight attendants reported that they had experienced WMSDs. More than half (65.8%) of the flight attendants experienced symptoms at some time during the past 12 months. Just over sixty-nine percent reported symptoms lasting less than one week and 43.9% indicated that the level was moderate. These results were in accordance with other studies (FSF, 2002; Lee et al., 2006).

Flight attendants' descriptions of psychosocial factors and health and well-being conditions

The mean and standard deviation of scores for the factors perception of job demands, job control, managerial support, colleague support, relationships at work, role conflict, organizational change, job satisfaction, general health, mental health, and vitality were obtained. The results of the reliability analysis are as follows: job demands (0.7), job control (0.7), managerial support (0.8), colleague support (0.7), relationships at work (0.6), role conflict (0.7), organizational change (0.6), job satisfaction (0.8), general health, mental health (0.8), and vitality (0.8). We considered the Chronbach's alpha higher than the acceptable level of 0.7 to be satisfactory. Thus, it appeared that all Chronbach's alpha values of the above-mentioned scales were acceptable. Except for role conflict (4.23), psychosocial factors levels were found to be moderate or low (Table 2). Colleague support (3.87) was moderately high, followed by job control (2.86), relationships at work (2.80), managerial support (2.75), and organizational change (2.64). Job demand (2.51) (pressured to work long hours, having unachievable deadlines, have to work very fast and intensively etc.) was low. Job satisfaction (2.81) was moderate, whereas general health (3.59), mental health (3.39), and vitality (3.33) were moderately high. The correlation between the psychosocial factors which were ranged from moderate to strong (0.17 to 0.65) was shown in Table 1. Table 2 shows that the correlation between the health and well-being conditions.

Table 1. *Mean, standard deviation and correlation between psychosocial factors (n=145)*

			Correlation						
	Mean	SD	1	2	3	4	5	6	7
1. Job demand	2.51	0.56	1						

2. Job control	2.86	0.69	0.17*	1					
3. Managerial support	2.75	0.78	0.19*	0.45**	1				
4. Colleague support	3.87	0.48	0.05	0.45**	0.50**	1			
5. Relationships at work	2.80	0.64	0.39*	0.04	0.01	0.26**	1		
6. Role conflict	4.23	0.51	0.07	0.31**	0.48**	0.45**	0.05	1	
7. Organizational change	2.64	0.81	0.13	0.49**	0.65**	0.41**	-0.03	0.36**	1

**correlation significant at the 0.05 level;* ***correlation significant at the 0.01 level*

Table 2. *Mean, standard deviation and correlation between health and well-being conditions (n=145)*

			Correlation			
	Mean	SD	1	2	3	4
1. Job satisfaction	2.81	0.73	1			
2. General health	3.59	0.89	0.33**	1		
3. Mental health	3.39	0.65	0.24**	0.33**	1	
4. Vitality	3.33	0.70	0.29**	0.51**	0.62**	1

**correlation significant at the 0.05 level;* ***correlation significant at the 0.01 level*

Bivariate associations between psychosocial factors and health and well-being conditions

Bivariate associations between psychosocial factors and health and well-being conditions are summarized in Table 3. Although psychosocial and health and well-being factors reflected different aspects, the findings indicated that conditions of job satisfaction and vitality are all highly interrelated with all psychosocial factors. General health and mental health have partial interrelationships with psychosocial factors.

Table 3. *Bivariate associations between psychosocial factors and health and well-being conditions (n=145)*

	Job satisfaction	General health	Mental health	Vitality
1. Job demand	0.38**	0.22**	0.29**	0.35**
2. Job control	0.29**	0.33**	0.09	0.25**
3. Managerial support	0.36**	0.27**	0.10	0.27**
4. Colleague support	0.25**	0.13	0.14	0.31**
5. Relationships at work	0.25**	0.15	0.30**	0.23**
6. Role conflict	0.23**	0.27**	0.16	0.28**
7. Organizational change	0.32**	0.14	0.06	0.20**

**correlation significant at the 0.05 level;* ***correlation significant at the 0.01 level*

Comparison of different demographic flight attendants

Table 4 shows the ANOVA (analysis of variance) results for the mean responses to psychosocial factors and health and well-being conditions from different personal backgrounds of flight attendants. Except for the variables of marital status and flight pattern, those of gender, age, tenure, job position, children number, block hour, and WMSDs were indicated to have significantly different responses to psychosocial factors and health and well-being conditions respectively. For example, male or female flight attendants' influence their own capacity to cope with job control (p = 0.023), social relationships at work (p = 0.015), general health (p = 0.003), and vitality (p = 0.025). Flight attendants with children, compared with those without children, had different perceived job demands (p = 0.029), job control (p = 0.037), and general health (p = 0.030). Moreover, the flight attendants with work-related musculoskeletal disorders, compared with those without WMSDs, had negatively perceived psychosocial job demands (p = 0.005), relationships at work (p = 0.007), general health (p = 0.005), mental health (p = 0.001) and vitality (p = 0.001).

Table 4. *The ANOVA results for the mean response to psychosocial factors and health and well-being conditions (n=145)*

	Gender	Age	Tenure	Job Position	Marital Status	Children Number	Block Hour	Flight Pattern	WMSDs
1. Job demand						4.41*			8.28**
2. Job control	5.28*			6.12**		4.86*			
3. Managerial support		6.07**	5.86**	9.13**					
4. Colleague support									
5. Relationships at work	6.06*								7.58**
6. Role conflict		2.91*	2.71*						
7. Organizational change		3.02*		4.05**					
8. Job Satisfaction				3.34*					

9. General Health	9.36**			4.80*		8.24**
10. Mental Health					3.59*	12.11**
11. Vitality	5.13*					12.68**

* *F statistics significant at the 0.05 level;* ** *F statistics significant at the 0.01 level*

*Regression analysis for predicting health and well-being and WMSD*s

We performed further multiple regression analyses to predict health and well-being condition, and work-related musculoskeletal disorders using psychosocial factors as explanatory variables. Table 5 shows the results of models to predict health and well-being conditions and WMSDs. For example, managerial support is the first and job demand is the second significant elements chosen from psychosocial factors to predict job satisfaction. Both factors are positive and can explain the variation of job satisfaction model by 21%. To predict the probability of having WMSDs, logistic regression analysis is used. The model shows that job demand factor is the only one positively significant factor which explains the variation of WMSDs by 9%. It is noted that managerial support is not much related to WMDS but job demand among psychosocial factors plays the most important role to predict health and well-being conditions and WMSDs.

Table 5. *Multiple regression analyses of health and well-being conditions and WMSDs (n=145)*

Model	β	Standard Error	*P* Value
Job satisfaction (adjusted R^2=0.21)			
Managerial support	0.32	0.10	0.000
Job demand	0.29	0.07	0.000
General health (adjusted R^2=0.14)			
Job control	0.22	0.11	0.008
Role conflict	0.19	0.14	0.024
Job demand	0.17	0.13	0.032
Mental health (adjusted R^2=0.12)			
Relationships at work	0.22	0.09	0.010
Job demand	0.21	0.10	0.015
Vitality (adjusted R^2=0.12)			
Job demand	0.35	0.10	0.000
Work-related Musculoskeletal disorders (WMSDs) (adjusted R^2=0.09)			
Job demand	1.10	0.40	0.006

The classification results in Table 6 present the correct percentage using logistic regression model with job demand as an explanatory variable to predict the probability of WMSDs.

Table 6. *Classification table of work-related musculoskeletal disorders (WMSDs)*

Observed		Predicted		
		WMSDs		Percentage Correct
		Yes	No	
WMSDs	Yes	115	0	100.00
	No	29	1	3.30
	Overall Percentage		80	

Discussion

One objective of this study was to identify what kind of psychosocial factors exist among flight attendants, which might affect WMSDs. The mean score for job demand is extremely low (2.51) followed by organizational change, managerial support. This means that the flight attendants perceived high job demand and low managerial support. There are undoubtedly many stressful aspects of the flight attendants, including evermore longer working hours in a long-haul flight, unachievable deadlines, working very fast in short-haul flight, working very intensively due to irregular working and rest patterns, etc. Although the prevention should be multidimensional (i.e. ergonomic interventions), the findings of this study suggested that matching workloads and job demands to flight attendants' capacities is critical.

This study also revealed the presence and severity of work-related musculoskeletal symptoms experienced by the flight attendants working for a commercial Taiwanese airline. The numbers of flight attendants that reported WMSDs are quite high. We suggest that work-related musculoskeletal symptoms are very common health problems in this special type of job and efforts to reduce the prevalence of WMSDs may be important to enhancing the well-being and satisfaction of flight attendants. Finally, our preliminary research has provided basic information regarding psychosocial factors related to health and well-being conditions and WMSDs among flight attendants. We found that there is an association between most of the psychosocial factors (i.e. job demand, job control, managerial support, relationships at work, role conflict) and the health and well-being factors. In particular, the research indicated the factor job demand is significantly related to job satisfaction, general health, mental health, vitality, and WMSDs. However, our findings did support MacDonald's (2004) ergonomics model and confirm the view that job and task demands are the main focus to minimize WMSDS.

Conclusions and Further Directions

In conclusion, this study showed that job demandis an important factor in predicting WMSDs. The findings from this study suggest a way for managers who want to prevent WMSDs or improve the work environment for flight attendants. The small sample size is one of the limitations. Therefore, a further larger study is recommended.

Acknowledgements

This research was partly supported financially by Chung Yuan Christian University, Taiwan. The authors wish to thank Ms Margaret Stewart, Business TAFE School, RMIT University, for her devoted work in academic consultation and assistance in editing the manuscript. Thanks also go to all participants for responding to the survey.

References

Bureau of Labor Statistics. http://www.bls.gov/opub/ted/2008/dec/wk1/art02.htm 〔Accessed January 2009〕

FSF Editorial Staff. (2002). Study of airline's flight attendants finds more than half of injuries affect muscles and bones in back, neck, shoulders. *Cabin Crew Safety, 37 (4),* Flight Safety Foundation.

Kao, L.-H., Stewart, M., Lee, K.-H. (2009). Using structural equation modeling to predict cabin safety outcomes among Taiwanese airlines. *Transportation Research Part E: Logistics and Transportation Review, 45,* 357-365.

Kelleher, C., McGilloway, S. (2005). Survey finds high levels of work-related stress among flight attendants. *Cabin Crew Safety, 40 (5)*, Flight Safety Foundation.

Kumar, R., Kumar, S. (2008). Musculoskeletal risk factors in cleaning occupation—A literature review. International *Journal of Industrial Ergonomics, 38,* 158-170.

Lee, H., Wilbur, J., Conrad, K.M., Moradam, D. (2006). Work-related musculoskeletal symptoms reported by female flight attendants on long-haul flights. *Aviation, Space, and Environmental Medicine, 77 (12),* 1283-1287.

MacDonald, L.A., Deddens, J.A., Grajewski, B.A., Whelan, E.A., Hurrell, J.J. (2003). Job stress among female flight attendants. *Journal of Occupational Environmental Medicine, 45 (7),* 703-714.

MacDonald, W. (2004). Workload, stress and psychosocial factors as hazards for musculoskeletal disorders. *Journal of Occupational Health and Safety, 20 (1),* 37-47.

NIOSH (1999). Stress at work. *NIOSH Publication No. 99-101*http://www.cdc.gov/niosh/stresswk.html〔Accessed January 2009〕

NOHSC (2003). How to identify and assess the determinants of work-related stress in physical handling work environments, and based on this, how to reduce workplace stress levels in such environments. *Issues Paper 6 for the review of the National Manual Handling Standard and Codes.* Work Environment Research Centre, La Trobe University.

OSHA (2002). Protecting workers from ergonomic injuries. http://www.osha.gov/Publications/JSHQ/spring2002/protwork.htm 〔Accessed February 2009〕

OSHA (2009). Ergonomics for the prevention of musculoskeletal disorders . Guidelines for nursing homes. http://www.osha.gov/ergonomics/guidelines/nursinghome/draft/index.html

Saastamoinen, P., Laaksonen, M., Leino-Arjas, P., Lahelma, E. (2008). Psychosocial risk factors of pain among employees. *European Journal of Pain*, Article In Press.

Siu, O.-L., Phillips, D.R., Leung, T.-W. (2004). Safety climate and safety performance among construction workers in Hong Kong The role of psychological strains as mediators. *Accident Analysis and Prevention, 36,* 359-366.

Walgren, L.G., Hanse, J.J. (2007). Job characteristics, motivators and stress among information technology consultants: A structural equation modeling approach. *International Journal of Industrial Ergonomics 37,* 51-59.

FATIGUE AND ITS EFFECT ON CABIN
CREW MEMBER PERFORMANCE

Stephanie Hide, MAS
PO Box 27203, Houston, TX 77027

Since 1993, the National Transportation Safety Board (NTSB) has stated *fatigue* was a contributing factor in eight airline catastrophes in the US resulting in 250 fatalities. Many proposals to mitigate fatigue as a safety issue in aviation have been suggested. Those on the NTSB List of Most Wanted Transportation Safety Improvements involve "hours of on-duty work" rules, which provide an essential set of limits on the work day for all transportation workers. However, most fatigue studies have focused on cockpit crew and not on the cabin crew. This report investigates cabin crew members, their scheduled work, rest and sleep times and the implications for aviation safety. A single case study is presented here, as well as a review of data suggesting why changes are necessary.

Keywords: fatigue, circadian, cabin crew

Background

Fatigue in aviation. Fatigue has been defined by John Caldwell, Ph.D., and Lynn Caldwell, Ph.D., who are both leaders in aviation fatigue research, as "the state of tiredness that is associated with long hours of work, prolonged periods without sleep, or the requirement to work at times that are 'out of sync' with the body's biological or circadian rhythms" (Caldwell, J. A. & Caldwell, J. L., 2003, p.15). Other contributing factors that create cabin crew member fatigue include early report times and breaks that are too limited to allow for eating or napping (Caldwell, J. A. & Caldwell, J. L., 2007). Cabin noise, vibration, turbulence and diminutive cabin quarters all add to increased stress levels and fatigue among Flight Attendants. Deficient crew rest space in operational areas or on aircraft, insufficient water supply or crew meals, commuting, sleep apnea or poor sleep habits also contribute to Flight Attendant weariness. Cabin crew member fatigue is predominantly thought of as a function of scheduling, workload requirements and many of the contributing factors mentioned above. As a result, this study looked at the affects of length and timing of work, off duty sleep quality and flight duty performance.

In the aviation environment, symptoms of fatigue include impaired mood, forgetfulness, reduced vigilance, poor decision-making, slowed reaction time, poor communication, or becoming fixated, apathetic, or lethargic (Conners et al., 2007). These symptoms result in performance errors and an unsafe environment during flight. Specifically:

> A person with a mental effectiveness of 70% has the same reaction time and cognitive ability as when he or she has a blood alcohol content (BAC) of 0.08 (the level corresponding to "legally drunk" in many countries). Studies have also shown an increase in human factors related accidents when people are fatigued and operating with decreased mental effectiveness (Sleep Performance, Inc., 2007, p.8).

Workload increasing the problem. "Between 1986 and 1999, the load factor for U.S. carriers serving domestic and foreign locations increased by about 13% and 21% respectively." Moreover, "from 1986 to 1998, the average U.S. domestic trip length increased from 767 to 813 miles, and the average foreign trip length increased from 2,570 to 3,074 miles" (National Research Council: Board on Environmental Studies and Toxicology, 2002). Current Flight Attendant duties reveal that their workload involves multiple tasks, consisting of walking, bending, lifting and pushing and being available to cope with numerous situations in the cabin. Juggling tasks, physical activity and dealing with the public are all stressful and increase the rate at which flight attendants are fatigued on these flights of increasingly long duration.

Solving the Problem. The FAA is in the process of deciding how to resolve and reduce risk of fatigue specifically as it falls under the Safety Management Systems (SMS) guidelines. Currently, the FAA is meeting with fatigue researchers, unions, airline management and the NTSB on how to resolve this issue.

The International Civil Aviation Organization (ICAO) is now requiring regulatory authorities worldwide to implement SMS. ICAO defines this as "an organized approach to managing safety, including the necessary organizational structures, accountabilities, policies, and procedures" (New, 2008, p. 1). Compliance with ICAO guidelines will be a step forward in operational safety, providing operators with a structure for recognizing and reducing the effects of universal hazards while constantly improving their programs. The program is based on a four-tiered model referred to as the "four pillars": (1) safety policy, (2) risk management, (3) safety assurance, and (4) safety promotion (New, 2008, p. 1).

The current NTSB approach to mitigate fatigue is twofold. On one hand, it recommends scheduling changes determined by using scientific-based computer models which consider circadian rhythms and the need for significant rest periods. In addition, the NTSB advocates educational programs for crew members and updating company attendance policies that discourage employees from calling in fatigued.

To show compliance with the ICAO and NTSB recommendations, the FAA may soon be more proactive in addressing the subject of fatigue. For example, Fatigue Risk Management Systems (FRMS) programs may be required in the near future at commercial airlines.

> FRMS is often understood to be a scheduling or rostering tool. It is actually a wider risk-management concept, which incorporates all mitigation strategies, training and education, and performance measures integrated to managing crew or operator fatigue in a manner that promotes safe operations (Graeber, 2008, p. 3).

Experiment

Limitations and Assumptions. Time and funding issues limited this study to collecting questionnaire data and Sleep Bracelet© results from only one flight attendant. This work assumes that the flight attendant in this study is representative. Because there was only one respondent, it was not possible to determine if the results were representative of a range of people and situations (e.g. psychometrically reliable). Future work should target collecting data from a pool of Flight Attendants.

Data Collection. Two sources of data were provided: (1) a quantitative source, data collected by a Sleep Bracelet© wrist monitor provided by Sleep Performance, Inc., and (2) a qualitative source of data, a questionnaire developed by the author.

The subject wore the Sleep Bracelet© for three consecutive trips and answered the questionnaire relating to these duty periods. The approximate time frame for acquiring the data was three weeks. This data was analyzed with descriptive statistics which describes the data by tables and graphs (Table 1, Figure 1 and 2). Additionally, the raw data is presented in a graphical format using the Sleep Bracelet© software detailing the changes that occurred in performance (Figure 2).

A two part questionnaire was based on a literature review and personal experience in the area of aviation safety and service. The first section was demographic information; the second section was a subjective questionnaire covering the time the Sleep Bracelet© wrist monitor was worn. The subjective questionnaire addressed sign in, layover and pick up times as well as crew rest length and passenger loads. Additional inquiries about noise levels of hotel rooms, crew break rest areas, passenger disruption issues, staffing of crew members and nutrition were posed. A panel of independent experts reviewed the questionnaire for face validity and found it acceptable, and the content was evaluated by the author.

Independent Data. Independent data was also collected from the participant's trips. Official airline records displayed the differences between the "scheduled" flying times and "actual" flying times, as well as significant ground delays. The "scheduled" flying time shows the trip's original planned time. The "actual" flying time is the final time it took to complete the trip (Table 1).

Results

Data Acquisition Time. The data acquisition time is broken down into baseline preset (calibration hours) working prep/working (time allocated to prepare the aircraft before boarding and flying time), sleep, and personal (free time) (Figure 1). During the acquisition time, discrepancies were mostly noted in the area of sleep time. There were 129 hours of sleep time recorded. These hours calculated into 35% of time sleeping; which averaged to 8.6 hours per day.

From the total sleep time, the participant was at rest for 13 hours during which the Sleep Bracelet© had noted the participant was sleeping. These discrepancies were due to minor time differences of sleep and wake periods reported by the participant. The inconsistency in total hours gives a 3.5% error rate. This error rate validates the accuracy of the Sleep Performance, Inc. sleep analysis since it falls within the error rate of 10% or less published by the manufacturer.

Mental Fatigue Analysis. As seen in Figure 2*,* the shaded lines indicating High Risk, Reduced and Normal within the bar graph along with the dotted line showing the 70% range of where cognitive impairment begins was interpreted 100% accurately and did correctly highlight mental effectiveness.

Discussion

The results of the quantitative data from the Sleep Bracelet© confirms that mental alertness is affected by long duty periods without a break as compared to long duty periods with a break. These results also highlight the affects of circadian rhythm on performance depending on time of day. Crew member's sleep is minimized the night before pick up because of the time change and a break in the body's circadian rhythm. Therefore, sleep is interrupted and is not restful. The Sleep Bracelet© was effective and accurate in demonstrating how the circadian rhythm controls our sleep patterns even when we cross time zones. Based on this study, the Sleep Bracelets© would be a useful tool in assisting with designing fatigue reducing schedules for cabin crew members.

The data from the scheduled trips reveals that frequently the actual flying time is longer than the scheduled flying time because of headwinds, routing due to weather, ground delays or air traffic. This is significant since it shows how actual total hours flown are often longer than scheduled flying hours. These actual scenarios demonstrate why layover times can be shortened or the drive home from a trip may be later than a cabin crew member anticipated.

Conclusion

The objective results of this study from the Sleep Bracelet© determine that it is consistent with the questionnaire and flight schedule data from the airline. The usefulness of the Sleep Bracelet© in identifying and predicting the fatigue risk of flight schedules with the aid of a computerized Fatigue Avoidance Scheduling Tool (FAST™) and a Sleep Activity Fatigue and Task Effectiveness Model (SAFTE™) model developed by the Institutes for Behavior Resources (IBR) is evident by the results shown in this study. "The FAST™ model is software which makes predictions about the levels of performance effectiveness that can be expected with specific work/rest schedules" (Caldwell, Jr. & Caldwell, 2003, p. 119). For example, based on this study, a commercial airline could implement earlier departures out of an East Coast Airport and later departures out of Europe to be more consistent with the East Coast time circadian rhythms of cabin crew members. Scheduling earlier take offs out of East Coast Airports to Europe would keep crew members on landing times that do not fall into the body's low circadian rhythm cycle.

The later departure times out of Europe would allow the body to adjust and be more rested for take off. A computerized program would be extremely beneficial in achieving this most important fatigue mitigating strategy. The benefits of creating timetables that are less fatiguing include improved attention and mental cognition and improved disposition and crew coordination (Caldwell, Jr. & Caldwell, 2003).

The technological advancement of computerized systems such as the FAST™ model would help aviation carriers identify which "city pairs" or trip combinations may cause fatigue. Schedule design principles of a computerized system would assist the FAA in reevaluating the scheduling and layover time regulations of 14 CFR 121.467 and 135.273 as they are currently written.

Table 1. *Actual versus scheduled flying time.*

TRIP	On Duty Layover (ODL)	Actual ODL	HOURS Scheduled Flying	Actual Flying	Difference Flying	Unscheduled Additional (%) Flight time	Duty time
1	25h 10m	24h 39m	15h 45m	16h 53m	1h 08m	7.2	14.2
2	25h 10m	24h 48m	15h 45m	16h 30m	1h 15m	4.8	n/a
3	24h 55m	25h 01m	17h 30m	17h 44m	0h 14m	1.3	n/a
Total	75h 15m	74h 28m	49h 00m	51h 07m	2h 37m	13.3	14.2

Note: There were 2 hours (h) and 37 minutes (m) of additional flying time calculated after all three round trips were completed. This averaged approximately 52 minutes per round trip of extra flying time based on head winds and airspeed. There was a 3 hour and 30 minutes ground delay with passengers on board the aircraft while a mechanical issue was being repaired on July 30 (Trip 1). This time was not included as additional flying time but was credited as "holding time" due to the extended period with passengers on board the aircraft while still parked at the gate. Here is another example of how a duty day can be longer than planned because of mechanical problems with an aircraft.
Trip 1: July 28 through July 30 (JFK BRU JFK)
Trip 2: August 4 through August 6 (JFK BRU JFK)
Trip 3: August 8 through August 10 (JFK MXP JFK)

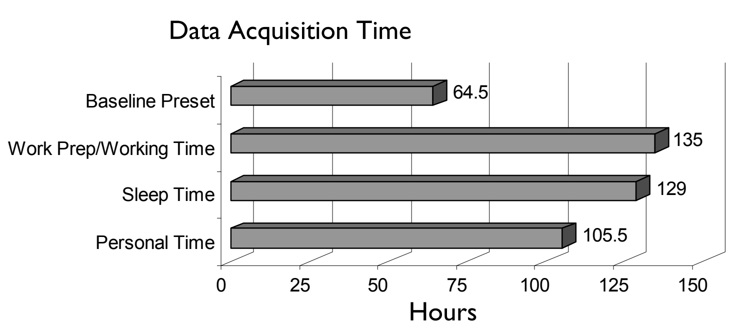


Figure 1. Data Acquisition Time.

There were 434 hours of data recorded by the Sleep Bracelet©. Of this recorded time, the only discrepancy was noted in the actual amount of sleep time. A total of 3.5 hours out of 129 hours were noted by the participant, which is a 3.5% error rate. This is below the 10% error rate published by the manufacture.

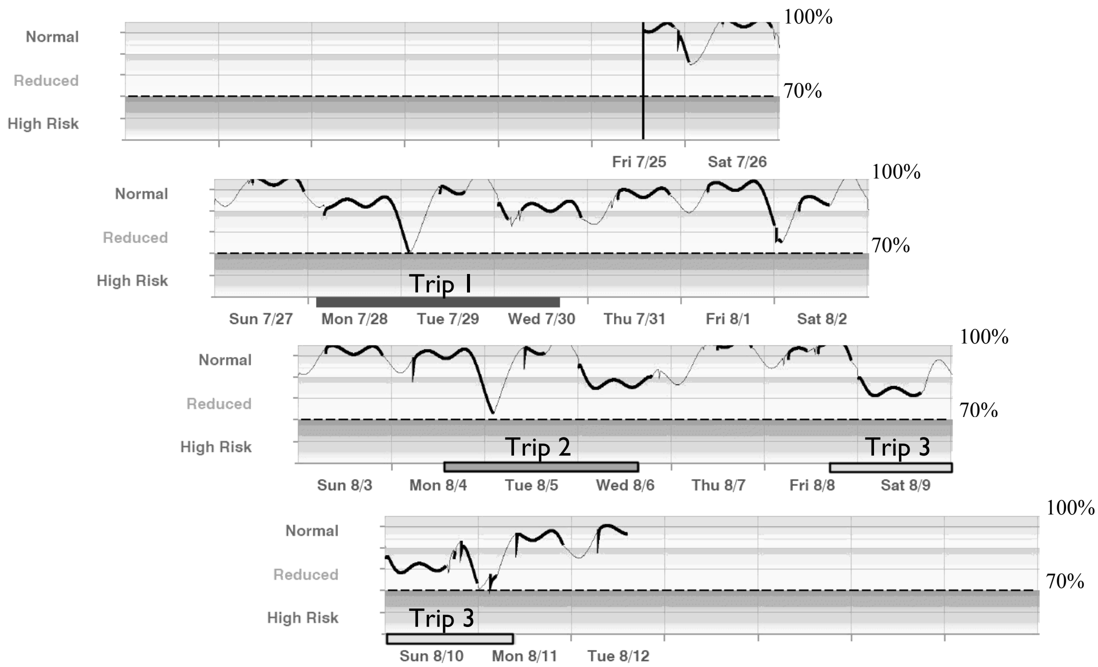


Figure 2. Mental Fatigue Analysis.

The areas indicating High Risk, Reduced and Normal within the bar graph along with the dotted line showing the 70% range of where cognitive impairment begins was interpreted 100% accurately and did correctly highlight mental effectiveness.

Scheduled breaks clearly reduced the possibility of mental effectiveness from falling into the high-risk zone during this study. Where the dark lines (period of work or time awake) have no break for long phases and dip to a lower level near high risk are times where no significant rest period could occur during long periods of wakefulness. When the dark lines stayed within the normal to slightly reduced range of mental effectiveness, breaks averaged at least one to two hours during extended periods of wakefulness. The circadian influence of arriving in the late night and pre-dawn hours along with the homeostatic factor of having been awake for a continuous period seems to coincide at these landing times and shows the most dramatic impact on the Sleep Performance, Inc. graph.

References

Caldwell, Jr., J. A. & Caldwell, J. L. (2003). *Fatigue in aviation: a guide to staying awake at the stick*. Burlington: Ashgate Publishing Co., 119.

Caldwell, J. A. & Caldwell, J. L. (2007, May). Understanding and managing fatigue in operational aviation contexts. *Aerospace Medical Association Proceedings* New Orleans, LA.

Conners, M., Rentmeister-Bryant, H., & De Roshia, C. (2007). *Flight attendant fatigue*. Washington, D. C.: Office of Aerospace Medicine, 2, 3, 11.

Graeber, R. C. (2008, June 17). *Top-down safety focus: fatigue risk management systems (frms); fatigue risk management system within sms* [Abstract]. Aviation Fatigue Management Symposium: Partnerships for Solution. Presented by Federal Aviation Administration, 3, 4.

References

National Research Council: Board on Environmental Studies and Toxicology. (2002). *The airliner cabin environment and the health of passenger and crew*. Washington, D. C.: National Academy Press, 15.

New, M.D. (2008, June 17) *Top-down safety focus: fatigue risk management systems (frms); sms: improving tomorrow… today* [Abstract]*!* Aviation Fatigue Management Symposium: Partnerships for Solution. Presented by Federal Aviation Administration, 1, 2.

Sleep Performance, Inc. (2007). *Sleep and fatigue analysis*. Retrieved October 3, 2008, from Healthy Sleeper Web site: http://www.sleepperformance.com/ web/misc/reports/SleepPerformanceReport_normal.pdf, 3, 7 8 & 12

Acknowledgements

I wish to thank Sleep Performance, Inc., including John Caldwell, Ph.D., and his entire staff, for the use of their Sleep Bracelet© and the processing of the data for this study. Also, I would like to acknowledge the contributions made by Amy Jurewicz, Ph.D., Margaret Watson, Ph.D., and Walter Protheroe, M.A.S., for taking time to review this report and provide helpful comments on the earlier version. The material used here was based on my Graduate Capstone Project at Embry-Riddle Aeronautical University.

POST-TRAUMATIC STRESS IN FLIGHT ATTENDANT'S LABOR

Filipieva Tatiana PhD
Moscow State University by M.V. Lomonosov, Faculty of Psychology
Air Company “Aeroflot-Russian Airlines”
Moscow, Russian Federation

Airline crashes, hijackings, turbulence incidents and other emergencies have the potential for inflicting severe emotional trauma in flight attendants involved either directly or indirectly. A critical incident is a part of the extreme professions in aviation. Although the aviation industry attempted to bring help following an air disaster, post-traumatic stress is still one of the most misdiagnosed and ignored illnesses that can beset an individual. In the article, which is based on the results of Russian and American aviation psychologists’ research, we try to answer the questions: How post-traumatic stress affects cabin crew’s work performance and flight attendants’ life? How to recognize the early warning physical, emotional and behavioral symptoms? What to do about negative effects of disaster-induced emotional injury on the flight attendants’ labor and lives? How to make more effective psychological support to eliminate emotional trauma?

Flight attendants’ names

Over the years the aviation industry has spent a great deal of money and time in exploring ways to reduce the number of aircraft accidents, but they continue to happen and the airlines must learn to recognize the very special needs of the people who survive or witness an incident, especially for the crew members. It’s not a secret that in many crashes the airport emergency plan only for injured and the dead is in effect. The air crash of Boeing-737 of air company “Aeroflot-Nord” in September 2008 (82 passengers, 2 pilots and 4 flight attendants died), a successful ditching of A-320 of US Airways (without victims) and other air crashes evoked us to describe in this article how a critical incident or traumatic event can affect the flight attendants (FA). After any crash of an aircraft the mass media says much about pilots, who are of cause responsible for flight safety, but it is still paid too little attention to the cabin crew members. The media forgets even to mention the FA’s names. Meanwhile the FA contact with the people on board an aircraft, and they are responsible for the passenger’s safety and even lives.

Stress in flight attendants’ professional life

The FA of all the airlines are told in the initial and the on going training programs that they are responsible for the passengers’ evacuation from the aircraft during an emergency. The FA must be ready to do this any time in any flight so they must always be in good physical and mental health condition. Professor d-r Maria Simonson (USA) investigated stress producing activities in the FA’s labor, both in flight and on the ground. She took into consideration the results of interpretation, differences in group administration, union rules and support, airline policy, age, sex, experience, time in service, personal factors and attitudes of the respondents. These are some of the areas affecting health and stress situations in the FA’s labor: 1. Flights, disruption of duties. 2. Career progress. 3. Working environment. 4. Health problems. 5. Sleep. 6. All types of communications with cabin crew management, colleagues, pilots. 7. Meal breaks and eating habits. 8. Schedules. 9. Family life. 10. Leisure activities. 11. Stress producing duties in family and flight activities. 12. Time changes, jet lag. 13. Fatigue. 14. Fears of flying. 15. Safety and emergency.
16. Passenger situations (all kinds). 17. Hygiene on board. 18. Group relations. 19. Job satisfaction. 20. Absenteeism. 21. Relations with management. 22. Lavatory and galley problems. 23. Air quality. 24. Pregnancy, PMS, menstruation. 25. Use the on board equipment (galley-pantry, emergency doors, etc.).

Many stressful factors in the FA’s work performance were described in the dissertation (Tatiana Filipieva, PhD, Moscow, Russia, 2006). The FA’s professional activity is carried out in a confined and narrow space, which is restricted by the design of fuselage of an aircraft flying at high speed. The specific character of the stress inherent in the FA’s labor is determined by: the need for wide-ranging technical knowledge concerning operation of on-board equipment; the unique psychological atmosphere aboard passenger aircraft (subconscious fear of death, underlying fear of flying); the fulfilling of multiple functions and difficult combinations of diverse professional roles; crowded conditions in a confined space; the need to deal with a wide diversity of passengers; constant exposure to the public; availability and openness to many people at the same time.

Other unfavorable factors of the working conditions (Anatoly Kochur, 1995) influence the FA's psychophysical condition and health. Some of them are (a) susceptibility to cosmic and solar radiation; (b) sharp barometric changes during takeoff and landing, as well as between airports up to 150 mm Hg in a short period of time (for example, the high level alpine airport in Katmandu [Nepal] or the airport in Amsterdam [Holland] which is located 4 m below sea level); (c) the drop in the portion pressure of oxygen (hypoxia – oxygen starvation); (d) the reduced level of humidity in pressurized cabins (6-8% of the norm); (e) noise levels exceeding maximum standards and unfavorably affecting the cardiovascular and nervous systems; (f) the vibration of resonance frequencies, causing deformation of organs and tissues; (g) frequent and rapid changes in weather-climatic conditions depending on the geography of the flight; changes in the time zones of up to 8 hours in the course of one flight; (h) stressful and conflict situations on board, etc. The above described unfavorable stressful factors of the working conditions are the FA's everyday professional life. Fortunately, the FA's management and unions of different airlines have already recognized these facts and start to realize that the crew members involved in accidents or incidents experience a number of psychological problems which may render the FA be unable to meet the exacting professional requirements, often through no fault of their own.

Critical incident

Professor d-r Simonson (USA) defines a critical incident as any situation that a person finds emotionally overwhelming and that attacks a person's ability to cope with it. Any critical incident is a tragic or traumatic event of such significance to the involved participants that it may cause a person to experience unusually strong emotional reactions which have the potential to overwhelm normal coping abilities. Types of critical incidents are as follows: (1) serious injury, unexpected death, or suicide of a colleague; (2) serious injury or death of a child, under tragic circumstances; (3) actual or perceived threat to physical safety/life; (4) actual or perceived threat to the organization with which you identify any disaster or major crisis that involves the organization (M. Simonson). A special questionnaire helped us in Aeroflot to get the following results (2000): the FA (n - 670) prioritized personal values as follows: health – 90%, family – 84.5%, personal safety – 6%.

Trauma Reactions

People recognize that they are all individuals and react to situations in different ways but in a critical incident there are a number of general reactions which are shared by many in common. If the FA understand the common and natural reactions which are the result of involvement in highly stressful situations, they will find themselves better prepared to deal with them. Reactions for critical incident can vary greatly according to the nature of the accident and is as unique as the persons involved. The FA's individual reactions to the traumatic events are highly personal and private. Severity of reaction is dependent upon: the individual's personality, current life situation, prior history, support systems. It's also important to take into consideration phases of psychological trauma. They are: (1) shock phase (24-48 hours), which may be characterized by shock, emotional numbness, disbelief, confusion and fear, impaired decision making and concentration are often present during this period; (2) impact phase or immediate aftermath (6-8 weeks after the incident, delayed stress reactions may occur); (3) long-term effects & adjustment (months to several years). Common reactions to traumatic events may include physiological responses which are beyond people's control such as nausea, profuse sweating, muscle tremors, crying, and urgent need for a lavatory. General reactions are practically similar to stress symptoms and manifest themselves in physical, emotional and behavioral symptoms.

Physical symptoms

Physical symptoms are: headaches, poor condition of skin, hair and nails; diarrhea; upset stomach; increased blood pressure and heart rate; dandruff; susceptibility to illness (colds, flu); chest pains; lung disorders, hyperventilation; changes in eating habits or gastric problems such as indigestion, sharp decrease or increase of appetite; undue fatigue, etc. Signs of physical tension also include insomnia or sleep disturbance – problems falling asleep, sporadic sleep patterns up or having repetitive dreams or nightmares. A FA may find himself exhausted for no particular reason – yet not be able to sleep well because of this state of hyper-vigilance and hyper-alertness. The fight-flight reaction is fully activated during the critical incident so that it can be difficult for a FA to calm down physically.

Emotional symptoms

Emotional symptoms include: worry, anxiety, absent-mindedness, aloofness, increased restlessness and tearfulness, a focus on disturbing subjects, sadness, grief, withdrawal, irascible reactions, pessimism, disappointment, feeling of weakness, apathy, easily offended, loss of self-control, insecurity, fear of flying, inability to pull oneself together. A lack of concentration and short term memory problems may interfere with carrying out daily tasks. Intrusive thoughts about the incident or flashbacks (relating to previous traumatic events) can also interfere. If there has been a personal threat or danger, the FA may feel unsafe or fearful that a similar event could happen again. Strong emotional shock, associated with unpleasant memories, evokes neurosis, which the FA try carefully to hide, but which can lead to illness, frustration and, eventually, to leaving the job. For example, an experienced cabin attendant A. (15 years of work for Aeroflot) felt fear when the plane was flying through the turbulent area over the Bermuda Triangle in flight from Moscow to Havana. Later, on subsequent flights at takeoff, on landing and in turbulence he felt a spasm in his throat and pressure at his chest, his forehead and palms became clammy. A psychologist happened to be beside A. during one such incident and gave him some helpful advice of coping strategy. Emotional stability in the face of adverse conditions aboard a passenger aircraft is developed through overcoming fear, acquiring professional skills, as well as through conscious and analysis of a FA’s actions and emotional states.

Other emotional reactions may include a feeling of powerlessness of not being in control, e.g., of something as important as the safety of life or some major aspect of it. When the FA dwell on the feeling that they can't do anything about this, they may start to feel depressed. Perception may be distorted so that hearing can seem muffled, time can be slowed up or vice verse seem to pass very quickly. Tunnel vision is another emotional distortion.

A very common emotional reaction is irritability and anger. The anger may be directed (a) at the organization (FA’s department or airlines) for not foreseeing and preventing the incident; (b) at the colleagues for procedures, errors or lack of support, particularly if someone professionally responsible for what happened; (3) at a close friend or relatives for being in the wrong place at the wrong time. The anger and even rage may be also directed at the person who caused the situation (passenger). According to the results of our research, the FA’s negative opinion (n – 350) of the passengers on board the plane is illustrated by the following figures: irritated (77%), captious (68%), rude (66%), upset (61%), annoying (45%), easily offended (40%), tiresome (37.5%), sex-minded (25.5%).

To improve the attitude to the passengers the FA should take into account the stress factors affecting air passengers in flight (A.Messer, H.Hock, Lufthanza, Frankfurt, Germany). On board the flying aircraft any passenger: a) is unable to stop the flight and leave the aircraft; b) has to obey the flight crew members; c) is compelled to follow the instructions and strict rules of behavior on board an aircraft; d) is limited in freedom of movement and action; e) is compelled to be in the company of strangers; f) has the experience of physical discomfort, causing fatigue, such as narrow space, prolonged inactivity, restricted movements in a fixed position in a passenger seat.

Behavioral symptoms

Behavioral symptoms are: general sluggishness and weakness, uncoordinated movements and actions, trembling in hands and legs, misconduct, rude behavior towards members of the flight crew, colleagues and passengers; disorganization, avoidance of responsibility, fussiness, constant moving things from place to place, deviation from standard procedures of passenger service, errors even in routine automated actions, self-isolation. The FA may feel “on guard” and alert to all kinds of possible threats in life. Checking over the shoulder, feeling apprehensive or becoming overprotective of the children are examples of this type of behavior. This alertness is common but it can be very unsettling and draining.

The FA may feel (unrealistically) responsible for what happened as if they somehow erred or misjudged a situation or failed in their professional responsibilities. Second guessing oneself (e.g., "What if I had … ", "If only I had not … ") is a common reaction if the FA are typically in a position of taking charge or looking after others' welfare. Sometimes the FA may think, that they get what they deserve. They may feel, that if something bad has happened, it might be because they've brought it on themselves or they may have helped it happen. Guilt and second guessing oneself may become prevalent. The FA may become indecisive for a time or lose usual confidence. Other signs of a critical incident stress reaction which may, on the surface, seem less directly connected with the traumatic experience, are: family problems or interpersonal conflicts; loss of interest in the job or previously enjoyed activities; increased use of alcohol, tobacco and other drugs (particularly to help sleep); increased accidents and illness. The FA must be psychologically prepared to experience some or all of the enumerated reactions.

Incorrect behavior while experiencing traumatic event is: (a) isolate yourself or think you are alone in your reactions; (b) get angry that you are experiencing unpleasant reactions for a time; (c) be afraid to arrange individual professional assistance; (d) mistrust your competence; (e) make any major irreversible life decisions with significant long-term implications, for several weeks, without consulting someone impartial – preferably a professional trauma-counselor; (f) self-medicate (caffeine, nicotine, alcohol). There are some troubling results indicated by an anonymous survey among the Aeroflot FA (n – 228): 75% respondents think that the FA use alcohol in business trips "to release stress and relax"; 70% – think that the FA have to take a sick list "in order to rest" not being ill.

A number of factors affect the degree of impact on an individual and should be considered in assessing an individual's condition such as: 1) the severity and nature of the disaster; 2) the impact of assigned/assumed responsibility for others; 3) physical and psychological proximity to the event; 4) the survivor's previous experience in personal crises; 5) the individual's life situation at the time of the event; 6) the nature and effectiveness of handling by others during the emergency; 7) the immediacy of psychological support and treatment. Sometimes there is an almost obsessive need to talk about the incident or bad experience. Again, this is not an uncommon reaction as stewardesses try to master their intense feelings.

The FA, by virtue of the specific rhythm and schedule connected with their work, are separated, frequently left to themselves and having to cope with distress without any help. Psychological trauma, especially if inadequately cared for, can result in behavioral and other effects that reduce the quality of life for the affected persons and lower their efficiency as employees. And no wonder that records of attempted and successful suicides are often a result of untreated or undiagnosed trauma. Research undertaken in 2005 at the institute Superiore di Sanita (Rome, Italy) has shown that the rate of suicide among stewardesses ages 23 to 44 is three times higher than among women of the same age in other professions. It is known that the problems of suicides, drug addiction and alcoholism among the FA exist practically in each air company and are in need of special attention.

Post-traumatic stress is one of the most misdiagnosed or ignored illnesses that can beset an individual. Although the aviation industry has attempted to bring help following an air disaster, it was alarming to count those who completely ignored the value of preventive measures in this area. A critical incident is a part of the extreme professions in aviation and no preventive coping training can dispel it magically. Anyone who has been in a disaster knows that no one is untouched by it. An air crash can reach out and touch any cabin crew member even though he is not in it. The psychological and emotional effects of an airline disaster extend well beyond those on board the aircraft and spread like ripples from a stone cast into a pond, generally diminishing with distance but present in significant amounts far from the point of impact. Experiencing traumatic event can be an emotional shock even when a FA is only indirectly involved in the critical incident. Many different incidents such as death of a loved one, witnessing a traumatic incident such as fire or accident, severe stress in marital discord, loss of friends or family, divorce etc., can contribute to the possibility and potential of post-traumatic stress in a vulnerable individual.

Psychological recommendations

It is helpful for the FA to think and to speak about what has happened. The following words can be used in a therapy conversation with a FA. First of all, you have experienced a very tragic event – an event which will have a very personal, private meaning to you. It may be that your sense of decency has been offended, you may be very sad about a loss, angry that it has happened, upset by the fact that you were unable to prevent the tragedy, or worried about others' and your own safety. It is very important to remember that the psychological reaction you have to this event is highly personal and will not be exactly the same as anyone else's reaction.

Secondly, it is very important to remember that whatever you are experiencing is an individual natural physical and psychological reaction to a very unnatural situation. The incident may have affected you less or perhaps more than you expected it would. You may be experiencing a reaction so intense that you have developed a post-traumatic stress reaction. You may be finding that you are remembering past painful events or personal memories. You may be having a difficult time with present life decisions that have to be made, or future work choices or relationship choices. Whatever your personal reaction, you should remember that you have been psychologically "wounded" and that your body and mind are going through their natural, effective process of psychologically "healing". Remember to allow yourself to go through your own, private, natural healing process, whatever it may be.

Third, you are one of many who are experiencing these reactions at this time. You are not less capable or less competent because of your reactions. Your own personal reaction is very similar to others' and that your co-workers are very accepting of your personal thoughts and feelings. You have also learned that when you experience some degree of confusion or distraction, you can depend upon your co-workers to understand this, to accept this, to assist and support you.

Fourth, you may have thoughts and memories about the incident which keep coming back to you. You may feel many emotional changes, or physical symptoms, or illness. You may find yourself thinking about what is important to you in your life, both at work and away from work, in your family and in your other friendships. You might begin to experience "rough spots" in your marriage or other important relationships. You should remember that these events are probably connected to your reaction to the incident and to your healing process. It is very important to understand this and to speak with someone about it. It is important that you do not make any major decisions about work, family life or any other irreversible life decisions without considering that they might be part of a post-traumatic stress reaction. Your sense of confidence and credibility will return in time. You will be able to resume your normal lifestyle in time.

Correct behavior

Correct behavior while experiencing post-traumatic stress is the following: 1) Understanding and accepting oneself. 2) Support of oneself as an individual. 3) Care of one's physical condition. 4) Active life and entertainment. Understanding and accepting oneself means: expect and accept a period of uncharacteristic thoughts, feelings and behavior; studying one's own priorities and values, analyze needs and desires, give yourself permission not to be "yourself"; permitting oneself to be natural, properly evaluate oneself and one's capabilities, identify reasons that lead to an increase of stress; be patient with your own process of healing; take self rating psychological tests for stress vulnerability and burnout.
Support of oneself as an individual means: share your thoughts about the incident, feelings and reactions with the co-workers, special friends and family members; make presents to yourself or to someone else, accept compliments from the passengers and colleagues and give them to others, allow oneself small indulgences, be able to relax completely when circumstances allow, ask the co-workers for help when needed; accept praise from others in the case of success, find the positive in failures, searching not for the reason something happened (why?) but for how it can assist future tasks (for what purpose?).
Care of one's physical condition means: take special care of your physical health, eat nutritionally (healthy food and a balanced diet) and avoid alcohol, caffeine and nicotine as much as possible, massage therapy, physical activity (running, jumping, walking, swimming), use physical exercise to help discharge the tension, sound sleep, and find time for good and sufficient rest. It is important to look after the physical health so that the FA could have sufficient reserves to deal with the emotional stress.

Active life and entertainment means: read interesting books, watch movies, go to the theatre and concerts, meet with friends, visit family and acquaintances, engage in favorite occupations (hobbies), taking pleasure trips, sight-seeing.

The airlines should start working on accident follow-up programs to ensure that the crew members get the help they need. This is to say that it is the right of every FA to have special training in the curriculum as a health benefit. This does not, however, negate the need for professional medical help, but FA's bruises and cuts are treated, bandaged neatly, yet no care is given or even thought about possible psychic wounds and distress. Disaster survivors need to have their emotional invisible injuries diagnosed and treated with the same care that is applied to physical injuries. They should be evaluated by a mental health professional soon after the emergency and have appropriate care prescribed and furnished just as for a physical injury. It may be very necessary to discuss the emergency situation with the survivor at a fairly early stage following the accident. The longer we wait, the more difficult it will be to convince the crew member of the need for therapy.

In order to help the survivors understand what is happening to them, it is often helpful to put them in touch with other FA who have experienced similar trauma. It is, of course, necessary to ensure that the "helper" is fully recovered from their own trauma before enlisting their help. The recovery will occur in stages. The FA may feel they've mastered their intense feelings, only to find they come back occasionally. With time the FA will become more detached from the event and they will be more freely able to choose to think, or not think about the incident. If the FA find that after 4 to 6 weeks they are still experiencing the stress reactions described above, it would be helpful to seek the assistance of a trained, professional trauma counselor.

There is no doubt that post-traumatic stress can influence the lives, health, safety and behavior as well as many other individual factors of a person's life who has been touched by stress, disaster or any particularly serious situation causing problems. These touch crew work performance, safety, colleagues, family members, and in retrospect the whole future progress.

Conclusion

1) Any aircraft emergency will cause some level of emotional trauma in a number of persons, some directly connected with the emergency, and some at a distance from it. Affected personnel can range from first line workers to top managers, depending on susceptibility and circumstances.
2) The effects of the emotional shock may appear immediately, or not until months or years later, depending on the degree of emotional suppression immediately following the incident.
3) Once the emotional effects appear, they may last for only a short while, or they may endure for years.
4) The emotional effects can appear as undesirable behavior that may manifest itself as irritability, efficiency or absenteeism. Certainly it reduces the quality of life for the survivors.
5) The immediate support of peers associates, and friends can often be all that is needed to help the survivor work through the emotional damage, and professional treatment may often not be necessary. However, this is much more likely to be true if helpers have been trained in psychological support techniques.
6) Appropriate training and crisis intervention programs sometimes including professional psychotherapy, can significantly reduce the adverse effects of this emotional trauma in a survivor, and can speed the FA's return to full effectiveness and enjoyment of life.
7) Helpers must themselves be cautious not to work themselves into "burnout".
8) If airlines' managers want to improve the cabin crew's work performance, they have to recognize the value and importance of training qualified, professional instructors, psychologically able to handle disaster and stress situations, protect and preserve individuals' health, render proper assistance and support to those involved in crashes or other disaster situations.
9) Proper care for the emotional injuries will help both the personnel involved and the company for which they work.
10) The knowledge of post-traumatic stress in FA's labor can promote the good health flight attendants deserve – in health care and in prevention. It will also benefit the air company's human recourses.

MODELING COCKPIT INTERFACE USAGE DURING LUNAR LANDING REDESIGNATION

Zarrin K. Chua
Georgia Institute of Technology
Atlanta, Georgia

Laura M. Major
Charles Stark Draper Laboratory
Cambridge, Massachusetts

Karen M. Feigh
Georgia Institute of Technology
Atlanta, Georgia

Fulfilling NASA's space exploration objectives requires precision landing to reach lunar sites of interest. During the approach and landing stages, a landing point redesignation (LPR) display will provide information to the crew regarding the characteristics of alternate touchdown points. Building on a previous study which examined crew tasks during LPR but did not account for the specialized behavior of experts, this investigation will present a new task sequence model, specific to expert decision-making. This analysis furthers the development of a predictive task execution model, which is used to test the efficacy of alternate information display and operator actuator design concepts. The task model and cockpit display recommendations presented in this study provide a significant improvement in LPR task execution time. This paper examines the task sequence during lunar landing, describes the predictive task execution process model, and recommends cockpit display requirements for effective decision making.

During Apollo missions the astronauts were required to land near a predetermined lunar site to achieve national and scientific mission objectives. Safe landing was achievable through certification of the area prior to flight, providing adequate safe landing areas. Future lunar missions place an even greater emphasis on complete lunar surface accessibility and precision landing at sites of interest (Brady, Schwartz & Straube, 2006). To successfully reach these goals, future lunar-bound astronauts will be aided by an Autonomous Flight Manager (AFM) and a set of displays to assist the crew in performing complex and critical tasks during landing. During the landing point redesignation (LPR) task, crucial information will be provided to the crew on the LPR display by translating raw sensor data (from a LIDAR) into information required to support crew decision making. The focus of this study is to improve the design of a display to facilitate crew cognitive processes during LPR. Specifically, this paper examines the task sequence during the LPR task, describes a predictive task execution process model, and recommends cockpit display design requirements to ensure effective decision making.

Model Description

Task Sequence

This work builds on a previous study (Chua & Major, 2009) which examined crew tasks during LPR but did not incorporate the behavior of experts. Experts react differently than novices, especially in time critical, high-stakes situations. Experts look for cues and patterns to determine the course of action to achieve an objective (Klein, 1998), rarely following a linear approach. Klein's theory of recognition-primed decision making (RPD) asserts that mental simulation is used to predict the outcomes of considered options. These options are evaluated as they are formed, rather than brainstormed *en masse* and eliminated individually. The decision making process usually ends once a satisfactory solution is obtained, rather than continuing until a perfect solution is reached.

The RPD theory applies to the initial stages of the LPR task. The previous task model included explicit cognitive steps for the crew to examine each set of hazards and landing aim point (LAP) recommendations individually (Chua & Major, 2009). However, applying RPD, one can predict that the crew would look for large scale patterns in the terrain and quickly determine if the location of the LAPs matched expected terrain patterns. The predicted behavior was further supported by a representative from the NASA Crew Office, who confirmed that the crew will most likely look for identifiable terrain markers (ITMs) to quickly assess the situation, and then make a rapid decision on the suitability of the LAPs given the situation. This behavior is also consistent with the Apollo missions, in which the astronauts were trained to identify specific patterns in the terrain, such as the "Snowman" configuration of several large craters during Apollo 12 (Manned Spacecraft Center, 1970). The recognition task during Apollo required more time than the anticipated execution of future landings because the Apollo decision aid (Landing Point Designator, LPD) was less sophisticated than what can be provided today. Apollo's LPD required several manual steps to first obtain the landing site location before identifying that site out the window.

Based on these insights, an LPR task model was developed for this study by building upon the generic model from Chua and Major (2009). The new LPR task model utilizes RPD theory, where the analysis is non-linear and includes attempting to match environmental cues to expectancies and only performing a detailed analysis if the cues are different, in orientation and in form, than what is expected. This task sequence is illustrated in Figure 1, with the RPD loop highlighted in green.

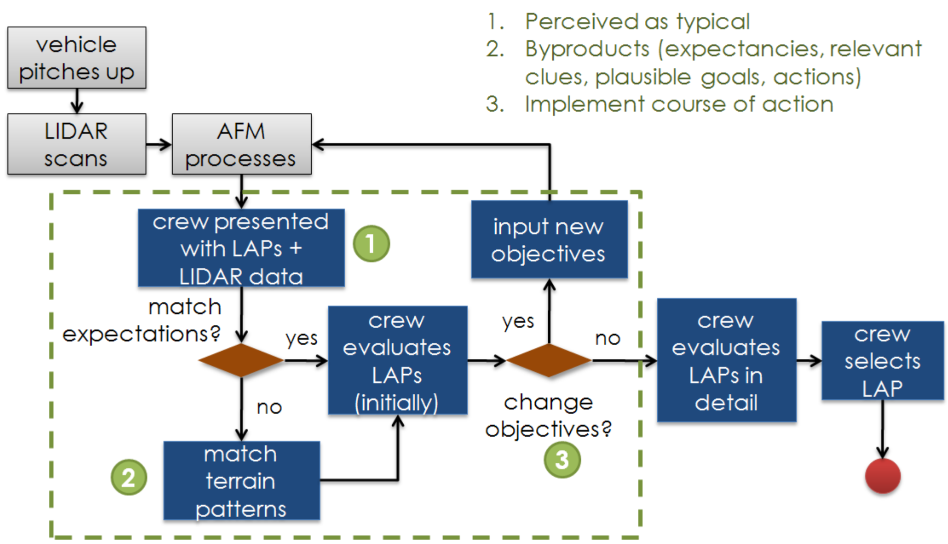


Figure 1. Landing point redesignation task sequence. The dark boxes are within the scope of this investigation.

Two major modifications were made to the LPR task sequence model described by Chua and Major (2009). First, a subtask is added prior to the initial evaluation of the LAPs. The previous model neglected the steps required to develop an overall understanding of the landing site and assumed that these steps would be completed prior to the LIDAR scan. This modification, based on the first phase of RPD theory, provides an estimate of the subtasks required to initially evaluate the landing site, especially in the event of unexpected terrain. Second, the detailed evaluation of the LAPs is revised to account for expert behavior. The previous model assumed astronauts would potentially evaluate every combination of alternative LAPs (based on the capabilities of the display design). While this calculation is acceptable for computing maximum times of expected task completion, RPD and the input provided by a Crew Office representative, are used to refine this assumption and enable more accurate time estimates. This modification places a limitation on the number of landing sites presented.

Once presented with the processed LIDAR data and the AFM information (hazard areas, LAP recommendations), the crew first performs a high level evaluation of the cues (terrain patterns and LAP recommendations), to determine whether the situation is what they expect or not. If the situation does not match their expectations, the crew will then further evaluate the terrain. If the scenario is nominal, the crew will typically proceed to evaluate the LAP options. Each of these distinct tasks (marked by dark boxes in Figure 1) is decomposed into smaller primitive tasks, using the KLM-GOMS (Card, Moran, and Newell, 1983; Olsen and Olsen, 1990). In addition to the tasks decomposed by the original model, the new model includes the new task of terrain pattern recognition, the location of LAPs, and a specific operator actuator for communicating objective changes to the AFM.

Predictive Task Execution Time Model

The predictive task execution process model provides an estimation of time to complete the LPR task based on the astronaut strategy and the mission scenario. This model is essentially a summation of the primitive operators associated with the LPR task sequence, as determined using the decomposition scheme described by Card, Moran, and Newell (1983) in the KLM-GOMS theory. This model also uses the secondary primitive operators in the Chua and Major (2008) study, as first presented by Olsen and Olsen in 1990. These primitive operators are based on the interactions of the operator with the LPR display as described in the previous section. This model is described the relationship presented in Equation 1:

$$f(\alpha, \varepsilon, \Pi, n, H) = 2.4\varepsilon + 4.8\alpha + 2.02n + \frac{5(n+1)(n+\Pi+H)}{2} + 5.72 \quad (1)$$

where α is the number of LAPs evaluated in detail (including the baseline point); ε is the training parameter, where ε is 0 if the training is correct, 1 if the astronauts are unprepared for the actual terrain; Π is the number of points of interest (POI); *n* is the number of objective changes; and *H* is the number of ITMs. A distinction must be made regarding hazards and ITMs. An ITM is a group of hazards that create such a shape or pattern that an astronaut regards this group as a single entity, rather than individual craters or rocks. While this formulation is capable of calculating most feasible lunar scenarios, equation 1 comes with restrictions. Input parameters such as the number of LAPs evaluated in detail and the number of objective changes is difficult to model, as they are dependent on the scenario and operator strategy. This model is better used to determine the range of potential task execution time. To generate this range of task times, the Crew Office is consulted regarding the most feasible astronaut behavior. The following three scenarios are utilized:

1. ***Minimum***: Best case scenario. Expectations are matched (the lunar terrain maps used during training correctly prepared the astronauts for what they actually see); no change in objectives, and only one alternative LAP is evaluated in detail. ($\alpha = 2$, $\varepsilon = 0$, $n = 0$)
2. ***Maximum***. Worst case scenario. Expectations are not matched (astronauts look for the ITM), objectives are changed twice, and all LAPs are evaluated in detail. ($\alpha = 4$, $\varepsilon = 1$, $n = 2$)
3. ***Nominal***. Expectations are matched, objectives are changed once, and only two other LAPs are evaluated in detail. ($\alpha = 3$, $\varepsilon = 0$, $n = 1$)

Equation 1 can then be computed over a range of points of interest and hazards to better understand the task execution time over a variety of scenarios.

Landing Point Redesignation Display Design Recommendations

A significant challenge of LPR is the balance between crew control and inherent time constraints. The crew must be given a means to refine the desired characteristics of a LAP, but too many options can overwhelm the crew and greatly increase mental workload (Smith, McCoy & Layton, 1997). Conversely, automating the decision of a final LAP is faster, but eliminates the crew's human advantages of adaptability and creativity (Wiener & Curry, 1980). The LPR display must seamlessly integrate the static inputs of the AFM (*a priori* mission estimations, etc.), the dynamic LIDAR data and dynamic goals of the crew (based on real-time data). The quality of this integration includes the presentation of AFM information to the crew, and correspondingly, the ability of the crew to communicate goals and intent to the AFM. Unfortunately, poor designs in either can result in bottlenecks, or localized increases in crew workload. Previously, two major bottlenecks were identified using the generic LPR task sequence model. These bottlenecks pertain to factors just presented - operator actuator design (to signal an objective change intent) and presentation of decision-making information to enable detailed evaluation and selection of a final LAP from several choices.

To mitigate the first bottleneck of communicating the intent to change objectives, several operator actuator designs are considered in this research. The previous design, a combination of three slider bars and two buttons (Forest, Cohanim, & Brady, 2008), granted maximum control by allowing the crew to manually set hazard tolerances and the weighting distribution between safety, fuel efficiency, and nearness to the Point of Interest (POI). However, this operator actuator design leads to the possibility of the crew extending more effort than necessary in calibrating the tolerances and weights during the landing. Two other designs are considered: a safety buffer dial and "hot keys", a series of buttons with predefined options regarding tolerances and weighting distribution.

The safety buffer dial changes only the safety tolerance remaining a safe distance from the hazards while also as close as possible to the POI. Turning the dial in one direction would communicate a desire to increase the safety tolerance, thus leading to safer LAPs (farther from hazards, defined as areas beyond a slope and roughness tolerance). Turning in the other direction would decrease the safety tolerance, potentially presenting LAPs closer to the POI. If the dial is designed to provide some static unit/radian and a representation of feedback, the operator could sufficiently fine tune to a specific tolerance. The safety buffer dial is modeled using the KLM-GOMS methodology as a "pointing mouse" primitive operator. While the safety buffer dial is simple to use, this operator actuator is limited in dimension – the specific slope and roughness tolerances and fuel efficiency are neglected. In addition, the specific weight distribution between the top three driving objectives (safety, fuel efficiency, nearness to POI) cannot be determined. Conversely, the hot key concept provides more dimensionality than the dial, but at the cost of tolerance and weight precision.

The hot key concept presents several distinct options to the crew. Each hot key represents a fixed set of tolerances on safety and fuel efficiency and a weight distribution between safety, fuel efficiency, and nearness to POI. The exact tolerances and weight distributions can be tuned based on the mission and crew preferences. The AFM would return alternate sites based on the objective function encoded for each hot key. During the LPR task the crew can toggle between the hot keys. A set of five hot keys are included in this study and are described in Table 1. The hot keys are modeled as a push button.

Table 1. *Hot keys used in the LPR task and their definitions.*

Hot Key	Definition
Safety	The safest landing sites (farthest from hazards, conservative tolerances on slope and roughness). Fuel efficiency and nearness to POI are held equal. Example of weight distribution (on a 100 point scale): 90/5/5 or 80/10/10
Fuel	Most fuel efficient sites (typically center and forward, aft of the LIDAR scanned landing area). Safety and nearness to POI are held equal. Example of weight distribution: 5/90/5, 10/80/10
POI	Nearest to POI. This objective could be interpreted in different ways, if there are multiple POIs presented: the AFM could find aim points nearest to all POIs, or the AFM could find the closest aim points to at least one POI. This research assumes the later interpretation. Tolerances on slope and roughness are less stringent and safety and fuel efficiency are held equal. Example of weight distribution: 5/5/90, 10/10/80
Balanced	Equal, or balanced weight distribution between safety, fuel efficiency, and nearness to POI. Weight distribution: 33/33/33
A Priori	This distribution is based on mission planning projections of the objectives deemed to be most critical during LPR. This distribution does not include any real-time data. The baseline aim point is based on this weight distribution. Examples of weight distribution: 10/25/65, 31/ 43/26, etc.

Of the three operator actuator designs investigated, the hot keys concept is considered the best due to robustness of input (both hazard tolerances and objective weights are communicated to the AFM) and speed of use (one button push). When modeled in usage with the LPR task execution model presented by Chua and Major (2009), the hot keys demonstrated a clear advantage in operator execution time (0.82 seconds vs. 2.35 seconds using the dial) while offering great breadth in objective function selection. Although precision and authority are generally viewed as favorable, especially in manned spaceflight, allowing astronauts to set the tolerance and weight distribution in mid-flight adds complexity to the task. This complexity deepens if the input-feedback loop is not immediate. Therefore, the preprogrammed tolerances and weights of the hot keys reduce operator workload by limiting the options to the crew, while providing adequate authority to the crew.

The hot keys are particularly effective when used in tandem with the new proposed method to rectify the second bottleneck of multiple LAP evaluation. The previous presentation in LPR information limits the crew to a cross-examination of three LAPs (nominal and two alternatives) and the terrain information of each LAP. This limitation increases visibility, as the display becomes difficult to comprehend if all information is presented. Thus, the crew must routinely select, evaluate, and reselect which LAPs to closely examine. To facilitate this evaluation process, the new display presents all of the detailed terrain information of four LAPs (nominal and three alternatives) concurrently, with each hot key selection. However, utilizing the same information presentation as that suggested by Forest, Cohanim and Brady (2008) may not leave all of the information apparent and readily accessible to the astronauts.

To mitigate this problem of information presentation, especially with regards to representation and location on the display, the LPR display is simplified and reorganized to maximize the amount of information seen within a person's field of view. This philosophy manifests in the form of simple symbols, narrowing the amount of data processed by the astronaut, and grouping necessary information in centralized locations. The astronaut can focus more on the degree of quality, rather than determining whether a LAP is within the acceptable envelope.
The following symbols are applied to the vehicle state and LAP terrain characteristics.

1. ***Information superimposed on map of landing area***. The necessary information for LPR is located on one display, with vehicle state and terrain data overlaid on the synthesized map of the LIDAR data. This arrangement allows the astronaut to efficiently focus attention on one main location, minimizing eye movement (Wickens & Carswell, 1995).
2. ***Fuel contour***. The Crew Office reported a fuel contour as critical information to execute the LPR task. This display utilizes a green ellipse superimposed on the photo of the landing area to divide the map into reachable and non-reachable fractions. All alternative landing sites are located within this ellipse. This ellipse also represents the relative fuel cost for each landing site. Landing aim points located closer to the center and along the major axis of this ellipse required less fuel than aim points located on the on the fringe and minor axis.
3. ***Vehicle Footprint Dispersion Error (VFDE)***. The VFDE is represented by a dashed purple circle proportional to the area encapsulated on the map. The diameter of the vehicle footprint plus errors is listed in a box in the lower half of this purple circle.
4. ***Vehicle cross-sectional area***. The vehicle cross-sectional area is represented by a green circle located in the center of the VFDE circle. The size of this cross-section is equivalent to the area on the map. Superimposing

this information on the landing map also allows the operator to quickly determine the relative distance from hazards and the POIs.

5. ***Terrain characteristics***. The terrain characteristics of each LAP are represented directly on the map. A modification of the four axis LAP information representation developed by Needham in 2008 is used for this study. Two of the four axes, slope and roughness margins, are utilized. The hazard and fuel margin axes are represented by other symbols. The slope and roughness margin information is displayed in the same manner prescribed by Needham (2009). Three marks along the axes are used to represent dangerous terrain characteristic (defined as, at the threshold), tolerable, and desired (far from the threshold). The arrows are desired to be as long as possible, hence representing a safe LAP.
6. ***Points of interest***. The points of interest are represented in blue and proportional to the size of this area. Circles and other geometric shapes can be used to represent lunar assets, or scientific spots of interest.

The final form of this display is illustrated in Figure 2. In this figure, the crater has been highlighted as the lone hazard. The nearest to POI hot key (labeled POI) has been selected, and the three alternative LAPs are shown. In accordance with the LPR algorithm formulated by Forest, Cohanim, and Brady (2008), the alternative LAPs are unique points and represent local optimums. Once the operator has chosen a final landing site, he can communicate this choice by pressing the identically labeled button in the lower left corner and immediately pressing the ARM button afterwards. At the conclusion of this action, the LPR task is formally finished.

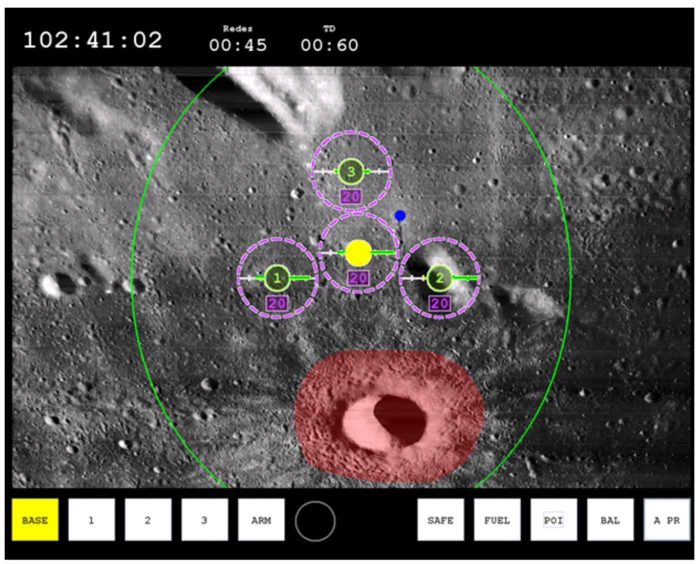


Figure 2. Landing point redesignation display.

Modeling Results and Discussion

The predictive task execution time model is applied to the full design space of potential lunar landing scenarios. These landing scenarios are defined by the number of hazards and points of interest defined in the previous section. The range of execution times across these scenarios is illustrated in Figure 3. The new LPR display design presented in this paper significantly expedites the LPR task. The recommendations, particularly with respect to LAP evaluation, dramatically reduce the time to perform the task by about 50% from the previous display. For the most feasible scenario, one POI, the LPR task is expected to conclude in approximately 31-36 seconds over the range of ITMs. The small variance with respect to the number of hazards is quite promising – denoting a near decoupling of task execution and terrain features. However, this analysis assumes a discrete number of hazards and does not examine hazard coverage of the landing area, or shape of the hazards.

The results from the predicted task execution time model are promising in the field of LPR display design. However, the models developed in this research are based on several key assumptions. The model assumes perfect human behavior. The operators used by the astronauts are completed the same time or less than those prescribed by Card, Moran, and Newell (1983) and by Olsen and Olsen (1990). This model also assumes the LPR algorithm is capable of presenting three alternative LAPs with every objective change. Finding three unique alternative LAPs may not be possible in extreme terrain conditions, and thus, the astronaut may make a quicker decision based on fewer points to consider. Lastly, this model assumes this is the first instance of LPR and first presentation of processed LIDAR scan results and the astronauts are making a decision based on this singular opportunity. To understand and determine the accuracy of this model, further verification and validation is necessary.

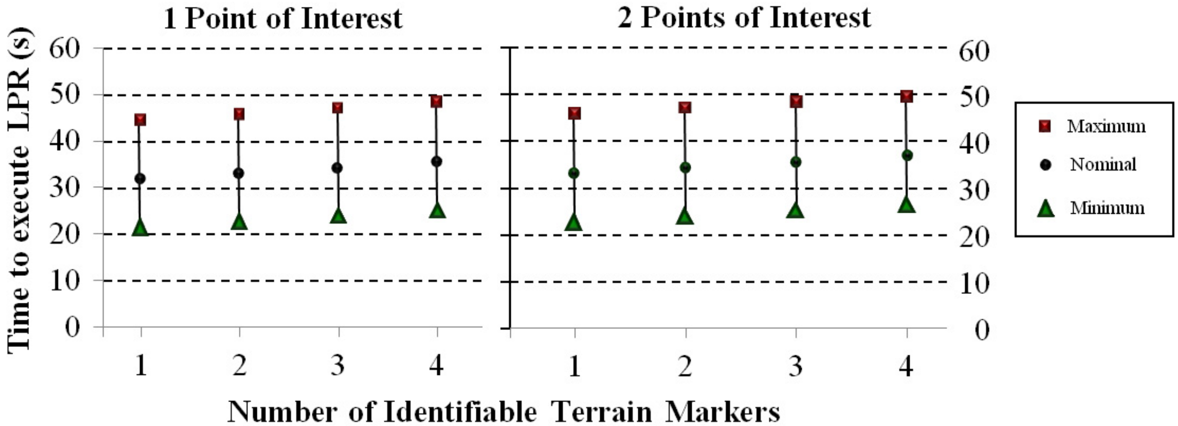


Figure 3. Landing point redesignation task execution times.

Conclusion

This paper presents the formulation of an improved landing point redesignation task model that accounts for the specialized behavior of expert decision-making. From this improved model, a predictive task execution time model was developed to estimate the minimum and maximum time range to complete the LPR task. The landing point redesignation display was also modified based on previously identified task bottlenecks. The hot keys concept was selected as the most effective operator actuator and the presentation of alternative landing site information was reorganized to focus on critical points. The task sequence model and cockpit display recommendations presented in this study provide a significant improvement in the time to execute LPR. For some scenarios, a time reduction of up to 50% is recorded. The most plausible scenario of one POI predicts an execution time of 31 - 36 seconds. While the results presented indicate effective LPR decision-making, the model developed in this study assumes perfect human performance. Further verification and validation is necessary, to quantify model accuracy.

Acknowledgements

The authors acknowledge the entire ALHAT project team consisting of outstanding and distinguished members from NASA's Langley Research Center, Jet Propulsion Laboratory, Johnson Space Center, C.S. Draper Laboratory, and the Applied Physics Laboratory. The program is led by Chirold Epp who deserves special thanks for his leadership and vision. The authors also thank Jennifer Needham and Robert Braun for their assistance and support on this paper.

References

Brady, T., Schwartz, J., & Straube, T. (2006, May). Operational Concept Description for the Autonomous Landing and Hazard Avoidance Technology (ALHAT). NASA Technical Report, ALHAT 2.0-002, Initial Release.

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

Chua, Z. & Major, L. (2009). Task Modeling for Lunar Landing Redesignation. InfoTech@Aerospace. In Press.

Forest, L. & Cohanim, B. & Brady, T. (2008) Human Interactive Landing Point Redesignation for Lunar Landing. IEEE Aerospace Conference.

Klein, G. (1998). *Sources of Power: How People Make Decisions*, MIT Press, Cambridge, MA.

Manned Spacecraft Center: Apollo 12 Mission Report. MSC-01855. March 1970.

Needham, J. (2008). Human-Automation Interaction for Lunar Landing Aimpoint Redesignation, S.M. thesis, Massachusetts Institute of Technology.

Olsen, J. & Olsen, G. (1990). The Growth of Cognitive Modeling in Human-Computer interaction since GOMS. Human-Computer Interaction, 5 (2). 221-265.

Smith, P., McCoy, C., & Layton, C., (1997) Brittleness in the Design of Cooperative Problem-Solving Systems. IEEE Trans. on Sys., Man., and Cybernetics, Part A. 27 (3)

Wickens, C. & Carswell, C. (1995). The proximity compatibility principle: Its psychological foundations and its relevance to display design. Human Factors, 37, 473-494.

Wiener, E. & Curry, R. (1980). Flight-deck automation: promises and problems. Ergonomics, 23 (10), 995–1011.

SYNTHETIC AND ENHANCED VISION SYSTEM FOR ALTAIR LUNAR LANDER

Lawrence J. Prinzel, III[1], Lynda J. Kramer[1], Robert M. Norman[2], Jarvis (Trey) J. Arthur[1], Steven P. Williams[1], Kevin J. Shelton[1], Randall E. Bailey[1]

NASA Langley Research Center[1]
Hampton, VA

Boeing Phantom Works[2]
Hampton, VA

Past research has demonstrated the substantial potential of synthetic and enhanced vision (SV/EV) for aviation (e.g., Prinzel & Wickens, 2009). These augmented visual-based technologies have been shown to significantly enhance situation awareness, reduce workload, enhance aviation safety (e.g., reduced propensity for controlled flight -into-terrain accidents/incidents), and promote flight path control precision. The issues that drove the design and development of synthetic and enhanced vision have commonalities to other application domains; most notably, during entry, descent, and landing on the moon and other planetary surfaces. NASA has extended SV/EV technology for use in planetary exploration vehicles, such as the Altair Lunar Lander. This paper describes an Altair Lunar Lander SV/EV concept and associated research demonstrating the safety benefits of these technologies.

Background

Synthetic Vision

Synthetic vision (SV) is a computer-generated image of the external scene topography from the perspective of the flight deck that is derived from vehicle attitude, high-precision navigation solutions and a database that includes terrain and may include obstacles, trajectory information, relevant cultural features, and other data (Figure 1). The SV display is unaffected by outside weather and environmental conditions (e.g., fog, clouds, or dust) and thus, provides a clear day view regardless of the available outside visibility and can be complemented by real-time enhanced vision (EV) sensors (Prinzel & Kramer, 2006).

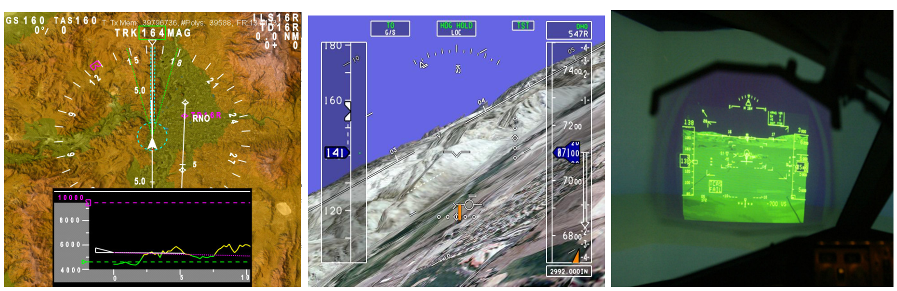


Figure 1. Commercial Aircraft Applications of NASA Synthetic Vision Display Technologies

Past research has demonstrated the substantial potential of synthetic and enhanced vision (SV/EV) for aviation (e.g., Prinzel & Wickens, 2009a; 2009b). These augmented visual-based technologies have been shown to significantly enhance situation awareness, reduce workload, enhance aviation safety (e.g., reduce propensity for controlled flight -into-terrain accidents and incidents), and promote flight path control precision. The issues that drove the design and development of synthetic and enhanced vision have commonalities to other application domains; most notably, during entry, descent, and landing on the moon and other planetary surfaces.

Altair Space Vehicle Synthetic Vision System

The Apollo lunar landings were an extraordinary achievement, requiring highly trained and skilled pilots. These astronauts were selected and trained to navigate a new vehicle in an unknown environment and adapt in the face of numerous potential failures and uncertain conditions with only basic flight instrumentation consisting mainly of "electro-mechanical" gauges (Figure 2). The lunar landing task relied on the pilots, coupled with large forward-facing windows and operational flight profiles which allowed the pilot to see the landing zone for extended periods of time, to perform a visual, manual landing approach or to redesignate and fly to a new landing area if the original landing site was not suitable. The Apollo Lunar Module windows provided almost 70 degrees nose-down visibility.

Though there is no weather on the moon, thrusters can create a dust cloud that can significantly reduce visibility during a critical phase of the flight. Further, sun angles can create visually powerful shadow effects which may cause the loss of depth cues, translational velocities, and landing zone awareness for the flight crew. The importance of pilot visibility was not only emphasized by trajectory design and window definition, but also, by conducting the Apollo landing task only at specific times and locations to provide optimal sun light on the landing site. The mission was designed around lighting conditions that would create shadows that would provide optimal depth perception. These optimal lighting opportunities typically lasted about a week. If the opportunity was missed, the next opportunity to land would not be for another month.

Preliminary concepts for the Altair Lunar Lander (Figure 2), in contrast, provide significantly less external visibility for the astronauts. The Altair design will be significantly larger than Apollo and the size of the windows and their location may be severely constrained and non-optimal for pilot visibility. Mission requirements for fuel-optimized trajectories to polar regions for maximum scientific benefit further compound this problem. As envisioned, future space operations will require frequent trips to less than ideally lighted lunar surface locations (e.g., lunar south pole) emphasizing the importance of providing technological capabilities for unfettered moon surface landings.

Synthetic vision may provide that needed technological capability to enable such future lunar operations. With SV, the designer controls the computer-generated scene lighting, terrain coloring, and virtual camera angles. The visual cues for the landing site in the SV are independent of the sun-angle. In addition, important vehicle state information such as forward and down velocities, altitude, and fuel remaining can be overlaid directly onto the terrain display to significantly ease pilot interpretation of the data and enhance situation awareness. Another advantage for using SV enhanced displays is that advanced precision guidance can be intuitively integrated into SV displays. With one of the US Space Exploration Policy goals to return to the moon, frequent missions to the moon will require precise landing of vehicles (possibly within 10 meters accuracy). It is currently planned to have several habitat modules, power generators, storage and surface mobility units. Therefore, the landing task will involve not only landing on suitable terrain but avoidance of man-made obstacles and approach procedures to avoid over-flight and potential contamination. SV can be complemented by EV, such as a Forward-Looking-Infrared (FLIR) or Light Detection and Ranging (LIDAR), to provide both a database and real-time sensor representation of the lunar terrain surface, man-made objects and obstacles, and hazards (e.g., boulders, craters).

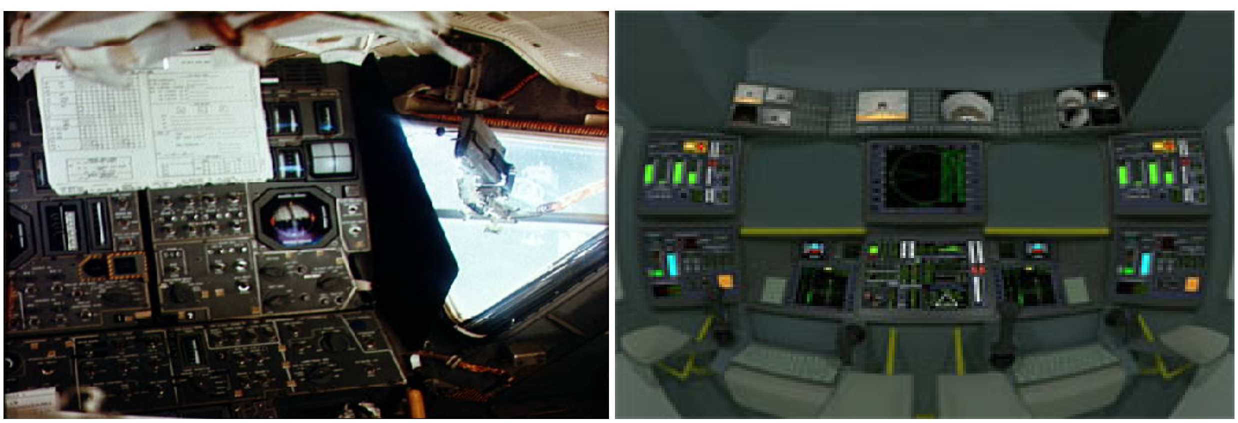


Figure 2. Apollo Lunar Module and Altair Conceptual Flight Deck

Flight Deck Design Development

Research is being conducted at the NASA Langley Research Center to evaluate "aeronautics-domain" technologies and expertise which might benefit the Altair mission. To achieve this goal, preliminary flight deck design concepts have been prototyped and piloted evaluations have begun to establish data which can be used for informed decision-making in the Altair design process. This research is on-going – focusing principally on SV/EV technologies first – and as such, these Lunar Lander flight deck concepts and potential technology applications continue to evolve.

Tactical Displays

Tactical displays are crucial for guidance and vehicle state information to aid the pilot in immediate navigation. SV and EV (see Figure 5) tactical display concepts were evaluated using a head-up and/or head-down display, primary flight display (PFD), and ego-centric perspective display (Figures 3-5). Research is on-going to develop and optimize these display concepts and is continually being modified from the examples shown here. For instance, the first challenge for the Lunar Lander application is the extreme attitude variations during the approach and its critical role on deceleration and precision landing. Display concepts with a velocity-vector centered display concept were initially drafted but were discarded and replaced with attitude-centered concepts. (The velocity vector display was difficult to interpret on approach and especially, near hover, with a potential loss of spatial awareness, energy awareness, and display functionality.) To date, a standard "eight-ball" is assumed as the head-down primary flight reference display to ensure uncluttered guidance and full spatial reference information. The conformal nature of the SV/EV information on the HUD (Figure 3) – even though attitude-referenced - provides critical head-out information especially useful during the final approach and hover transition maneuvers. Automatic transition of HUD imagery from SV to EV is used for optimal performance.

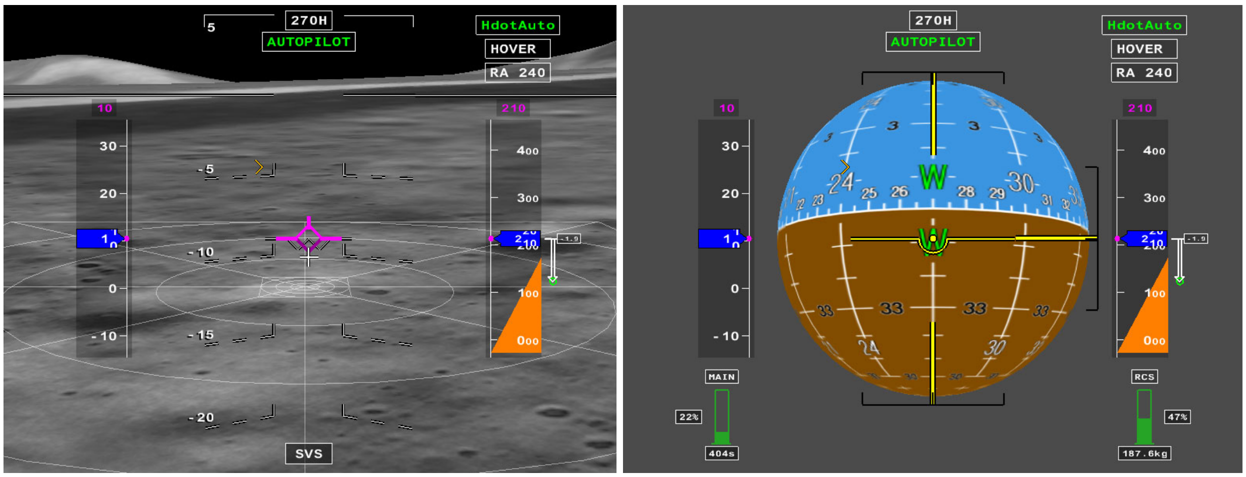


Figure 3. HUD (shown with SV) and Primary Attitude Flight Display

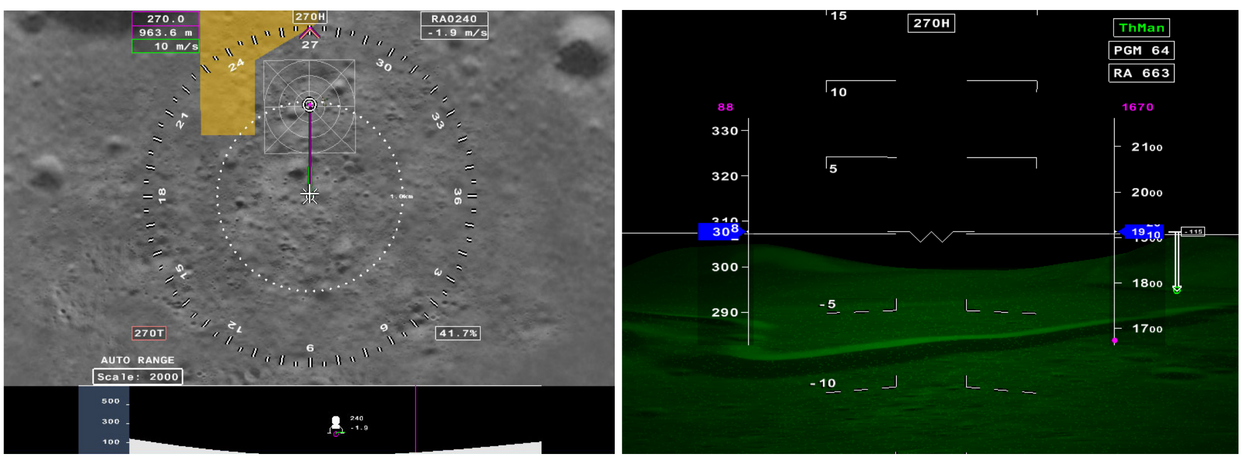


Figure 4. Navigation Display and HUD (shown with EV)

Two companion displays to the PFD and HUD appear to be ideal for introducing valuable SV/EV information for the Lunar Landing mission – the Auxiliary Display (AD) and Head-Worn Display (HWD). The AD (Figure 5, Left) provides ego-centric SV/EV information with a pilot-controllable reference frame. Attitude-reference, velocity vector-reference, or slewable views are provided. By using the AD and its variable viewpoint references, critical terrain and obstacle awareness is provided on a head-down display during the approach phase without the potential loss of spatial awareness or critical guidance information. The PFD is not compromised to accommodate the introduction of SV/EV and the attitude extremes of the planetary descent. Critical SV/EV information is provided by the AD with selected (i.e., minimal) symbology to provide visual momentum between the AD and PFD and other displays. Symbology tailoring and control features are still being defined for the AD. The other important display is a head-tracker HWD. The HWD promises conformal, unlimited field-of-regard information. The formats for SV/EV on HWD displays are now being developed.

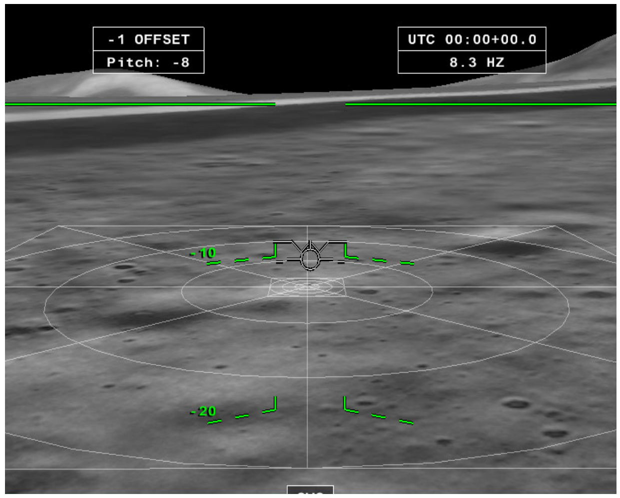


Figure 5. Ego- Centric Auxiliary Displays

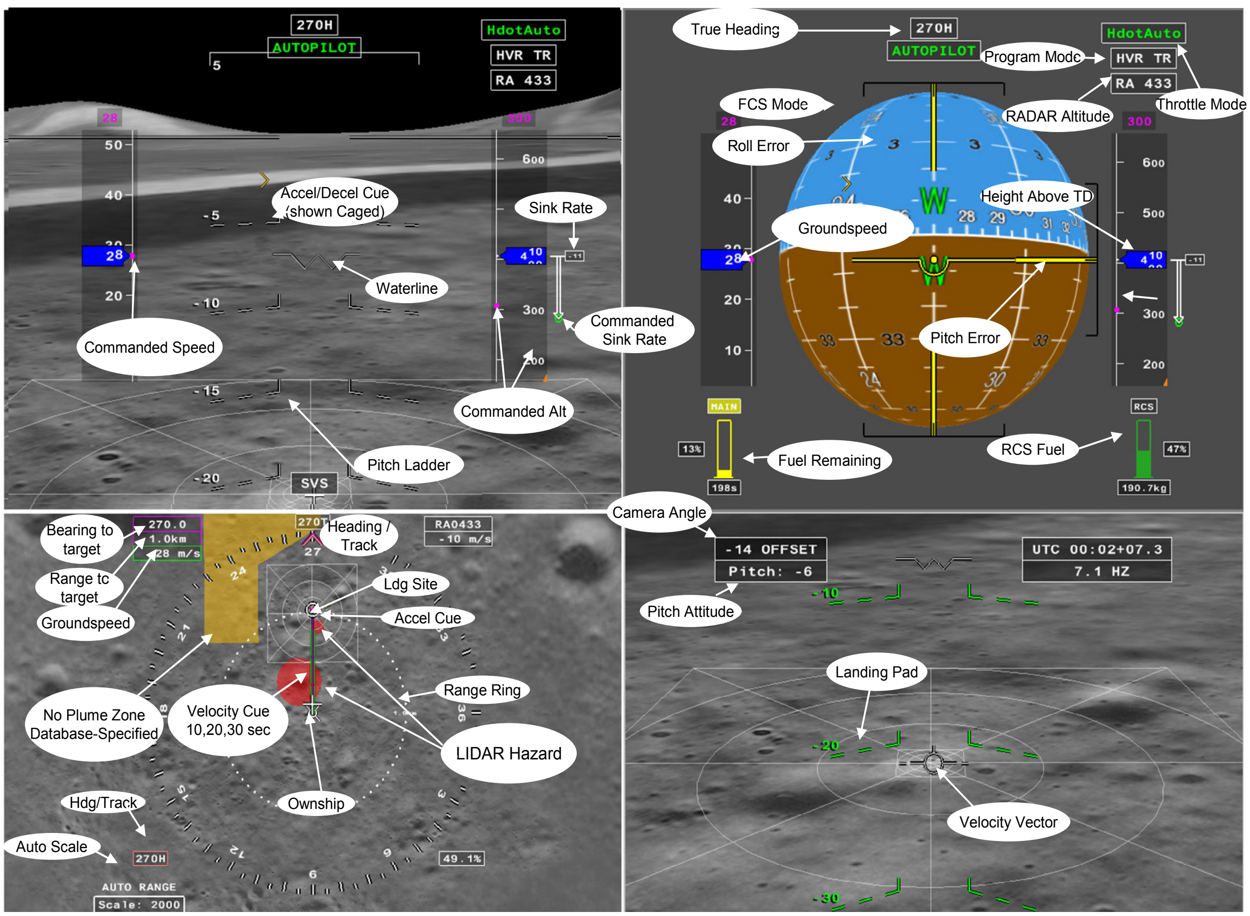


Figure 6. Select Symbology Elements for HUD (Upper Left), PFD(Upper Right), ND (Lower Left) and AD (Lower Right)

Navigation Display.

Initial navigation display (ND) concepts are also being developed using the "aviation-domain" as a point of departure in our investigations. In Figure 4, a simple two-dimensional, top down view with own-ship position located at the center of the display is shown. The synthetic lunar terrain is layered under the symbology. A touchdown zone is drawn centered on the designated touchdown site with a pilot-controllable range selection (500 meters is shown). The navigational display can be oriented track-up or heading-up as required by the pilot. A vertical situation display shows a companion vertical profile of the lunar terrain surface, vehicle trajectory, and longitudinal and vertical rates (shown in Figure 4). Conformal overlay of critical task information, such as inhabited "no-plume" areas, hazardous landing areas, and critical descent abort areas are indicated by various symbology elements.

Figure 6 (lower right-half quadrant) presents the lunar navigation display symbology which is modeled after modern aircraft displays with the exception of the velocity vector and guidance information, critical to the Lunar Landing task.. During the initial part of the approach, a zero horizontal velocity symbol (white circle) is used to indicate the point where the vehicle's horizontal velocity will reach zero, based on its current lateral and longitudinal accelerations. This symbology provides a rough, first-order awareness of and verification for the pitch and thrust guidance law (computed by the guidance, navigation, and control system) by use of conformal symbology. The display concept is the subject of current research to refine the symbology to enhance the utility of display for navigation and guidance precision to desired touchdown location during the lunar approach. The symbology is being modified to include tailored hover symbology to show ground track velocities and a acceleration/guidance cue, similar to successful helicopter and Vertical/Short Take-Off and Landing (V/STOL) vehicle programs (e.g., Schroeder and Merrick, 1992). Work is also progressing toward the use of exo-centric display concepts using SV/EV information as well as guidance data. The display concepts and symbology descriptions are sketched in Figure 6.

Research Evaluations

Aviation research associated with SV/EV has demonstrated that these intuitive displays enhance situation awareness and detection and avoidance of hazardous terrain in high workload situations. To begin building a research database for Lunar Landing, these hypotheses were tested. The landing phase was identified by Neil Armstrong as being the most difficult part of Apollo 11 and, therefore, research has focused on the descent stage to landing. The Apollo 15 landing site was chosen for study because of the interesting terrain features and the availability of higher resolution lunar terrain data, for this landing site.

Off-Nominal Testing

Eight participants were asked to fly 20 approaches to the Apollo 15 landing site with four display concepts (SV only; EV only (FLIR + LIDAR); baseline; SV + EV) which were factorially ordered and randomly presented to the pilots. The pilots were selected on the basis of having both fixed-wing commercial aircraft and V/STOL or helicopter flying experience which reflected the necessary piloting task skills. The scenarios required the pilots to either: (a) monitor the autopilot beginning at 170km from landing site to pitch-over maneuver (at 25km range) to clear mountainous terrain (Mt. Hadley); or, (b) monitor the autopilot from 15km to a hover transition point at 50m height-above-touchdown wherein the autopilot transitioned the lateral/horizontal position task to manual control (vertical sink rate was controlled by autopilot) for descent to the lunar surface. Each of these trials were flown with either: (a) an Apollo-like visibility condition which presented a larger out-the-window view (large field-of-view); or, (b) Altair-like visibility condition which presented a smaller field-of-view.

During the 20 experimental trials, six off-nominal situations were unknowingly presented to the pilots (3 off-nominal conditions X 2 visibility conditions): (a) guidance failure during landing, (b) guidance failure on initial approach, and (c) navigation failure during landing combined with either (a) low visibility condition or (b) high visibility condition. The guidance failures led the vehicle on a trajectory to a touchdown point on a hazardous terrain location. The navigation failure (a more difficult failure to detect) resulted from a poor navigation solution in which the vehicle calculated its position incorrectly. The failures were detectable by the pilot by either recognizing terrain cues (e.g., velocity vector tracking toward hazardous terrain), symbology dissociations, or heading/track errors (see Figure 8). Both synthetic vision and enhanced vision were correctly shown if present on the display condition for that trial. The navigation failure, however, resulted from erroneous vehicle data. Therefore, during this off-nominal condition, the synthetic vision was in error because it is a database-derived terrain representation based on the vehicle's erroneous navigation solution. The EV and EV+SV conditions, however, correctly showed the terrain because the EV (EV and LIDAR operative at 340m and 1000m, respectively in this experiment) was sensor-derived and independent of navigation solution

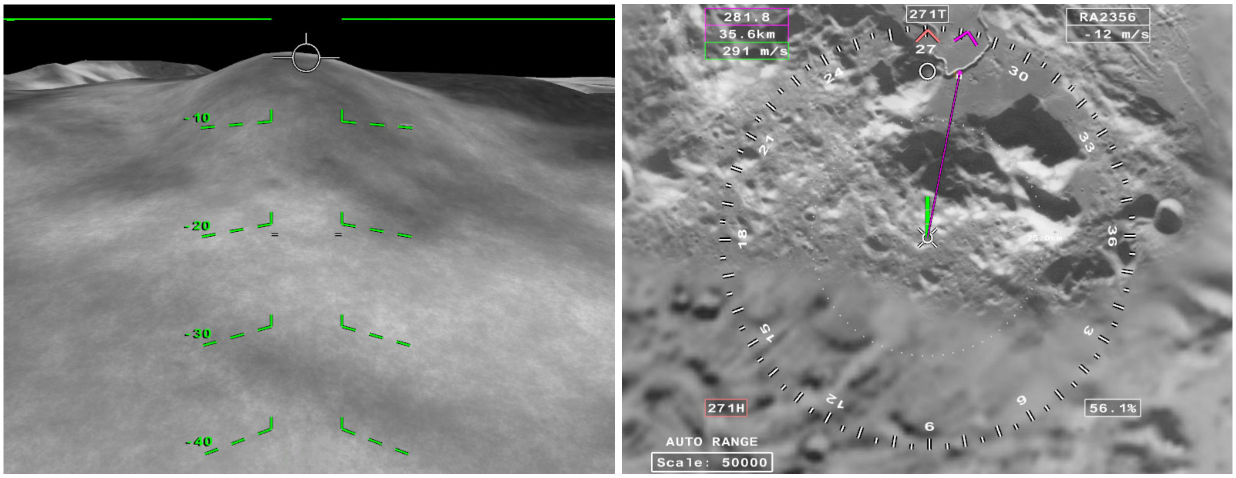


Figure 8. Example Guidance Failure on Approach 36 km from Landing Site (shown with SV)

The results evinced that the EV and EV+SV displays were significantly better for detection of navigation failures. All participants who experienced a navigation failure with either the SV-only or baseline display conditions either failed to recognize the failure or re-designated but landed on another hazard. None of the pilots with the EV or EV+SV display concepts failed to recognize the failures and were able to re-designate and land safely and accurately (M = 0.1m from designated landing touchdown point). Paired comparison (SA-SWORD; Vidulich & Hughes, 1991) and Situation Awareness Rating Technique (SART; Taylor, 1990) results demonstrated that pilots significantly preferred the SV+EV display concept for both situation awareness and mental workload compared to the other three display conditions, [$F(3,32) = 36.158$, $p < 0.001$ for SA-SWORD; $F(3, 48) = 3.262$, $p \leq .05$ for SART]. Pilots also reported better awareness of lunar surface and hazards with SV+EV display during navigation failure, ($F(3,16) = 8.080$, $p < .01$), and landing guidance failure, $F(3,16) = 5.140$, $p < .01$. For the off-nominal approach guidance failure, the baseline display concepts was rated significantly poorer for detection of hazards and awareness of lunar surface, $F(3, 16) = 6.78$, $p < .01$. These results support the requirement of SV and EV for Lunar Landing, matching our expectations based on aviation-domain research with these technologies. These data also emphasize the importance of having a real-time database or navigation integrity monitoring system, real-time sensors, or other suitable verification method to complement SV.

Future Directions

NASA research on SV and EV systems for space vehicles leverages on aeronautics research which has repeatedly demonstrated both safety and operational benefits of the technologies. Although there are many commonalities between aeronautics and space applications, there remain numerous areas of research and opportunities to further refine and optimize the concepts for the Altair Lunar Lander. Future areas of research include further off-nominal testing, improved symbology, accuracy assessment for lunar terrain databases, fusion of database and external sensors, mission rehearsal visualization, and unlimited field-of-regard HWD concepts.

References

Prinzel, L.J., & Kramer, L.J. (2006). Synthetic vision systems. In W. Karwowski (Ed.), *International Encyclopedia of Ergonomics and Human Factors* (pp. 25-1 – 25-10). Philadelphia, PA: Taylor & Francis

Pirnzel & Wickens (2009). Synthetic Vision Systems. In L. Prinzel & C. Wickens (Eds.), *International Journal of Aviation Psychology (Special Issue), 19(1-2)*.

Williams, S. P., Arthur III, J.J., Shelton, K. J., Norman, R. M. & Prinzel III, L. J. (2008). Synthetic vision for lunar and planetary landing vehicles. SPIE. Synthetic and Enhanced Vision 2008.

Schroeder, J.A. and Merrick, V. K.: Control and Display Combinations for Blind Vertical Landings, *Journal Of Guidance, Control, And Dynamic,s* Vol. 15, No. 3, May-June 1992.

Taylor, R.M. (1990). Situational awareness rating technique (SART): The development of a tool for aircraft system design. In *Situational awareness in aerospace operations* (AGARD-CP-478; pp. 3/1-3/17). Neuilly Sur Seine, France: NATO-AGARD.

Vidulich, M.A., & Hughes, E.R. (1991). Testing a subjective metric of situation awareness. In *Proceedings of the Human Factors & Ergonomics Society 35th Annual Meeting* (pp. 1307-1311). Santa Monica, CA: Human Factors & Ergonomics Society

Attitudes Toward Automation and Information Requirements of Experienced Predator Operators

William P. Marshak, John Hudson, Jeff Collier
SYTRONICS, Inc.
Michael C. Findler
SYTRONICS, Inc. and Wright State University

Two simulation studies were conducted using the Multiple Uninhabited Air Vehicle Agency (**MAGE**) software with nine Predator experienced pilots and sensor operators. In one study, MAGE's situation awareness and decision support elements were employed as supplemental information displays to a Predator crew station simulation within the Air Force's SIMAF facility. In the second simulation, each pilot controlled two autonomous UAVs during a complex multi-target mission scenario using the MAGE software suite. After the simulations were complete, the crewpersons answered questions about their attitudes concerning automated features and what kinds of non-vehicle information (weather, air spaces, intelligence, etc.) they felt they needed during operational missions. The results reflected their experience with the Predator system, to some degree their age, and the particulars of the MAGE user interface. Their responses provided valuable insights about the information requirements and evolution of the UAV Ground Control Station's user interface.

In response to operational needs observed in the Predator operations community, ***Air Force Research Laboratory (AFRL)*** identified two specific areas of deficiency and directed that research to address the needs. The first need was to enhance ***Uninhabited Air System (UAS)*** performance by integrating net-centric information from beyond the ***Uninhabited Air Vehicle (UAV)*** into its ***Ground Control System (GCS)***. Operations units had already added additional LCD displays to the Predator GCS to provide supplemental information into the GCS. The second operational need was to reduce Predator manning requirements by allowing a pilot to control more than one UAV at a time. MAGE was created to study how to accomplish these needs (Figure 1).

Figure 1. MAGE Hardware Configuration

"Operational" Test and Evaluation of MAGE as a Supplemental Display in SimAF

The installation and evaluation of MAGE as a supplemental display in the SimAF MQ-9 Simulator was performed for two reasons: (1) to evaluate technical integration issues and (2) to evaluate human system interface issues. Only the human system evaluation will be discussed here. Transitioned MAGE technology would most likely be as supplemental displays to the UAV manufacturer's GCS. The integration of MAGE into the SimAF MQ-9 Simulator would provide valuable insights into the integration strategies and problems associated with such integration. Since the opportunity to integrate into an operational GCS did not present itself, the integration into an existing high fidelity simulation was the next best thing.

The second reason for the integration had to deal with evaluating how operators would employ MAGE supplemental displays. The three ***Subject Matter Expert (SME)*** operators were given training in the MAGE displays and employed the MAGE displays during the SimAF simulation and a series of stand-alone target engagements during lulls in the SimAF simulation. Observers noted the operations and comments during the exercise to provide insights into both the MAGE capabilities and their use during mission simulation. One of the SME's pointed out there had not been any systematic requirements analysis for supplemental displays in the Predator GCS that he was aware of.

At the conclusion of the day, an exit interview was taken from each of the three crew persons concerning their experience with the MAGE displays in the Reaper crew station simulation. Question 1 simply asked whether supplemental information was of value to Predator crews. The answer was a unanimous yes among the SMEs. Questions 2 through 11 asked the three SMEs to rate the importance of different information sources and is shown in Figure 2.

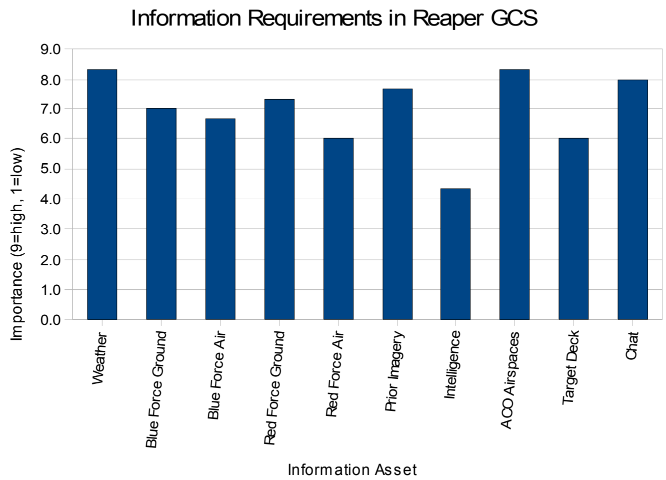


Figure 2. Average rated importance of information sources in Reaper/Predator operations (N=3)

The two highest rated categories (>8) were weather and Air Operations Center airspaces, which relate to the safe operation of the UAV. The next most important categories (7-8) included red force ground tracks, prior imagery, and chat. These are all related to prosecution of ground targets. The next highest categories (6-7) include other tracks (air and ground), and the overall target deck. Intelligence rated the lowest 4.3. This is due to the fact that target assignment is not the crews job; they only have veto over execution and then only if the rules of engagement or commander's intent would be violated by the strike.

All three SMEs agreed that there is a very real need for supplemental information in the GCS. Only one suggested an additional source of information, which was fuel states and range information when dealing with requests to retask the vehicle they were flying. This information would require coupling of the tracker fuel information and the FalconView-Mission Planner combination. When asked about database investment (Question 13), two of the three SMEs were aware of efforts to create mission and intelligence databases. Specifically called out were air tracks they shared the airspace with and imagery. When specifically asked about the TD, the SME's liked the augmentations of FalconView with qualifications. They thought the MAGE TD was an improvement over the current augmented FalconView displays, but thought it needed to be tuned more toward the Predator mission specific needs. There was interest in real-time ground track information to maintain situation awareness of friendly forces. There was concern expressed about clutter and clutter management. Clearly, fielding of the augmented TD will require revisiting operational units to further define the kinds of information displays to satisfy their needs.

Voice recognition was deemed useful, but our implementation was somewhat problematic. Even though interviews with one of the three SMEs were used to define the vocabulary, there was enough within user variation not to mention between user variations to create dissatisfaction. Clearly, the brief training opportunity was insufficient to familiarize the SMEs with the specific vocabulary. Broader vocabulary definition and greater recognition accuracy were both requested. A desire for dialog recognition was stated for use in chat operations since keyboard use demands skill and attention. The younger SME, the SO, was much more receptive to voice input and asked that coordinate definition function be added. Since all three operators were knowledgeable in FalconView's manual control, they often resorted to those manual controls instead of voice control.

Subjects were asked specifically about the effectiveness of adding additional special purpose supplemental displays, as is the case in the operational GCSs which various reports now place at six. There was consensus that the current configuration was getting the job done, but at some expense of operator workload. There seems to be a trade-off between cluttering fewer displays and increasing the display surface area. When asked about the effectiveness of the MAGE displays in reducing display count, the response was an endorsement, but not a resounding one. One SME found the information merge worked, but in a limited fashion. The second thought it was potentially useful, but needed to be better tuned to operational requirements. The last, and youngest, liked the ability to bring information together from disparate sources in one display.

SMEs seemed to like the enhanced chat function. They thought that it worked well, enhanced the ability to use the information within the different chat rooms, and liked the tagging of information and the ability to search by tags. Automated mission planning was considered useful, but in a limited fashion. There was little interest in it for normal mission execution; there simply wasn't much known beyond the next destination. The value was perceived in the timely and continuous replanning of the emergency mission.

Crews endorsed integration of MAGE into the GCS only if it were tailored and tuned to operational needs. In its current form they thought it would be useful, but not a "must have." However, they felt that careful tuning could turn the MAGE software into a valuable tool for Predator crews. They also thought MAGE function could benefit other levels of the "kill chain" above the GCS. One suggested the ability to share "screens" so the higher authority could more easily share information and make decisions more collaborative. It was pointed out that the Air Combat Command needs to see MAGE and for them to decide where such information is needed. MAGE was seen as a means of propagating a common view of the battlefield.

MAGE Stand-Alone Simulation Evaluation

An experimental, performance-based evaluation of the MAGE system was impossible without a baseline with which to compare. Self-baselined studies, using the same system with features turned on and off, are difficult to interpret. This is especially true for MAGE, in which monitoring, navigation, and mission execution are highly integrated. Thus, the approach taken was to do a subjective usability evaluation of MAGE using experienced Predator SMEs. Using these experienced operators, we were able to shed light on how well MAGE operated, which features were effective, and where future efforts would be best spent.

The study was conducted in a facility of USI, who employs and provided the Predator operator/subjects. The conducted survey asked the participants to critically evaluate each of the MAGE component technologies separately and each question had space for written comments. This took approximately a half hour to complete. Each subject took approximately 2.5 hours for training, simulation, and survey. There were four subjects on each of two days for a total of eight. All the participants were experienced pilot or sensor operators in the Predator system and are now civilians. Most of the participants had military Predator experience, though one was a civilian flight test engineer and no military experience. Questions alternated in their scale orientation, worded such that positive may tend toward both the positive and negative poles of the Likert question form. All questions were reoriented for analysis such that 1= most unfavorable, 4=neutral, and 7=most favorable; higher response means reflect more favorable disposition.

Tactical Display: Every UAV GCS we have seen at site visits, trade shows, or professional meetings has some form of spatially-oriented Tactical Display. The MAGE system is no exception and employs an augmented version of the Georgia Tech Research Institute FalconView mission planner as the basis for its Tactical Display (Figure 3). The FalconView map database serves as the backdrop for display of the UAV position, the position of other air traffic, friendly and enemy ground unit locations, weather, and restricted air spaces. The Tactical Display also displays UAV imaging tasks, weapons targets, and hosts the mission planning displays of current and proposed route alternatives.

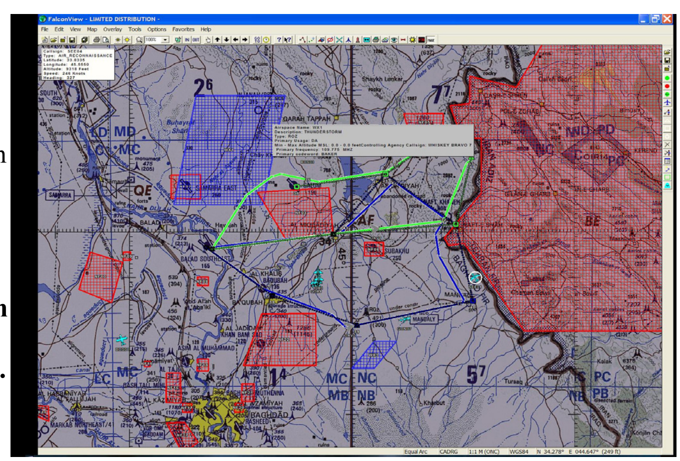


Figure 3: Tactical Display used to display UAV position, planned routes, weather, and airspaces in a spatial-map context. Information augmentation is seen in rollover and hooked textual information.

Operators were generally favorable impression of the MAGE Tactical Display. The overall assessment of the display was very effective (M_{Q11} = 5.9, SD_{Q11} = 1.6). Comments asked for a larger display to deal with clutter, better contrast control to separate symbology from the map underlay, icon management to control clutter, and a "restore" function to return to earlier configurations.

Next a series of features and representations were specifically called out for evaluation. Rollovers, which are details triggered by mouse cursor proximity, was the highest rated feature ($M_{rollovers}$ = 6.8, $SD_{rollovers}$ = 0.5). Data presentation as represented by flags, constantly visible tags that follow an icon, and hooks (where the flag is displayed to one side and the icon is marked with a halo). These were still seen as favorable (M_{flag} = 5.6, SD_{flag} =1.9; M_{hook} =, SD_{hook}), it is somewhat less enthusiastic than for the rollovers. The features assessed included airspaces, weather, blue force tracking, and other air traffic. Respectively, their means were $M_{airspace}$= 5.9, $M_{weather}$ = 5.3, $M_{blueforce}$ = 6.1, and $M_{air\ traffic}$ = 6.1 with SDs ranging from 0.6-2.1. Again, we see the dislike for automated mission planning manifest itself. One suggestion was the ability to display registered imagery from weather satellites under the symbology. Generally, the SMEs liked the rollover option and were favorable to hooks if they provided more flexibility in text placement. A preference was expressed for symbology that at least reflected types of aircraft.

Voice Recognition System: Performance of the voice recognition system was somewhat lower than had been achieved with earlier uses. The reason may have been because of the headset hardware, or because of the relative inexperience of the test users. Overall the participants were favorable to the voice recognition system. The system was graded as nearly moderately accurate (M_{Q26} = 5.8, SD_{Q26} = 0.5), better than moderately effective in context tracking (M_{Q27} = 6.3, SD_{Q27} = 0.7), and screen control was nearly moderately effective (M_{Q28} = 5.9, SD_{Q28} = 0.9). Most of the operators were familiar with FalconView's manual screen controls and preferred to use them in lieu of the voice control.

Earlier, ***SYTRONICS*** used one of the SMEs to define the MAGE fixed vocabulary. The lowest rating in voice control was in vocabulary quality, just barely appropriate (M_{Q29} = 5.3, SD_{Q29} = 2.0). This clearly indicates that with fixed vocabulary systems, enough alternative phrasing must be defined to capture more users' normal spoken language. The voice system is capable of alternative definitions, but this is time consuming and the research nature of the system did not justify the time investment. An operational system would require wider vocabulary and range of phraseology. Whatever the SME's concern about the vocabulary, they uniformly were enthusiastic about employing voice technology in the GCS (M_{Q30} = 6.3, SD_{Q30} = 0.9) and most strongly felt it could aid the Predator system (M_{Q31} = 6.8, SD_{Q31} = 0.5). This is probably in response to the manually intensive function in the Predator GCS, with operators looking to technology to reduce the manual typing workload.

Timeline Display: Timeline Displays are uncommon in operational GCSs, but a popular topic in UAV advanced development or research stations Cummings & Mitchell (2007) (Figure 4). The MAGE system is a simple temporal representation of planned mission events displayed on a linear timeline. The Timeline Display allows operators to detect co-temporal demands from the multiple vehicles. When mission replanning is requested due to task management, the alternative timelines are displayed beneath the current mission timelines so they may be compared with the current mission and other vehicles before approval of the alternative plans.

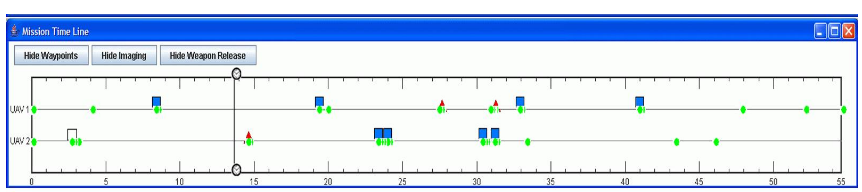


Figure 4. Timeline Display for Two UAVs with Waypoints (Green Dots), Reconnaissance Tasks (Blue Boxes), and Weapons Releases (Red Triangles)

A key issue to understanding the SME reaction to the Timeline Display is that the display assumes execution of a complete preplanned mission. This is an alien concept to Predator/Reaper crew persons, who usually operate with just a heading to maintain or do not know where or when the next object of surveillance will be directed. The Timeline Display's depiction of mission events was the lowest of any of the technologies assessed (M_{Q21} = 5.4, SD_{Q21} = 1.5). Timeline's greatest value was seen for planning an entire mission (M_{Q24} = 6.0, SD_{Q24} = 1.7). All but one SME felt it moderately or most effective; while a lone dissenter found it moderately useless. Moderately favorable averages were observed for the mission alternative comparison (M_{Q22} = 5.5, SD_{Q22} = 1.1), and value to the Predator mission ($M_{Q25} = 5.5$, $SD_{Q25} = 1.3$). The least favorable response concerned the scaling and scrolling controls of the Timeline ($M_{Q23} = 4.8$, $SD_{Q23} = 1.9$).

Chat System: The MAGE research team observed Predator crews and learned of their heavy reliance on IRC for communicating with mission essential elements around the world. However, chat continues to be used because it does not require immediate attention, it leaves a written record of communication, and reduces the chance of misunderstanding. The ***Extended Instant Messaging (xIM)*** client was developed for MAGE that integrates automatic extraction of information from the chat content, entering information into a database for later search, extraction and plotting of coordinates in a map, and highlighting high interest participants Collier, Hudson & Marshak (2007). Queries can be made through the voice recognition interface, and information can be moved from chat to other functions via an implementation of a "clipboard " shown in Figure 5.

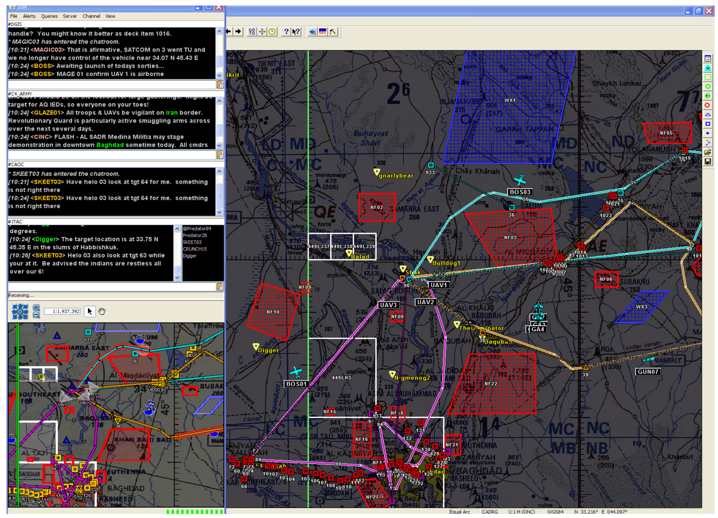


Figure 5. Extended Instant Messaging (xIM) at Left of Tactical Display with Four Chat Room Windows and Geospatial Content Display (lower left)

Again, our subject matter experts (SMEs) were largely favorable toward the chat enhanced MAGE. They found the features overall moderately useful ($M_{Q32} = 6.3$, $SD_{Q32} = 1.0$). The geo-location features of the chat system were not exploited by the test scenario, which is probable cause for the low effectiveness rating of that feature ($M_{Q33} = 4.9$, $SD_{Q33} = 1.6$). As it turns out, there was no mission demands to correlate chat content with the map, so the SMEs found the feature somewhat superfluous to the scenario. There was strong utility in highlighting of chat room call signs ($M_{Q34} = 6.4$, $SD_{Q34} = 0.7$) and making the content available to other system components via the clipboard ($M_{Q35} = 6.5$, $SD_{Q35} = 1.1$). The SMEs saw moderate utility for enhanced chat to UAS crews ($M_{Q36} = 5.9$, $SD_{Q36} = 1.2$). SME comments on xIM were entirely positive. It was judged to be a boon for situation awareness, seen as a way to recover history information, and was judged to be a time saver.

Decision Support Interface: The DSI employed Cognitive Engineering in its design and had many advanced technology components. Its design was based on a theory from cognitive psychology. Klein (1989) proposed that experts employ Recognition Primed Decision Making when faced with a decision. Experts look at the relevant information and based on their experience, "recognize" the best decision or course-of-action. The ***Decision Support System (DSS)*** uses the MAGE intelligent agent architecture to collect relevant information from Web centric sources and assemble them in a windowed workspace. One example of a DSI template is shown in Figure 6.

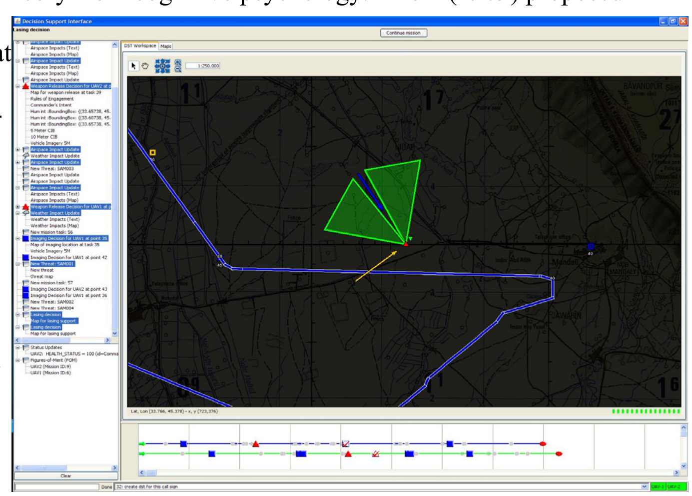


Figure 6. An Advanced DSI Display Depicting a Weapons Release Engagement with Laser Funnels (Green), Sun Direction (Yellow Arrow), Target (Red Triangle), Safe Weapons Dump Point (Green Triangle), and Approach Heading (Blue Arrow)

Information overflow and other pending decisions are displayed textually in a tree structure to the workspace's left. Additionally, users can query the system for additional information not presented by the automation. At the top of the display, the instigating event requiring intervention and the decision alternatives are presented. The decision is left to the human, but accelerated by the collection, representation, and decision expression process.

One question probed the usefulness of event triggered decision support. This fundamental principle of the DSI is that software agents would sense the mission conditions requiring operator intervention and trigger the assembly and presentation of a template. Our SMEs found this moderately useful (M_{Q37} = 6.4, SD_{Q37} = 0.9). Next, we began to evaluate the format design features individually. Identifying the instigation or trigger event at the top of the format was judged moderately useful as well (M_{Q38} = 6.1, SD_{Q38} = 1.1), eliminating any doubt about the antecedent conditions. On the left side of the format is the presentation of the question cue and data sources in tree format. Our SMEs found this only barely to moderately useful (M_{Q39} = 5.4, SD_{Q39} = 1.7). The workspace area faired slightly better, closer to moderately useful (M_{Q40} = 5.8, SD_{Q40} = 0.9). The DSI information window manager was judged to be moderately effective (M_{Q41} = 5.9, SD_{Q41} = 0.3). The SMEs did judge the query system to be moderately effective. This feature allowed operators to further information from other sources (M_{Q42} = 6.0, SD_{Q42} =0.0). Graphical representations like map mash-ups were moderately useful (M_{Q43} = 5.7, SD_{Q43} = 2.1) while text representations were rated slightly higher (M_{Q44} = 6.3, SD_{Q44} = 0.8).

Another of the new technologies evaluated was the "Sprocket" display format. This format supported choosing between mission alternatives based on a Visual Thinking design (McKim, 1972). Due to an error in the morning data collection, the format was not presented to the first four subjects and only the last three or four subjects responded in the afternoon. The three responding subjects were markedly different; two thought highly of the format rating it "highly effective" and one rated it "moderately ineffective." Individual differences are large with regard to the Sprocket format (M_{Q45} = 5.3, SD_{Q45} = 2.9). SMEs were asked to evaluate the effectiveness of the decision support format and they graded it "barely-to-moderately" effective (M_{Q46} = 5.4, SD_{Q46} = 2.0). Everyone but one SME gave it a 6-7; with one SME giving it a 1 (not at all useful). However, when asked whether the functionality could aid the Predator GCS they graded it closer to moderately useful (M_{Q47} = 5.9, SD_{Q47} = 1.0).

The response to the DSI was largely positive, though there were some suggestions to improve the implementation. Screen manipulation could be made easier by interacting with the whole data window instead of the edges as with the current implementation. Imagery should be put in track-up orientation and be centered on the target. SMEs liked the ability to mix text, graphics, and imagery in the DSS workspace. They praised the concept, indicating it would facilitate management of multiple UAVs that it brought important and high relevance information to the forefront and with the Timeline Display allowed precise management of workload.

Conclusions

Although these findings are favorable to the MAGE approach to net centric information integration, they highlight how user experience shapes acceptance of new technology. MAGE was built on the assumption of mostly autonomous UAVs, which is in contrast to the mostly manually flown Predator system. Older SME's seem biased against increasing automation. Web mash-ups like Google Maps seem to be paving the way for greater information integration in displays. Perhaps there will be greater acceptance of cognitive systems engineering of information requirements and integrated display integration.

References

Collier, J. Hudson, J. and Marshak, W. (2007). Integrating instant messaging as a net-centric information source. AUVSI's Unmanned Systems North America, August 6-9, Washington, DC.

Cummings, M.L., & Mitchell, P.J. (2007). Operator Scheduling Strategies in Supervisory Control of Multiple UAVs. *Aerospace Science and Technology,* Vol. 11, p. 339–348.

Klein, G.A., (1989) Recognition-Primed Decisions in *Advances in Man-Machine Systems Research*, W.B. Rouse, Editor., Greenwich CT: JAI Press.

McKim, R.H. (1972). Experiences in Visual Thinking. Monterey CA: Brooks/Cole.

DETERMINANTS OF CONFLICT RISK JUDGMENTS IN AIR TRAFFIC CONTROL

Stéphanie Stankovic
The University of Queensland
Brisbane, Australia
Esa Rantanen
Rochester Institute of Technology
Rochester, NY, USA
Nicolas Ponomarenko
French Civil Aviation Authority
Toulouse, France
Shayne Loft
The University of Western Australia
Perth, Australia

The aim of the study was to identify the determinants of conflict risk judgments in air traffic control. Fourteen expert controllers made conflict risk judgments about air traffic situations in which three variables (conflict geometry, time of closest point of approach or TCPA, and vertical separation between aircraft) were manipulated. The results indicate that conflict geometry, TCPA, and the vertical separation between aircraft significantly influenced judgments of conflict risk. In addition, there was a significant interaction between these three variables. Risk perception was largest under conditions where aircraft were on the same headings, with short TCPA, and no minimum vertical separation. The study was successful in determining what factors of air traffic scenarios increased subjective risk judgments in air traffic control.

Understanding how experts make risk judgments in safety critical work contexts is a major issue in cognitive ergonomics. One prototypical example of an applied work context where individuals are required to make risk judgments under conditions of uncertainty and time pressure is in air traffic control (ATC; Loft, Sanderson, Neal, & Mooij, 2007). Air traffic controllers make judgments about the future relative positions of aircraft in order to assess whether they will lose minimum separation. In approach control, aircraft are in conflict if they are projected to violate both lateral (3 nautical miles) and vertical (1,000 feet) separation simultaneously.

Several theoretical accounts of how controllers make conflict judgments have recently been published (Loft, Bolland, Humphreys & Neal, under review; Rantanen and Nunes, 2005; Stankovic, Raufaste, & Averty, 2008). The Stankovic et al. (2008) model uses three horizontal distance metrics to predict controllers' judgments of conflict risk for pairs of converging aircraft; the distance between the crossing point of the aircraft pair trajectories and the closest aircraft to that point (Dt_0), the distance between the two aircraft when they are horizontally closest (Dt_h), and the horizontal distance between the two aircraft when their growing vertical distance reaches 1,000 feet (Dt_v). The Stankovic et al. model could account for up to 50% of the variance in expert controller risk judgments. However, a significant limitation of the Stankovic et al. study was that it presented aircraft pairs with limited sets of geometry features.

In addition to aircraft converging on common intersection points (crossing headings), pairs of aircraft in ATC often follow each other on the same flight path (same headings), or travel toward each other on the same flight path (opposite headings). In the current study we examine this more representative set of conflict pairs. In addition, we present aircraft pairs that are descending through the altitudes of each other. Judgment of conflict status when aircraft are converging both vertically and horizontally can be challenging due to the difficulty of trajectory estimation in the vertical plane (Boag, Neal, loft & Halford, 2006). The changing difference in altitudes between two aircraft is not directly visually perceptible, but has to be deduced from the numerical altitude readouts in aircraft data blocks, and the estimation of vertical separation at some point in the future is a result of combining these altitude calculations with estimation of groundspeed and future lateral separation. In the current study we examined the relationship between vertical separation and conflict risk judgment by varying the minimum vertical separation of aircraft whilst keeping the minimum distance of lateral separation constant. Finally, we examined the effect of time to closest point of approach (TCPA). We expressed TCPA as the time in minutes for the aircraft to reach minimum horizontal separation. In summary, the study reported here examined the effect of conflict geometry, TCPA and vertical separation on conflict risk judgments.

Predictions

The purpose of the present study was to identifying the factors that determine conflict risk judgments. Our first hypothesis concerned the effect of conflict geometry. For all three conflict geometries (same headings, opposite headings and crossing headings) controllers need to extrapolate the minimum horizontal and vertical separation between aircraft. A simple heuristic used to achieve these separation assessments may consist of mentally moving the two aircraft velocity vectors along the horizontal plane and inferring the time when they will come closest on that plane. Then, the aircraft positions at this point (which can be roughly associated with the crossing point between the aircraft trajectories) have to be maintained mentally in order to estimate the horizontal and vertical separation between them (Stankovic et al., 2008). Conflict geometries should influence the ability to apply such heuristic. In the case of same headings, minimum horizontal and vertical separation extrapolations are particularly difficult. The crossing point between the two aircraft is difficult to estimate in this situation especially because of the lack of perceptual cues of aircraft positions at each side from the crossing point. In the crossing or opposite heading scenarios, controllers' horizontal and vertical separation predictions are facilitated by the presence of these perceptual cues. Thus, controllers may be more uncertain about the future relative positions of same heading aircraft. Increased controllers uncertainty of the future relative positions of aircraft is associated with increased probability that controllers label situations conflicts (Loft et al., under review). We expected to obtain higher conflict risk judgments in same heading situations than in opposite and crossing heading conditions.

In the model proposed by Stankovic et al. (2008), two horizontal distances, minimum horizontal separation (Dt_h) and the horizontal distance between the two aircraft when their growing vertical distance reaches 1,000 feet (Dt_v), both significantly predicted conflict risk judgments. It has been demonstrated by numerous studies that controllers are more likely to consider aircraft to be in conflict as minimum horizontal separation decreases (e.g., Stankovic et al., 2008). Thus, we fixed minimum horizontal separation at the conventional threshold of 3 nautical miles used in approach control. By the fixing minimum horizontal separation we were able to more precisely examine the influence of minimum vertical distance. We defined Dt_v as the minimum vertical distance between aircraft when lateral separation was 3 nm. Loft et al. (under review) found experts always interned to vertical problems when lateral separation was constant at 0 nm, even when minimum vertical separation was up to 4,000ft. This is not consistent with the finding of Stankovic et al. (2008) where Dt_v significantly predicted variation in controller risk judgments. However, it is noteworthy that Loft et al. (under review) used a dichotomous rating scale (intervene vs. not intervene). In comparison, Stankovic et al. (2008) investigated risk judgments about conflict by using an 8-point scale. This rating scale should be more sensitive to detecting differences in controllers' perceptions of conflict status with changes in minimum vertical separation. We expected to find differences in risk judgments as function of vertical distance, controllers judging situations as riskier as vertical separation decrease.

We coded Dt_0 in time (as advised by a subject matter expert) to express our TCPA variable. We expected to replicate the results obtained by Stankovic et al. (2008), with conflict risk judgments increasing as time to crossing point increased.

Method

Participants

Fourteen air traffic controllers (12 men and 2 women) from the Toulouse-Blagnac airport volunteered to take part in the experiment. Their ages ranged from 25 to 59 years ($M = 43.29$, $SD = 10.94$). Their average length of experience as an air traffic controller ranged from 2 to 37 years ($M = 20.36$, $SD = 10.13$). Their experience length since sector certification ranged from 0 to 34 years ($M = 10$, $SD = 9.96$).

Variables

The three independent variables were manipulated across 36 static scenarios. These variables are the time to point of closet approach (TCPA), the geometry of conflict, and the vertical separation between aircraft at the moment of the crossing point (VS). Values of TCPA correspond to the time given in minutes that the aircraft would take to reach the crossing point and took the values of 3 or 6 minutes. Geometry of conflict corresponds to the relative headings of the two aircraft. Three conflict geometries were manipulated: (1) situations were a faster aircraft followed a slower aircraft (same headings), (2) aircraft heading directly toward each other on the same flight path (opposite headings), and (3) aircraft converging at 90-degree angle on a common intersection point (crossing headings). The vertical separation variable corresponds to the minimum vertical separation between two aircraft computed at the moment when the aircraft reached their minimum lateral separation of 3 nm. These values of vertical separation were 0, 2,000 or 4,000 feet. The horizontal distance computed at the moment of the crossing point was fixed at 3 nm which corresponds to the conventional minimal separation used in French approach control. Two versions of each of the 18 scenarios were presented. The scenarios were chosen from historical flight data observed from the Toulouse-Blagnac airport.

Each participant was presented the 36 experimental trials in a random order. The main dependent variable was the conflict risk judgment, provided on a 12 points scale from no risk at the far left (1) to extreme risk at the far right (12).

Materials and Procedure

The experiment lasted about 25 minutes. Each experimental situation was displayed on a white sheet of paper. In each situation a pair of aircraft converged toward the same point: one was cruising in altitude and the other one was descending. In all cases the aircraft pair reached their minimum horizontal (3nm) and vertical (0, 2000, or 4000ft) distances of separation in 3 or 6 minutes. A 3 nm scale marker was presented on this display, and the rate of descent (V_z) was set at 1,000 feet per minute. Each aircraft had a data block that displayed its speed in knots, its current flight level (altitude in hundreds of feet), a sign "=" for the cruising aircraft or a down arrow followed by a cleared level for the descending aircraft. In each situation, the aircraft was descending through the level of the cruising aircraft. Two-minute velocity vectors were displayed for the aircraft. Participants were instructed to judge the risk of conflict for each pair of aircraft. Three other judgments were also requested relating to strategies used to ensure separation between two aircraft, but these results are beyond the scope of this paper.

Results and Discussion

A conflict geometry × TCPA × vertical separation (3 × 2 × 3) within-subjects ANOVA was conducted, with conflict risk judgments as the dependent variable. The values for small, medium and large effect sizes are .10, .25 and .40 respectively (Cohen, 1988). Tukey post hoc comparisons were conducted to follow up significant omnibus effects.

Descriptive statistics are given in Table 1. As predicted, geometry of conflict had a significant effect on conflict risk judgment, $F(2, 13) = 5.18$, $p = .013$, $\eta_p^2 = 0.28$. Post hoc tests showed that controllers indicated higher conflict risk assessments for same heading air traffic scenarios ($M = 8.17$, *SE*

= 0.47) than either opposite heading (*M* = 6.95, *SE* = 0.65) (*p* = .04) or crossing heading (*M* = 6.77, *SE* = 0.66) (*p* = .018) scenarios. There was no significant difference between the opposite heading and crossing heading conditions.

Vertical separation had also a significant effect on conflict risk judgment, $F(2, 13) = 14{,}33$, $p < .001$, $\eta_p^2 = 0.52$. Post hoc tests showed that controllers indicated higher conflict risk assessments when the vertical separation between aircraft was 0 feet (*M* = 9.14, *SE* = 0.36) compared to when it was 2,000 feet (*M* = 6.96, *SE* = 0.72) (*p* = .006) or 4,000 feet (*M* = 5.79, *SE* = 0.77) ($p < .001$). There was no significant difference between the 2,000 feet and 4,000 feet minimum vertical separation conditions. Thus, in contrast to the findings of Loft et al. (under review), controllers conflict status judgments were sensitive to the minimum vertical separation between aircraft. Controllers were making calculations regarding the future vertical separation between aircraft when making conflict decisions.

Finally, TCPA had a significant effect on risk judgments, $F(1, 13) = 7.90$, $p = .015$, $\eta_p^2 = 0.38$. The situation was judged more risky when aircraft were at 3 minutes from the crossing point (*M* = 7.66, *SE* = 0.51) compared to when they were at 6 minutes from this point (*M* = 6.93, *SE* = 0.58) (*p* = .015). In contrast to the findings of Stankovic et al. we found that conflict risk judgments increasing as time to crossing point decreased. This result is more compatible with the interpretation that situations in which TCPA has high values simply offer more time before a clear decision about conflict needs to be made.

Table 1. *Means of judgments of conflict risk as a function of conflict geometry, TCPA, and minimum vertical separation.*

Geometry	TCPA	Vertical Separation	Means	SE
Same Headings	3 min	0 ft	10.00	0.46
		2000 ft	9.43	0.46
		4000 ft	8.54	0.70
	6 min	0 ft	8.57	0.52
		2000 ft	6.46	0.78
		4000 ft	6.04	0.92
Opposite Headings	3 min	0 ft	9.46	0.61
		2000 ft	6.57	1.02
		4000 ft	4.68	0.91
	6 min	0 ft	9.04	0.56
		2000 ft	7.04	0.97
		4000 ft	4.89	1.03
Crossing Headings	3 min	0 ft	9.54	0.57
		2000 ft	5.71	1.02
		4000 ft	5.04	1.04
	6 min	0 ft	8.21	0.86
		2000 ft	6.57	0.91
		4000 ft	5.57	0.87

The main effects were qualified by a 3-way interaction between conflict geometry, TCPA and vertical separation variables on conflict risk judgments $F(2, 13) = 7.73$, $p =. 038$, $\eta_p^2 = 0.17$ (Figure 1). As indicated in Figure 1, TCPA only had an effect on conflict risk judgments for same heading aircraft geometries, and not for opposite or crossing heading geometries. The increase in risk judgments with increased vertical separation was reasonably consistent across the three conflict geometry types when TCPA was 6min. In comparison, the pattern of increase in risk judgments with increased vertical separation when TCPA was 3 min was less for same headings geometries than opposite or crossing heading geometries. Risk judgments for TCPA 3 min problems were higher for same heading geometries than the other geometries when vertical separation was 2,000 ft or 4,000 ft (at 0 ft they were similar).

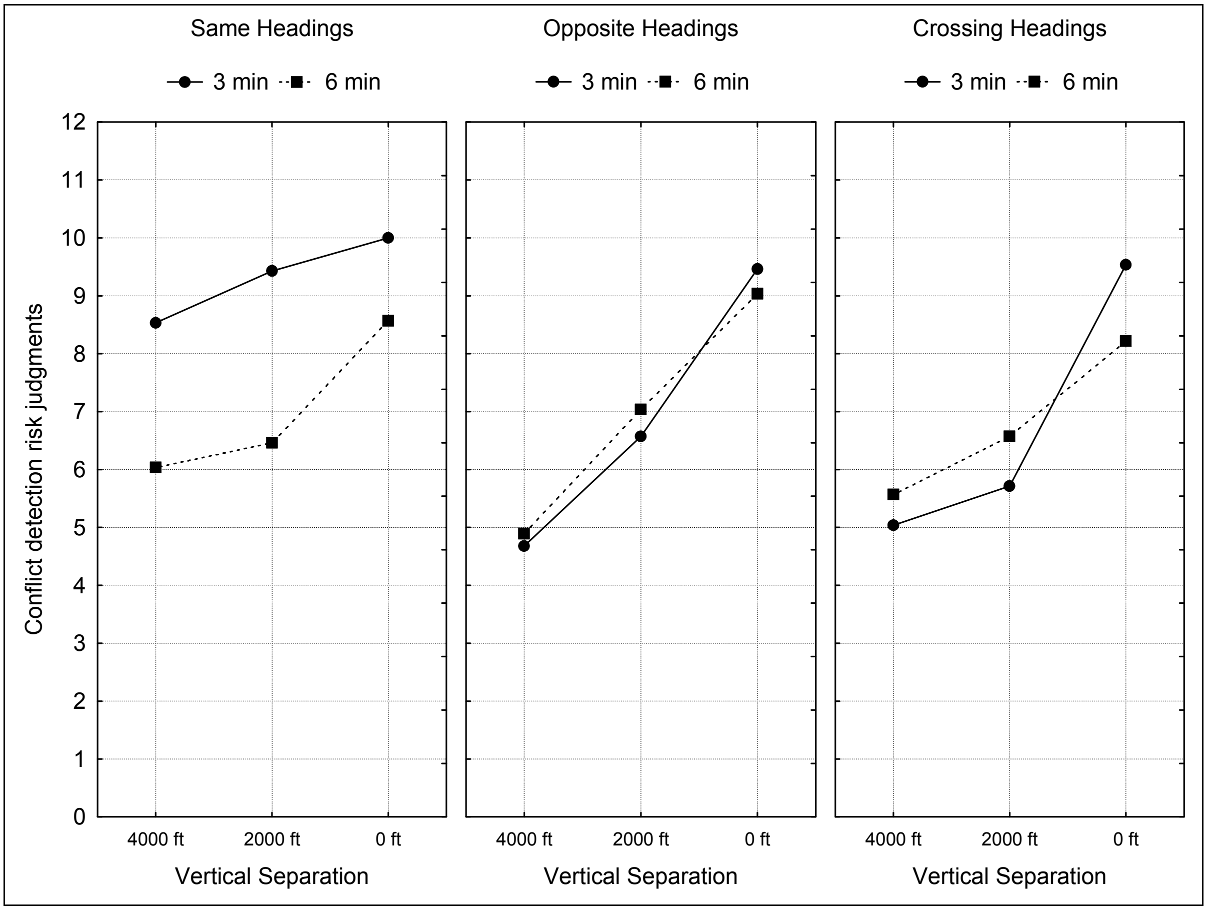


Figure 1. Conflict detection judgments as function of conflict geometry, TCPA and vertical separation.

Conclusion

The study was to examine the effect of three factors on judgments of conflict risk; conflict geometry, minimum vertical separation and time of closest point of approach (TCPA). Controllers judged air traffic situations more risky with decreased vertical separation, when aircraft was traveling on the same heading as other aircraft, and when TCPA was short. To our knowledge, this is the first demonstration in the literature that the respective headings of aircraft can influence controller's perceptions of conflict risk. The paper extends the work of Loft (under review) by demonstrating that vertical separation can indeed affect conflict status judgments. The effect of TCPA is compatible with the fact that controllers are sensitive to the time available to perform various control tasks (Loft et al., 2007; Wickens, Mavor, & McGee, 1997; Payne, Bettman, & Johnson, 1993).

The study is limited by the fact that the air traffic situations presented to controllers were static. At the same time, we see no reason why these patterns of effects would not be replicated using a dynamic simulation (e.g., 5 sec update rate) (Boag et al., 2006). In addition, the air traffic scenarios themselves had sound validity to the extent that they were selected by subject matter experts at Toulouse-Blagnac airport. In conclusion, the study was successful in determining what factors of air traffic scenarios may increase subjective judgments of conflict risk made by air traffic controllers.

References

Boag, C., Neal, A., Loft, S., & Halford, G. (2006). An analysis of relational complexity in an air traffic control conflict detection task. *Ergonomics, 14*, 1508-1526.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (under review). A theory and model of conflict detection in air traffic control: Incorporating environmental constraints. Manuscript under review at *Journal of Experimental Psychology: Applied.*

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors, 49,* 376-399.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision Maker*. Cambridge: Cambridge University Press.

Rantanen, E. M., & Nunes, A. (2005). Hierarchical conflict detection in air traffic control. *International Journal of Aviation Psychology,15*, 339–362.

Stankovic, S., Raufaste, E., & Averty, P. (2008). Determinants of Conflict Detection: A Model of Risk Judgments in Air Traffic Control. *Human Factors, 50*, 121-134.

StatSoft (2005). Statistica (Version 7.1) [Computer software]. Maisons-Alfort, France: Author. (www.statsoft.fr)

Wickens, C. D., Mavor, A. S., & McGee, J. P. (1997). *Flight to the Future: Human Factors in Air Traffic Control*. Washington, D.C.: National Academy Press.

EFFECTS OF TIME PRESSURE ON THE USE OF AN AUTOMATED DECISION SUPPORT SYSTEM FOR STRIKE PLANNING

Yves Boussemart, Birsen Donmez, M. L. Cummings, Jonathan Las Fargeas
Humans and Automation Laboratory
Massachusetts Institute of Technology, Cambridge MA

This paper describes the results of an experiment designed to examine the effects of time pressure on behavioral patterns. The main research hypothesis is that people under time pressure tend to increasingly rely on automation in order to cope with the added workload. The context is that of a missile strike planner having to create a set of matches between resources (missiles) and requirements (missions). We introduce time pressure by changing the temporal requirements towards the end of the mission. Overall performance, calls to automation and qualitative strategies are recorded and analyzed using ANOVA and other non-parametric tests. The main finding of this study is that while the number of calls to the automation did significantly increase under time pressure, there did not seem to be a statistically significant shift in problem solving strategies under time pressure. The experimental results show the importance of good automation-human interface design so as to gain maximum benefit from the use of an automated decision support systems.

Introduction

Missile strike planning is a complex example of multivariate optimization, where a set of resources must be paired with a set of goals in a manner that meets constraints and achieves a certain level of quality. In terms of strike planning for Tomahawk Land Attack Missile (TLAM), our representative domain, the missiles and associated missions are characterized by a series of variables. A strike planner's task consists of making sure that no hard constraint on these variables is violated when a specific missile is assigned to a specific mission. In addition, the final solution, that is the set of mission/missile assignments, should optimize soft constraints: it should be as "good" as possible along potentially subjective or dynamic references that may change in the planning process. In addition to such constraints, strike planners usually have to operate under temporal pressure and have a limited amount of time to finalize a strike plan.

Payne defines time pressure as "changing the time available to make a decision" [1]. The effects of time pressure on decision making have been described in the literature extensively and so have the resulting operator coping processes used. The most frequently cited coping processes for dealing with temporal stress are acceleration, filtering and omission [2]. Acceleration is probably the most obvious effect of time pressure and denotes an increased information processing rate. It has been shown, however, that with an increasingly stringent deadline subjects were less likely to rely uniquely on acceleration [3]. Filtering refers to processing some parts of the information more than others; the research has consistently shown that the attributes seen as less important tend to be filtered out first [4, 5]. Omission, also referred to as "shallower search for information" [6], implies ignoring particular parts of the information. In contrast to these coping processes, research has also shown that a common cognitive strategy shift is a tendency to lock into one problem solving strategy under time pressure even if it is suboptimal, a process also known as regression to learnt behaviors [7].

Previous work [8, 9] investigated the creation of decision-support tools aimed at leveraging human-automation collaboration to enhance the quality of the strike mission planning process. The current experiment builds on this previous work by adding temporal constraints to mission planning in order to examine the effects of time pressure on the use of automation during the strike planning process. The main research hypothesis we address is that people under temporal stress will rely more on automated tools in order to cope with the added workload. Time pressure is central to the context of Command and Control (C2) since theaters of operations are inherently dynamic; the conjunction of changes and fixed deadlines tend to put operators under considerable stress due to the time-critical nature and the importance of the decisions they have to make.

Method

Apparatus: StrikeView

StrikeView (Figure 1) is an interface designed to facilitate the process of planning strikes by decreasing the overall workload and improving the quality of the strike [8]. This interface allows the operator to solve the problem, i.e., build a set of mission/missile assignments, either manually or with the help of the computer.

The matching task consists of pairing a set of pre-planned missions with missiles available on different ships. This constitutes a complex, multivariate resource allocation problem, where a human operator must not only satisfy a set of matching constraints, usually part of the rules of engagement (ROE), but also optimize the mission-missile assignments to minimize operational costs or enhance the quality of the overall plan. Generally it is left to the strike planner to manually assign missiles to missions, taking into account the different mission and missiles characteristics, as well as the constraints *du jour* included in the ROE. Given the scope of this experiment, we consider two hard-constraints based on the features of the missiles: navigation equipment (GPS[1], DSMAC[2] or both) and warhead type (penetrating, unitary, submunition). We also consider three so- called "soft-constraints": mission priority (low, medium, high), firing rate (probability of hitting a target) and days to port (number of days until the ship is due back to the harbor). The automated decision support provides the user with a heuristic-based computer-generated solution that only takes into account hard constraints along with a limited set of additional criteria. The solution provided usually is not optimal, but always exhibit correctness with respect to hard constraints. Finally, a time bar gives subjects a visual indication of how much time they have left to generate their solution. There also is a message box where information from Central Command can be relayed.

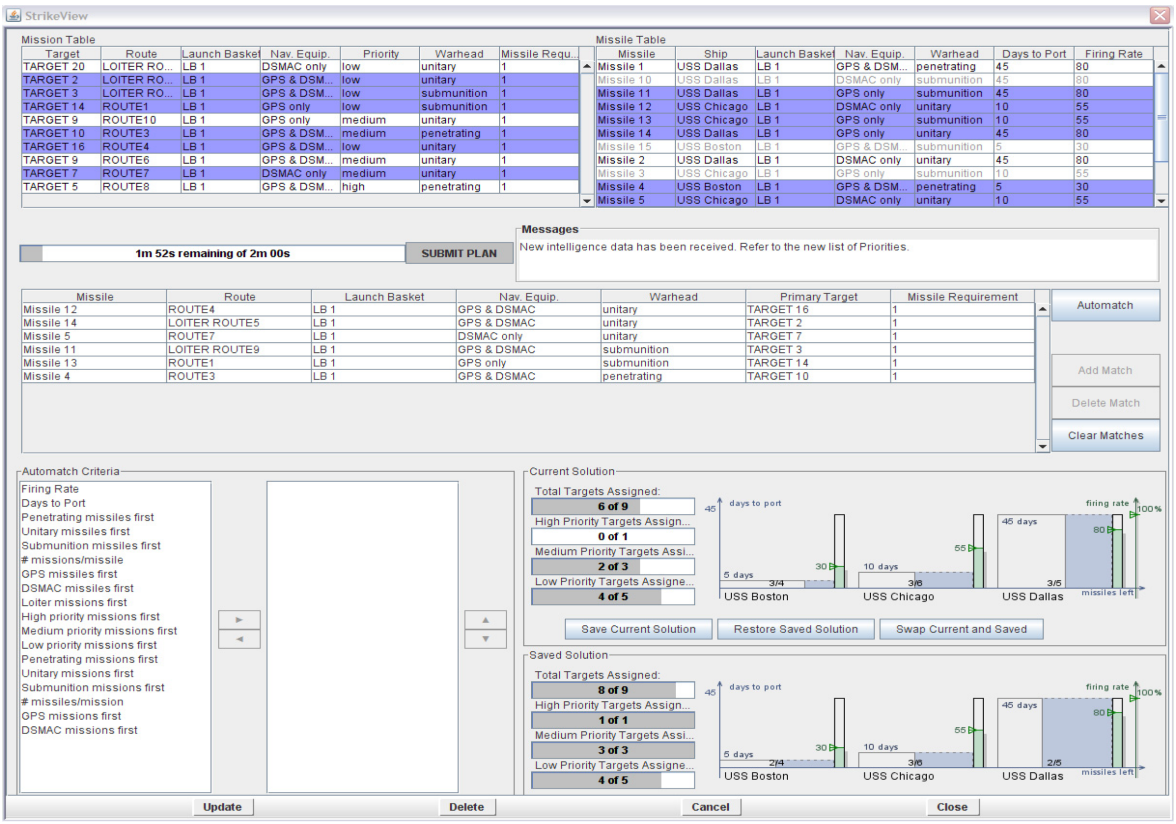


Figure 1. StrikeView Interface

Experimental Design

The experiment was a 2x2 mixed factorial design with time pressure (Low Time Pressure – LTP, High Time Pressure - HTP) as a within subjects variable and the order of presentation as a between subjects variable (LTP first, HTP first). The order of presentation was counterbalanced and randomly assigned to subjects.

Participants

18 participants were recruited, mostly from the MIT student population. Due to software glitches, the data obtained from two participants had to be dropped, therefore data from 16 participants were analyzed. In these 16, the male/female split was 11/5, 9 started with the LTP scenario and 7 with the HTP scenario. Finally, 8 were undergraduates and 8 were postgraduates (either graduate students of professionals). Each participant was paid $10 for the hour-long experiment, with a prospect of earning

---

[1] GPS: Global Positioning System

[2] DSMAC: Digital Scene-Mapping Area Correlator, is a high resolution satellite radar image of the target area which the Tomahawk follows to within feet of the intended target

an additional $60 gift certificate awarded to the best performer on the task. This monetary incentive was used to promote participants' involvement along with a drive to achieve an optimal solution.

Experimental Scenarios

In the LTP scenario, participants were given five minutes to complete the matching task, a duration that was determined to be comfortable during pilot studies. The HTP scenario started just like the LTP scenario with a five minute deadline. However, three and a half minutes into the experiment, the participants received new orders from Central Command to invert the priorities of the missions (that is low priority missions should now be regarded as high priority and vice versa). The participants had to re-plan the strike in the remaining one and a half minutes: this corresponds to the increased time pressure phase of the experiment. For both scenarios, the number of missions were greater than the number of missiles available, thus, it was not possible to assign a missile to every single mission.

Experimental Procedure

After signing consent forms, the participants were given a quick overview of the interface and experimental procedure via a slide-based presentation. A short training phase with the real interface and a mockup scenario was then provided in order to allow users to familiarize themselves with the task. There was no specific time limit on this hands-on training, and participants were asked to tell the experimenter when they felt they had achieved sufficient level of proficiency with the experimental setup. During the subsequent testing phase, each participant was presented with LTP and HTP scenarios in a random order. Before starting each scenario, the participants were given an identical set of ROE, which stated that the missions should be treated in the "normal" order of priority (high, normal, then low), and that they should try to maximize the firing rate attribute while minimizing the days to port attribute. In the HTP case, these priorities were reversed three and a half minutes into the scenario. After completing the two scenarios, the subjects were debriefed orally while a screen capture of their behavior was replayed. Questions were specifically asked to determine what type of strategy was used to solve the problem, how they reacted to the change of ROE and if they felt that time pressure affected their decision making process.
The data was recorded through built-in non-intrusive logging of the user interactions. The data consisted of mouse clicks, hovers and other interactions with UI features. We used TRACS2.5D to record participants' behavioral patterns. TRACS2.5D takes each triplet of successive mouse actions which was then fed through a parser which determines what category of action and what level of information detail was involved [9, 10]. Finally, as a failsafe, all trials were recorded using screen capture software.

Dependent Variables

The first dependent variable is a performance metric based both on the percentage of missions covered by a missile and on the optimization of the soft constraints (eq. 1).

$$perf = \sum_{all\ priorities} \Pr\left(\frac{1}{2}pm + \frac{1}{2n}\left(\frac{FR - \frac{DTP}{5 * DTP_{\max}}}{100}\right)\right)$$

$\Pr(H, M, L) = \{60,30,10\}$: *priority factor*

$pm \in [0,100]$: *percentage of matches per priority* Eq.1.

n : *number of missiles per priority*

FR : *Firing Rate*

DTP : *Days to Port*

$DTP_{\max}$: *Max. Days to Port*

$0 < perf < 100$

The second dependent variable is the number of calls to the automated help. Finally, the last dependent variable is aimed at providing a finer grained analysis of the participant's behavior by examining specific sequences, or chains, of user events. This provides information regarding the succession of actions that are most likely to be undertaken by the strike planner. The specific features of sequences of interest were determined by using the TRACS2.5D tool [9, 10] and manually noting the most strongly-recurring chains of events. Three recurring chains were identified as strongly recurring using this method. Following the TRACS2.5D nomenclature, the patterns of interest were: (1) browse, evaluate, select, (2) browse, select, create a match, and (3) select criterion, call automatch, evaluate match. These patterns covered on average about 85% of all interactions.

Results

Number of Calls to Automation

Figure 2 shows the observation frequency for different number of automation calls. Number of automation calls greater than or equal to two are grouped under one category as there were a few observations that were greater than two. The figure shows a general trend of increased use of automation for the high time pressure condition. An ordered logit model, specifically proportional odds, was developed to compare the level of automation calls for the two different time pressure conditions adjusted for order of presentation, and order – level of time pressure interaction. A proportional odds model takes into account the ordinality of the data [11], in this case the three bins for the number of automation calls. Repeated measures were accounted for by creating a population-average model. Because the data consists of repeated measures, generalized estimating equations (GEE) was used for estimation.

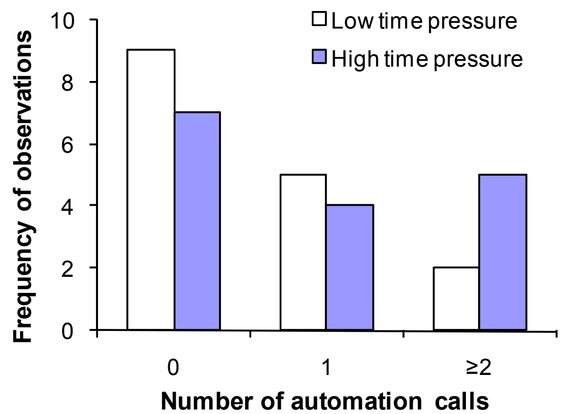
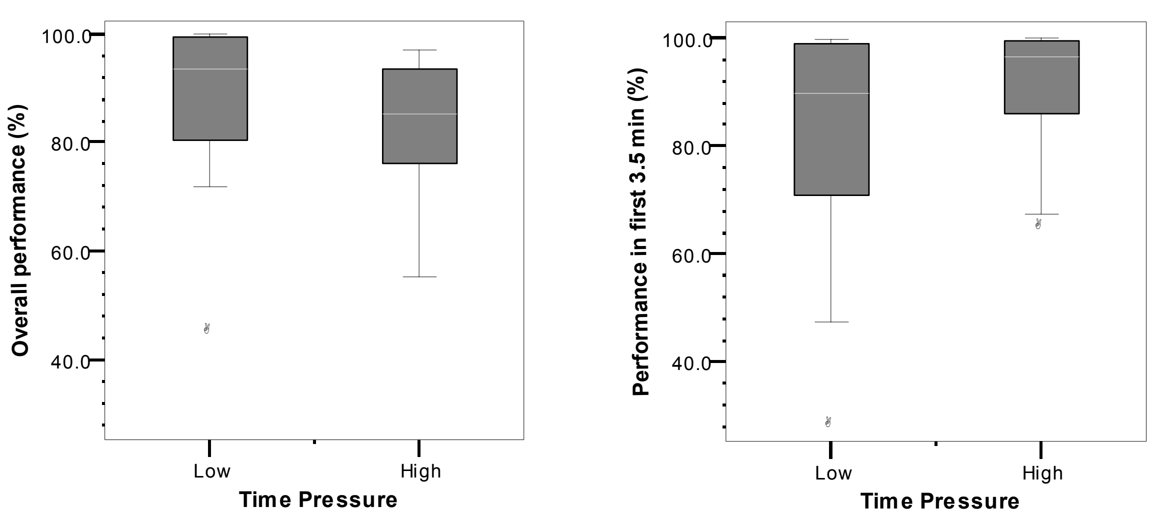


Figure 2. Observation Frequencies for Different Number of Automation Calls

Wald statistics for type GEE analysis revealed that time pressure ($\chi^2(1)=6.26$, $p=.01$) was statistically significant. Order, and order – level of time pressure interaction were not significant ($p>.05$), and hence were dropped from the model. High time pressure had 2 times higher odds of automation call than low time pressure (95% CI: 1.16, 3.42).

Performance



(a) (b)

Figure 3. Performance for Different Time Pressure Conditions

A repeated measures ANOVA was conducted on overall performance (Figure 3a). Time pressure, order, and their interaction were not significant (p>.05). This suggests that the time pressure may not have been severe enough to affect the overall performance, or the increased use of automation under higher time pressure may have compensated for the otherwise diminished performance.

Homogeneity of Scenario Difficulty

LTP and HTP scenario were designed to be as similar as possible in terms of difficulty. However, because the scenarios were not precisely identical, we wanted to ensure that the previous results were due to the difference in time pressure and not to uncontrolled variation in scenario difficulty. In order to show that the increased odds of automation use under the high time pressure condition was in fact due to the dynamic taskload increase rather than the different Strike View tasks performed in the two conditions, number of automation calls and performance in the first three and a half minute portion of each condition was analyzed (Figure 3b). No difference was expected since the first three and a half minute of each condition had similar taskload. The results showed a significant order effect ($F(1,28) = 4.74$, $p = .04$), but non-significant time pressure effect. The interaction between time pressure and order interaction was also not significant (p>.05). The number of automation calls (zero, one, or more) were analyzed with an ordered logit model. As expected, time pressure, and time pressure-order interaction were not significant (p>.05). These results suggest that the significant increase in automation use reported in the previous section is indeed due to the higher time pressure induced as opposed to variation in scenario difficulty.

Overall Strategy Switch

The question of whether time pressure would have impacts other than usage of automation in the participant's behavior was approached by measuring the presence of three strongly recurrent patterns in the user's behavior. Each pattern can be seen as a chain of three sub-events and can be scored by using pattern matching algorithms that count the number of time a specific pattern appears in the data. As discussed previously, the three patterns of interest were: (1) browse, evaluate, select, (2) browse, select, create a match, and (3) select criterion, call automatch, evaluate match. The analysis of the scores for the different pattern did not show any significant difference between the different scenarios or between the different phases of each scenario.

Discussion

The results of this experiment show that time pressure did lead to an increase use of automated help and thus verify our main research question; however, temporal stress did not seem to produce any significant differences in the way the participants chained their actions to solve the problem. The results obtained highlight the difficulty of measuring the impact of time pressure on human behavior because time pressure does not always lead to measurable changes in cognitive strategies: coping processes can balance out the effects of time pressure and maintain the same output [2].

Learning Effect and Training Issues

Performance measures revealed that there were significant order effects, with the second scenario yielding a higher performance. This trend suggests that the training may not have been sufficient to get participants at a reasonably stable level of proficiency, and that the first scenario might have had effects akin to an additional training session. The lack of shifts in strategies between phases of the low time pressure scenario could be a possible consequence of the unsatisfactory level of proficiency achieved by the participants with the interface. At the end of the experiment, multiple participants reported that they realized they should have used the automation after the change of ROE, but that they had been under too much pressure to think straight and actually implement what they recognized to be, a posteriori, the best solution. Some participants were clearly overwhelmed by the additional workload and the time stress engendered by the change in ROE. It is likely that had the experiment been repeated, the subjects would have been more ready to respond to a change in operational parameters simply because they would have had seen one already. It is also worth mentioning that Navy strike planners are trained officers and might therefore exhibit a different type of behavior.

Still, it is difficult to get the right balance between training the user to use an interface and biasing them by overemphasizing a given strategy. Over constraining the user into a pre-determined behavior does not help make an assessment of the cognitive strategies.

(Dis)Trust in Automation and Satisficing

Trust in automation [12, 13] is a vast topic that is barely touched in this experiment. During the training it was specifically mentioned that the automation feature would provide a correct, albeit likely sub-optimal, solution. The MIT population tends to be biased towards technically-inclined and detail-oriented personalities. As a consequence, we had some participants who refused to use the automation because they didn't like to use an algorithm they were not familiar with and that was described as suboptimal. This behavior was usually linked to the feeling that they could do better by using a fully manual strategy. In essence, this means that such participants were really trying to optimize the solution as much as possible, and would not settle for sub-optimality. Although not addressed by the data analysis, experimental observations tend to suggest that such people were the most affected by the change of ROE because they had to change their optimal search model, and were not trying to rely on simplifying heuristics. On a related note, the majority of people in this situation mentioned that they did not see it worth the effort to get familiar with the automated feature. There was thus a clear cost-benefit analysis that was made regarding the effort needed to understand the automation and the potential advantages it could bring. This conclusion could therefore have impact on the procedure used to train the strike planners.

At the other extreme of the spectrum, one of our participants was a very experienced US Air Force officer used to designing flight plans with the aid of a computer. His experience and training had taught him to trust the automation, and, according to the subject, even though the solution wasn't perfect it was considered to be "good enough". Such users were still trying to optimize the solution based on the automated one. However, because they usually weren't very familiar with the data, such optimization efforts were usually limited. Conversely, their strategy changed little under time pressure since they were able to leverage the automation to instantly create a plan that they accepted as being good enough. The best performance observed on the high time pressure scenario was actually from a participant who explicitly mentioned that he trusted the algorithm and thought it was good enough. After receiving the new ROE, this participant made use of the automation and changed the assignment of a pair of missile, thereby gaining an edge and outperforming all the other participants.

Conclusions

Experimental results supported our main research hypothesis, namely that, under time pressure, subjects tended to use more automation than in a baseline, low temporal stress, situation. Conversely, the experiment did not exhibit statistically significant changes of cognitive strategies between the two conditions. Based on post-experiment verbal reports, however, this result might be attributed to an insufficient amount of training. Further studies should be performed in order to satisfactorily answer this question. Still, the overall conclusions of this experiment highlight the need for a thorough understanding of the nature of the human-automation collaboration, especially in contexts such as command and control where time-critical decisions must be taken contingent on dynamic environments.

Acknowledgements

This research was sponsored by Boeing. The authors also wish to acknowledge the help of Kostas Speridakos and Edward Twigg for their help in coding and running the experiment.

REFERENCES

[1] J. W. Payne, J. R. Bettman, and E. J. Johnson, *The Adaptive Decision Maker*: Cambridge University Press, 1993.

[2] A. Edland, "Attractiveness judgments of decision alternatives under time stress," *Cognition and Decision Research Unit,* vol. Report No. 21, 1985.

[3] A. J. Maule and G. R. J. Hockey, "State, Stress and Time Pressure," in *Time Pressure and Stress in Human Judgment and Decision Making*, O. S. a. A. J. Maule, Ed. New York: Plenum Press, 1993.

[4] P. Wright, "The harassed decision maker: Time pressures, distractions and the use of evidence," *Journal of Applied Psychology,* vol. 59, pp. 555-561, 1974.

[5] O. Svenson, Edland, A., & Slovic, P., "Choices between incompletely described alternatives under time stress," *Acta Psychologica,* vol. 75, pp. 153-169, 1990.

[6] I. L. M. Janis, L., *Decision making: A psychological analysis of conflict, choice, and commitment*. New York: The Free Press, 1977.

[7] A. S. Luchins, "Mechanizations in problem solving: The effect of Einstellung," *Psychological Monographs,* vol. 54, 1942.

[8] S. Bruni and M. Cummings, "Human Interaction with Mission Planning Search Algorithms," in *Human Systems Integration Symposium*, Arlington, VA, 2005.

[9] S. Bruni and M. L. Cummings, "Tracking Resources Allocation Cognitive Strategies for Strike Planning," in *COGIS*, 2006.

[10] S. Bruni, J. Marquez, A. Brzezinski, and M. L. Cummings, "Visualizing Operators' Cognitive Strategies In Multivariate Optimization," in *50th Annual Meeting of the Human Factors and Ergonomic Society*, San Francisco, CA, 2006.

[11] M. E. Stokes, C. S. Davis, and G. G. Koch, *Categorical data analysis with the SAS System*, 2nd ed. Cary, NC: SAS Institute Inc., 2000.

[12] J. D. Lee and N. Moray, "Trust, Self-Confidence, and Operators' Adaptation to Automation," *International Journal of Human-Computer Studies,* vol. 40, pp. 153-184, 1994.

[13] L. J. Skitka, K. L. Mosier, and M. D. Burdick, "Accountability and automation bias," *International Journal of Human-Computer Studies,* vol. 52, pp. 701-717, 2000.

CUE-BASED TRAINING EFFECTS ON VISUAL SCANPATHS DURING WEATHER-RELATED DECISION MAKING

Michael W. Sawyer
Clemson University
Clemson, South Carolina
Scott A. Shappell
Clemson University
Clemson, South Carolina

Flight into adverse weather remains a leading cause of fatal accidents in general aviation. The situation assessment hypothesis suggests that pilots continue into adverse weather because they fail to accurately recognize the weather conditions present. In this study 20 participants' eye movements were tracked as they viewed various weather scenes before and after training. The results showed that after training participants made decisions using fewer visual fixations and less total gaze time. Further, the average time until first fixation on critical weather features was decreased after training. Participants were effectively taught what weather features are important, thus allowing participants to make quicker, more efficient decisions. Eye tracking was found to offer objective evidence of cue-based training's ability to affect and improve the decision making process.

Adverse weather is consistently cited as a cause in general aviation (GA) accidents; especially in fatal accidents. The majority of these fatal weather accidents, up to 90%, occur when pilots who are certified to fly only according to visual flight rules (VFR) continue flight into instrument meteorological conditions (IMC) (Coyne, Baldwin, & Latorella, 2008). VFR allows pilots to fly only in conditions which allow the pilot to control the aircraft by visual reference to the environment outside the cockpit. Specific weather minimums have been established in terms of visibility and cloud ceiling to ensure that pilots can control their aircraft. Regardless of these regulations, VFR pilots continue to fly into IMC, often with tragic results. VFR into IMC accidents have been shown to be result in at least one fatality 75% of the time, compared to 18% for other types of GA accidents (Wiegmann & Goh, 2000). What's more, these accidents should technically never happen as they involve pilots flying in conditions they are not certified for.

The majority of GA accidents are a result of pilot performance, as opposed to mechanical or structural failures (O'Hare, Wiggins, Batt, & Morrison 1994). There is strong evidence suggesting that pilots continue into IMC because they fail to accurately recognize the severity of the conditions ahead of them (Goh & Wiegmann, 2001; Wiegmann, Goh, & O'Hare, 2002). Further, several studies have shown that given the same written information about a flight, pilots tend to make consistent, predictable decisions (Driskoll, 1998; Hunter, 2003). This suggests that pilots must be having trouble gathering and identifying weather conditions accurately.

Training

Research suggests that training is the best method for improving situational assessment (Gaba, 1995). Cue-based training in particular has established itself as an effective method for training decision-making in many industries, including: emergency response (Ash & Smallman, 2008, aviation (Wiggins & O'Hare, 2003), cognitive rehabilitation (Hampstead, Sathian, Moore, Nalisnick, & Stringer, 2008, medicine (Jenkins, Shields, Patterson, & Kee, 2007), and law enforcement (Santarcangelo, Cribbie, & Hubbard, 2004). Cue-based training identifies and teaches specific cues that signify a change in system state which require a specific response (Smith, Giffen, Rockwell, & Thomas, 1986). One such program, WeatherWise, has been developed specifically to teach weather-related decision making (Wiggins & O'Hare, 2003).

WeatherWise was developed through a series of research studies aimed at understanding how pilots make weather decisions. Interviews with expert pilots were used to identify the key features of weather available to pilots as they made weather related decision (Wiggins & O'Hare, 2003). An online survey was then created to allow pilots to rate the importance of each cue. From the results of the study it was identified that the presence of three or more cues indicated that pilots should divert from their flight path. These studies thus served as the basis for the creation of a training program. An initial test of the program was completed by a group of pilots in 2003. The study found the program increased the pilots' subjective importance ratings of all nine weather cues. Further, pilots who received the training condition outperformed the control group on a decision making task involving a simulated flight. While the program increased the pilots' subjective importance ratings for weather cues, a more thorough analysis of its affect on pilot behavior is needed.

This study, therefore aims to analyze the effects of cue-based training on weather-related decision making. Eye movements will be utilized as the primary measure to study the effects of training. Eye movements are strongly correlated to interest (Starker, 1990), and have been shown to provide insights into a person's decision making process (Land, 2007). Eye movements, therefore should provide a novel method of assessing the training program's effectiveness.

Methodology

Participants

This study involved 20 participants recruited from Clemson University and surrounding areas. There were nine males and eleven females participants. The participants were an average age of 25.5 years old with a standard deviation of 4.8 years (min = 21, max = 42). None of the participants had accumulated any flight hours or certifications. All subjects in the study reported having normal or corrected to normal vision.

Apparatus

A Tobii ET-1750 eye tracking monitor was used to collect all eye movement data. The ET-1750 has a 17" monitor and samples at a rate of 50 hz with 0.5º accuracy. The resolution of the monitor was set at 1280 x 1024 pixels. Eye movement data was collected using the software program ClearView 2.7.0 developed by Tobii Technology.

Training Program

The training program WeatherWise was used in this experiment to teach weather-related decision making. WeatherWise is a cue-based training program developed by Dr. David Hunter, Dr. Mark Wiggins, and Dr. David O'Hare. The program was produced by the Federal Aviation Administration, Office of Aerospace Medicine for the Aviation Safety Program of the Flight Standards Service, with the assistance of The Ohio State University, The University of Western Sydney, The University of Otago, and King Schools. The program is available free in the public domain. WeatherWise and its development are well documented (Wiggins & O'Hare, 2003).

Weather Images

In preparation for this study a group of 120 royalty-free or creative commons weather images were gathered from various photo websites. The pictures collected showed a variety of weather scenes a pilot might encounter, ranging from relatively clear skies to severe thunderstorms. From the original 120 images, the 36 best images were chosen. These images were then sent out for external validation by a panel of five certified flight instructors with over 5,000 flight hours. The panel rated the conditions of each image as either VFR or IFR conditions. From their ratings, the 10 images which best represented each set of weather conditions were chosen. These 20 images were then randomly divided into two even groups.

Procedures

Participants were first given an initial briefing about the nature and goals of this study. They then read and signed an informed consent form. Each participant then completed a brief demographic questionnaire.

Participants were then given an introduction to GA and weather decision making. Participants were given a description of VFR including the specific requirements for daytime flight in class G airspace. Participants were told to assume they were on a cross-country VFR daytime flight in class G airspace. Participants were then asked if they had any questions or needed any clarification about VFR. The introduction was read from a script to ensure consistency.

The first task consisted of showing participants a series of 10 randomly ordered weather images on the eye tracker. For each image, participants verbally responded to the question "could you continue along your current flight path while staying above VFR minimums?" The image was displayed until a response was stated. This process was continued for each of the 10 images.

Participants then completed the training program WeatherWise. Participants were allowed as much time as needed to fully complete the training program.

Participants then returned to the eye tracker to view another series of weather images. Participants first viewed a randomly ordered set of new weather images, again responding whether or not they could continue on their flight path while staying above VFR minimums. Participants then viewed and responded to the original 10 images from the first part of the study, again in a random order.

Subjects were then thanked for their time, compensated and dismissed.

Data

The experiment was setup to capture differences resulting from training. The two types of data collected in this experiment consisted of the verbal responses to each weather scene and the eye movement data collected during each trial. The accuracy of each decision, both overall and in terms of signal detection, was calculated by comparing participants' responses to the responses identified by the panel of experts of each image.

Eye tracking data was continuously collected during each of the 1,800 weather decisions made in the study. Areas of interest (AOIs) were defined in each image prior to testing. In each image the following AOI features were identified: terrain, clear sky, clouds, horizon, cloud base, and cloud darkness. These features were chosen to be consistent with previous research findings (Wiggins, 1999). The eye tracking data was analyzed in terms of fixations, dwell time, and time until first fixation.

Results

Response Accuracy

Correct trials occurred when participants' responses to a weather scene matched the responses of the expert pilot panel. When the scene conditions were below VFR minimums (i.e. adverse weather was present) correct responses were coded as hits, while incorrect responses were misses. When conditions were above VFR minimums (i.e. adverse weather not present), correct responses were correct rejections (CRs), while incorrect responses were false alarms (FAs). An analysis of the responses to each weather scene is shown below in Table 1.

Table 1. *Response accuracy for all trials*

	n	Correct	Incorrect	Hit	Miss	CR	FA	d'	c
Pre-Training	200	143 (71.5%)	57 (28.5%)	66	34	77	23	1.151	0.163
Post-Training New	200	149 (74.5%)	51 (25.5%)	82	18	67	33	1.355	-0.238
Post Training Repeat	200	147 (73.5%)	53 (26.5%)	81	19	66	34	1.290	-0.233
All Trials	600	439 (73.2%)	161 (26.8%)	229	71	210	90	1.241	-0.096

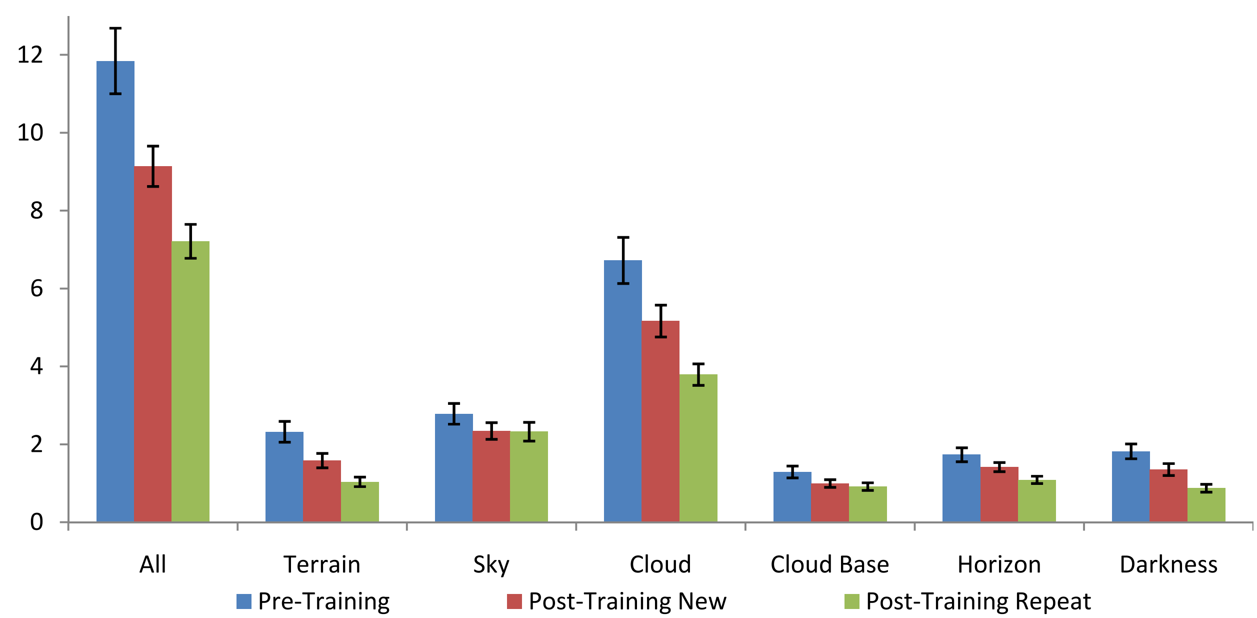
While an increase in response accuracy was observed after training, a 2-proportion z-test found the change was not significant between the pre-training condition and either the new image condition (z = 0.68, ns) or the repeated image conditions (z = 0.45, ns). Therefore, based on these data, it cannot be said that training improved decision accuracy.

It was found, however, that training did significantly reduce the number of flights into adverse weather as seen in the improved hit rate in both post-training conditions (new: z = 2.62, s, repeated: z = 2.44, s). No significant change was seen in the false alarm rate in either conditions (new: z = 1.58, ns, repeated: z = 1.74, ns).

An overall shift in bias was observed after training, as seen by the change in c, from a positive value (a liberal bias) to a negative value (a conservative bias). Further, an increase in discriminability was also seen in the increased d' value in both post-training conditions.

Eye Tracking Data

Eye tracking data was continuously collected over all trials in this study. Of the 600 trials, there were 22 (3.6%) in which no fixations were recorded. Those trials were excluded from the analysis. The average numbers of fixations per trial are shown below in Figure 1. The *All* category represents the average number of total fixations per trial and is not simply a summation of other categories due to overlapping AOIs.



As can be seen, there was a large decrease in the number of fixations per trial in both post-training conditions. Differences between the pre-training and post-training new images best show the effects of training as the participants were not familiar the weather scenes. As participants had seen the images in the repeated condition before, a large decrease in the number of fixations would be expected regardless of training condition. There was a significant reduction in the number of visual fixations used to make a decision after training. The largest decrease in fixations was seen in the cloud group, followed by the terrain group.

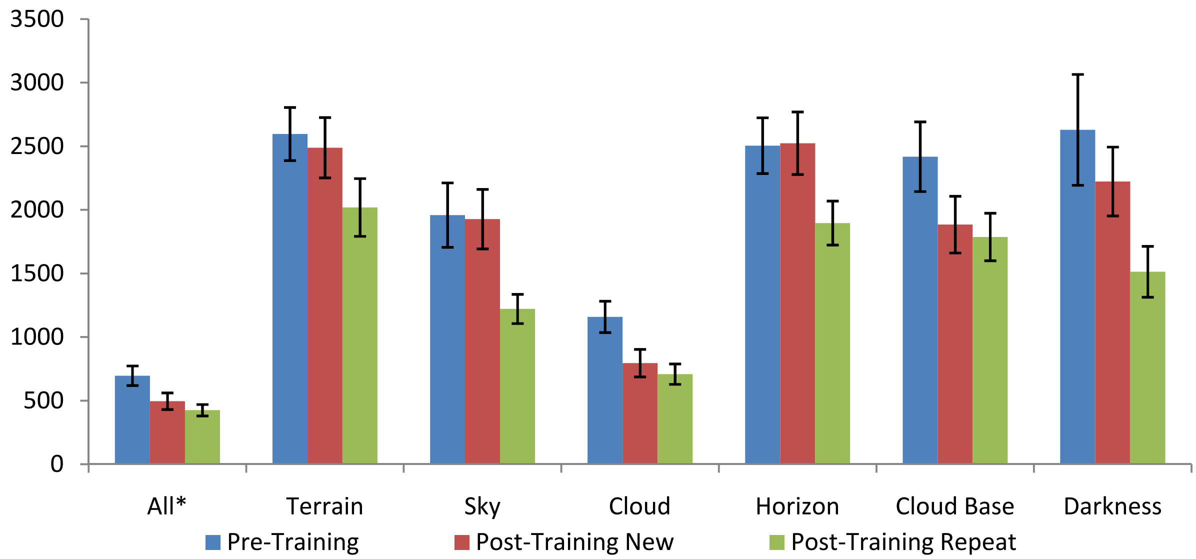


Figure 1 Average Time Until First Fixation

A decrease in the time until the participants' first fixation was seen after training. This signifies that after training participants scanned the image less before fixating on key features. This suggests that after training participants better knew what they were looking for in each image. Comparisons of the time until the first fixation within each category showed differences in the cloud and cloud base category.

Discussion

The training program WeatherWise was able to reduce the number of simulated flights into adverse weather, while also changing participants' visual scan behavior. While overall decision accuracy was not improved by the training program, the increased hit rate represents a successful decrease in inadvertent flights into adverse weather. In high-risk environments, this shift towards a conservative decision making bias should be considered a success.

After training, participants required less visual information to make decisions with the same degree of accuracy. Participants' fixations tended to be spread more evenly among the features after training, reducing the number of redundant feature fixations. This indicates that after training participants learned how to interpret the significance of each weather feature more efficiently. Eye tracking was found to offer objective evidence of cue-based training's ability to affect and improve the decision making process. Eye tracking would be a suitable tool for assessing a training program's ability to affect behavior in many different environments. Further testing of the program's long-term effects on decision making bias and scanpaths are needed, however.

References

Ash, J. S., & Smallman, C. (2008). Rescue Missions and Risk Management: Highly Reliable or Over Committed? Journal of Contingencies & Crisis Management, 16(1), 37-52.

Driskill, W. E., Weissmuller, J. J., Quebe, J. C., & Hand, D. K. (1998). Evaluating the Decision-Making Skills of General Aviation Pilots. In O. o. A. Medicine (Ed.) (pp. 43): Federal Aviation Administration.

Gaba, D. M. H., Stephen K.; Small, Stephen D. (1995). Situation Awareness in Anesthesiology. Human Factors, 37(1), 12.

Goh, J., & Wiegmann, D. (2001). Visual Flight Rules Flight Into Instrument Meteorological Conditions: An Empirical Investigation of the Possible Causes. THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY, 11(4), 21.

Hampstead, B. M., Sathian, K., Moore, A. B., Nalisnick, C., & Stringer, A. Y. (2008). Explicit memory training leads to improved memory for face?name pairs in patients with mild cognitive impairment: Results of a pilot investigation. Journal of the International Neuropsychological Society, 14(5), 883-889.

Hunter, D. R., Martinussen, M., & Wiggins, M. (2003). Understanding How Pilots Make Weather-Related Decisions. THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY, 13(1), 15.

Jenkins, J., Shields, M., Patterson, C., & Kee, F. (2007). Decision making in asthma exaverbation: a clinical judgement analysis. Archives of Disease in Childhood, 92(8), 672-677.

Land, M. F., Van Gompel, R. P. G., Fischer, M. H., Murray, W. S., & Hill, R. L. (2007). Fixation strategies during active behaviour: A brief history. In Eye Movements (pp. 75-95). Oxford: Elsevier.

Santarcangelo, M., Cribbie, R. A., & Hubbard, A. S. E. (2004). IMPROVING ACCURACY OF VERACITY JUDGMENT THROUGH CUE TRAINING. Perceptual & Motor Skills, 98(3), 1039-1048.

Smith, P. J., Giffin, W. C., Rockwell, T. H., & Thomas, M. (1986). MODELING FAULT-DIAGNOSIS AS THE ACTIVATION AND USE OF A FRAME SYSTEM. Human Factors, 28(6), 703-716.

Starker, I. B., Richard A. (1990, April). A Gaze-Responsive Self-Disclosing Display. Paper presented at the CHI.

Wiegmann, D., Goh, J., & O'Hare, D. (2002). The Role of Situation Assessment and Flight Experience in PIlots' Decisions to Continue Visual Flight Rules Flight into Adverse Weather. Human Factors, 44(2), 9.

Wiegmann, D., & Shappell, S. (2000). Human Error Perspectives in Aviation. THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY, 11(4), 341-357.

Wiggins, M. (1999). The development of computer assisted learning systems for general aviation. In D. Ohare (Ed.), Human Performance in General Aviation (pp. 153-172). Aldershot, England: Ashgate.

Wiggins, M., & O'Hare, D. (2003). Expert and Novice Pilot Perceptions of Static In-Flight Images of Weather. THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY, 13(2), 15.

Wiggins, M. W., & Bollwerk, S. (2006). Heuristic-based information acquisition and decision making among pilots. Human Factors, 48(4), 734-746.

Wiggins, M. W., & O'Hare, D. (2003). Weatherwise: Evaluation of a Cue-Based Training Approach for the Recognition of Deteriorating Weather Conditions during Flight. Human Factors, 45(2), 9.

USING MICROSOFT FLIGHT SIMULATOR X TO DEVELOP AERONAUTICAL DECISION-MAKING SKILLS IN THE CLASSROOM

Wendy S. Beckman, Mark Callender, Paul A. Craig, Steve Gossett, Wayne Dornan
Middle Tennessee State University
Murfreesboro, Tennessee

In the Aerospace Department at Middle Tennessee State University, Microsoft Flight Simulator X (MSFSX) is being utilized in the classroom to develop the aeronautical decision-making skills future pilots will need. The utilization of this inexpensive software to create realistic scenarios is discussed and a variety of examples are provided. While working through a scenario, students view either pre-recorded segments of a virtual flight, or participate in real time decision-making as a flight segment is "flown" by the instructor. In either case, students see an aircraft instrument panel and the view outside the windscreen, as they would in flight. The emphasis of each scenario is making appropriate flight management decisions given a variety of circumstances. Aerodynamics, systems failures, flight into deteriorating weather, and cross country skills are all areas with MSFSX scenario-based learning applications.

The Microsoft Flight Simulator software series was first made available in 1980, and over the past 28 years there have been ten editions released (Grupping, 2007). In the early days of the software, both the graphics and processing capabilities of computers and the level of sophistication of the software resulted in the program not being able to portray flight in a very realistic manner. This caused certificated pilots to view the software as solely a game; an entertaining and fun diversion, but not something that could be used for training or proficiency purposes. However, in the last decade, both the software and the capabilities of relatively inexpensive computers have evolved to the point of being able to provide a fairly realistic flight experience. This improvement has led to the use of the MSFSX package by certificated pilots both for training and proficiency purposes, even though it is not possible to "log" flight time in the conventional sense of the word.

While use of the program for practicing specific flight maneuvers or procedures is the most common application of the program, Middle Tennessee State University (MTSU) has found that the software is very helpful in teaching the concepts of aeronautical decision making in Private Pilot ground school classes. The idea that aeronautical decision making should be taught to Private Pilot students, along with the traditional technical aspects of flight, dates back at least as early as 1991, when the Federal Aviation Administration (FAA) published Advisory Circular AC 60-22, Aeronautical Decision Making (Federal Aviation Administration, 1991). The idea that pilot judgment can be taught, and not just acquired as a by-product of flight experience, was first expounded upon in that publication. From that time, and continuing through the current day, myriad efforts have been made to identify how exactly to teach aspiring pilots to implement effective aeronautical decision making from their earliest flight training activities.

With the advent of the FAA Industry Training Standards (FITS) approach in 2004, the scenario-based approach to flight training entered the general aviation training paradigm (FAA, 2004; Glista, 2003). Unlike traditional approaches, scenario-based flight training utilizes "real life" situations in training to afford training pilots an environment in which they can make decisions, and see the impact of those decisions, while still in a safe environment (i.e., under the supervision of their flight instructor). Research has shown that student immersion in and retention of lessons learned in scenario-based learning exceeds that of students trained using conventional methods (Ayers, 2006; Beckman, et al, 2008; Craig, et al, 2005a, 2005b; Dornan, et al, 2007a, 2007b, 2006). MTSU was one of the early users of the scenario-based FITS training concepts in flight training, and the success of this approach has led to the inclusion of scenario-based concepts in Private Pilot ground training. Given the physical constraints of a classroom, the use of MSFSX as a method of bringing a flight scenario into a class was identified as an innovative solution.

Discussion

There are many different areas of Private Pilot ground training that can be enhanced by the use of MSFSX. In most cases, it is still necessary to spend preparation time in conventional, fact-based learning modes to enable students to make the most of the scenarios that will be presented. For each topic area discussed below, examples of how MSFSX can be used to assist in the initial acquisition of functional knowledge will be provided, followed by how the software package can be used in scenario-based learning.

Aerodynamics

The software is immediately useful for flight students at the earliest stages of training, as by using the MSFSX 'spot plane' view of the aircraft being flown, students can see how the movement of the flight controls effects the movement of the aircraft about the three axes. Students can also see what an aerodynamic stall looks like to both an outside observer and to the pilot of the stalling aircraft. Once a basic understanding of aerodynamics has been achieved, including the concepts of coordinated flight, stall speeds, and the effect of turns on stall speeds, there are two scenarios which can be demonstrated. The first is a flight on which a newly certificated Private Pilot decides to perform a steep turn to demonstrate to his passenger the exciting things an aircraft can do. The passenger, who was ready for an adventure, has not seemed to be impressed by the uneventful takeoff, climb out, and cruise flight that he has experienced to this point. In his haste to show his passenger a good time, the pilot rolls into a 50 degree angle of bank turn, and forgets to add power. As the airspeed quickly bleeds off, students should be directed to keep an eye on the airspeed indicator. At the point of both the stall warning horn and the actual stall, the airspeed should be noted. Class discussion after this demonstration can center around: 1) What the problem was (students often do not even realize that additional power was not applied), 2) The actions that could have been taken to remedy the problem, and 3) Whether or not the load factor from the turn did indeed increase the stall speed as forecast in the aircraft manual.

The second aerodynamic scenario involves another newly certificated Private Pilot, on a flight to show her family (husband, two kids) their house from the air. After locating the house, she begins to circle around it so her family can see. The flight occurs in a low wing aircraft, and there is considerable consternation by the passengers about their inability to see the house properly. In an attempt to improve their view, the pilot begins to use her rudder controls to "get the wing out of the way," resulting in uncoordinated flight. Due to the attention being paid to viewing the house, she also inadvertently pitches up a bit, allowing airspeed to slow. This is continued to the point of an uncoordinated stall. Students can be shown the view inside the cockpit alternating with the view looking out of the side window at the house as the scenario progresses, and can also be shown the view of an outside observer at the moment the stall occurs. This view demonstrates quite dramatically how fast a spin can occur from an uncoordinated stall condition. Discussion after this scenario centers around: 1) The importance of coordinated flight, 2) The appropriate division of attention in this type of situation, and 3) The authority of the pilot in command in explaining to passengers that a particular maneuver is not possible.

Aircraft Systems

The area of aircraft systems and instruments is full of possibilities for enacting scenarios. The MSFSX software allows various systems and instruments to be set to fail within either a specified window of time, or randomly. The software is useful as an introduction to such basics as the throttle, mixture, and elevator trim. For example, instead of just talking about the throttle controlling RPM, the throttle can be advanced and the increase in RPM demonstrated. While the possibilities for scenarios with various system failures are numerous, one that has been done to great effect is an alternator failure. The scenario involves a night cross country flight, and midway through the flight, the alternator fails. This affords the class the opportunity to grapple with the decision-making process about whether to attempt to continue to the destination or to land at a nearby, suitable airport. If the decision is made to continue, the battery power becomes exhausted prior to reaching the destination. The ensuing total electrical failure allows students to experience, in a more realistic manner than reading in a textbook, the effect of losing the entire electrical system at night. After this scenario, a number of items can be discussed, including: 1) Aircraft night VFR equipment requirements, 2) The need to discontinue a flight when a system fails, 3) How to reduce electrical load to conserve battery power, 4) What aircraft systems are lost in a complete electrical failure, 5) Procedures for arriving at either towered or non-towered airports with no communication, 6) No flap and no landing light approach and landing procedures, 7) The need to carry a flashlight for night flights.

Another example of a system failure scenario that may be demonstrated is the loss of the vacuum pump on a moonless night flight over a sparsely populated area. Because students often seem to regard the loss of a vacuum pump as somewhat inconsequential on a VFR flight, seeing how important this system is when visual reference is compromised is a good learning experience. As with the alternator failure, the decision-making exercise revolves around whether to continue the flight to the destination, or to land as soon as practicable. Issues such as contacting ATC for assistance should also be addressed as the class works through the decision-making process.

Weather

An interesting item to demonstrate to students, when discussing VFR weather minimums, is what various visibilities look like. Simply showing a class what unlimited, 10 miles, 7 miles, 5 miles, 3 miles, and 1 mile visibility look like during both day and night operations is enlightening for them. Students are surprised to discover how little they can see with 3 miles visibility, particularly at night. In addition to this demonstration, to build a scenario, the instructor can either pre-record on MSFSX a VFR flight into deteriorating weather conditions which the class can watch and discuss; or, if actual current weather conditions exist that will allow demonstration of this problem, MSFSX can display real-time weather data from the internet to be used instead. In either case, a cross country flight to a destination airport should be started, with progressively worse weather encountered than what had been forecast. It is particularly instructive to have looked at, as a class, the weather briefing for the route of flight before beginning the demonstration, and to then note how conditions are not meeting expectations. If the flight is being flown in real time, decisions will constantly need to be made about whether or not to continue, and what alternatives are available once aloft. If a pre-recorded version of the flight is being viewed, the instructor should pause the replay at strategic points to discuss what the next possible steps might be for the flight. There are advantages to both methods, but the necessity of finding the appropriate weather conditions to do the exercise in real time may dictate the need to pre-record this lesson. When pre-recording the flight, the instructor may simply pause the flight being recorded at various points to change the weather being experienced to what is desired for the demonstration. These pauses will not be seen when the recorded segment is viewed by the class. Discussion during and after the scenario revolve around: 1) The importance of setting personal weather minimums, 2) Determining when weather conditions warrant a new course of action, 3) Obtaining updated weather information, 4) Generating a suitable course of action if diversion becomes necessary, and 5) Determining what will be necessary to enact a new course of action.

Navigation/Cross Country Flight

As students learn the skills to plan cross country flights, their flight planning for a particular route can be checked by actually flying the planned leg. After the class has planned a relatively short cross country leg and completed a navigation log using dead reckoning and pilotage skills, the leg can be flown to practice both finding check points and making groundspeed and time calculations. This exercise is also helpful in teaching students to read and understand VFR charts. To be most effective, the winds in MSFSX should be set to something slightly different than what was forecast, so that students are forced to make adjustments to their flight planning during the exercise. Calculations of a new heading and times to succeeding checkpoints and the destination based on this information should be performed. Once students are adept at the basic skills, a scenario may be used that makes fuel an issue. Students should be assigned flight planning, including weight and balance and aircraft performance determination, for a flight in which the aircraft must use reduced fuel due to passenger and baggage weight constraints. Based on performance data, the scenario should be designed so that the flight has a 45 minute fuel reserve. When the scenario flight begins, the winds aloft should be set such that the headwind component experienced is considerably greater than what was forecast, such that fuel exhaustion will occur before the planned destination airport is reached. The class should be required to make speed and time calculations as the flight progresses, which should lead to the realization that insufficient fuel is available. When this is discovered, decision-making to determine an appropriate course of action should ensue. The discussion items for this scenario include: 1) The importance of keeping up with groundspeed and time issues, instead of simply relying on the GPS, 2) The identification of appropriate courses of action given the situation, and 3) The importance of identifying suitable alternate airports before a flight segment is started.

Once basic navigation skills are mastered, MSFSX is also valuable for both introducing VOR and GPS navigation and for using these skills in a cross country scenario. To introduce the VOR, the tuning and identification of the navigation radios, the concept of flying to or from a station and how to set the OBS for each, and the ability to determine a position from a station are all concepts that can be demonstrated far more effectively in MSFSX than by using a PowerPoint presentation. For GPS operations, the avionics set up and usage procedures can be demonstrated to students on the actual equipment. Several of the available aircraft in MSFSX are equipped with a Garmin 500-series GPS, and two single-engine aircraft, including a C-172, are equipped with a G-1000. Once the basics of VOR and GPS navigation have been covered, a cross country leg can be planned and flown using one of the devices. For these scenarios, circumstances should be set that will require the class to make decisions throughout the flight. For instance, the students will learn a lot from the initial set up of the avionics, as well as in the airport area departure and establishment on the course line. But, once the flight has been established at cruise for a length of time, it is time to introduce some difficulties to the flight. For instance, the need to quickly divert to a nearby airport due to an ill passenger is effective. This will require students to think through a number of issues, such as determining the nearest suitable airport, re-setting the navaids as necessary, gathering alternate airport information, and communicating with Air Traffic Control (ATC) as appropriate.

Beyond basic navigation, there are a number of other cross country skills that may be practiced as a class using MSFSX. For instance, when Airport/Facility Directory use is taught, students can first review the information about an airport layout and then can be shown on MSFSX how that airport will appear when approached from various directions. Discussion and demonstration of a variety of issues, including how to enter the traffic pattern at a non-towered field, the communication requirements when entering various types of airspace, airport signs and markings when taxiing, and the view of an airport when arriving at night are all good areas for exploration. Short scenarios can be developed using all of these ideas. For example, the class can research the airport layout at a nearby Class C facility. Then, in MSFSX, the aircraft can be positioned as if it has just completed its landing roll and taxied off the active runway. The instructor can give the class taxi instructions to the general aviation ramp, which the class is responsible for copying, reading back, and following, so the aircraft moves correctly to its parking destination. During and after this exercise the need for using airport diagrams at unfamiliar airports, the interpretation of airport signs and markings, the use of ATC for progressive taxi instructions when necessary, the requirement to read back all hold short instructions, and the necessity of being absolutely sure of one's position on the airport surface, are all useful discussion areas.

Summary

The scenarios which can be developed using MSFSX are limited only by the user's imagination. However, it is not necessary for every ground school instructor to spend the time and effort to think of these situations. In addition to the ideas presented above, there are two books available which address using MSFSX in variety of aspects of pilot training, Microsoft Flight Simulator X for Pilots: Real World Training and Microsoft Flight Simulator as a Training Aid (see reference list). These texts both present a number of other possibilities for the use of MSFSX.

Although MSFSX was long thought of as primarily a game, it is in fact an inexpensive yet very valuable tool which can be used to bring real-life scenarios to the classroom. As the industry expects pilots to become better acquainted with good aeronautical decision making in early flight training, it is necessary to bring elements of the decision making process to the classroom. Simply discussing items related to good decision-making such as hazardous attitudes, "I'M SAFE," or "PAVE," while valid conceptual knowledge, does little to prepare pilots to implement successful decision-making in an aircraft. Working through scenarios on the ground prior to experiencing them in flight seems to be both effective and well enjoyed by students. The only drawback to the scenario-based approach in teaching a Private Pilot ground school is that it is challenging to cover all of the FAA required topics in a regular college semester when the time is taken to experience scenarios. To work most effectively, students must read and come to class prepared to work in a particular area so class time can be spent on scenario training. The scenario-based training method has been proven effective in flight training, and MSFSX offers a means to bring scenario-based training to the classroom. By engaging students in scenarios from the earliest stage of their training, sound aeronautical decision making can be developed from the start.

References

Ayers, F. H. (2006). The application of scenario based recurrent training to teach single pilot resource management (SRM) under the FAA Industry Training Standards (FITS) Program. *Journal of Aviation and Aerospace Research, 3*, 13-25.

Beckman, W.S., Callender, M., Gossett, S., Dornan, W.A., Craig, P.A. (2008). Moving FITS Training From the Laboratory to the Flight Line. *Collegiate Aviation Review,* 26(1), 1-8.

Craig, P. A., Bertrand, J. E., Dornan, W., Gossett, S., & Thorsby, K. K. (2005a). *Ab Initio training in the glass cockpit era: New technology meets new pilots.* Paper presented at the International Symposium on Aviation Psychology.

Craig, P. A., Bertrand, J. E., Dornan, W., Gossett, S., & Thorsby, K. K. (2005b). Scenario-based Private/Instrument syllabus versus traditional maneuver-based syllabi: A preliminary descriptive analysis. *Collegiate Aviation Review, 23*, 23-27.

Dornan, W. A., Beckman, W., Gossett, S., & Craig, P. A. (2007b). A FITS Scenario-Based Training Program Enhances GPS Pilot Proficiency in the General Aviation Pilot. *Collegiate Aviation Review, 25*(1), 19-27.

Dornan, W. A., Beckman, W., Gossett, S., Craig, P. A., & Mosey, P. (2007). *The Implementation Of The FAA Industry Training Program In Technically Advanced Aircraft (TAA): Lessons Learned.* Paper presented at the International Symposium on Aviation Psychology, Dayton, Ohio.

Dornan, W. A., Craig, P. A., Gossett, S., & Beckman, W. (2006). Best evidence for the FAA Industry Training Standards program for pilot training in technically advanced aircraft. *Collegiate Aviation Review, 24*(1), 58-66.

Federal Aviation Administration. (2004). *FAA-Industry Training Standards (FITS) Program 2004*. Retrieved, from the World Wide Web: http://www.faa.gov/education_research/training/fits

Federal Aviation Administration. (1991). *Aeronautical Decision Making.* (AC No: 60-22). Washington, DC: Government Printing Office.

Glista, T. (2003, March/April). FAA/Industry Training Standards (FITS): Times (and training requirements) are a changing. *FAA Aviation News,* 1-4.

Grupping, J. (2007). Flight Simulator History. Retrieved February 23, 2009 from Flight Simulator History Web Site: http://fshistory.simflight.com/fsh/timeline.htm

Van West, J. & 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.