16TH INTERNATIONAL SYMPOSIUM ON AVIATION PSYCHOLOGY

MAY 2 – MAY 5, 2011
WRIGHT STATE UNIVERSITY, DAYTON, OHIO

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

Table of Contents

Aviation Safety Culture

Improving Communications

Workload and Situation Awareness

NextGen Flight Deck

Crew Resource Management Training

International Organizational Factors

NextGen Air Traffic Control

Crew Resource Management: Individual & Organizational Factors

Risk Taking Behavior, Decision Making, and Performance

Tuesday Posters

Panel: Human Factors Challenges Facing UAS Integration into the NAS

Runway Incursions

Transitioning between Technology

Commercial Crew Selection

Psychophysiology in Aviation

Training and Selection

Maintenance

Improving Traffic Separation

Symposium: Operator Interface Concepts for Supervisory Control of Remotely Piloted Aircraft

Threats to Alertness

Advances in Selection

Training for Threats

Flight Deck Displays

Wednesday Posters

Cognitive Engineering

General Aviation

Investigative Methods

Cognitive Modeling in Modern Aviation

Aeronautical Decision Making and Weather

Collegiate Aviation Programs

Symposium: High-Fidelity Simulation and Training to Improve Coordination between Aerospace Specializations

Automation

Cognition in Aviation

Aviation Professional and Cultural Issues

Simulation

SINGLE-ENGINE FIREFIGHTING AIR TANKER HUMAN FACTORS

R. Small, C. Wickens, A. Sebok, R. Sargent
Alion Science and Technology Corporation
Boulder, Colorado USA
M. Bickham
Boise, Idaho USA
C. Kemper
Queen Bee Air Specialties
Rigby, Idaho USA

This paper describes the unique environment and safety challenges for pilots who fly single-engine air tanker (SEAT) firefighting aircraft, and the highlights from a human factors (HF) analysis of SEAT operations conducted in 2009. The HF analysis used the FAA's Human Factors Analysis and Classification System (HFACS) because of its broad examination of HF issues and prior use in aviation. We examined operations, training, and the applicable safety reports, and we interviewed or received anonymous surveys from 71 stakeholders – pilots, supervisors, and managers. The analysis yielded 63 recommendations, many of which were adopted for the 2010 fire season. Despite a busier fire season in 2010 compared with 2009, SEAT safety improved. The Bureau of Land Management (BLM) attributes this improvement, in part, to the adopted recommendations.

From 1997-2007 the single-engine air tanker (SEAT) accident rate was three times higher than the general aviation (GA) accident rate in the U.S. (BLM, 2009; NTSB, 2009). While the SEAT firefighting mission is inherently more challenging than typical GA flights, the BLM, the agency responsible for the SEAT program, determined that the predominant contributing factor for the SEAT program's poor safety record was human factors (BLM, 2009).

Unique Environment

The majority of SEAT firefighting aircraft are AT-802s, a turbo-prop aircraft designed for crop-dusting (Figure 1). Because aerial firefighting is predominantly a summer activity, many SEAT pilots earn their living during the rest of the year in agricultural operations. Agricultural operations are typically flown in the early morning with light or no winds, over flat terrain, with little or no radio communications, and few, if any, other aircraft in the vicinity. Agricultural operations and training are fairly unstructured under FAR Part-137.

Figure 1. Air Tractor AT-802 SEAT aircraft, a turbo-prop designed for dropping chemicals from low altitudes.

In contrast, SEAT firefighting operations are often flown in the afternoons over rugged terrain (e.g., mountainous with trees), with variable winds due to the local effects from growing fires, and with extensive radio communications. Radio calls are with a dispatch agency, aerial supervision, and sometimes other aircraft in the fire traffic area, including helicopters. Another important distinction is that crop-dusting is most effective when chemicals are applied as low to the crops as practical; whereas with fire retardant, the ideal drop height is 85 feet, plus or minus 5 feet, above the foliage. SEAT training and operations follow FAR Part-135 which provides more structure, including simulator training (starting in 2011; Figure 2), and operations and training manuals for pilots and maintainers. Firefighting pilot tasks and responsibilities, therefore, are significantly different from crop-dusting tasks.

Figure 2. AT-802 prototypical cockpit simulator to be used for more rigorous initial and recurrent pilot training.

Methods

We began the analysis by reviewing some SEAT background information, NTSB SEAT accident reports, and with the desire to cast as wide a net as practical to understand the SEAT firefighting environment and contributing factors to the poor safety record. We decided to use HFACS because it not only prompts analysts to study pilot issues, but also supervisory and organizational issues (Shappell & Wiegmann, 2000). For example, in the "unsafe acts" category, we examined pilot skill-based errors, decision errors, perception errors, and violations. Preconditions for these unsafe acts included adverse mental states, adverse physiological states, limitations, and personal readiness. Supervisory issues that impact safety included inadequate supervision, inappropriate operations, failure to correct known problems, and supervisory violations. And, finally, organizational issues that contribute to safety problems are resource management, organizational climate, organizational processes, and poor oversight.

We also thought that a survey, with the promise of confidentiality, would enable us to receive inputs from as many stakeholders as practical. The BLM customers supported this goal by providing a cover letter that promised confidentiality.

After reading SEAT program documents (e.g., operations guides, training materials, vendor contracts), we attended SEAT pilot training which clearly illustrated the many challenges faced by pilots. During the training program, we also interviewed stakeholders and distributed our survey. After training, we visited SEAT bases to observe operations, conduct interviews, and distribute more surveys. We obtained a video of actual firefighting operations, with radio calls, which further illuminated the operational tempo and numerous opportunities for errors and distractions.

Another key step in our method was to analyze accident, incident and safety reports from the SEAT program safety database. This information gave us a historical perspective, since our personal observations were limited to a single fire season (2009). We used HFACS to categorize these reports. For more details about our methods, and the results and recommendations, below, interested readers should see Small et al. (2009).

Results

Results from background reading indicated the complexity of aerial firefighting operations. Training and SEAT base observations reinforced these results. On the one hand, the AT-802 is well-suited to the role of initial attack on a wildfire because of its range from bases located in fire-prone areas, and because of its ability to efficiently carry a load of fire retardant. It can be dispatched quickly to keep fires from growing, and it affords the pilot good visibility from a crash-worthy (9 g) cockpit.

On the other hand, a single-pilot aircraft precludes crew coordination and sharing responsibilities for cockpit tasks when operations, including radio calls, become complex. Also, the AT-802 is highly maneuverable and has a high-torque engine which puts it closer to its flight envelope limits than other firefighting aircraft (AMD, 2009). Plus, the drop gate and radio controls are both fairly complex (Figure 3). For pilots with only agricultural experiences, the radio panel presents a big challenge.

Figure 3. Typical AT-802 drop gate control panel (left) and radio control panel (right). While most SEAT pilots are already familiar with the drop gate controls due to their crop-dusting experiences, few have experience with the radio controls until they begin firefighting training.

Results from surveys and interviews indicated that the Top Three hazards were: (1) weather & visibility; (2) terrain (unfamiliar, mountainous, or low-level flight over it); and (3) long periods of no operations followed by high activity. The survey also asked about flying distractions, which focused our attention on: radio communications and the difficult radio panel interface, other traffic in the fire area (e.g., helicopters dropping water on the fire), and diversions to other fires or bases. We also noted that safety is emphasized in different ways by the various companies who provide SEAT services. In particular, some companies pressure pilots to fly against a fire in marginal conditions because operational flights are how the companies earn more money under their BLM contracts.

The survey results revealed that training, operational support and procedural safeguards are sufficient and effective. Also, maintenance and aircraft reliability were not identified as problems, even though some safety reports indicated otherwise. This disparity was likely due to the timing of the safety reports from years past compared to the more current information gathered from interviews and surveys. In the prior four years, before our analysis began, older aircraft were retired in favor of the newer AT-802.

Results from the safety report database analysis included a total of 19 reports from NTSB and federal SEAT accident and incident summaries (1996-2009). In these 19 accidents or incidents, there were 5 fatalities. Table 1 lists the probable causes for these 19 SEAT accidents and indicates that engine failures and controlled flight into terrain (CFIT) are the primary causes of accidents, responsible for 80% of fatalities and for 68% of the accidents. It is unclear whether the engine failures were due to maintenance or pilot performance issues. CFIT accidents are distinguished from *loss of control in flight* (LOCIF) accidents, so the CFIT accidents reported here are inferred to be due to loss of pilot situation awareness. Similarly, *failure to see obstructions* is also related to pilot situation awareness deficiencies. Therefore, we may conclude that 80% of fatalities and 42% of accidents resulted from a loss of situation awareness, which may be due, in part, to the distractions mentioned above.

Further analysis of the 19 reports by phase of flight indicates that the fire retardant dropping phase is associated with the most accidents (63%) and fatalities (80%). Within this phase, accidents and incidents are evenly distributed among approach to drop, drop, and climb after drop.

Table 2 identifies the HFACS category for the 19 accidents and incidents. Some accidents involved multiple categories and subcategories, so that the total number of accidents and incidents for a given category does not equal the sum of the subcategories. This table indicates that deficiencies were identified across the SEAT system: pilots, supervisors, and the SEAT organization all contributed to these 19 SEAT accidents and incidents.

Table 1.

Summary of accident and incident reports (1996-2009).

Probable Cause	**Number of accidents**	**Fatalities**
Engine failure, loss of power, or throttle failure	7 1	
Controlled flight into terrain (CFIT)	6 3	
Failure to see obstruction (dead trees, wires)	2 1	
Loss of control or stall	1 0	
Blown tire	1 0	
Environmental conditions	1 0	
Overweight	1 0	

Table 2.

HFACS categories and subcategories for the 19 accidents & incidents.

HFACS Categories	***Number of accidents or incidents with these factors identified***
Unsafe acts	12
Errors	9
Violations	7
Preconditions for unsafe acts	8
Substandard conditions of pilots	6
Substandard practices of pilots	5
Unsafe supervision	9
Inadequate supervision	4
Planned inappropriate operations	5
Failed to correct problem	8
Supervisory violations	3
Organizational influences	10
Resource management	2
Organizational climate	2
Organizational processes	7

Summarizing all results from our data gathering and analysis methods, the top 12 SEAT safety problems that we identified, in order, are below. To achieve this ordering, we used five main sources: survey responses (weighted most heavily), interview notes, our personal observations, safety reports, and the video of actual operations. We noted frequencies of problems and severity. We placed more weight on survey responses due to the relatively high response rate (N=48) and the high correlation between the number of times a specific hazard was mentioned and its order in the Top Three list. We assume that anonymous respondents did not collaborate on their responses, so the survey's results were, in a sense, self-validating.

1. Weather, visibility (e.g., winds over rugged terrain, smoke)
2. Terrain awareness, CFIT (controlled flight into the terrain)
3. Pilot skills, CTM (cockpit task management)
4. Energy management, engine failure, loss of power, LOCIF (loss of control in flight)
5. Air traffic
6. Safety culture
7. Radio *chatter* (i.e., radio communications not directly relevant to the SEAT pilot)
8. Difficult radio interface (Figure 2)
9. Poor communications (e.g., unclear drop location description from the incident commander)
10. Workload transition (prolonged period of no flying, then significant operations)
11. *Cowboy mentality* (risky pilot behaviors)
12. Maintenance

Recommendation Highlights

While our initial sense was that in such a dangerous business the accident rate might be higher, we learned that many accidents and incidents were preventable, and that there are hazards that can and should be addressed as resources and a consensus to act dictate. We strongly supported the BLM's decision to transition the SEAT program from an FAR Part-97 operation to one that follows Part-135 as much as practical – a transition that provides more structure for training and evaluating pilots and maintainers due to the requirement for formal operations, training, and maintenance manuals. Our recommendations focused on the categories of personnel selection and training, cockpit equipment and procedures, communication impediments, maintenance, and organizational issues. Table 3 highlights our recommendations; its third column indicates which of the 12 SEAT safety problems (above) are addressed by the recommendation category. The recommendations are ordered by the number of higher ranking problems addressed and the relative ease (lower cost or lower effort) for adopting a recommendation.

Table 3.

Recommendation categories, description, and safety problems addressed.

Recommendation Category	***Brief description of recommendation category***	***Problems addressed***
Personnel selection	Identify and use screening tools to ensure that SEAT pilots (and other SEAT personnel, e.g., SEAT managers) have the necessary skills and personality characteristics.	3,6,11
Training	Targeted training can be developed to address a wide variety of issues, including: energy management; allocating attention (e.g., visual scan patterns) to maintain awareness of the terrain, surrounding aircraft and obstructions; practicing radio procedures; handling sudden workload transitions; adopting a safety- and procedure-conscious attitude; and refining piloting skills.	1-11
Cockpit equipment and procedures	Select equipment to support improved SEAT operations. Examples of equipment include: radios with improved user interfaces for easier interaction; an angle of attack display for energy management; a modified TCAS to increase awareness of surrounding traffic; and, a geographic display with interaction capabilities to be used by pilots and incident commanders (to facilitate shared situation awareness).	1,2,3,4,5,8,9
Communication protocol and procedures	Improve radio communications protocol and procedures. Train SEAT personnel (dispatchers, pilots, incident commanders) on improvements that should provide a consistent, efficient and minimalist way of transmitting information. Minimize dispatch calls when automatic flight following (AFF) is working properly.	7,9
Improved maintenance procedures	Develop and implement more stringent maintenance inspection procedures to ensure that all vendors are using the proper equipment and performing preventative maintenance as needed (on aircraft and support vehicles) in accordance with FARs and DoT regulations. Adopting applicable portions of FAR Part-135 should help here. For incidents and accidents resulting from maintenance errors, perform more detailed HFACS-type analyses to obtain more specific maintenance-related contributing factors to the error (e.g., documentation, supervision, lack of available equipment, etc.).	6, 12
Consistent rule enforcement	Implement and enforce rules that target key safety violations (e.g., flying below 60' AGL except for takeoffs and landings) with severe penalties. Consistency and due process are important, too.	6, 11
No-blame safety culture	Encourage the use of safety reports and do not use them to punish offenders, except for willful or egregious violations. Include safety discussions (especially of recent accidents or incidents) in all training, briefings, and debriefings; reward safe behaviors.	2-12

Conclusion

The BLM adopted several of our recommendations for the 2010 fire season, including: more thorough and formal mission pre-briefings and debriefings; video-taping proper approach-to-drop, drop, and post-drop maneuvers for training purposes; requiring a training flight after 10 days of inactivity during fire season; and, discussing past accidents and incidents with pilots, and how they could have been avoided.

Initial assessments of the effectiveness of these adopted recommendations follow. More formal briefings are not yet widespread, so their effectiveness is still to be determined. Cockpit videos of drops over varied terrain have anecdotally helped newer pilots "get the proper picture." The 2010 fire season was also fairly slow, so training flights about every 10 days seemed to help with pilot proficiency, according to anecdotes from pilots and from aerial supervisors. Discussions with pilots about past accidents and incidents also seemed to help pilots reflect upon what they would have done differently in similar circumstances, according to the recently retired SEAT program manager (fifth author). These discussions covered pilot stressors, such as extended time away from home and financial pressures to fly in marginal conditions. These stressors seemed to play a minor role, according to those pilots familiar with the accident or incident pilot, although simple awareness of such stressors may have a beneficial mitigating effect.

The work reported herein was accomplished in 2009. Subsequently, the 2010 firefighting season was one of the safest for the SEAT program. While this improvement is due to the hard work of many stakeholders, we are proud to have played even a small part in such an outcome. Time will tell if such improvements can be sustained.

Acknowledgements

The SEAT HF analysis was done under Dept. of the Interior, BLM contract L09PX00178 for the National Interagency Fire Center (NIFC). We thank Alion intern M. Ziarnick who helped with the original report's safety database analysis and classification. We are indebted to Dr. Scott Shappell, co-creator of HFACS, for his advice during the conduct of the study, and to Matt Jurgenson who provided a video of actual aerial firefighting operations. NIFC staff members who provided extensive background information and context during the SEAT HF analysis were Joe Bates, Kevin Hamilton, Glen Claypool and Rusty Warbis. The simulator photo (Figure 2) is courtesy of Tony Peckler at ATC Flight Simulators, Inc. Lastly, we could not have completed our analysis without inputs from the 71 anonymous interviewees and survey respondents.

References

AMD (2009). *Air tanker performance & risk: A brief analysis*. PowerPoint presentation given to Alion on 12/15/09 in Boise by Aviation Management Directorate (AMD) personnel.

BLM (2009). *Single Engine Air Tanker Human Factors* (solicitation L09PS00023, posted Jan 28, 2009). Retrieved from the Federal Business Opportunities web site https://www.fbo.gov/?s= opportunity&mode=form&id= a56bd72528cac7bd00e9353a91991106&tab=core&_cview=0.

NTSB (2009). *Accidents, Fatalities, and Rates, 1997 through 2007, U.S. General Aviation*. Retrieved from http://ntsb.gov/aviation/Table10.htm.

Shappell, S.A., & D.A. Wiegmann (2000). *The Human Factors Analysis and Classification System – HFACS* (DOT/FAA/AM-00/7). Washington, DC: FAA Office of Aviation Medicine.

Small, R., C. Wickens, A. Sebok, R. Sargent, & M. Ziarnick (2009). *Single Engine Air Tanker Human Factors Analysis Final Report* (under BLM contract L09PX00178). Boulder, CO: Alion Science & Technology.

SAFETY CULTURE MATURITY ASSESSMENT IN AIR TRAFFIC MANAGEMENT: READY-TO-GO?

Initial validation and correlations with related concepts

Michaela Heese
Karl-Franzens University
Graz, Austria
K. Wolfgang Kallus
Karl-Franzens University
Graz, Austria
Werner Artner, Thomas Marek
Austro Control GmbH
Vienna, Austria

Safety is the number one priority in Aviation and it has long been recognized that a Safety Management System (SMS) cannot be effective without an appropriate safety culture. The Civil Air Navigation Service Organisation (CANSO, 2009) in line with international regulatory requirements has defined safety culture as the 'enabler that integrates the various SMS elements into a coherent system.' This paper reports findings and recommendations following a pilot implementation of a 'ready-to-go' safety culture survey developed by CANSO at a local Air Navigation Service Provider. In addition correlations with related concepts such as resilience, organisational citizenship behavior as well as recovery and stress are investigated. The results will be used to develop a shorter and improved version of the survey including related concepts.

Safety culture as enabler for the integration of safety management system elements

Safety is the number one priority in aviation and stakes are high compared to other industries. The introduction of the 'Single European Sky' (SES) is an ambitious initiative, launched by the European Commission in 1999, to reform the architecture of European Air Traffic Management (ATM) to meet future capacity and safety needs. The European Organisation for the Safety of Air Navigation (EUROCONTROL) expects that today's traffic will have doubled by 2020. Current systems, with ongoing improvements, should be able to handle this increased load until the middle of the next decade. After that, more radical measures are called for in order to avoid serious congestion.

The International Civil Aviation Organization (ICAO) has mandated the implementation of a safety management system (SMS) as an organized approach to managing safety, including the necessary organisational structures, accountabilities, policies and procedures. However, a formal SMS alone is not enough to ensure safety as number one priority in light of the new requirements and challenges the aviation industry is facing. Effective safety management requires a genuine commitment to safety on the part of everyone in the organisation. This implies that organisations are not exempted from cultural considerations and that the success of a SMS is completely dependent on the development of a positive and proactive safety culture in the Air Navigation Service Provider (ANSP) organisation. The Civil Air Navigation Service Organisation (CANSO, 2009) defines safety culture as 'enduring value, priority and commitment placed on safety by every individual and every group at every level of the organisation. Safety culture therefore reflects the individual, group and organisational attitudes, norms and behaviors related to the safe provision of air navigation services.' National and international regulators like the European Aviation Safety Agency (EASA) also support this ICAO mandate by looking into the development and assessment of safety culture within an organisation. CANSO has developed a new SMS Standard of Excellence placing safety culture as key enabler that integrates the various SMS elements into a coherent system.

Assessing safety culture

CANSO has formed a Safety Culture Working Group that is dedicated to helping CANSO members with developing and measuring safety culture in their organisations as fundamental basis for implementing their SMS. CANSO emphasizes a phased step-by-step implementation of SMS from an initiating level to one of continuous improvement. The achievement of the highest level of SMS is a long-term process focusing on developing predictive measures to foster a proactive approach to safety management rather than waiting for the system to 'fail' in order to identify weaknesses and take remedial actions.

Safety culture research has identified many different models and dimensions underlying safety culture. One of the most common approaches is based on the five component model by James Reason (1997) which includes ‘just’, ’reporting’, ‘learning’, ‘informed’ and ‘flexible’ cultures. A review of the existing literature on the subject of safety culture showed that, although there is some consensus on the subject, definition and characteristics often do not correspond with each other depending on the author and the field of application (Wiegmann et al. 2002; Montijn et al. 2009). Mearns et al. (1999, p.1) explored the concepts of safety culture and safety climate in an attempt to determine which is the more useful for describing an organisation's ‘state of safety’. They argue that, ‘*although the two terms are often interchangeable, they are actually distinct but related concepts and should be treated accordingly.* ‘*Safety climate’ best describes employees' perceptions, attitudes, and beliefs about risk and safety, typically measured by questionnaire surveys and providing a ‘snapshot’ of the current state of safety. ‘Safety culture’ is a more complex and enduring trait reflecting fundamental values, norms, assumptions and expectations, which to some extent reside in societal culture.'* Therefore most research refers to assessing ‘safety culture’ as ‘cultural’ elements can be seen through safety management practices which are reflected in the safety climate. For the purpose of this paper the term ‘safety culture survey’ will be used when talking about assessing safety climate in an organization. In 2004 EUROCONTROL started a scientific study of safety culture in ATM resulting in a Safety Culture Measurement Toolkit assessing safety based upon five main elements ‘Safety Management Commitment’, ‘Trust in Organisation Safety Competence’, ‘Involvement in Safety’, ‘ATCO Safety Competence’ and a ‘Just, Reporting and Learning Culture’ outlining 18 sub-categories (Mearns et al., 2009).

Based on the available research the CANSO Safety Culture Working Group developed a safety culture model founded on the following eight components (as shown in Table 1).

Table 1

CANSO Safety Culture components (CANSO, 2009).

1. Informed Culture
2. Reporting Culture
3. Just Culture
4. Learning Culture
5. Flexible Culture
6. Attitudes towards Safety
7. Safety-related Behavior
8. Risk Perception

The first five components are equivalent to Reason’s (1997) five elements. The sixth component ‘Attitudes towards Safety’ is following the EUROCONTROL component ‘Safety Management Commitment’. The seventh component ‘Safety-Related Behavior’ has been included based on the assumption that there is a linkage between shared organisational beliefs and safety behaviors. The indication is that if safety beliefs are poor this will be reflected in poor human performance. Finally ‘Risk Perception’ was chosen according to findings from Mearns and Flin (1995, p. 1) demonstrating that ‘subjective perceptions of risk form the basis for risk acceptance (…) and as such are important for understanding feelings of safety, attitudes to safety, risk-taking behavior and accident involvement amongst the workforce.’ For detailed meanings and descriptions of the eight safety culture components the interested reader is referred to the CANSO website (http://www.canso.org/safety/documents).

In a second step the CANSO Safety Culture Working Group chose a total of 40 questions from the CANSO safety culture database for their ‘ready-to-go’ survey. The database consists of 777 questions obtained from various organisations such as ANSPs, airlines and universities. All questions and statements that were incorporated in the database have been applied in safety culture surveys in those organisations previously. Finally the 40 questions were assigned to one of the eight underlying safety culture components based on a content analysis. These 40 questions then formed the ‘ready-to-go’ survey proposed by CANSO.

Safety Culture and related concepts

Over the past decades research has primarily focused on validating safety climate as a robust leading indicator or predictor of safety outcomes across industries and countries and only little effort has been placed on relationships with other established constructs (Zohar, 2010). Whereas there has been some significant progress in this direction over the last 30 years such as ‘leadership as a climate antecedent’, ‘safety culture and overlaps with risk management or safety performance’ (Guldenmund, 2000; Sorensen 2002; Hoffmann, 2003) much more work is needed. Literature also suggests relationships of safety culture with ‘Resilience’ (Chevreau, 2006), ‘Organisational Citizenship Behavior’ (Organ, 2006) and ‘Stress’ (Fogarty, 2005).

Hollnagel (2006) puts a proactive safety management and a positive safety culture into perspective with the resilient organisation. He suggests that the first step towards business resilience is to analyse, measure and monitor the resilience of organisations in their operative environments. Resilience as the ability of an organisation to cope with unexpected events and dangers and to bounce back after untoward events is influenced by its way of managing safety and the overall safety culture of an organisation.

Organisational Citizenship Behavior (OCB) is a concept introduced by Organ (1988) as 'individual contributions that exceed the minimum role requirements of the job and improve organisational effectiveness.' According to recent research OCBs have a number of important efficiency and effectiveness benefits for an organisation and contribute to competitive advantage, although it is not formally rewarded by an organization (Organ et al. 2006). It is therefore suggested that OCB should correlate with behavior related to safety and safety culture. Creative and innovative actions in solving safety related issues for example increase individual and team performance, coordination between teams and the ability to cope with change (Podsakoff et al., 2000).

Fogarty (2005) investigates psychological strain as mediator on the impact of safety climate on errors in Aviation Maintenance. His findings support safety climate acting primarily on the psychological health of individual workers and that psychological strain is a primary determinant of maintenance errors. Furthermore he found that individual level variables, including safe behavior and general health, mediated the indirect effects of the organizational variables. Stress, in particular, was an important mediator of both organizational and environmental variables on the impact of safety climate on errors.

Scope and objective

This paper presents the results and recommendations following the pilot implementation of the 'ready-to-go' safety culture survey developed by the CANSO Safety Culture Working Group at a local ANSP. In addition correlations with related concepts such as resilience, OCB as well as recovery and stress are presented. Results will be used to develop a shorter and improved version of the survey including these related concepts.

Method

The pilot study consisted of two data collection phases. The 'safety culture maturity questionnaire' (SCM-Q) was administered in December 2010 to 80 licensed (tower and approach) air traffic controllers (ATCOs) in two ATM units, followed by voluntary 'safety-related reconstruction interviews' in association with the previous shift in the same units in January 2011. Both questionnaire and interviews were run during working hours in scheduled breaks ensuring enough time for the ATCOs to relax and that operational duties were not interrupted. Eighty ATCOs were invited to complete the survey, while 25 ATCOs were randomly drawn for the interviews. However, participation was voluntary and ATCOs on rostered non-operational office duty or night shifts were exempted. In order to facilitate analysis interviews were recorded on digital voice recorders with the permission of the participants.

Detailed descriptions of the CANSO components, the CANSO 'ready-to-go' survey and the complete data analysis and results including full references are available in the final report upon request from the authors.

Measures

The CANSO 'ready-to-go' survey consisted of 40 questions using a four-point Likert scale (0 strongly disagree to 3 strongly agree). Two original CANSO questions were split in two items and a couple of frequently used terms were adapted to facilitate better understanding (e.g. the term 'staff' was replaced by 'employees'). One question on the general understanding of the term 'just culture' and an entry question were added. These changes resulted in a total of 44 safety culture questions. In addition to the safety culture questions nine questions from the 'Resilience Instrument' (Mallak, 2006) and nine questions from the 'OCB Questionnaire' (Organ et al., 2006) were selected based on their corrected item-total correlations. Finally seven questions on a seven-point frequency scale (0 never to 6 always) were selected from the 'Recovery-Stress-Questionnaire for Athletes' (Kellmann & Kallus 2001). The final version of the 'Safety Culture Maturity Questionnaire' (SCM-Q) consisted of 69 questions with a total duration of approximately 12-15 minutes. In addition an information sheet explaining the data collection process and the use of the six-digit personal code to ensure data confidentiality was handed out. Upon the request of the organisation and to facilitate the translation process of the English items for the survey participants, the entire questionnaire was also presented in the national language. However, ATCOs were asked to complete the English version of the questionnaire and only refer to the questions in national language for translation purposes, to allow future data comparisons with other CANSO member states.

In phase 2 safety-related reconstruction interviews based on the 'Reconstruction Interview for the Integrated Task Analysis (ITA) for Air Traffic Controllers' (Kallus, Barbarino & VanDamme, 1998) associated with the previous shift were used to assess qualitative aspects of safety culture maturity and to gain information about safety-relevant concepts and behaviors in the organisation. In addition ATCOs were invited to rate safety-relevant situations during their previous shift on the 50-point 'Subjective Critical Situations' (Kallus et al., 2008) Scale (0=routine situation, 50=critical incident). The interview ended with questions relating to 'adhering to procedures',' team quality' and 'leadership commitment', as well as two questions on feedback about the previously administered questionnaire and the communication and implementation of the pilot study. The interview took about 40-50 minutes. After the interview ATCOs were invited to find out about their individual stress levels by providing saliva samples using 'Salivettes® Cortisol'.

Sample

In total 50 out of 80 licensed tower and approach ATCOs (62.5%) returned the safety culture maturity questionnaires. Twenty-five ATCOs were randomly assigned for the interview and 21 ATCOs agreed to participate in the voluntary stress measurements. Table 2 provides an overview of the available sample.

Table 2

Sample of the Safety Culture Maturity Pilot Study.

De-identified ATM Units	Questionnaire	Interviews	Cortsiol
A	n=19	n=11	n=7
B	n=31	n=14	n=14
TOTAL	N=50	N=25	N=21

The majority of participants was in the 31-40 age group and had an average of 7-14 years of experience as a licensed ATCO. More than half of the participants were supervisors and on the job instructors. Some of the participants were also members of the 'local safety committees', the 'local competence assessment' teams or were Critical Incident Stress Management (CISM) peers.

Analysis

The analysis followed a stepwise questionnaire development procedure, as outlined in Kallus (2010). The data were subject to an examination of their internal consistency (reliability), principal component analysis (PCA) including Varimax rotation and, Pearson's correlation analyses using SPSS (Statistical Package for Social Sciences) Version 17.0.

Results

The first step was a listwise exclusion of four items each with more than two missing values. This was executed to keep the total sample size constant for further analyses. As more than 15 items were missing one or two responses, the available sample was reduced 49 responses on 44 CANSO items.

In a second step a reliability analysis of the eight CANSO components was performed (see Table 1). Components with a Cronbach's Alpha (α) higher than .70 were retained; components with items with corrected item-total correlations ($r_{i\ (T-i)}$) smaller than .30 were excluded and underwent further analysis. Low corrected item-total correlations indicate that the concerned item is not measuring the same construct as the rest of the items. Exclusion of those items generally increases the internal consistency of a component. Only component 7 'Safety-Related Behavior' reached the Cronbach's $\alpha > .70$ cut-off. As a result of high correlations between scales '1.Informed Culture' and '2. Reporting Culture' ($r=.563$ $p<.01$) as well as '4. Learning Culture' and '5. Flexible Culture' ($r=.723$ $p<.01$) were merged for a second analysis. Scales '6. Attitudes towards Safety' and '8. Risk Perception' demonstrated very low internal consistencies having no items with $r_{i\ (T-i)}< .30$. These scales were therefore subject to further analysis.

A principal component analysis (PCA) was performed on all the items excluded from the earlier analysis. The PCA grouped three items from the original ‘Just Culture’ component with four items from different components. With these items a new ‘Just Culture’ component was constructed. Table 3 reports reliabilities before and after item exclusions.

In the next step all items excluded on the basis of the earlier reliability analysis were correlated with all the available components in CANSO scales 1-8. This suggested one item that was originally associated with ‘Flexible Culture’ to be moved to the new ‘Just Culture’ component and one original ‘Risk Perception’ item to be placed into the new combined component 4+5. It was not possible to group the items previously excluded based on their poor reliability into any of the proposed categories.

Table 3

Report on statistical Reliability of the revised eight CANSO Scales.

#	CANSO Component	No. of items	Initial Cronbach‘s α	Final Cronbach’s α
1	Informed Culture	5	.614	.771
2	Reporting Culture	6	.676	
3	Just Culture	6	.490	.757
4	Learning Culture	5	.577	.788
5	Flexible Culture	5	.585	
6	Attitudes towards Safety	5	.158	tbd
7	Safety-related Behavior	6	.702	.702
8	Risk Perception	5	.294	tbd
	Items excluded based on poor reliabilities	7	.300	tbd

Finally previously excluded items were correlated to components from related concepts such as ‘Resilience’, ‘OCB’ and ‘Recovery/ Stress’. Some significant correlations were found between five of these items with the ‘Resilience’, the ‘OCB’ and the ‘Recovery/ Stress’ scale. When looking at correlations between the CANSO components and related concepts the data showed significant correlations between ‘Just Culture’ and ‘Recovery’ (r=. 328 p<.05) and ‘Just Culture’ and ‘OCB’ (r=.406 p<.01). ‘Learning and Flexible Culture’ had some small correlations with ‘OCB’ (r=.294 p<.05). The results also indicated a negative relationship between ‘Stress’ and ‘OCB’ (r=-.380 p<.01). Finally some small correlations between ‘Resilience’ and ‘Recovery’ (r=.308 p<.05) and ‘Resilience’ and ‘OCB’ (r=.343 p<.05) were identified.

Discussion and Outlook

The initial validation of the eight scales (table 3) developed by CANSO (2009) demonstrated high reliabilities for component ‘7. Safety-related Behavior’ and the revised ‘3. Just Culture’ Component. Components ‘1. Informed Culture’ and ‘2. Reporting Culture’ and ‘4. Learning Culture’ and ‘5. Flexible Culture’ were merged together based on their inter-correlations suggesting that they may be measuring the same underlying construct. The results also indicated that seven questions were not reliably assessing safety culture and should therefore be removed from the ‘ready-to-go’ survey. Components ‘6. Attitudes towards Safety’ and ‘8. Risk Perception’ did not reach the criterion level for internal consistency. However, as current literature (Mearns & Flin, 1995, Mearns et al. 2009) suggests, as these items are correlated with safety culture, it is recommended to retain them for further validation in future developments of the instrument.

Next correlations between safety culture and related concepts were looked at. Data indicate some significant correlations between the CANSO scales and ‘Resilience’ scales, following Hollnagel’s (2006) concept of putting a positive safety culture into perspective with organizational resilience. However, in contrast to earlier work, no significant correlations between CANSO component ‘7. Safety-related Behavior’ and ‘OCB’ (Podsakoff et al., 2000) could be found. Moderate correlations between ‘3. Just Culture’ and ‘7. Safety-related Behavior’ indicate that these scales might be inter-related, which explains the significant correlations found between ‘3. Just Culture’ and ‘OCB’. The authors recommend reviewing the ‘OCB’ items used in terms of their reliability and amending these items in the instrument. Further items in the new ‘3. Just Culture’ component should be subject to another validation round. Fogarty (2005) demonstrated that ‘Stress’ is an important mediator of both organizational and environmental variables on the impact of safety climate on errors. No significant correlations between the CANSO safety culture components and ‘Stress’ could be found. One possible explanation is that the safety culture components proposed by Fogarty significantly differ from the CANSO components. Furthermore the study was undertaken in a different operational environment (aviation maintenance as opposed to ATM). It is therefore recommended to perform a gap analysis between the CANSO components and the components proposed by Fogarty and to look further into replicating his findings.

Further validation of the CANSO ‘ready-to-go’ survey and related concepts is planned in the course of the main study based on a larger sample size. The main study will also further investigate correlations between safety culture and related concepts such as resilience, OCB and stress. In the course of this research it is also planned to look into comparing results with other international ANSPs aiming at a standardized approach in assessing safety culture maturity in ATM.

Acknowledgements

Special thanks go to the local ANSP for funding the pilot study. Research reported is provided by the University of Graz based on the available data and does not reflect the views of the employers or granting organisation.

Selected References

CANSO (2010). Safety Excellence. The CANSO Standard. *Airspace Magazine*, Quarter 3. Retrieved from: http://www.canso.org/safety/documents

Fogarty, G. (2005). Psychological strain mediates the impact of safety climate on maintenance errors. *International Journal of Applied Aviation Studies* 5 (1), 53–63.

Hollnagel, E., Woods, D.D & Levenson, N. (2006). *Resilience Engineering. Concepts and Precepts*. Ashgate Publishing UK.

Kellmann, M. & Kallus, K.W. (2001). *Recovery-Stress Questionnaire for Athletes*. User manual. Champaign, IL: Human Kinetics.

Kallus K.W. (2010). *Erstellung von Fragebogen*. Facultas, Wien.

Kallus, K.W., Hoffmann, P. & Winkler, H. (2008). The concept of individual critical situations (SCS). In: *Proceedings of the 28th Conference of the European Association for Aviation Psychology*, 27-31 October 2008, Valencia, Spain.

Kallus, K.W., Barbarino, M. & VanDamme, D. (1998). *Integrated Task and Job Analysis of Air Traffic Controllers (HUM.ET1.STO1. 1000-REP-03).* Brüssel: Eurocontrol.

Mallak, L.A. (2006). *Resilience Instrument*. Engineering Management Research Laboratory, Western Michigan University, Kalamazoo, MI. One-time royalty-free.

Mearns, K., Kirwan, B. & Kennedy, R.J. (2009). Developing a Safety Culture Measurement Toolkit (SCMT) for European ANSPs. *Eighth USA/Europe Air Traffic Management Research and Development Seminar (ATM2009)*.

Podsakoff, P.M., MacKenzie, S.B., Paine J.B, and Bachrach, D.G. (2000). Organizational Citizenship Behaviors: A Critical Review of the Theoretical and Empirical Literature and Suggestions for Future Research. *Journal of Management*, 26(3), 513-563.

Wiegmann, D.A., Zhang, H., von Thaden, T.L., Sharma, G. and Mitchell, A. (2002) A Synthesis of Safety Culture and Safety Climate Research, Technical Report ARL-02-3/FAA-02-2, Institute of Aviation, Savoy(IL), United States.

Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions. *Accident Analysis and Prevention* 42 (2010) 1517–1522.

APPLICATION OF A HUMAN ERROR TAXONOMY FOR THE IDENTIFICATION OF AIR TRAFFIC CONTROL ERRORS AND CAUSAL FACTORS

Katherine Berry and Michael Sawyer
TASC, Inc.
Washington, DC USA

With the complexity of the National Airspace System, a comprehensive taxonomy is needed to classify incident causal factors within the air traffic control (ATC) industry. These causal factors must not be limited to the individual involved in the incident, but must extend to the role of management and the organization. The Human Factors Analysis and Classification System (HFACS) is a human error taxonomy when integrated into the incident investigation process permits for causal factors to be classified at various system levels. Utilizing HFACS, this study investigates human performance within the ATC industry and classifies ATC incidents cases using HFACS, investigates the effects of on-the-job training, and further examines the causal factor networks present. These findings can be utilized to better understand the system-wide impacts on ATC incidents and assist in targeting mitigations towards latent conditions, which have the potential for a greater impact.

The Federal Aviation Administration is currently engaged in an effort to modernize the National Airspace System. While the stated goals of the modernization includes increasing safety, the magnitude and nature of the proposed changes create the opportunity for the introduction of a variety of new active and latent error modes not present in the current system (Zemrowski & Sawyer, 2010). The ability to identify errors relevant to controllers is essential to the safety of the National Airspace System. An integrated taxonomy is needed to ensure that these error modes and their casual factors can be identified to allow for detailed analysis. Such a taxonomy would allow an analyst to make comparisons between different sets of conditions in order to determine the different types of causal factors present. For example, the presence of on-the-job training could affect the types and causes of errors occurring in ATC incidents, and an integrated error taxonomy would allow an analyst to identify the specific issues present during on-the-job training.

Human Factors Analysis and Classification System

Arising from a need for a common accident investigation taxonomy, the HFACS taxonomy was modeled on Reason's (1990) Swiss cheese model of active and latent conditions (Wiegmann & Shappell, 2003). Initially designed for aviation and in particular, the flight deck environment, the HFACS taxonomy (Figure 1) consists of one tier of active errors – unsafe acts – and three tiers of latent conditions – preconditions for unsafe acts, unsafe supervision, and organizational influence. The taxonomy provides a methodological approach for investigating both accidents and near miss incidents (for more information, see Wiegmann & Shappell, 2003).

Due to its origins, the HFACS taxonomy has been applied to the many facets of the aviation industry, including commercial (Wiegmann & Shappell, 2001), military (Li & Harris, 2006), and general aviation (Shappell & Wiegmann, 2004). Additionally, the application of the taxonomy has extended beyond the aviation industry to include maintenance (Berry, Stringfellow, & Shappell, 2010), mining (Patterson, 2009), and rail (Baysari et al., 2008). The HFACS taxonomy was also retrospectively applied to the ATC industry to assess operational errors attributed to controllers (Scarborough, Bailey, & Pounds, 2005).The sub-section of the operation error form detailing controller actions was mapped to HFACS causal categories. Since the portion of the operational error form being analyzed only examined the acts of the controller, only HFACS causal categories at the unsafe acts tier – skill-based errors and decision errors – were identified. It should come as no surprise that as the ATC operational errors mapped only to HFACS errors the study lacked the necessary detail to fully examine human performance. Therefore, the present study presents a more robust examination of ATC human performance by extending beyond the unsafe acts tier to examine latent failures throughout the entirety of the HFACS taxonomy.

Causal Factor Associations

Typically, errors violations are not random events, but can be attributed to a combination of causes and contributing factors (Senders & Moray, 1991). The combination of causes interact in unique and varying ways allowing for Reason (1990) to establish the existence of interactions between active error and latent conditions. It is important to note that the identification of causal factors does not equate to the identification of a single, absolute cause, but rather to the identification of causal chains or networks (Senders & Moray, 1991). The incident investigator or safety manager should identify not only active errors, but should expand their investigation to incorporate these causal networks. Doing so would permit for the ability to target interventions and even to predict when and under what circumstances future accidents may occur (Alkov, 1997).

In an effort to extend beyond the traditional frequency-based analysis, many accident and incident causation studies have investigated the associations and linkages between contributing factors and errors. A more recent use of the HFACS taxonomy has been to explore error pathways, or the relationship among HFACS causal factors. In aviation, linkages have been identified among the crew resource management category and both skill-based error and decision error causal categories have been established (Li & Harris, 2006; Li, Harris, & Yu, 2008). Similar pathways were also exhibited in a non-aviation analysis (Berry, Stringfellow, and Shappell, 2010). Examining the associations among HFACS causal categories is still at the beginning stages of research and has yet to be expanded to the ATC industry.

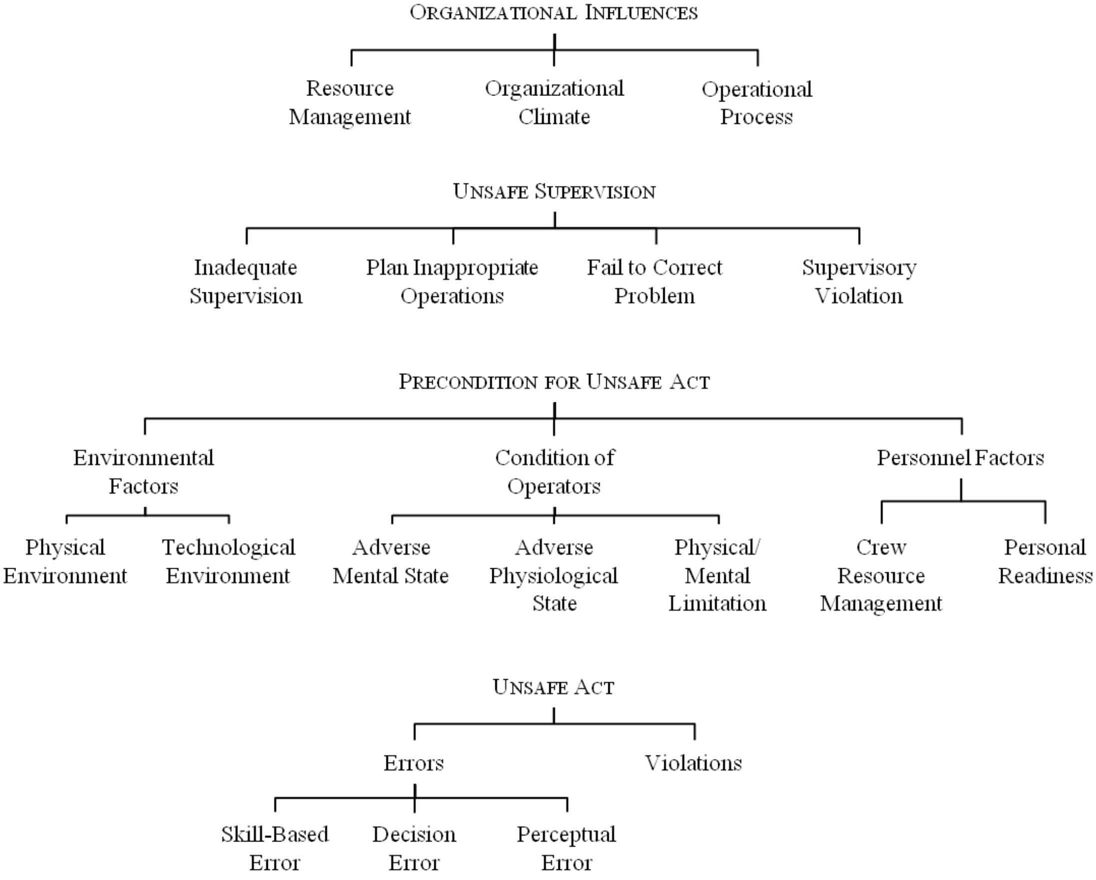


Figure 1. *HFACS Taxonomy (Wiegmann & Shappell, 2003)*

Purpose

The purpose of this study is three-fold. First, the traditionally HFACS analysis will be expanded to the ATC industry to identify HFACS causal categories. Unlike previous analyses, this study will not be limited to the unsafe act tier, but will incorporate the HFACS taxonomy as a whole and will identify causal categories at all four tiers. Second, the study examines the effects of on-the-job training in regards to incident data and causal factors. Finally, this analysis will identify associations between active errors and latent conditions. Ideally, association analyses would permit for the shifting targeted mitigation areas from active errors to latent conditions.

Methodology

This analysis utilized data from NASA's Aviation Safety Reporting System (ASRS) database. Spanning 2007 through 2009, 1,324 ASRS cases that identified ATC as a factor were collected. The cases represented a variety of controller types and positions. 35.60% of the cases involved ARTCC controllers, 35.91% of the cases involved TRACON controllers, 27.86% of cases involved tower controllers, and 0.63% of cases involved the flight service station. A causal factor analysis was then performed using the HFACS taxonomy. The ASRS cases were classified using HFACS by one of two methods – independent coders or consensus (Berry, Stringfellow, and Shappell, 2010). The first method involved multiple analysts and a quality assurance individual. At least two analysts independently classified a case and if a disagreement existed, an additional analyst performed a quality assurance check to solve the disagreement (Shappell, et al. 2007). A secondary method of consensus coding was adapted for efficiency purposes. This method required multiple analysts to review a case and come to a consensus or agreement on the causal factors contributing to the case. This technique does not require a quality assurance individual to moderate the process.

Statistical Analysis

Each ASRS case was evaluated across all four HFACS tiers, and the presence or absence of each HFACS causal factor was recorded. It is important to note that the HFACS categories are not mutually exclusive. For example, an individual case can include a skill-based error, violation, and adverse mental state. For each HFACS causal factor, the percentage of ATC cases containing at least one instance of the causal factor was calculated. Additionally since ATC utilizes on-the-job training frequently, the effect of on-the-job training on HFACS causal factors was examined. If on-the-job training was occurring during the time of the case, it was noted. Cross tabulations of the various HFACS causal factors and the presence or absence of on-the-job training were created and displayed in contingency tables. In order to estimate the statistical strength of the effects of on-the-job training, a Pearson's Chi-Square analysis for independence was performed and the impact of that effect was determined utilizing a relative risk analysis. The relative risk value was assessed in a hypothesis test, where the null hypothesis was the relative risk value is equal to one, which is considered to be neutral, and the alternative hypothesis was that the value does not equal one (Sheskin, 2004).

The HFACS causal factor association analysis was conducted under the assumption that causal factors within both adjacent and non-adjacent tiers may be associated. For example, associations among unsafe acts causal factors were not limited to only the preconditions for unsafe acts causal factors, but were extended to both the unsafe supervision and organizational influence causal factors as well (Berry, Stringfellow, & Shappell, 2010). Cross tabulations of HFACS causal factor pairings were created and displayed in contingency tables. A Pearson's Chi-Square analysis for independence was performed. While the Pearson's Chi-Square test determined the significance of an association, the direction of the associations cannot be determined from the test. If a pair of causal factor was found significant ($p \leq 0.001$), the odds ratio was calculated. Since the HFACS classification incorporates casual factors at various levels, an asymmetrical measure of association strength is also necessary. In addition to odds ratio, the relative risk was also calculated. Relative risk, in addition to be asymmetrical and directional, also has the benefit of being more intuitive than odds ratios (Sheskin, 2004). Due to the asymmetrical nature, two different relative risks were calculated – 1) relative risk of the higher tier given the lower tier occurs and 2) relative risk of the lower tier given the higher tiers occurs. Finally, the odds ratio value and both relative risk values were assessed in hypothesis test.

Results and Discussion

The results from the ATC HFACS causal factor analysis can be viewed in Table1. The percentages in Table 1 do not add up to 100% since incidents typically are classified with more than one causal factor. The results of the causal factor analysis can be interpreted as 43.49% of ATC cases had at least one occurrence of a skill-based error causal factor. Many of the causal factors findings are similar to HFACS findings in other industries (Berry, Stringfellow, & Shappell, 2009). Of interesting note are those causal factors (skill-based error, decision error, technological environment, adverse mental state, and crew resource management) with high percentage of cases values. Many of the skill-based error cases can be attributed to communication errors, which occurred in 17.67% of cases. Much of the ATC position involves communication with various actors and therefore, it is not a surprise that the communication error causal factor is a main contributor. Many of the decision error cases can be attributed to knowledge errors, which occurred in 17.07% of cases.

The ATC job involves varying software and technology, which explains technological environment being a large contributor in the causal factor analysis. Many of the technological environment cases can be attributed to checklist/procedure design (8.08% of cases) and communication equipment /tools (6.19% of cases). Additionally, controllers are required to communicate, coordinate, and dynamically plan with other controllers, pilots, dispatchers, and other actors explaining why crew resource management (CRM) is a main contributor to the causal factor analysis.

Table 1. *ATC HFACS Causal Factor Analysis – Percentage of Cases.*

HFACS Causal Factor	**ATC (n=1324)**	**HFACS Causal Factor**	**ATC (n=1324)**
Unsafe Acts		*Unsafe Supervision*	
Skill-Based Error	43.96%	Inadequate Supervision	6.65%
Decision Error	26.81%	Plan Inappropriate Operations	5.36%
Perceptual Error	3.25%	Fail to Correct Problem	6.80%
Violation	0.91%	Supervisory Violation	0.45%
Precondition for Unsafe Act		*Organizational Influences*	
Physical Environment	5.29%	Resource Management	3.17%
Technological Environment	24.09%	Organizational Climate	1.21%
Adverse Mental State	33.16%	Operational Process	6.12%
Adverse Physiological State	1.51%		
Physical/Mental Limitation	4.68%		
Crew Resource Management	31.80%		
Personal Readiness	1.06%		

On-the-Job Training

In a similar manner to CRM, on-the-job training plays an important role in ATC. At the beginning stages of their career as a developmental, controllers participate in on-the-job training to initially reach an acceptable level of proficiency, and once that same controller becomes a certified professional controller (CPC), the controller will mentor a developmental by administering on-the-job training. To assess its effects, the presence or absence of training being conducted during the time of the incident was identified. As shown in Table 2, three HFACS causal factors resulted in significant Chi-Square values ($p \leq 0.001$) and significant relative risk values ($p \leq 0.01$).The relative risk values indicate that for ATC incident cases, the likelihood of a decision error occurring while training is being conducted is 2.35 times higher than when training is not being conducted.

Table 2. *HFACS Causal Category Relative Risk Values for On-the-Job Training*

HFACS Causal Category	**Relative Risk**
Decision Error	2.35
Physical/Mental Limitation	9.23
Crew Resource Management	2.64

During many training ATC incidents, the CPC would partially transfer the responsibility to the developmental by allowing the developmental to actively control traffic while the CPC monitors the traffic and the developmental's actions. In some instances, an operational error would occur when the developmental was actively controlling traffic and the CPC failed to take over the position in a timely manner to prevent the operation error. The CPC would acknowledge the poor choice (decision error), but identified the event as a learning experience for both the developmental and the CPC. In other incidents, the developmental would be in a situation where his or her current knowledge base did not have the necessary information to adequately perform in a unique situation (physical/mental limitation – knowledge limitation). Also, since the developmental and CPC would be working as a team controlling traffic, it is not surprising that crew resource management causal factors would be present.

HFACS Associations

The HFACS causal factor category association findings can be found in Table 3. Only those causal factor pairings that were found significant from the Chi-Square analysis ($p \leq 0.001$) were reported. The relative risk of the higher tier indicates that an ATC incident case with a decision error is 4.32 times more likely to have a physical/mental limitation than an incident case that does not have a decision error. The relative risk of the lower tier indicates that an ATC incident case with a resource management causal factor is 5.00 times more likely to have a plan inappropriate operation causal factor. Many causal factor pairs were identified and reported a significant Chi-Square value ($p \leq 0.001$). Additionally, it should be noted that all association findings reported in Table 3 also have an odds ratio that was determined to be statistically significant ($p \leq 0.01$) indicating a non-neutral odds ratio. However, many relative risk values did not succeed in rejecting the null hypothesis. Most of the associations incorporated either a precondition for unsafe act causal factor and/or an unsafe act causal factor, and this may be attributed to the nature of these two tiers. The incident factors correlated with these lower HFACS tier causal factors are more readily and easily identifiable than causal factors at the higher two tiers.

Table 3. *ATC HFACS Associations Findings*

HFACS Causal Categories	Pearson's Chi Square*	Odds Ratio Value	Odds Ratio p-value	Relative Risk Higher Tier Value	Relative Risk Higher Tier p-value	Relative Risk Lower Tier Value	Relative Risk Lower Tier p-value
HFACS Tier 4 - Organizational Influences							
Resource Management X Adverse Mental State	28.67	5.34	**	5.04	**	2.24	
Resource Management X Personal Readiness	134.18	50.04	**	22.02	**	40.70	**
Resource Management X Plan Inappropriate Operation	29.08	6.26	**	5.51	**	5.00	**
Organizational Climate X Plan Inappropriate Operation	21.39	8.55	**	8.02	**	6.19	**
Organizational Climate X Physical Environment	21.80	8.69	**	8.14	**	6.29	**
Organizational Climate X Adverse Physiological State	32.35	17.52	**	15.05	**	14.43	**
Organizational Climate X Personal Readiness	48.46	27.21	**	21.59	**	22.30	**
Operational Process X Technological Environment	11.26	0.28	**	0.30	**	0.34	**
HFACS Tier 3 - Unsafe Supervision							
Inadequate Supervision X Skill-Based Error	15.45	0.38	**	0.40	**	0.53	
Plan Inappropriate Operation X Adverse Mental State	14.04	2.45	**	2.32	**	1.67	
Fail to Correct Problem X Skill-Based Error	22.50	0.30	**	0.32	**	0.44	**
Fail to Correct Problem X Technological Environment	15.29	2.36	**	2.20	**	1.80	**
HFACS Tier 2 - Precondition for Unsafe Act							
Technological Environment X Skill-Based Error	62.84	0.33	**	0.42	**	0.49	**
Technological Environment X Decision Error	28.09	0.41	**	0.50	**	0.50	**
Adverse Mental State X Skill-Based Error	23.28	1.76	**	1.46	**	1.36	**
Physical/Mental Limitation X Skill-Based Error	15.98	0.29	**	0.31	**	0.43	**
Physical/Mental Limitation X Decision Error	39.40	4.72	**	4.32	**	2.44	**
Crew Resource Management X Skill-Based Error	17.38	0.60	**	0.71	**	0.74	
Crew Resource Management X Decision Error	11.20	1.54	**	1.33		1.36	

* indicates significant at alpha of 0.001 ** indicates significant at alpha of 0.01

The skill-based error causal category produced in six significant association pairings. However, all but one of the pairings resulted in either a non-significant relative risk value or a relative risk value less than one. This can be attributed to the large number of occurrences of skill-based errors. As seen in Table 1, the skill-based error causal category is the most widely reported and identified HFACS causal category for ATC incident cases and is therefore associated with a plethora of causal categories. The only skill-based error pairing with a significant relative risk values greater than one is that pairing with the adverse mental state causal category. Since the adverse mental state causal category is the second most reported and identified HFACS causal category for ATC incidents, this pairing should be taken into consideration and further examined. Of interesting note is the higher number of organizational influence associations in comparison with previous findings (Berry, Stringfellow, & Shappell, 2010; Li & Harris, 2006; Li, Harris, &Yu, 2008). In air traffic control, the organization or FAA is easily identified and its role and responsibilities more clearly defined than other industries (e.g. general aviation). The organizational climate HFACS casual category incorporates safety culture causal factors and could explain many of the associations with that particular causal category. For example, if safety culture is a causal factor, a controller may not regard safety-related personal readiness causal factors (e.g. adhering to rest requirements, reporting to duty in a manner fit to control traffic) as important.

Seldom are incidents an outcome of one single event or unsafe act, but are culminations of various factors including, but not limited to environmental, task-related, situational, and organizational factors. Since latent conditions have the ability to remain passive in a system, the unpredictable combinations of latent conditions and active errors result in continuing and more hazardous gaps (Reason, 1990). These findings can be utilized in targeting interventions and mitigations for current operations and can help to guide safety considerations and assessments for future operations. Also, interventions targeted towards latent conditions (organizational influence tier, unsafe supervision tier, and preconditions for unsafe act tier) rather than active errors (unsafe acts tier) have the potential to result in the greater gain and have the greater impact. Safety managers should take into consideration both the ATC HFACS findings (Table 1) and the associations findings (Table 3) when determining areas for improvement. Not only should the most frequency causal category mitigated, but the causal category associations should be accounted for in interventions. For example, if a mitigation is targeted towards the resource management causal category, the potential exists to impact not only the resource management causal category, but also the plan inappropriate operations, personal readiness, and adverse mental state causal category.

Conclusion

In the present study, 1,324 ATC incident cases were collected and classified using the HFACS taxonomy. The effects of on-the-job training were also examined, and three HFACS causal categories – decision error, physical/mental limitation, and crew resource management – were found to be impacted by on-the-job training. Associations between HFACS causal categories were reported, and significant causal factor pairings emerged from data. Findings from the HFACS classification and the association findings should be utilized when determining mitigations targeted towards latent conditions. Future work in this area should further examine the organizational influences causal categories in more detail.

References

Alkov, R. A. (1997). Human error. *Aviation safety: The human factor* (pp. 75-87). Casper, WY: Endeavor Books.

Baysari, M. T., McIntosh, A. S., & Wilson, J. R. (2008). Understanding the human factors contribution to railway accidents and incidents in Australia. *Accident Analysis and Prevention*, 40(5), 1750-1757.

Berry, K.A., Stringfellow, P.F., & Shappell, S.A. (2009). Considering trends among industrial accidents: a preliminary meta-analysis of HFACS causal factors across industries. In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Berry, K.A., Stringfellow, P.F., & Shappell, S.A. (2010). Examining error pathways: an analysis of contributing factors using HFACS in non-aviation industries. In *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Li, W. C., & Harris, D. (2006). Pilot error and its relationship with higher organizational levels: HFACS analysis of 523 accidents. *Aviation, Space, and Environmental Medicine*, 77(10), 1056-1061.

Li, W. C., Harris, D., & Yu, C. S. (2008). Routes to failure: Analysis of 41 civil aviation accidents from the republic of china using the human factors analysis and classification system. *Accident Analysis and Prevention*, 40(2), 426-434.

Patterson, J. M. (2009). *Human error in mining: A multivariable analysis of mining accidents/incidents in Queensland, Australia and the United States of America using the human factors analysis and classification system framework*. (Doctor of Philosophy, Clemson University). (Clemson University Electronic Theses and Dissertations)

Reason, J. (1990). *Human error*. New York: Cambridge University Press.

Scarborough, A., Bailey, L., & Pounds, J. (2005). *Examining ATC operational errors using the human factors analysis and classification system* (No.DOT/FAA/AM-05/25) Federal Aviation Administration, Office of Aerospace Medicine.

Senders, J. W., & Moray, N. P. (1991). *Human error: Cause, prediction, and reduction*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Shappell, S. A., & Wiegmann, D. A. (2004). HFACS analysis of military and civilian aviation accidents: A North American comparison. In *International Society of Air Safety Investigators*. Queensland.

Shappell, S. A., Detwiler, C., Holcomb, K., Hackworth, C., Boquet, A., & Wiegmann, D. A. (2007). Human error and commercial aviation accidents: An analysis using the human factors analysis and classification system. *Human Factors, 49*(2), 227-242.

Sheskin, D. J. (2004). Handbook of parametric and nonparametric statistical procedures (3rd Ed.). Boca Raton: Chapman & Hall.

Wiegmann, D. A., & Shappell, S. A. (2003). *A human error approach to aviation accident analysis: The human factors analysis and classification system*. Burlington, VT: Ashgate Publishing, Ltd.

Wiegmann, D. A., & Shappell, S. A. (2001). Human error analysis of commercial aviation accidents: Application of the human factors analysis and classification system (HFACS). *Aviation, Space, and Environmental Medicine*, 72(11), 1006-1016.

Zemrowski, K.M, & Sawyer, M.W. (2010). Impacts of increasing reliance on automation in air traffic control system. *Journal of Air Traffic Control*, 52(4).

The Effects of a Net-Centric Communication Tool on Communication Monitoring and Threat Detection

Kelly Satterfield[2], Victor Finomore[1], & Courtney Castle[2]
Air Force Research Laboratory, Wright-Patterson AFB[1];
Oak Ridge Institute for Science and Education, Wright-Patterson AFB[2];

Command and Control (C2) operators rely heavily on radio and chat communication as well as tactical displays to efficiently plan, direct, coordinate, and control assets. The integration of information for the C2 operator is critical for mission success. This environment, which is communication intensive, imposes a high degree of workload on operators thus resulting in failures of detection or comprehension of critical messages. Multi-Modal Communication (MMC) is a net-centric communication management suite with advanced tools to better equip the warfighter in managing their communication. This study examined performance associated with monitoring communication channels, while also monitoring a dynamic tactical visual display for potential threat of enemy fire. Operators monitored and responded to the occurrence of critical signals presented during the 12-min monitoring task. Performance was analyzed in regard to signal detection for both tasks while response accuracy and time were collected for the communication task. An overall detection score reflected a combined score of the number of critical messages and threats detected. Data showed that with the MMC tools operators detected more critical events in the communication and threat detection tasks as compared to a standard radio. Perceived mental workload as measured by the NASA-TLX was also rated as lower in the MMC than radio condition. MMC can be a beneficial communication tool in its ability to aid C2 operators in communication management while allowing for greater performance on the monitoring of their tactical display.

DESIGN AND EVALUATION OF A CONTROLLER-PILOT DATA LINK COMMUNICATIONS (CPDLC) INTERFACE

John Fugedy*
Emmanuel Letsu-Dake, PhD
Honeywell International
Golden Valley, MN
Dave Pepitone, PhD*
Joe Rakolta*
Craig Schimmel*
*Honeywell International
Albuquerque, NM

Data link communications will, for the most part, replace voice communications between aircraft and air traffic control (ATC) systems in Next Generation Air Transportation System (NextGen). One goal in implementing data link communications is to establish Controller Pilot Data Link Communications (CPDLC) procedures, logic and page designs with existing equipment that will maximize total system-pilot efficiency. A new message menu tree index was developed to provide quick and easy access to all CPDLC messages (uplinks and downlinks as defined by RTCA SC-214). Page formats were also developed to minimize errors and enhance the efficiency of pilots in data link communications. Four pilots evaluated the MCDU datalink prototype under realistic flight conditions. Subjective and objective measures from the evaluation suggests that using existing flight deck technology for data link operations is feasible but with some limitations. Workload when using the data link system was reasonable. However, the pilots focused on performing the data link tasks at the expense of monitoring aircraft flight path.

Background

Controller Pilot Data Link Communications (CPDLC) is aimed at improving safety and efficiency by taking over many tasks now accomplished through voice communications. While voice communications has many benefits, it is also prone to certain disadvantages such as frequency congestion, missed calls, misunderstood clearances or communications, etc. Ultimately, data link communications will be the primary means of communication between the flight crew and the Air Navigation Service Provider (ANSP) for airspaces that require data communications capability for clearances and 4-dimensional trajectory amendments. For these aircraft, voice communications between the flight deck and ANSP will be used only for extraordinary purposes (JPDO, 2007). Implementing data link communications especially through the establishment of CPDLC infrastructure that takes advantage of existing equipage can introduce human factors issues such as excessive heads down time, entry error and recovery, excessive task time with finding, composing and sending messages and standard operating procedures.

Previous research shows that textual data link modality results in higher workload and increased interaction time compared to other candidate forms e.g. synthesized speech, digitized speech and text/synthesized speech (Lancaster, 2008). Data link communications in different flight phases will also have different workload and situation awareness demands and challenges. Additionally, data-link communication could lead to increased head-down time and attention issues. Pilots will also be required to monitor multiple data link transmissions at a time requiring pilot attention. Increased head down time implies decreased time looking outside the cockpit and the time available to properly monitor other flight deck displays. This is generally detrimental to flight safety (Wiener and Nagel, 1988).

CPDLC design should consider fundamental human limitations (e.g., memory, computation, attention, decision-making biases, and task timesharing) which should not be exceeded for effective use of the system. For example, in air traffic control (ATC) communication clearance negotiations, the current method of using voice to mediate ATC communications allows for flexible and elastic negotiation of the clearances. It also allows a closed loop verification of ATC or pilot intent. CPDLC may force crews to find message strings that are not inherently intuitive or represent the routine way or method of communicating. Searching, composing and sending datalink messages may place excessive memory demands on the pilot because of the inherent functional cognitive differences between ‘hearing and responding’ and ‘reading and responding’ to datalink message sets. Datalink interfaces that do not consider these limitations run a higher probability of increasing task time, workload, heads down time and errors. These human factors issues which have been identified in previous research and literature call for an efficient use of human factors design to develop CPDLC flight deck interfaces that meet the demands of NextGen requirements and accommodate pilots and industry’s needs for compatibility with existing flight deck displays, controls and procedures.

This paper outlines a pilot-in-the-loop part-task human factors evaluation that was part of an iterative design process to develop a CPDLC display based on human factors issues. The evaluation was directed at scrutinizing pilot performance under a broad range of operational conditions (to ascertain operational suitability).

Human-Centered Design

The pilot-in-the-loop human factors evaluation was part of an iterative design process aimed at establishing the functionality and human factors issues relating to the new CPDLC design. The primary human factors considerations for the design were: crew acceptability, perception and information processing enhancement, and interface design features. Using human-centered design principles, a message menu tree index was developed to provide quick and easy access to over one hundred and fifty standard data link messages currently defined (RTCA, 2009). Page formats were also developed to minimize errors and enhance the efficiency of pilots in data link communications. The design philosophies were focused on the goal of allowing the pilot quick, simple and accurate data link communication with ATC. The menu system incorporates human factors design philosophies that have been researched and proven valid for effective pilot-controller communications.

System Description. The menu system accommodates message element standards envisioned for NextGen operations (RTCA, 2009). The design covers over one hundred and fifty uplink and downlink RTCA SC-214 pre-defined standard message elements which are structured so that finding and sending the intended messages can be accomplished in a timely and accurate manner (RTCA, 2009). Previous designs (EUROCONTROL, 2009), industry standards (RTCA, 2000) airline standard operating procedures (UAL, 2009), as well as practical pilot experience were all considered in developing the design concept. The menu system design consists of the menu pages and the menu logic. The menu pages are designed using current 24x14 character Multi-Function Cockpit Display Unit (MCDU) page formats. The pilot is made aware of his location in the page tree index by displayed navigation options to go to any page. The prime initiative for the menu page organization is to reduce the number of menu navigation levels to locate a required page. Design of the menu organization and control logic was based on two considerations: (1) frequency of use and (2) match the way a pilot would typically prioritize/organize messages in categories. The SC-214 Message sets were divided into two logical groups: Message Type and Message Category. The message type and menu category logic are shown in Figure 1. This grouping approach minimized page clutter by organizing screen contents and eliminating unnecessary text to create a clean and refined design.

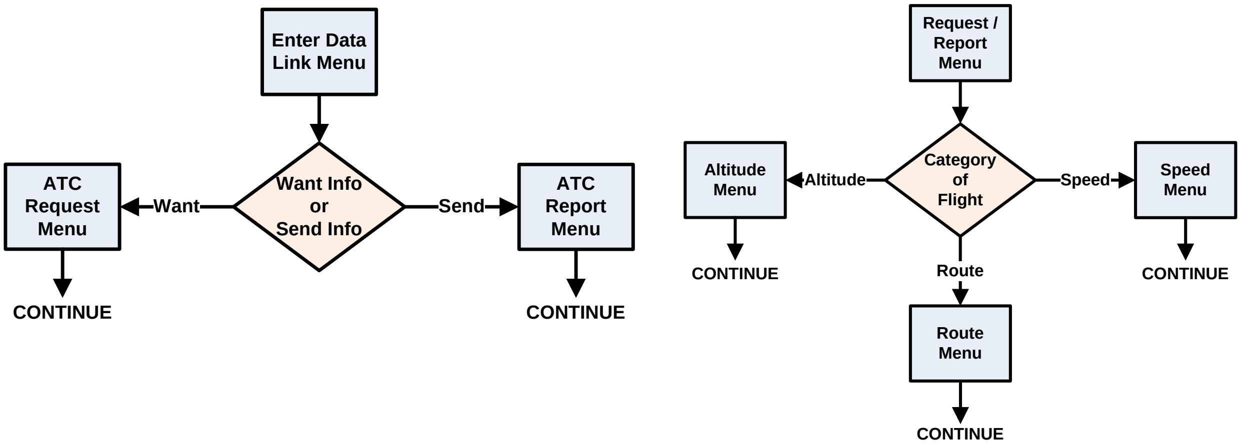


Figure 1. Data link message type and menu category logic.

An embedded help feature was implemented as a quick way for the pilot to “jog” his memory for the appropriate syntax to be used for data input. The embedded help feature that allows the display of proper input syntax (for a particular SC-214 message) while remaining on the same menu page. Based on analysis, it was found that the majority of datalink clearances were heading, speed, altitude or DIRECT TO clearances. Because a pilot would spend the majority of time responding such messages, an MCDU page was designed to handle this communication traffic. Thus a pilot could handle the majority of ATC communications using to a single page, thus mitigating a lot of the antecedent issues contributing to total task or heads down time. This single menu page enables the pilot make quick, easy and accurate routine requests. In addition, the menu system visually and aurally alerts the pilot of new uplink data link messages and displays the number of active messages in the queue on each menu page. This method of alerting the pilot to a new message de-conflicts the use of the scratchpad as an ATC alerting area.

Evaluation

Method

Simulation Facility. A part task crew station was used to evaluate the CPDLC system. Each pilot had a ‘glass’ large screen Primary Flight Display (PFD) and shared Multi-Functional Display (MFD) and a center console with throttles, flaps, speed brake, gear handle, radios, and other controls to control the simulation. The simulation had an outside moving visual scene. The simulation was controlled with various parameters such as initial position, weather, pause and simulation acceleration factors.

The MCDU was simulated using a stand-alone Tablet PC computer which was networked with a desktop computer that simulated the ATC controller. The MCDU was mounted just adjacent to the throttle quadrant on the center console. As closely as possible, the reach envelope was approximately the reach required in a 737 series aircraft. The ATC station was a PC station located behind the cockpit station and oriented so that the ATC controller could view subject pilot actions with the MCDU. The ATC station was capable of sending and receiving datalink communications with the MCDU. Although data link communications shared the same MCDU as the Flight Management System (FMS), pilots were not required to modify the FMS route during the experiment.

Scenario. To evaluate the HMI, a simulated real time flight was planned from Albuquerque, NM to Phoenix, AZ. Subject pilots evaluated a prototype of the Honeywell CPDLC menu design under realistic flight conditions in a flight simulator. The test set-up as previously described included a part-task crew station and an ATC station for sending and receiving ATC messages. The scenario tested data link communications from approximately 10,000 ft. out of Albuquerque to approximately 10,000 ft. inside the Phoenix TRACON area as shown in Figure 2. This figure shows the different points in the experiment were complex data link communication exchanges are simulated.

Experimental Design. A within-subjects design was used for the study. All subjects completed the same scenario conditions. The experiment assessed (i) the CPDLC interface design; (ii) pilot errors and error recovery; (iii) pilot task completion data; (iv) pilot workload measures; and (v) unprompted pilot comments. Prior to designing the test scenario research questions were formulated based on the human factors issues identified so that scenario elements could be designed to elicit data around each research issue. The scenario played a fundamental role in ‘steering’ the pilot into situations or events that could be measured. However, the experiment design also needed to balance the need for data against introducing artificial workload factors (e.g. unrealistic communications or too many communications for that phase of flight). Subjects were provided with a training scenario to familiarize them with the simulator and CPDLC system operation. Pilots were furnished with Standard Operating Procedures and Checklist for the scenario.

Subjects. Four professional pilots were recruited for the study. The pilots had between 800 and 10000 total hours of flying experience (median = 2850 hours). The median age of the pilots was 44 years (range= 28-67 years). Participation in the study was voluntary and pilots were not paid. The experiment time ranged between 60 and 80 minutes.

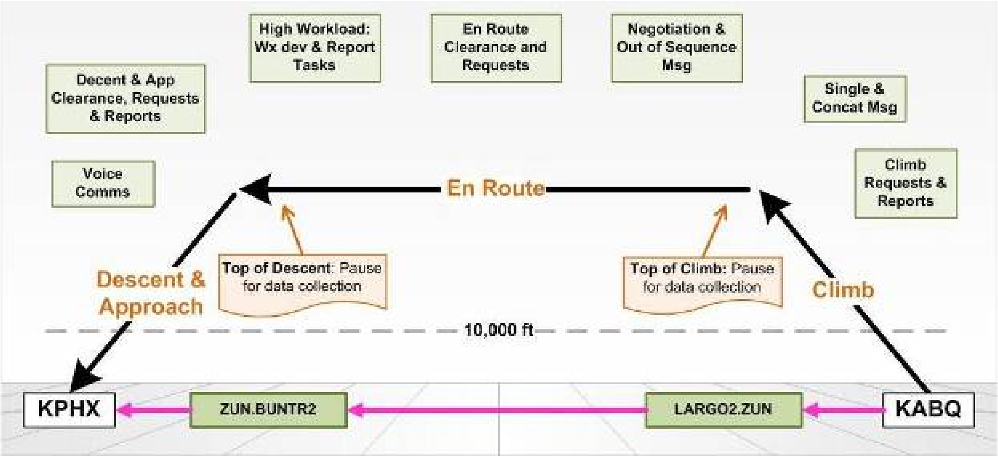


Figure 2. Simulation evaluation scenario flight profile.

Results and Discussion

Acceptability. Subjective ratings by subject pilots confirm the overall acceptability and usability of the page format and page navigation scheme and logic. Three of four pilots rated the CPDLC system as GOOD (4 on a 5-point scale). The remaining pilot rated the CPDLC system as FAIR. Pilot comments also indicated that the CPDLC system will be highly acceptable to pilots. Pilots rated eleven interface usability attributes (for example comprehension, readability, clutter, page labels, message formats and navigation) of the menu design on a dichotomous scale (ACCEPTABLE/UNACCEPTABLE). There was a unanimous ACCEPTABLE rating for eight of eleven evaluation design attributes. Two UNACCEPTABLE ratings were due to specific issues identified by the pilots. These issues were: (i) data format for the “fuel remaining” parameter, and (2) meaning of “NEXT DATA AUTHORITY” downlink message. There was no consistently repeated design comment that was critical to the functionality or operational design of the CPDLC Menu system. Supporting pilot comments suggest that the system is usable and is acceptable to the representative group of users. Pilots also reported that the CPDLC menu structure and page formatting facilitated error detection and enabled pilots to recover from a majority of their errors.

Errors. Error rate percentages were calculated as the ratio of the number of errors of a particular category to the number of “opportunities for error” for the scenario. The overall error rate (mean error rate for all categories of error) was 8.7%. The highest category of error was missed altitude callouts (14%). This is because there was direct visual competition between the datalink and other visual monitoring tasks. The relatively high observed altitude missed call out error rate suggests that datalink communications may impact current crew procedures such as monitoring aircraft flight path, navigation or systems. This is most likely due to the amount of head’s down visual attention commanded by the MCDU position which is exacerbated by increased task time caused by excessive search time and message complexity (number of required keystrokes).All non-callout errors were detected and corrected. Pilot recovery from all non-callout errors means that the CPDLC design adopted for the experiment was effective in helping pilots recover from errors.

Workload. Subject pilots were asked at specific points in the scenario to provide a workload rating number (instantaneous self-assessment) that best describes their workload on a five-point scale ranging from 1-VERY LOW to 5-VERY HIGH. Half scale units were acceptable (e.g. 2.5). Figure 3, is a plot of reported workload for each phase of flight: Climb, Enroute and Descent. Although the majority of workload ratings were LOW-MEDIUM (2-3 rating), all pilots during debriefing mentioned some loss of situation awareness during the flight. Pilots did not focus attention on navigation, flight path or out of the window view due to the visual attention required for the CPDLC tasks. Both the median and inter-quartile ranges for each phase of flight show that the workload ratings incremented during the flight. There is no statistically significant difference in workload ratings between the three phases of flight (p = 0.24). The results of the self assessment workload ratings suggest that the there is no significant high workload effect on pilot workload as a result of using datalink.

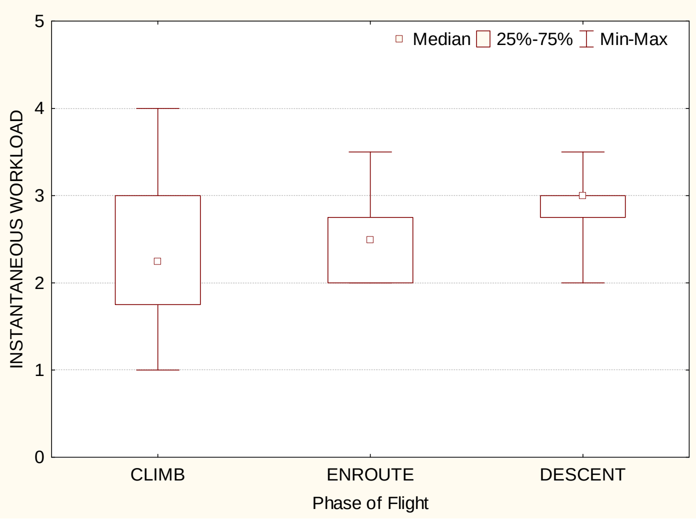


Figure 3. Box-Plot of subject pilot's impression of workload during Climb, Enroute and Descent.

Pilot Transaction Times (Head Down Time). The head down time for each phase of flight is an approximate percentage of visual attention the subject pilot focuses on the CPDLC menu system for both uplink and downlink messages. The percentage of time spent head-down for each phase of flight is calculated as the ratio of the cumulative sum of downlink and uplink message times for a phase of light to the total flight time for the phase of flight. The rationale for this calculation is that downlink and uplink message time give an approximation of the head down time on the MCDU. The percentage of head-down time for each phase of flight is shown in Figure 4. There is no statistically significant difference in head down time between the different phases of flight. The mean head down percentage time for all phases is 45% (range 30-76%). This represents the percentage of total flight time during which subject pilots allocated visual attention to the MCDU for CPDLC related tasks. For the scenario, pilots spent approximately 45% of the flight time visually focused on the MCDU. This leaves the remaining 55% of the time available for other required flight desk tasks.

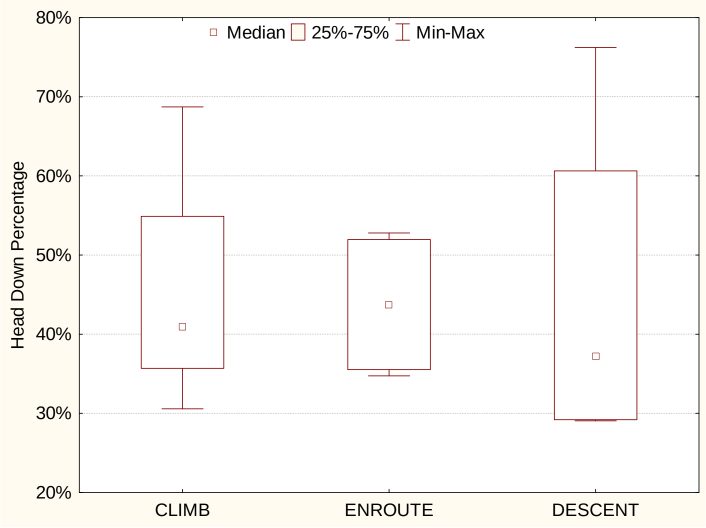


Figure 4. Box-Plot of calculated percentage of pilot head- own time for Climb, Enroute and Descent Flight phases.

Conclusions

Subjective and objective measures from the evaluation suggest that using existing flight deck displays for data link operations is feasible.

- All four subject pilots thought that the operation of the CPDLC system was easy to learn and easy to operate. In general, the subject pilots rated the CPDLC interface and menu design features to be intuitive and of an acceptable good design. With the given design, all pilots thought that minimal training would be required.
- Low workload ratings suggest that pilots had spare cognitive capacity for other flight deck tasks when using the CPDLC display.
- Visual attention, measured by head down time, suggests that pilot situational awareness could be reduced by heads down data link tasks. This does not necessarily suggest a page format, or page navigation design issue. Rather, the head down time is more of a reflection of an MCDU based implementation that is located out of the pilot's primary field of view.
- Because of the competition for the visual channel, the pilots believed that CPDLC would take visual attention away from other tasks during critical phases of flight (i.e. altitudes below flight level (FL) 180 where pilots must practice SEE and AVOID). In the future, it is recommended that shortcut devices, better placement of CPDLC components, better alerting for incoming or open messages be investigated so that the operational use of CPDLC can be expanded into other flight phases.
- Because of visual focus required (head down time) the pace and frequency of transmitting ATC messages during different phases of flight should also be a consideration when developing CPDLC procedures and SOPs. Lengthy or complex communications or messages may require too much head down time during periods requiring visual attention elsewhere.
- Overall, a clear picture emerged that confirms that pilot performance with a single MCDU CPDLC system was acceptable and improved over time even under high workload conditions. The results support the Honeywell implementation of an MCDU based datalink communications system for the proposed use. Additional research on the mitigating effects associated with primary-field-of-view annunciation should be considered before determining appropriate critical phase-of-flight limitations.

Acknowledgement

The presented work was sponsored by the FAA under contract DTFAWA-10-A-80003. Findings contained herein are not necessarily those of the FAA.

References

EUROCONTROL (2009) ATC Data Link Operational Guidance for LINK2000+ Services. Version 5.0

Joint Planning and Development Office (JPDO) (2007). Concept of Operations for the Next Generation Air Transportation System. Version 2.0, Washington, DC, 226 pp.

Lancaster, J. (2008). Investigating pilot performance using mixed-modality simulated data link. *Human Factors*; 50(2):183-93.

RTCA SC-214/EUROCAE WG-78 (2009) Data Communications Safety and Performance Requirements, Draft E.

RTCA (2000). Minimum human factors standards for air traffic services provided via data communications utilizing the ATN, Builds I and IA (RTCA/DO-256). Washington, DC: RTCA, Inc.

United Airlines (UAL), (2009) B-777 Flight Manual Chapter 13 Additional Procedures, Section 40 Communications; CNS Enroute Operating Guide.

Wiener, E. L., and Nagel, D. C. (Eds.). (1988). Human factors in aviation. San Diego: Academic Press.

ZAGREB, TENERIFE AND COVE NECK: REVISITING THE ASSUMPTIONS UNDERLYING ICAO'S LANGUAGE PROFICIENCY PROGRAM

Simon Cookson
J. F. Oberlin University
Tokyo, Japan

In 1976 a DC-9 and a Trident 3B collided over Zagreb in the former Yugoslavia; in 1977 two Boeing 747s collided on the ground in Tenerife; and in 1990 a Boeing 707 ran out of fuel after a missed approach and crashed at Cove Neck. Because they involved language issues and resulted in 832 deaths, these three accidents have been cited by the International Civil Aviation Organization (ICAO) in justification of a worldwide program to improve the language proficiency of pilots and air traffic controllers. This paper shows that: (1) both linguistic and non-linguistic causal factors contributed to each accident; (2) a range of linguistic causal factors were involved, such as code switching and L1 interference, with each accident featuring different factors; and (3) the linguistic factors were in all three cases exacerbated by the effects of high workload, stress and fatigue.

The International Civil Aviation Organization (ICAO), the UN agency that oversees international air transport, has over the last decade been implementing a program to improve the English language proficiency of pilots and air traffic controllers around the world. This program has seen the development of language proficiency requirements (LPRs) and a six-level proficiency rating scale. ICAO initially intended for all pilots and controllers involved in international flights to demonstrate proficiency at Level 4 or higher on this scale from 2008, but difficulties in implementing the changes resulted in the deadline for compliance being put back to 5 March 2011.

In justifying the new program, ICAO has cited a number of airline accidents that were at least partly caused by language factors. Seven accidents, for example, were listed at the Language Proficiency Implementation Plan Workshop held in 2008 at the ICAO Asia and Pacific Office (Lamy 2008). This paper examines three accidents cited by ICAO: the mid-air collision over Zagreb in 1976; the runway collision at Tenerife in 1977; and the crash of Avianca Flight 052 at Cove Neck in 1990. Table 1 summarises the three accidents.

The following questions are addressed by this paper. What were the language factors that contributed to the three accidents? Was it simply that non-native speakers had inadequate English language proficiency? Or were more complicated factors involved, and if so, can any patterns be identified? The paper examines each accident, with a summary of the main events followed by analysis of the linguistic factors. Sources of data include accident reports, air traffic control (ATC) transcripts, cockpit voice recorder (CVR) transcripts and ICAO documents. The paper concludes with a summary of the lessons that can be drawn from the analysis.

Table 1. *Summary of the accidents analysed in this paper.*

	Accident 1	Accident 2	Accident 3
Location of accident	Zagreb, Yugoslavia	Tenerife, Canary Islands, Spain	Cove Neck, Long Island, New York, USA
Date of accident	10th September 1976	27th March 1977	25th January 1990
Number of fatalities	176	583	73
Accident type	Mid-air collision	Runway collision	Fuel exhaustion
Aircraft type / operator & flight number	(1) Trident 3B / British Airways Flight 476 (2) McDonnell Douglas DC-9-32 / Inex Adria Airways Flight 550	(1) Boeing B-747-206B / KLM Flight 4805 (2) Boeing B-747-121 / Pan Am Flight 1736	Boeing B-707-321B / Avianca Flight 052
L1 of flight crew	(1) English (2) Serbocroatian	(1) Dutch (2) English	Spanish
L1 of air traffic controllers	Serbocroatian	Spanish	English

Accident 1: Zagreb 1976

On 10th September 1976, British Airways Flight 476 was flying from London Heathrow to Istanbul when it collided in mid-air with Inex-Adria Airways Flight 550, en route from Split to Cologne. The crash occurred in daylight and in fine weather conditions at an altitude of 33,000 feet over Zagreb in the former Yugoslavia. All 176 passengers and crew on board the two aircraft died in the collision.

This tragedy occurred because of the co-incidence of a number of factors (AAIB 1977). Firstly, in the ATC centre at Zagreb a flight progress strip was not handed over as Flight 550 climbed into the upper sector. The controller of the upper sector was overloaded with other work, and the assistant controller was absent for several minutes prior to the collision, an absence that went unnoticed by the chief of shift. In addition, the first call from Flight 550 to the upper sector controller was delayed by radio traffic. As a result of these factors, the upper sector controller did not follow the vertical movement of Flight 550 and did not provide adequate separation between it and the other plane, British Airways Flight 476. When he finally did give a warning, the controller gave an incorrect flight level for Flight 476 and he also switched from English to Serbocroatian, his L1, which meant that the British Airways flight crew could not understand a critical part of the dialogue. The two aircraft collided moments later.

Two of the causal factors involved language issues: the code switching of the upper sector controller from English to Serbocroatian and the numerical slip that meant he gave an incorrect flight level for the British Airways aircraft. Considering the first issue, it is not clear why the controller switched from English to Serbocroatian, but perhaps he realised a collision was imminent and deliberately switched to his L1 (and that of the crew of Flight 550) to ensure he would be immediately understood. Or perhaps the code switching was not deliberate, but brought about by a combination of time and workload pressure. Ganushchak and Schiller (2009) note that unintentional code switching may occur under conditions of psychological stress. They suggest that when speakers using L2 are subjected to time pressure there may be intrusions from the ‘dominant native language’, on account of the higher cognitive workload required to speak a second language. The controller was under time pressure to resolve the problem, and workload pressure from handling the upper sector alone for most of the previous ten minutes, when two or three controllers should have been assigned to each sector. In addition, he was quite possibly suffering from fatigue as a result of working three twelve-hour shifts in three days (Stokes and Kite 1994).

Whatever the reason, the code switching unfortunately prevented the pilots of British Airways Flight 476 from monitoring this critical part of the dialogue as they could not understand Serbocroatian. With hindsight it has been suggested that had the controller only spoken English, the British Airways flight crew might have realised that the two aircraft were at the same altitude and tried to prevent the collision (Beaty 1995). However, the accident report points out that even if this had happened there may not have been enough time for the pilots to take avoiding action (AAIB 1977). Furthermore, the suggestion that only English should have been used fails to take account of all the causal factors listed above; specifically, it ignores the pressure the controller was under, which may have rendered him incapable of accurately communicating the vital message in English, his L2.

The second language factor was the numerical slip: the controller told Flight 550 to stop climbing because another plane was approaching at flight level 335 (= 33,500 feet). The other plane, Flight 476, was actually at flight level 330 (= 33,000 feet), its assigned altitude. The controller later claimed that his radar indicated flight level 335, but ten minutes before he had spoken with the crew and confirmed they were at flight level 330. It has been speculated that the controller, believing that disaster was inevitable, ‘may have attempted to mask the terrible reality by “adding” that extra 500 feet to the flight level’ (Weston and Hurst 1982). The mistake may be attributable, though, to other factors. As Dismukes et alia (2007) note, high workload can cause recently acquired information to be forgotten or not remembered correctly, and both Cushing (1994) and Monan (1986) list numerous instances of miscommunication involving numbers. In fact, the upper sector controller made two other numerical slips in the two minutes preceding the crash: first he gave an incorrect radio frequency, which he immediately corrected, and then he incorrectly read back a flight number. Those two mistakes were of no consequence, but the third was catastrophic as it directed two aircraft to the same flight level just as their flight paths crossed.

Accident 2: Tenerife 1977

On 27 March 1977, KLM Flight 4805, flying from Amsterdam, and Pan Am Flight 1736, from Los Angeles via New York, were bound for Las Palmas Airport in the Canary Islands when a bomb explosion caused the airport to be closed. The aircraft, both Boeing 747s, were diverted to Los Rodeos Airport on Tenerife Island. After Las Palmas Airport reopened, the Pan Am crew wanted to leave Tenerife as soon as possible, but had to wait until the KLM aircraft, which was blocking access to the runway, finished refuelling. Weather conditions deteriorated as thick cloud enveloped the airport. Finally both aircraft were instructed to taxi along the active runway, the taxiway being blocked by other aircraft. Flight 4805 reached the end of the runway, turned around, and began its takeoff roll without receiving clearance from the control tower, and with Flight 1736 still on the runway. The KLM aircraft had just started to lift off when it collided with the other plane. A total of 583 people were killed in the collision.

The report produced by Spain's Civil Aviation Accidents and Incidents Investigation Commission cited four actions by the KLM captain that caused the accident, plus nine contributory factors (CIAIAC 1978). This accident, like Zagreb, was the result of multiple factors and featured communication problems between pilots and controllers. However, unlike Zagreb, it was preceded by a bomb attack that caused diversions and delays. Los Rodeos was, moreover, a small regional airport at an elevation of 2,000 feet, subject to rapidly changing visibility caused by wind-blown clouds, and not used to handling international traffic or aircraft as large as the Boeing 747.

The communication problems occurred in the final five minutes before the collision. The first problem was that the Pan Am crew had difficulty understanding instructions from the control tower 'because of the heavy Spanish accent of the controller as he spoke English' (Roitsch et alia 1978). As they taxied down the runway the Pan Am crew struggled to find the correct exit because they were not familiar with the airport, visibility was poor, and there were no signs marking exits. At 17:02 GMT the controller said, 'Affirmative, taxi into the runway and – ah – leave the runway third, third to your left.' The Pan Am captain thought the controller had said 'first', so the first officer asked for confirmation and another controller replied, 'The third one Sir, one, two, three, third third one.' This solved one problem but created another because the third exit required the aircraft to make two 148-degree turns. In fact it was 'a practical impossibility' for a Boeing 747 to negotiate these turns (ibid 1978). In the event, either by mistake or because they thought it easier, the Pan Am crew continued towards the fourth exit, which they were approaching at the time of impact.

The second language problem occurred as the KLM crew prepared for takeoff. At 17:05 the aircraft rolled forward slightly until the first officer warned the captain that they had not yet received ATC clearance. The tower then issued ATC clearance, giving permission to fly the first part of the route. Confusingly, this clearance included the word 'take-off', but it was not the takeoff clearance. The first officer repeated the ATC clearance to the tower: 'Ah– Roger sir, we are cleared to the Papa beacon flight level nine zero, right turn out zero four zero until intercepting the three two five. We are now at take off.' Just after the first officer had started this read-back, the KLM captain released the brakes and began the takeoff roll.

There has been much debate since the accident about the last sentence in the first officer's read-back, underlined above. Some commentators have posited linguistic interference from the first officer's L1: in Dutch a preposition may be used with the infinitive form of a verb to indicate an action currently being performed (ICAO 2004). Hence, the first officer may have meant the phrase 'at take off' to mean 'in the process of taking off'. His words, though, were 'hurried and the voice tremulous', indicating that he was under stress and making transcription of the voice recorder tapes so difficult that a report by the Airline Pilots Association (ALPA) suggested that he might actually have said, 'We are, uh, taking off.' (Roitsch et alia 1978).

According to the Spanish accident report, the controller interpreted the first officer's last sentence as meaning 'We are now at take-off position' (CIAIAC 1978). In other words, the controller assumed an elliptical construction and, from the available context, inferred the missing element was 'position'. He believed, after all, that the KLM aircraft was waiting for takeoff clearance. The controller replied at 17:06, 'O.K.', and continued, 'Stand by for take-off ... I will call you.' The Pan Am first officer made a radio transmission at almost the same time saying, 'and we are still taxiing down the runway, the Clipper one seven three six.' Either message could have alerted the KLM crew to the imminent disaster, but unfortunately the near-simultaneous transmission caused interference – another communication problem – so that neither could be heard clearly by the KLM crew.

The controller then requested the Pan Am aircraft to report when they were clear of the runway. This message could be heard in the KLM cockpit, but instead of referring to the Pan Am plane as 'Clipper' the controller used the phrase 'Papa Alpha', which was less likely to catch the attention of the pilots. The KLM flight engineer presumably heard this message, for he twice asked his colleagues whether the other plane had cleared the runway. The KLM captain replied emphatically (but mistakenly), ‘Oh, yes.’ Seconds later the collision occurred.

The Tenerife accident involved the tragic co-incidence of many factors, and the ALPA accident report, which focused on human factor issues, highlighted the effects of stress and fatigue on those involved. In the tower, the two controllers had been on duty since 10:00 and had to handle three frequencies, with an unusually large amount of traffic on account of diversions from Las Palmas, compounded by the fear of a further bomb attack at Tenerife (Roitsch et alia 1978). Workload pressure and fatigue may explain why one controller used the wrong flight number three times, saying KLM 8705 instead of KLM 4805, and it may also have prevented the controllers from realising that a Boeing 747 could not negotiate the third taxiway.

The Pan Am crew had been on duty for 11 hours, having to deal with the frustration of diverting to Tenerife when they wanted to hold at altitude, and being delayed on the ground until the KLM aircraft finished refuelling. Meanwhile, after 9 hours on duty the KLM crew knew they had to depart soon on account of strict regulations on duty time limits introduced by the Dutch Government the previous year. Within the KLM cockpit there was an additional problem caused by a steep ‘trans-cockpit authority gradient’. Wiegmann and Shappell (2003) note that ‘when very senior, dictatorial captains are paired with very junior, weak co-pilots, communication and coordination problems are likely to occur.’ The KLM captain was the Head of the Flight Training Department, while the first officer was a junior pilot who had recently received his type qualification check from the captain. This made it difficult for the first officer to question or challenge the captain’s decisions. Furthermore, the KLM captain had not flown as a line pilot in the previous 12 weeks due to his work as a simulator instructor, in which role he would have issued clearances to training crews. As Roitsch et alia (1978) point out, ‘There is never a need for the crew to hold the simulator in position awaiting takeoff clearance.’

Accident 3: Cove Neck 1990

On 25 January 1990, Avianca Flight 052 was scheduled to fly from Columbia to John F. Kennedy International Airport (JFK) in New York. Poor weather in the north-eastern part of the United States meant the aircraft had to entered three separate holding patterns for a total of 77 minutes. During the third holding period, the flight crew notified ATC that they could only hold for about five more minutes and could no longer reach their alternate airport in Boston because they were running out of fuel. As the aircraft finally descended towards JFK it encountered wind shear and the crew executed a missed approach. While trying to return for a second approach, all four engines suffered a loss of power and the aircraft crashed at approximately 21:34 EST at Cove Neck, Long Island. Of 158 passengers and crew on board the plane, 73 died as a result of the crash.

The National Transportation and Safety Board report stated one probable cause of the accident and a number of contributory factors (NTSB 1991). The probable cause ‘was the failure of the flightcrew to adequately manage the airplane’s fuel load, and their failure to communicate an emergency fuel situation to air traffic control before fuel exhaustion occurred’. This accident was caused by multiple factors, like Zagreb and Tenerife, and might have been averted if any of the factors had been absent. The report notes, for example, that the accident would not have occurred if the crew had not been prevented from successfully completing the first approach by a combination of wind shear, stress and fatigue.

Two language factors were identified in the NTSB report: the crew’s failure to communicate an emergency fuel situation to ATC, and the lack of standardized, understandable terminology for minimum and emergency fuel states. At 20:44, during the third holding period, the New York Air Route Traffic Control Center (ARTCC) told Flight 052 to expect further clearance information at 21:05. The first officer, who was handling communications with ATC while the captain flew the plane, read back the time and said, ‘I think we need priority we’re passing [unintelligible]’. He then reported they could only hold for five minutes more and, when asked to repeat the alternate airport, he said, ‘It was Boston but we can’t do it now we, we, don’t, we run out of fuel now.’ The first officer thus informed ARTCC about the fuel problem more than 45 minutes before the crash. Crucially, though, he did not declare a fuel emergency. Moreover, ARTCC did not notify approach control about the fuel problem, and the aircraft subsequently received routine vectors including a 360° turn for spacing.

Between 21:03 and 21:06 the flight engineer briefed his colleagues on the go around procedure for a low fuel situation. It is not clear from the CVR data whether this briefing was a continuation of an earlier conversation because the recording only covered the final 40 minutes of the flight. The crew did not contact ATC again about the fuel problem until they had executed the missed approach at 21:23. From then, during the final 10 minutes of the flight, the captain five times declared in Spanish ‘we don’t have fuel’ or ‘we are in emergency’, and repeatedly instructed the first officer to notify ATC. Instead, the first officer three times said in English ‘we’re running out of fuel’. He finally requested ‘priority’ handling at 21:32 after two engines had flamed out, but even at this point did not declare an emergency. Why not? Was this simply a language proficiency problem?

Describing the first officer’s English as ‘excellent’ and ‘unaccented’, Helmreich (1994) hypothesises that the crew’s actions can be attributed at least partly to cultural factors, and discusses the role played by collectivism, power distance and uncertainty avoidance. Coming from a strongly collectivist culture, the Columbian crew may have been reluctant to declare an emergency and push ahead of other crews they perceived to be in difficulty. Their reluctance may have been reinforced by a transmission from an American Airlines crew at 21:02 giving a minimum fuel advisory and warning they would soon declare a fuel emergency. Gladwell (2008) develops the hypothesis, noting that authority is highly respected in Columbia and explaining that the first officer – 28 years old and lacking flight experience – would have seen himself as subordinate to the captain and the ‘domineering Kennedy Airport air traffic controllers’. Gladwell (2008) suggests that the first officer instinctively used mitigated speech ‘to downplay or sugarcoat the meaning’ of his messages in deference to the authority of the captain and controllers. Hence the softening of the captain’s statement ‘we don’t have fuel’ to ‘we’re running out of fuel’, and of ‘we are in emergency’ to ‘we need priority’.

Concerning the words ‘priority’ and ‘emergency’, Krause (2003) points out that the first officer may have thought them interchangeable. According to the testimony of an Avianca captain cited in the NTSB report, training provided by Boeing gave the impression that ‘the words priority and emergency conveyed the same meaning to air traffic control’. The report also states that a Boeing manual used to train Avianca crews advised that ‘priority handling from ATC should be requested’ during operations with very low fuel quantities. This is at odds with the advice of controllers questioned during the investigation, who stated that flight crews should only use the terms ‘Mayday’ or ‘Pan-pan-pan’ or ‘Emergency’ when declaring a fuel emergency (NTSB 1991).

Another reason for the communication problems was that the crew, especially the captain, were suffering the effects of high workload, stress and fatigue. The flight time of 6 hours 26 minutes was much longer than the scheduled 4 hours 40 minutes. Furthermore, maintenance problems with the autopilot and flight director led investigators to conclude ‘that the aircraft might have been flown manually from Medellin to JFK’ and that the captain flew the first approach ‘without the aid of a flight director’ (NTSB 1991). In a Boeing 707 – built in the 1960s and designed in the 1950s – this would have been exhausting. Dismukes et alia (2007) state that both stress and fatigue narrow an individual’s focus of attention and impair cognitive processing. As a result, decision-making ability and communication may be adversely affected. Communication degradation may manifest itself not only in impaired speech production but also in a ‘decreased ability to receive and interpret messages’ (Stokes and Kite 1994). This was clearly the case with the captain: analysis of the CVR data reveals a discourse pattern whereby the captain would inquire about ATC communications or the configuration of the aircraft and then ask several times for the same piece of information – usually a number – to be repeated. At 21:17 he even said, ‘tell me things louder because - - I’m not - - hearing it’ (NTSB 1991).

The communication issues in this accident were complex, and they brought about a critical difference in situation awareness between the flight crew, who believed they had informed ATC about their fuel problem, and the controllers, who did not realise that an emergency situation existed. Noting that the first officer twice asked for ‘priority’ and four times advised ATC that the plane was low on fuel, Krause (2003) states that ‘it would seem reasonable and logical’ for the controllers to have asked for clarification. None of them did so, however, and the aircraft crashed before the second approach could be completed.

Conclusion

The aims of the ICAO language proficiency program are laudable, and it goes without saying that clear communication between pilots and air traffic controllers is essential for the smooth and safe operation of the international air transport system. However, the claim by ICAO (2004) that ‘inadequate language proficiency has played a role in accidents’ is a somewhat misleading simplification when it comes to the accidents analysed in this paper: Zagreb, Tenerife and Cove Neck. The analysis underscores the complexity of each of these disasters, and shows that a range of language issues were involved, including numerical slips, code switching, poor pronunciation, L1 interference, and ambiguity caused by ellipsis. The only factors present in more than one accident were the numerical slips, which featured in all three (but were critical only in the collision over Zagreb). Each accident was tragically unique, but nevertheless the following common features have been highlighted by the analysis:

- All three accidents were complex and occurred as the result of multiple causal factors.
- Each accident involved a combination of linguistic factors and non-linguistic factors. (In other words, none of the accidents were caused solely by language problems.)
- The pilots and air traffic controllers involved were in each case a mixture of native English speakers and non-native speakers. (None of the accidents involved only non-native speakers.)
- While all three accidents featured oral communication problems, there is a suggestion that one of them – Cove Neck – also involved a problem relating to written language.
- In all three accidents the linguistic factors were exacerbated by the effects of high workload, stress and fatigue, and in one case – Cove Neck – cultural factors were also significant.
- Finally, in each accident the language factors contributed to the setting up of a critical difference in situation awareness between the flight crews and the air traffic controllers.

References

Air Accidents Investigation Branch (AAIB) (1977). British Airways Trident G-AWZT, Inex-Adria DC-9 YU-AJR: Report on the Collision in the Zagreb Area, Yugoslavia, on 10 September 1976. Aircraft Accident Report 5/77. London: Her Majesty's Stationery Office.

Beaty, D. (1995). *The Naked Pilot: The Human Factor in Aircraft Accidents (new edition)*. Marlborough, UK: Airlife Publishing.

Civil Aviation Accidents and Incidents Investigation Commission (CIAIAC) (1978). *KLM, B-747, PH-BUF and Pan Am B-747, N736, Collision at Tenerife Airport, Spain, on 27 March 1977.* Report Number CIAIAC A-102/1977. Madrid: Secretary of Civil Aviation.

Cushing, S. (1994). *Fatal Words: Communication Clashes and Aircraft Crashes.* Chicago: The University of Chicago Press.

Dismukes, R. K., Berman, B. A., & Loukopoulos, L. D. (2007). *The Limits of Expertise: Rethinking Pilot Error and the Causes of Airline Accidents*. Aldershot: Ashgate Publishing Limited.

Ganushchak, L. Y., & Schiller, N. O. (2009). 'Speaking one's second language under time pressure: An ERP study on verbal self-monitoring in German-Dutch bilinguals.' *Psychophysiology*. 46/2, 410-419.

Gladwell, Malcolm (2008). *Outliers: The Story of Success.* London: Allen Lane, Penguin Books Ltd.

Helmreich, Robert L. (1994). Anatomy of a System Accident: The Crash of Avianca Flight 052. *The International Journal of Aviation Psychology.* 4/3, pp. 265-284.

International Civil Aviation Organization (ICAO) (2004). *Manual on the Implementation of ICAO Language Proficiency Requirements*. ICAO Document 9835, November 2004. Montreal, Canada: ICAO.

Krause, Shari S. (2003). *Aircraft Safety: Accident Investigations, Analyses, and Applications*. 2nd edition. New York: McGraw-Hill.

Lamy, P. (2008). *Module 1 – Language Proficiency Requirements Update.* Presentation by the Acting Deputy Director of ICAO's Air Navigation Bureau at the Language Proficiency Implementation Plan Workshop held at the ICAO Asia and Pacific Office, 29-31 January 2008.

Monan, William P. (1986). *Human Factors in Aviation Operations: The Hearback Problem.* NASA Contractor Report 177398. Moffett Field, CA: NASA Ames Research Center.

National Transportation Safety Board (NTSB) (1991). *Aircraft Accident Report: Avianca, The Airline Of Columbia, Boeing 707-321 B, HK 2016, Fuel Exhaustion, Cove Neck, New York, January 25, 1990*. NTSB/AAR-91/04. Washington, D.C.: NTSB.

Roitsch, P. A., Babcock, G. L., & Edmunds, W. W. (1978). *Human Factors Report on the Tenerife Accident.* Washington, D.C.: Air Line Pilots Association.

Stokes, A., & Kite, K. (1994). *Flight Stress: Stress, Fatigue, and Performance in Aviation.* Hampshire, England: Avebury Aviation.

Weston, R., & Hurst, R. (1982). *Zagreb One Four: Cleared to Collide?* New York: Jason Aronson, Inc.

Wiegmann, D. A., & Shappell, S. A. (2003). A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System. Aldershot: Ashgate Publishing Limited.

SITUATION AWARENESS, WORKLOAD, AND PERFORMANCE IN MIDTERM NEXTGEN: EFFECT OF VARIATIONS IN AIRCRAFT EQUIPAGE LEVELS BETWEEN SCENARIOS

L. Paige Bacon[1], Thomas Z. Strybel[1], Kim-Phuong L. Vu[1], Joshua M. Kraut[1],
Jimmy H. Nguyen[1], Vernol Battiste[2,3], and Walter Johnson[3]

[1] Center for Human Factors in Advanced Aeronautics Technologies
California State University, Long Beach
Long Beach, CA
[2] San Jose State University Foundation
San Jose, CA
[3] NASA Ames Research Center
Moffett Field, CA

We investigated the impact of mixed equipage traffic on air traffic controller (ATC) performance, workload and situation awareness in an airspace simulation of an en route sector. Nine retired ATCs and seven student ATCs managed traffic and responded to probe questions designed to measure workload and situation awareness. ATC performance was measured with the AT-SAT OTS Performance Rating Form. Although workload decreased with increasing mixtures of equipped aircraft, the number of LOS appeared to be related more to the percentage of alerted conflicts in the scenario. The probe measure of workload and situation awareness was related to AT-SAT ratings.

The Next Generation of Air Transportation System (NextGen), is a program being developed under the guidance of several government agencies in order to expand the capacity of the National Airspace System (NAS), reduce the negative environmental impacts on the environment, improve efficiency and safety, and reduce delays on the ground (FAA Implementation Plan, 2010). To accomplish these goals, new tools, technologies and procedures for NextGen operations are being developed (e.g., Prevot, Homola, & Mercer, 2008). Because these tools, technologies, and concepts will change how air traffic controllers (ATCs), perform their tasks, it is important to evaluate NextGen developments in terms of their impact on human operator performance, workload, and situation awareness (SA), and investigations of these impacts have appeared in the literature (e.g., Willems and Schulz, 2010).

Technologies such as Data Comm are expected to reduce ATC workload and enable controllers to manage higher traffic densities by reducing the need for ATC-pilot voice communications. Willems and Schulz (2010) determined that the number of voice communications, an indicator of ATC workload, was reduced only when the percentage of equipped AC exceeded 50%; at lower equipage levels, no change in workload was observed. In addition, tools such as automated conflict detection and resolution (CD&R), and trial planners should further reduce controller workload and improve their performance. However, because these tools may be useful only for separating pairs of appropriately equipped aircraft (AC), ATCs will need to rely on their current-day air traffic management skills for detecting and resolving conflicts involving pairs unequipped ACs in addition to using the advanced tools. Therefore, in the present study we determined the impact of air traffic containing different percentages of ACs equipped with both Data Comm and CD&R capabilities on ATC performance, workload, and situation awareness.

Evaluations of NextGen tools such as Data Comm and CD&R on operator performance require reliable and valid measures of performance. The need for reliable, valid, and diagnostic measures of human operator performance is not new, and remains a problem for the assessment airspace operators in both current-day and NextGen environments because the relationships between objective measures of airspace system performance and operator performance have yet to be determined, yet objective measures of sector outcomes are easily extracted from digital systems. Rantennan (2004), in a recent review of ATC performance metrics, summarized the problem as follows: “The problem is therefore not in the availability of data, but in derivation of valid, reliable, and meaningful measures from the abundance of data (p. 2).” This problem becomes more pronounced for measuring cognitive constructs such as workload and SA because they are not directly observable, but are considered important determinants of performance. Workload and SA have been shown to be related to operational errors (e.g., Jones & Endsley, 1996). Workload is commonly measured with standardized self-rating methods such as ATWIT (Stein, 1985) and NASA TLX (Hart & Staveland, 1988) both of which have been extensively tested and validated. Consensus has not been achieved, however, on the best approach to SA measurement (e.g., Stanton, et al., 2006).

In the present investigation, we attempted to identify the individual performance metrics that contribute the most to overall ATC performance. We first measured overall ATC performance using two standard ATC assessment tools, the Behavioral and Event Checklist (BEC) and a modified AT-SAT High Fidelity Simulation Over the Shoulder (OTS) Rating Form (Ramos, Heil, & Manning, 2001). The modifications to the AT-SAT were needed to include an evaluation of ATC's use of Data Comm. Both instruments have shown to be reliable and valid predictors of ATC performance (e.g., Manning and Stein, 2005). The overall measures then were used to determine a set of performance, workload, and SA measures that are significantly associated with the overall measure. We assessed operator SA, or the operators' understanding of the task environment, by using a variant of the Situation Present Assessment Method (SPAM; Durso & Dattel, 2004) because a real-time probe method can measure SA when it is distributed in the task environment. (Chiappe, Strybel & Vu, submitted).

Method

Participants

Sixteen participants served as ATCs in the present study. Nine were radar-certified, retired air traffic controllers (6 TRACON and 3 ARTCC) and seven were students enrolled in the Aviation Sciences Program at Mount San Antonio College, an FAA CTI institution.

Apparatus

The simulation was run using the Multi Aircraft Control System (MACS), Aeronautical Datalink and Radar Simulator (ADRS), and VoiceIP software, developed by the Airspace Operations Laboratory at NASA Ames Research Center (e.g., Prevot et al., 2006). ATCs used MACS configured as a Digital System Replacement (DSR) display with integrated Data Comm, conflict alerting, and conflict probing for all aircraft (AC) designated as "equipped." Online SA probes were presented on a touch-screen workstation located adjacent to the DSR display.

Scenarios

Six experimental trials were run, each lasting approximately 50 minutes, with two each containing 25%, 50%, or 75% equipped AC. The experimental scenarios were originally constructed with eight planned traffic conflicts, and the number of conflict alerts was manipulated in the scenario by setting AC equipage. Table 1 shows the intended number of conflicts and conflict alerts for each percentage of equipped AC. However, it was determined after data collection was finished that not all scenarios contained 8 conflicts (see Table 1 for the actual number of conflicts). Therefore, the scenario containing 2 conflicts (one of the 50% Data Comm scenarios) was not analyzed. Moreover, the variables Percent Equipped AC and Number of Alerted Conflicts were combined into one variable, that of Percentage of Alerted Conflicts, also shown in Table 1.

Table 1

Traffic Characteristics (Percent Equipped AC, Percent Alerted Conflicts) of the Scenarios

Percent Equipped AC	Planned Number of Alerted Conflicts (out of 8 conflicts)	Number Conflicts	Number Alerts	Percent Alerted Conflicts
75 6		8	6	75
75 4		7	4	57
50 5		5	2	4
50 3		7	2	28
25 3		2	0	--
25 1		8	1	13

Procedure

Participants were given one day of training and practice prior to the experimental trials. During the experimental scenarios, 14 SA probe questions were presented, one every three minutes, beginning four minutes into the scenario. The sequence of events for each online probe question was as follows. The ATC received a "Ready for

Question" prompt on the touch screen accompanied by an audio alert. The participant was instructed to respond by touching a button on the screen only when he/she had sufficient capacity to answer a probe question. Once the participant responded affirmatively, a probe question and response alternatives were immediately displayed on the panel and the participant responded via the touch screen. If the participant did not respond to the "Ready" prompt within two minutes, the query was withdrawn.

The probe questions were developed beforehand, to ensure that equal numbers of questions were asked from each of four task- information categories: conflicts, sector status, command-communication and subjective assessment. Conflict questions queried information regarding current and future conflicts between an aircraft pair. Sector status questions were regarding the current state of the sector, such as number of equipped aircraft. Questions categorized as command-communications asked for information on previous or future commands and communications (both voice and Data Comm). Subjective assessment questions asked participants to rate their concern (1= low, 5=high) regarding a future conflict event. At the end of each scenario, participants completed the Situation Awareness Rating Technique (SART) and the NASA TLX workload measure.

Expert Evaluation of Performance

The performance of each participant was evaluated by two of three expert raters who reviewed screen recordings of a simulation run. Three experts (two retired en route ATCs, one retired TRACON ATC) were trained on the use of the BEC and AT-SAT OTS Rating. These raters independently watched a set of identical test recordings and recorded events on the BEC. At the end of the recorded scenario, the raters completed the AT-SAT Rating Form. The AT-SAT Rating Form consists of 8 rating dimensions. The evaluator assigned a number between 1 and 8 based on information from the BEC. The evaluators trained until an acceptable level of agreement between the raters was reached. Subsequently, each rater independently assessed 32 of the possible 96 recordings. Raters also completed the NASA TLX at the end of each scenario.

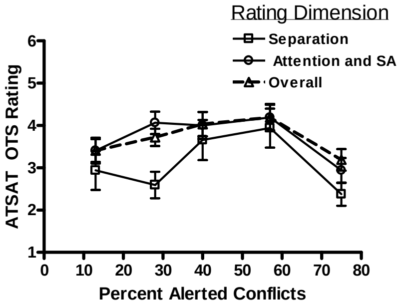


Figure 1. Mean AT-SAT OTS Rating as a function of Percent of Alerted Conflicts and Rating Dimension

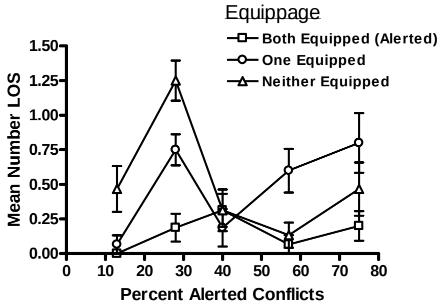


Figure 2. Mean Number of LOS as a function of Percent of Alerted Conflicts and Conflict Pair Equipage.

Results

All 8 Rating Dimensions on the ATSAT-OTS form were significantly correlated with each other (*r*s = .34 to .90; *p*s< .003). A mixed-design ANOVA was run on ratings for each dimension, with the within-subject variable of percentage of alerted conflicts (4%, 13%, 28%, 57% and 75%) and between subject variable expertise (student vs. expert). A main effect of percentage of alerted conflicts was obtained on three ATSAT OTS dimensions (*p*s <.06): Maintaining Separation, Maintaining Attention and Situation Awareness and Overall Performance, as shown in Figure 1. Maintain Attention and SA and Overall Performance Ratings were affected similarly by percentage of alerted conflicts. The highest average rating was given when 57% of the conflicts were alerted; lower ratings were made for higher and lower percentages. The effect of percentage of alerted conflicts on Maintaining Separation did not follow this pattern, however. AT-SAT ratings were lowest for scenarios having 13%, 28% and 68% alerted conflicts. The NASA TLX ratings were not significantly correlated with the OTS ratings. However, the TLX composite score was significantly affected by percentage of alerted conflicts.

A battery of performance metrics were also computed for each ATC and scenario. These were used to measure safety, efficiency, and workload. Measures of safety included number LOS, and average lateral and vertical distance between ACs. The number of LOS was counted separately for conflicts involving two equipped

AC, one equipped AC, and no equipped AC. The latter two conflict types were not alerted. For each variable, a mixed ANOVA was run with the factors percentage of alerted conflicts and expertise. A significant effect of percentage of alerts was obtained only for the number of LOS involving at least one unequipped AC (p <.002). In Figure 2, the average number of LOS for each equipage is shown. For scenarios having low percentages of alerted conflicts, most LOS occurred between two unequipped AC, which is not surprising because these scenarios contained relatively few equipped aircraft. At the highest percentages, more LOS occurred between equipped - unequipped AC pairs. A main effect of expertise was obtained for conflicts between equipped and unequipped ACs only (p <.001.) Students, on average had nearly double the number of LOS compared with experts (student: $M = .64$; ATC: M=.25) when one AC was equipped and one AC was unequipped. Although experts had slightly more LOS for the remaining AC pairs, the differences were not significant.

The average lateral and vertical separations were significantly affected by percentage of alerts, and an interaction was obtained between expertise and percentage of alerted conflicts, was found for average vertical separation between AC (p<.02). As shown in Figure 3, expert ATCs typically maintained greater vertical separation between AC when the percentage of alerted conflicts was either very low or very high. In the mid range, the differences were minimal. Students maintained greater vertical separation in the scenario containing the 28% alerted conflicts. Performance measures of efficiency included mean and standard deviation of handoff latency, distance traveled, and time through the sector. All but mean handoff time were affected by the percentage of alerts (ps <.01). No effects of expertise were observed on these metrics.

Two performance measures of workload were computed, total number of voice communications and percentage of time on voice. A significant effect of percentage of alerts on total number of voice communications was obtained, most likely because the percentage of alerted conflicts increased with percentage of equipped aircraft. The proportion of voice communications decreased from .85 when the percentage of alerts was 13% (and 25% of AC were equipped) to .38 when the percentage of alerts was at least 57% (and 75% of AC were equipped). Thus, the total number of voice communications was related to the percentage of equipped AC. Unlike the OTS-TLX composite scores, participant self-ratings of TLX workload were not significantly affected by percentage of alerts.

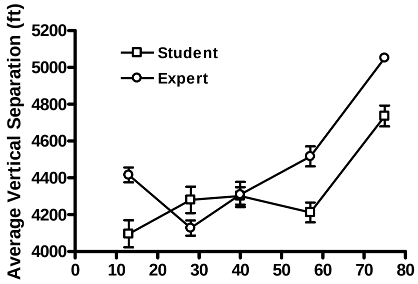


Figure 3. Mean Vertical Separation as a Function of Percent Alerted Conflicts and Expertise

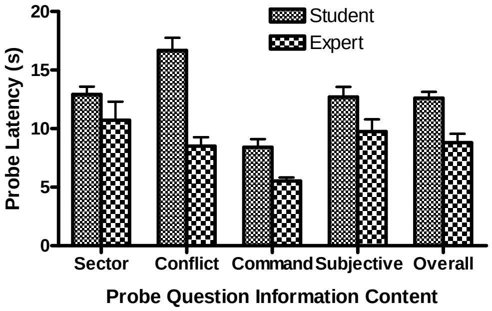


Figure 4. SA Probe Latency as a Function of Information Content and Expertise

Accuracy and response latency for probe questions overall, and by information category, were analyzed with the same mixed design ANOVA. All effects on probe accuracy were nonsignificant. Overall, response latency was affected expertise (p <.01). When probe latencies were evaluated for each probe information category, differences in response latencies between expert and student ATCs were significant for conflict and command-communication probes (ps <.01) and marginally significant for subjective and sector-status probes (ps <.07.) As shown in Figure 4, experts consistently responded faster than students, with the largest difference observed for conflict questions. Response latencies to the ready prompt, assumed to measure workload, was significantly affected by expertise (p <.03) and marginally affected by the percentage of alerts (p <.10.).

Predicting AT-SAT OTS Ratings

Hierarchical regressions were run to determine the combination of performance metrics that best describe ATC overall performance. A three level model was used. The experimental variables of percentage of alerted conflicts and expertise were entered first, followed by objective performance metrics, and then SA and Workload measures. The number of LOS variable was combined into two categories, alerted LOS (both Equipped AC) and non-alerted LOS (at least one AC unequipped). For situation awareness, overall probe latency and accuracy were entered, and for workload the number of voice communications, participant rating of TLX, and latency to the probe ready prompt were entered. Regressions were performed on the three rating dimensions, in which percentage of alerted conflicts significantly affected expert ratings, shown in Figure 1.

AT-SAT OTS ratings of Overall Performance was predicted by experimental variables (percentage of alerted conflicts and expertise), objective performance metrics (number non-alerted LOS), situation awareness (probe latency), and workload (ready latency), as shown in Table 2. This model was significant *F*(10,65) = 10.92; *p*<.0001) and accounted 46% of the variability in overall performance ratings. The model for the Rating Dimension "Maintaining Attention and Situation Awareness" was identical to Overall Performance rating with the exception that the Ready Latency variable was not significant.

Table 2

Significant predictors of AT-SAT OTS Overall Performance Rating (ps<.15).

Variable	**Parameter Est.**	**Partial r^2**
Percent Alerted Conflicts	-1.87	.07
Expertise -	0.82	.03
Number non-alerted LOS	-0.43	.21
Handoff Latency	-4.17	.07
Probe Latency	-0.04	.03
Ready Latency	-0.05	.01

Table 3 shows the regression model for the Rating Dimension "Maintaining Separation." Several additional factors were significant to the ones mentioned above. Neither expertise nor percentage of alerted conflicts entered into this model. Instead, the most highly predictive variables were the number of non-alerted LOS which accounted for 31% of the variance in ratings. The number of alerted LOS accounted for an additional 6%. This was followed by variables reflecting safety (e.g., average vertical distance) and efficiency (average handoff latency). Lastly, both situation awareness and workload entered in the model, with probe latency accounting for 4% of the variance in ratings, and participant TLX accounting for 2%.

Table 3

Significant predictors of AT-SAT OTS Rating "Maintaining Separation" (ps<.15)

Variable	**Param-eter**	**Partial r^2**	**Variable**	**Param-eter**	**Partial r^2**
Number non-alerted LOS	-0.83	.31	Average Lateral Distance	0.19	.02
Number alerted LOS	-1.09	.06	Std Dev Distance Traveled	-5.94	.02
Handoff Latency	-6.31	.06	Std Dev Time Thru Sector	0.71	.03
Probe Latency	-0.07	.04	Std Dev Handoff Accept Latency	-0.05 .01	
Average Vertical Distance	-0.001	.02	TLX – Participant Rating	-0.02	.02

Discussion

This preliminary investigation indicated that ATC workload decreased with increased percentages of equipped AC, similar to the finding of Willem et al. (2008). However, in our scenarios, we included CD&R tools for the management of equipped AC pairs, and the percentage of conflict alerts affected the number and types of LOS made: more conflicts between equipped and unequipped aircraft were found as the percentage of equipped AC increased. Moreover, experienced ATCs were more capable of detecting and resolving these non-alerted conflicts in a mixed equipage environment compared with student ATCs, suggesting that students may have relied more heavily on automated CD&R tools.

AT-SAT OTS Rating dimensions of Maintaining Attention and SA and Overall Performance were significantly affected by the percentage of alerted conflicts, with the lowest ratings made for 75% alerted conflicts. Considering that LOS has the most impact on the rating of overall performance (i.e., overall performance rating cannot exceed 3.0 when an LOS occurs), these ratings are consistent with our findings regarding LOS. However, our SA measure of probe latency significantly reduced the variance in AT-SAT ratings for overall performance, maintaining attention and situation awareness, and maintaining separation. We believe this provides strong support for the validity of an online probe technique as a measure of SA and workload. Probe latencies also discriminated between expert and student ATCs. However, the probe latencies to questions categorized by task information did not predict AT-SAT ratings.

Acknowledgements

This simulation was partially supported by NASA cooperative agreement NNA06CN30A: *NRA: Metrics for Situation Awareness, Workload, and Performance in Separation Assurance Systems* (Walter Johnson, Technical Monitor). Preparation of this paper was supported by NASA cooperative agreement NNX09AU66A, *Group 5 University Research Center: Center for Human Factors in Advanced Aeronautics Technologies* (Brenda Collins, Technical Monitor).

References

Chiappe, D.L., Strybel, T.Z., & Vu, K.-P.L. (2010). Review of situation awareness theories from the perspective of cognitive science. Submitted for publication.

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

Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.) *Human Mental Workload* (pp. 139-183), Amsterdam: North-Holland.

Jones, D.G. & Endsley, M.R. (1996). Sources of situation awareness errors in aviation. *Aviation, Space and Environmental Medicine, 67* (6), 507-512.

Manning, C. and Stein, E. (2005). Measuring air traffic controller performance in the 21st Century. In B. Kirwan, M. Rodgers and D. Schafer (Eds.) Human Factors Impacts in Air Traffic Management. Vermont: Ashgate.

NextGen Implementation Plan, March, 2010 [PDF File]. Retrieved from http://www.faa.gov.

Prevot, T., Homola, J., Mercer, J. (2008). Human-in-the-loop evaluation of ground-based automated separation assurance for NextGen. *ICAS 2008-11.4.4, 6th International Congress of the Aeronautical Sciences (ICAS), and AIAA-ATIO-2008-8885*, Anchorage, Alaska, Sept 15-19, 2008.

Ramos, R.A., Heil, M.C., & Manning, C.A. (Eds.) (2001). *Documentation of validity for the ATSAT computerized test battery: Volumes I and II.* (Report No. DOT/FAA/AM-01/6). Washington, DC: FAA Office of Aviation Medicine.

Rantanen, E.M. (2004). *Development and validation of objective performance and workload measures in air traffic control* (AHFD-04-19/FAA-04-7). Savoy, IL: Aviation Human Factors Division.

Salmon, P., Stanton, N., Walker, G., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics, 37* (2), 225-238.

Stein, E.S. (1985). *Air traffic controller workload: An examination of workload probe* (DOT/FAA/CT-TN84/24). Atlantic City International Airport, NJ: DOT/FAA, William J. Hughes Technical Center.

Willems, B., Hah, S., & Schulz, K. (2010). *En route data communications: Experimental human factors evaluation* (DOT/FAA/TC-10/06). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

ASSESSING CREWWORKLOAD ON AN INSTRUMENT METEOROLOGICAL APPROACH INTO A NON-RADAR AIRPORT

Marilyn French-St. George, Ph.D.

Transportation Safety Board of Canada

Gatineau, Canada

European Air Traffic Management Program (CARE, 2003) recommendations for a 3- phased approach to workload assessment provided Transportation Safety Board investigators insight into how operating conditions for approaches into a non-radar airport under instrument meteorological conditions impact crew workload.
It was possible to develop and use secondary task questions for three of four cognitive task domains. Qualitative assessment of verbal responses illustrated how crews use verbal information to support mental models. A trend towards longer response times for the cognitively more demanding questions supported the hypothesis that maintaining situational awareness of flight status within the approach sequence is cognitively more demanding than monitoring flight control indicators. Changes in heart rate variability could be linked to changes in task demands. NASA Task Load Index data provided quantitative and qualitative indicators of overall workload and demonstrated that high workload conditions can be triggered by a variety of operational conditions.

In the course of an occurrence investigation, A09W0037 (2009), Transportation Safety Board investigators were interested in quantifying crew workload on an approach to a non-radar airport under instrument meteorological conditions. The approach included an unusual hold configuration. The CRJ705 aircraft was hand flown from glide-slope intercept using a Head-up Guidance System (HGS). Investigators were interested in gaining insight into the relative crew workload in this condition compared to a standard ILS + autopilot approach into the same airport.

In 2003, the European Organization for the Safety of Air Navigation recommended an integrated approach to the assessment of crew workload including performance-based measures, subjective ratings and physiological arousal measures (CARE-Integra-TRS-130-02-WP2, 2003). The rational for developing a 3-step approach to workload assessment acknowledged that workload cannot be implied from task analysis alone as the cognitive resources applied to a task will differ markedly between experienced and novice operators. Similarly, measures of individual effort or arousal in response to task load may also not uniquely reflect cognitive workload as the operator may not increase effort level to meet operational demands. Finally, accuracy of primary task performance will not provide evidence of the cognitive reserve available to handle unexpected events.

Methodology

Simulation trials were conducted using the CRJ simulator at the CAE training facility in Toronto, Canada. Two volunteer crews (matched in age and experience with the occurrence crew) were instructed that they would be flying simulated approach and landings into Whitehorse airport under Instrument Meteorological Conditions. They were given time to review the Jeppesen plates prior to entering the simulator. The crews were assigned one of two simulation trial sequences: HGS followed by Auto-pilot +ILS or Auto-pilot +ILS followed by HGS.

Based on CARE, 2003 recommendations three sets of measurements were taken:

1. Secondary Task Performance
   During each trial, a series of probe questions were presented to each Captain and First Officer. The probes were designed to challenge the crew's cognitive awareness. Each probe challenged one of three cognitive tasks domains based on the categorization strategy described by Anding (2008). Domain 1 probes challenged the crew's awareness of operation conditions that are considered to be attributes of tasks (primary tasks) within the current focus of attention. Domain 2 probes challenged the crew's awareness of events that are within the crew's current operational condition. Domain 3 probes challenged the crew's current situational awareness of

the flight within the greater context of the approach (table 1). The time interval between the end of the question and the start of the response defined the response time.

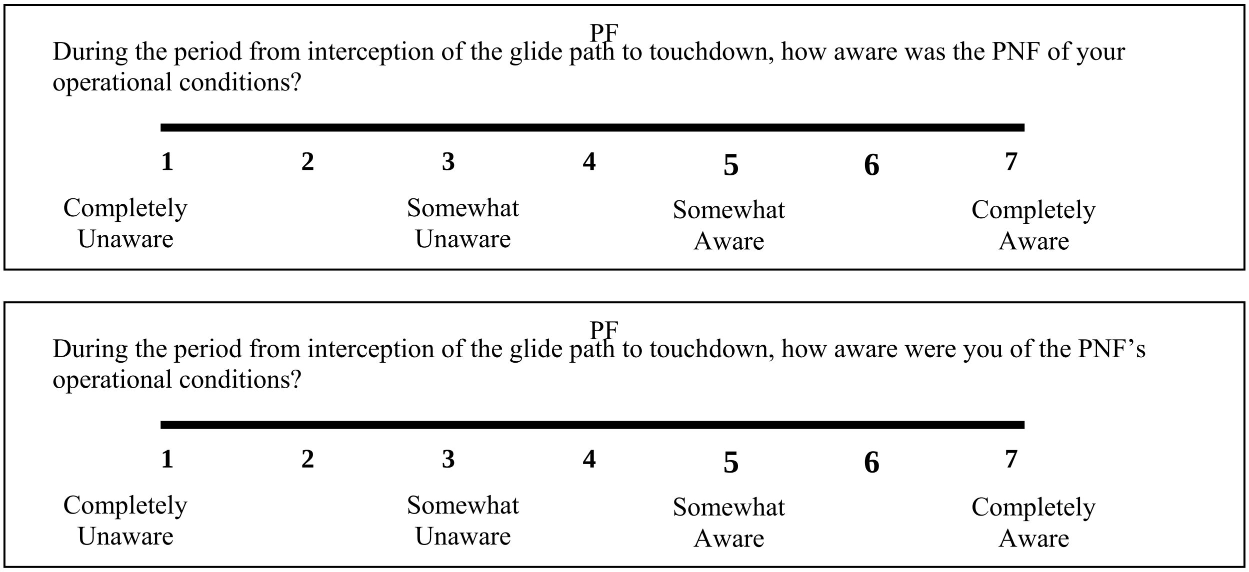
2. Heart Rate Variability
   The low frequency spectral power of all NN intervals between 0.04 and 0.15 Hz (LF) is the recommended Heart Rate Variability measure. Sampling was performed over successive 5 minute intervals via portable Holter monitors fitted by a trained technician.
3. NASA Task Load Index (TLX)
   The TLX was administered in a pencil and paper format in the cockpit immediately following each trial (http://humansystems.arc.nasa.gov/groups/TLX/). The task context for the TLX ratings was specified as after intercepting the glide slope. The TLX were estimated by each crew member for both primary and secondary tasks. PF: primary task: Establish and maintain stabilized flight, PF secondary task: Error trap miscommunications and missed communications. PNF primary task: Perform all radio and intercom tasks. PNF secondary task: Monitor aircraft performance and provide feedback to PF re departures from stabilized flight

Table 1.
Sample Cognitive Domain questions.

Secondary Task Probes		**Cognitive Domain**
PF	**PNF**	
Captain, what is your current altitude?	First Officer, what is the current wind direction and speed at this altitude?	1
Captain what was the FO's last call back to ATC	First Officer, what is your current altitude and estimated time to missed approach?	2
Captain what is the traffic ahead of you	First Officer, what is the timing from Robinson inbound to the missed approach point?	3

Note. Domain 1 challenges crew's awareness of operation conditions that are considered to be attributes of tasks (primary tasks) within the current focus of attention. Domain 2 probes challenged the crew's awareness of events that are within the crew's current operational condition. Domain 3 probes challenged the crew's current situational awareness of the flight within the greater context of the approach.

In addition to these three measures, a fourth subjective assessment of flight deck mutual awareness was administered after each simulation trial (Figure 1).



PF
During the period from interception of the glide path to touchdown, how aware was the PNF of your operational conditions?

PF
During the period from interception of the glide path to touchdown, how aware were you of the PNF's operational conditions?

Figure 1. Mutual Awareness Assessment scales for the PF. The upper scale allows the PF to estimate how aware he thought the PNF was of the PF's operational conditions. The Lower Scale allows the PF to estimate his awareness of the PNFs operational conditions. A similar set of scales were presented to the PNF.

Results

Secondary Task Analysis.

Figure 2 illustrates the average response times to cognitive domain questions by the PFs and PNFs in both HGS and ILS conditions. Responses to Cognitive Domain level 3 questions were significantly longer than level 1 questions (p=0.00125). These data suggest that we were successful in designing probe questions that challenged different cognitive demand levels. There was no significant difference between responses in the HGS compared to the ILS operational condition indicating that both conditions provided similar cognitive challenges.

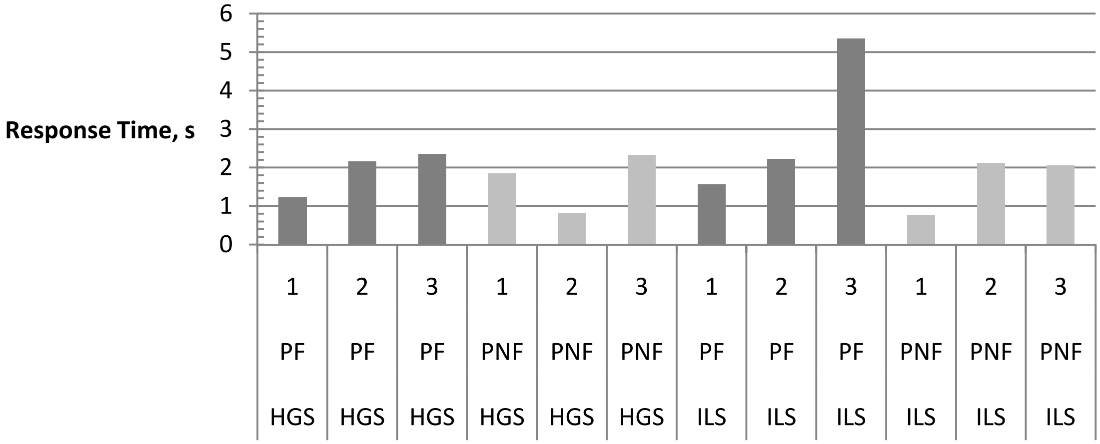


Figure 2. Average response time in seconds to cognitive domain questions (levels 1, 2, and 3) by PFs and PNFs in HGS and ILS operational conditions

Content analysis (Table 2) of the responses revealed similar recall deviations to level 2 and 3 cognitive domain questions in HGS and ILS conditions. The PFs typically did not provide all components of tower communications and confused the sequencing of Tower and Center communications. PNFs delayed call-back request by Center and miscalculated missed approach timing. Whitehorse timing for missed approach should be referenced to the Robinson NDB which is 17 miles from the airport.

Table 2
Sample responses from PFs in response to level 2 and 3 cognitive domain questions

Probe Question	**Response**	**Interpretation**
Captain please recall as many details as you can about the most recent ATC call	C1 ah The call was to contact Tower... cleared for the approach #2 contact Tower and second radio call him back at 9000 feet.	The very last communication from ATC was that a CRFI report would be available when contact with tower is made. The response given was for the second last call which contained the most recent clearance and instructions. Captain recalled 4/7 items
Captain what was the FO's last call back to ATC	Ah… He acknowledged the maintain this heading	FO also read back the altitude of 11000 and that the hold clearance would be reissued. Captain remembered approximately 1/3 of what FO read back.
Captain what does the tower know about the status of your approach	C1 ah knows that we are 9000ft and does not know that we have crossed the final approach fix yet	The tower knows that they are on the ILS31L and are to call 10 miles back. His response about 9000 feet is related to the centre controller.
	C2 ahhh we called him 10 miles and he needs an extra call at five and we are on the ILS	Captain did not recall that FO made call at 6 miles out

Heart Rate variability

Figure 3 illustrates the Low Frequency (LF) heart rate variability measures for the first and second crew for both HGS and ILS trials. As mental workload goes up, the LF heart rate variability measure goes down.

The first trial for Crew 1 was curtailed due to a SIM malfunction that induced an abrupt missed approach response from the crew at 7:40 pm. A sharp downward dip occurs in the heart rate data at approximately 7:40 pm for both the captain and the first officer indicating that work load increased quickly just before the trial was stopped. Additional dips can be seen in the Captain's trace between 7:05-7:10 and 7:25-7:30. These time intervals correspond to Edmonton Center initially updating the hold sequence and the communications transfer between Edmonton Center and Tower respectively.

The second trial (ILS) starts with both pilots indicating relatively low mental workload compared to the first trial. This is not unexpected as the crew is now more familiar with what will be expected of the trial. However, the captain's HF curve dips sharply between 8:20 and 8:25 pm which corresponds to the time when the crew realized that they intercepted the localizer above the glide path. This required manipulating the Flight Management System to increase the rate of descent to intercept the glide path from above.

For the second crew, the upward trending of the heart rate variability measures appears to indicate an easing of the workload for the second trial. In this trial, it would appear that the Captain's workload is high and is maintained throughout the trials. There are two possibilities to explain the relatively flat heart rate variability data demonstrated by the second captain. Firstly, this captain provided significantly more verbal expression of his thought strategy which he shared with his first officer. It was clear that he was thinking ahead out loud and maintaining awareness of the FO's understanding of the flight status. Secondly, the act of talking itself can serve to disrupt the heart rate variability measure.
The apparent lowering of workload for the first officer between 9:05-9:10 and 9:55-10:00 corresponds to being in revised hold patterns at Robinson and ELTAG respectively.

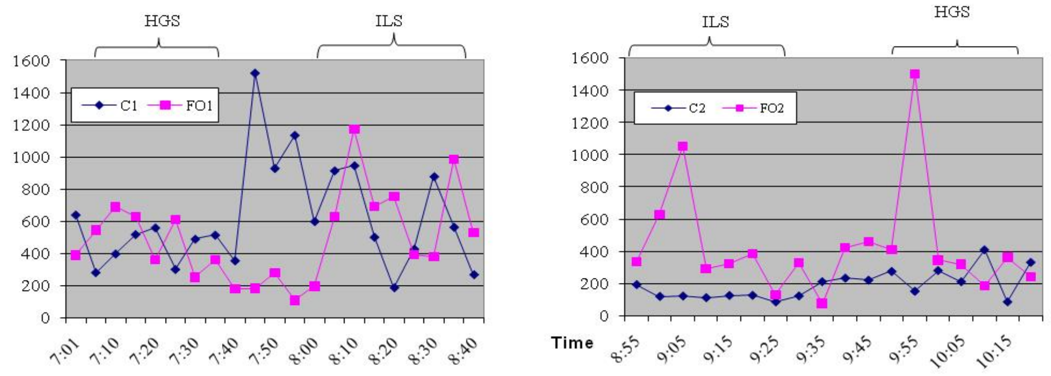


Figure 3. Heart Rate variability measures for C1/FO1 and C2/FO2 as a function of time

NASA Task Load Index measures

Tables 5 and 6 tabulate the overall TLX ratings for the first and second crews respectively. The first Crew generally assessed the task load to be somewhat higher in the ILS condition than the HGS condition. The captain volunteered that the increased workload was largely attributable to the fact that he took Whitehorse clearance above glide slope and had to correct for it on landing.

The second crew rated the overall Task Load slightly higher under the HGS condition compared to the ILS condition. These findings are consistent with the heart-rate variability measures where the second crew captain clearly showed reduced heart rate variability consistent with a higher mental task load.

Table 3

Overall TLX Ratings in percent for Crew 1 and 2

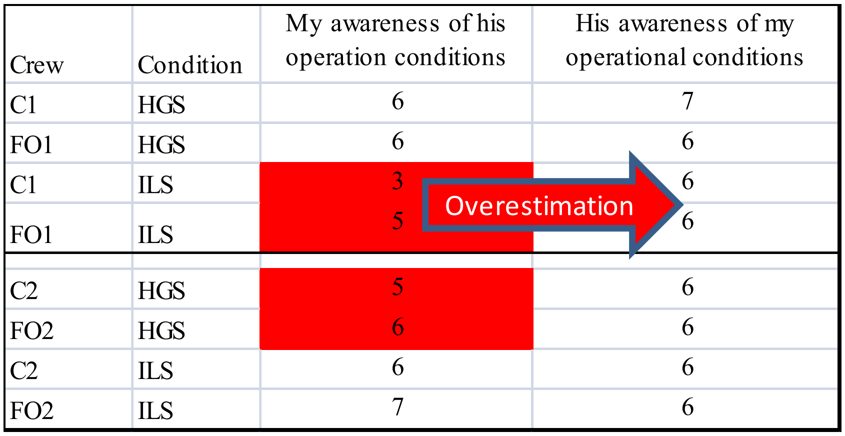
Overall TLX Rating, %		
	HGS	ILS
FO1P	66	73
FO1S	49	51
C1P	57	72
C1S	65	79

Overall TLX Rating, %		
	HGS	ILS
FO2P	75	69
FO2S	74	68
C2P	72	64
C2S	70	67

The second crew did demonstrate a small, systematic shift towards higher workload ratings in the HGS mode. Their HGS scores were at or above 70 which is considered to be the threshold for high workload (Hancock, 2009). According to Hancock there are no guidelines as to how long high workloads can or should be sustained. While this crew did show a small effect for the HGS mode, the first crew did not. Their ratings were far more influenced by the operational demands of intercepting the glide slope at an altitude somewhat higher than optimum.

Mutual Awareness Ratings

The final assessment component of mutual awareness indicated that crews appear to have sufficient self awareness of their ability to monitor their own operational conditions. In moments of high workload there appears to be a significant risk that crews will overestimation the ability of the other pilot to maintain situational awareness levels.



Crew	Condition	My awareness of his operation conditions	His awareness of my operational conditions
C1	HGS	6	7
FO1	HGS	6	6
C1	ILS	3	6
FO1	ILS	5	6
C2	HGS	5	6
FO2	HGS	6	6
C2	ILS	6	6
FO2	ILS	7	6

Overestimation

***Figure** 4* Mutual awareness ratings for both crews in each condition

Observations

The current data suggest that use of specific cockpit technologies per se is unlikely to be intrinsically associated with high workload. Rather, there are likely to be significant inter crew, and possibly inter-pilot differences in perceived work load based on experience and comfort level with the technology. Furthermore, operational decisions such as descending to meet the glide path after intercepting the localizer produce similarly high work load conditions as unfamiliarity with particular cockpit technologies. Given that the occurrence captain volunteered that he was "locked-on" to the Heads-Up Guidance System display, and the occurrence flight was the First Officers first live HGS approach, it is possible that they were both approaching a performance-based maximum workload for the duration of the flight after intercepting the glide slope. Finally, mutual awareness rating data indicate that when high workload conditions arise, crew members may become aware of their colleague's more channeled attention but may not understand the impact that it has on mutual awareness. As a consequence there appear to be no strategies to facilitate the restoration of divided attention behaviors necessary to maintain optimum situational awareness.

References

A09W0037 (2009) Air Investigation Report: Risk of Collision between Air Canada Jazz Bombardier CL-600-2D15, C-FDJZ and Airport Maintenance Vehicles (Snow Sweepers) Whitehorse International Airport, Yukon 06 March 09.

Anding, P.N. (2008) A generalized model of cognitive task categorization, Retrieved from http://www.kmice.uum.edu.my/kmice08/listofPapers.html

CARE-Integra-TRS-130-02-WP2 (2003). Review of Workload measurement, Analysis and Interpretation Methods, retrieved from http://www.eurocontrol.int/integra/gallery/content/public/documents/expt_def_wp2_final.pdf

Hancock, Peter (2009). Personal Communication. Provost's Distinguished Research Professor, Applied experimental Human Factors, University of Florida

NEXTGEN FLIGHT DECK SURFACE TRAJECTORY-BASED OPERATIONS (STBO): SPEED-BASED TAXI CLEARANCES

Deborah L. Bakowski[1], David C. Foyle[2], Christina L. Kunkle[1], Becky L. Hooey[1], Kevin P. Jordan[1]
[1]San Jose State University at NASA Ames Research Center, Moffett Field, CA
[2]NASA Ames Research Center, Moffett Field, CA

A pilot-in-the-loop simulation was conducted that required pilots to taxi following acceleration and speed profiles under two Speed-conformance conditions (Defined and Undefined). Pilots were given a commanded speed in both conditions, however, in the Defined Speed-conformance condition, air traffic control (ATC) issued alerts when the aircraft speed exceeded a +/- 1.5 kt speed range. A current-day, baseline trial with no required speed profile was also included. While pilots achieved required time of arrival (RTA) errors of less than 10 sec in each condition, both Speed-conformance conditions produced more visual fixation time on the speed tape, located head-down on the primary flight display (PFD), compared to the baseline condition. Fourteen out of eighteen pilots reported that the demand of maintaining the required speed conformance range in actual operations would compromise safety. These results indicate the need for advanced flight deck displays to enable pilots to safely comply with runway RTAs during taxi.

The present study investigated the taxi-out departure environment (from the ramp area to the runway) of the next generation (NextGen; JPDO, 2009) of the National Airspace System. Current research efforts are aimed toward the development of surface traffic management (STM) systems for air traffic control (ATC) to provide optimized taxi clearances that eliminate active runway crossing delays and enable more efficient use of runways. These taxi clearances would have a speed- or time-based component with which the pilot must comply. These NextGen taxi operations have been referred to as “4-D taxi” (with the fourth dimension referring to the time component), or surface trajectory-based operations (STBO). STM systems are envisioned to use dynamic algorithms to generate speed- or time-based taxi clearances for aircraft to calculate the most efficient movement of all surface traffic and enable precise surface coordination (Cheng, Yeh, Diaz, & Foyle, 2004; Rathinam, Montoya, & Jung, 2008). To accomplish the required precision, the STM system provides speed/time commands to pilots throughout the taxi route, such that they arrive at certain airport “traffic flow points” (e.g., traffic merge points, active runway crossings, etc.) at specific times. The aircraft’s speed may need to be adjusted if the pilot is unable to conform to the STBO command, or if traffic is unable to comply creating a reduction in separation, or to meet the needs of the dynamic airport surface.

Previous Research

Foyle, Hooey, Kunkle, Schwirzke, and Bakowski (2009) investigated the impact of speed commands on pilots’ ability to meet a required time of arrival (RTA) at traffic flow points between a ramp departure spot and the runway. The goal of the study was to drive aircraft to a specified location (runway end or traffic flow point) at a specific time by having ATC provide the pilot a taxi clearance with a commanded speed to be followed. (Note that speed is the parameter that pilots actually control via throttle inputs – arrival time derives from that speed control. If ATC provided clearances with a time requirement, pilots would have to transform that into speed, taking route distance into account.) In the ‘Limited’ NextGen condition[1], the Primary Flight Display (PFD) presented the commanded and current ground speeds. Trials consisted of 1, 3, or 5 segments, where each segment had a commanded speed of 10, 14, 18, or 22 kts. Because the total distance of all trials was similar, segment distance in one-segment trials was longer than in the three- or five-segment trials. Pilots were instructed to comply with the commanded speed on straight segments, accelerate/decelerate “aggressively”, and, for commanded speeds of 18 and 22 only, slow to 15 kts for turns. The RTA was originally calculated using the taxi route segment length and the ATC-commanded speed for the straight segments. The primary measure of performance was time of arrival (TOA) error, calculated by subtracting the RTA from the observed arrival time at each segment transition. The results indicated TOA errors of -24 sec (early) to 53 sec (late) for one-segment trials. However, it was noted that the RTA

---

[1] The study examined two conditions of speed commands, which varied according to the complexity of the flight deck avionics, however only the ‘Limited’ NextGen condition is discussed here.

calculation used in Foyle et al. did not account for turns or acceleration/deceleration, which may have exaggerated true TOA errors. To correct this, and provide a suitable comparison to the present study, the RTA for each segment was recalculated to account for the commanded speed in turns and an assumed underlying speed profile of 2 kts/sec acceleration/deceleration, commonly considered the maximum acceleration/deceleration limit (Cheng, Sharma, & Foyle, 2001). TOA scored against the recalculated RTA, is shown in Figure 1, and as can be seen, mean TOA errors ranged from -28 sec (early) to 27 sec (late) for one-segment trials. That is, even with a more precise RTA calculation method, there is still considerable TOA error.

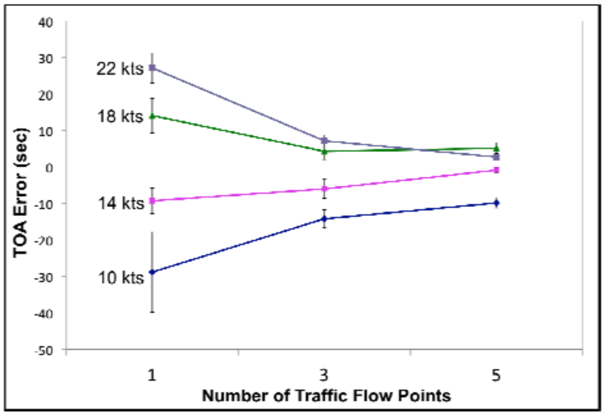


Figure 1. TOA Error in Limited NextGen Condition (Foyle et al., 2009) using adjusted RTA calculation. Negative TOA errors indicate that the pilot taxied too quickly and therefore arrived early. Positive TOA errors indicate that the pilot taxied too slowly and therefore arrived late. (Error bars are +/- 1 Standard Error).

The goal of the present study was to investigate the possibility of improving TOA performance beyond what was observed in the Limited NextGen condition through the manipulation of procedural instructions. Two flight deck procedures were implemented in the present study with the expectation of reducing TOA error: 1) Requiring pilots to follow a specified speed profile with acceleration/deceleration rate of 2 kts/sec, and specific speeds on straightaways and in turns; and, 2) Specifying a required speed conformance level (specifically +/- 1.5 kts) from the commanded speed. The purpose was to evaluate the impact of these procedural manipulations on pilots' ability to comply to taxi clearances with a speed requirement (resulting in an RTA).

Method

Participants

Eighteen commercial pilots (both current and recently retired) participated in the study. The mean pilot age was 45 years with a range of 25 to 65 years. Thirteen of the pilots were Captains and five were First Officers. The mean flight hours logged was 3,832 hours (range of 300 – 11,000 hours).

Flight Simulation

The study was conducted in a medium-fidelity, part-task simulator in the Human-Centered Systems Laboratory (HCSL) at the NASA Ames Research Center. The airport environment was the Dallas/Fort Worth International Airport (DFW) with high visibility and distant fog/haze conditions. Aircraft controls included a tiller side-stick control with left/right rotation for nose-wheel control, non-differential throttle, and toe brakes. A B737 aircraft simulation control model was used. Participants wore an Applied Science Laboratory Mobile Eye with EyeHead Integration eyetracker.

The forward out-the-window scene was rear-projected on a 2.44 m horizontal (53.13 deg visual angle) by 1.83 m vertical (41.11 deg) screen located 2.44 m in front of the pilot's eye point. The side window scenes were presented on two 48.26 cm (19-in diagonal) monitors, one on each side of the pilot, at a viewing distance of 0.91 m (29.57 deg visual angle). The simulator flight deck included a Primary Flight Display (PFD), Navigation Display (ND), Taxi Navigation Display (TND), Datalink Display, and an Electronic Checklist.

Primary Flight Display (PFD). The PFD was modified for taxi operations by expanding (doubling) the speed scale (left) from 0-60 kts to support taxi operations. Current ground speed (14 kts in Figure 2 left) was shown as a white sliding indicator with a digital, whole number value inside. Commanded speed was issued verbally by ATC and was not displayed on the PFD. The PFD was identical in both conditions (Undefined and Defined).

Taxi Navigation Display (TND). To assist in navigation of the airport, the simulator flight deck included a TND that depicted the airport layout in a track-up perspective during taxi (see Figure 2 middle). The ownship aircraft's position, shown as a white chevron, and other aircraft traffic within the ownship's 1,250 ft declutter circle, shown as yellow aircraft icons, were updated in real time. The taxi clearance, presented graphically as a magenta route, indicated a positive cleared-to-cross runway clearance.

Navigation Display (ND). The ND was used to present a graphic view of the tailored departure path, representing what had been loaded into the Flight Management System (FMS). When the departure route was first datalinked to the cockpit, the pending departure path was shown as a white dashed route. The departure route of the lead aircraft was represented by a solid green line. Once accepted by the pilot, the ownship's route was shown as a solid magenta path (see Figure 2 right).

Figure 2. PFD (left), Taxi Navigation Display (middle), and Navigation Display (right).

Datalink Display. The datalink display was used to present a text description of the tailored departure path, representing what had been sent from ATC. An auditory chime accompanied the delivery of the datalink, which pilots viewed by alternating between the TND and the datalink text. Pilots were responsible for verifying that the 4-D departure path sent by ATC was the same as that received and loaded in the FMS.

Experimental Design

The taxi task portion of the experiment consisted of three within-participant factors: Speed-conformance condition (Undefined and Defined); Number of traffic flow points (1, 3, and 5); and, Commanded speed (14, 18, and 22 kts). These three factors were crossed factorially to create nine nominal trials in each of the two Speed-conformance conditions. For all participants, the Undefined Speed-conformance condition trials were tested first, followed by the Defined Speed-conformance trials. Testing was done in this order so that performance in the Undefined Speed-conformance condition represented the pilots' "natural", uninstructed speed conformance level. A single "current-day taxi" trial with no commanded speed was presented at the midpoint of the study. In addition, three off-nominal trials were tested, but they are not discussed here.

Procedure

Three familiarization trials were presented prior to the start of the Undefined condition and a fourth was presented prior to the start the Defined condition. In each of the experimental trials, pilots taxied from a ramp departure spot to the departure runway. Pilots received a speed-based taxi clearance during each trial. A verbal ATC command (e.g., "NASA227, taxi at 14 kts.") accompanied each taxi segment transition (14, 18, or 22 kts) at the traffic flow point location. Segment distances and speed changes were not depicted on the TND. Two specific speed profile instructions were given, and, with the exception of the baseline trial, applied to all trials in both implementations: Taxi all turns at 14 kts and accelerate/decelerate at a rate of 2 kts/sec (e.g., a 0 kt to 14 kt initial acceleration should take 7 sec, a 22 kt to 18 kt speed change should take 2 sec, etc.). In the baseline (current-day) trial, pilots were not given a commanded speed and were instructed to taxi as they normally would in actual operations. With the exception of the absence of a speed command, all other requirements of this trial were the same as the other experimental trials (e.g., checklist task, departure clearance verification task, navigation, maintaining safe separation, etc.).

In the Undefined Speed-conformance condition, pilots were instructed to taxi as close to the commanded speed as was reasonable. No required speed conformance range or performance feedback was provided in this condition. However, in the Defined Speed-conformance condition, pilots were instructed to taxi within +/- 1.5 kts of the commanded speed. When ground speed exceeded the +/- 1.5 kt range for more than a continuous five-sec period, ATC delivered a verbal alert, "NASA227, check speed". The verbal alert repeated every 10 sec until the pilot's speed returned to within the +/- 1.5 kt range from the commanded speed. The computer-activated ATC "check speed" alert was disabled for 2-4 sec (depending on the size of the speed change for acceleration/deceleration to the new speed) after a new speed command was issued by ATC, and in the area around the turns.

In every trial, a departure clearance was datalinked to the cockpit and auto-loaded to the FMS and delivered at one of three possible times: 1) prior to the start of taxi; 2) in the first half of the taxi route; or, 3) in the second half of the taxi route. Pilots were required to crosscheck the altitude, heading, and speed in the departure datalink text message against the route displayed graphically on the ND (Figure 2 right). As a secondary task to emulate pilot workload, pilots were required to monitor a simplified electronic checklist on the instrument panel to ensure that all items were checked; a specified button press on the tiller was required when an item became unchecked according to a randomly timed schedule. A paired departure task was also presented following the completion of the taxi task in each trial. This departure task was examined independently of the taxi task and is not discussed here.

Following each taxi trial, pilots supplied subjective ratings of situation awareness, workload, and impact of the departure clearance on taxi performance. At the completion of each block of Speed-conformance (Undefined and Defined) trials, pilots also completed questionnaires that pertained to display usage, safety, and acceptability of the speed conformance condition.

Results

The taxi TOA error analyses included nine nominal trials in each condition, Undefined Speed-conformance condition and Defined Speed-conformance condition. The primary measure of pilot performance on the taxi task was TOA error, which was calculated by subtracting the RTA from the observed arrival time. The RTA for each segment was calculated using the taxi route segment length, the ATC-commanded speed for the straight segments, 2 kts/sec acceleration/deceleration, and a turn speed of 14 kts.

As seen in Figure 3, TOA error is quite good, and is much better than found in Foyle et al. (2009, see Figure 1) where no specific speed profiles or speed conformance requirements were placed on the pilots. A 2 (speed-conformance condition) by 3 (number of traffic flow points) by 3 (commanded taxi speeds) within-participants ANOVA showed that there was an interaction between number of traffic flow points and commanded taxi speed, $F(4,68)=3.44$, $p=.013$. A trend analysis revealed a significant Linear by Linear interaction, $F(1,17)=8.15$, $p=.011$. This suggests that TOA error increased linearly as commanded speed increased from 14 kts to 22 kts, and decreased as the number of traffic flow points increased from one to five.

There was also an interaction between Speed-conformance condition and number of traffic flow points, $F(2,34)=4.44$, $p=.019$. Post hoc tests showed a simple main effect of number of traffic flow points in the Defined Speed-conformance condition, $F(2,34)=7.75$, $p=.002$. TOA error for one-segment trials ($M=2.28$, $SD=3.12$) was significantly higher than for three-segment trials ($M=.71$, $SD=1.75$), $t(17)=2.70$, $p=.015$, and five-segment trials ($M=.39$, $SD=1.15$), $t(17)=3.21$, $p=.005$. This was consistent with the Foyle et al. (2009) results (see Figure 1), in which pilots exhibited more difficulty maintaining a commanded taxi speed for a long distance (as in the one-segment trials), than for shorter distances (three- or five-segment trials). The simple main effect of number of traffic flow points in the Undefined condition was not significant.

A main effect of speed was also found, $F(2,34)=21.83$, $p<.001$. A significant Linear effect, $F(1,17)=39.57$, $p<.001$ suggests that TOA error increased as the commanded speed increased.

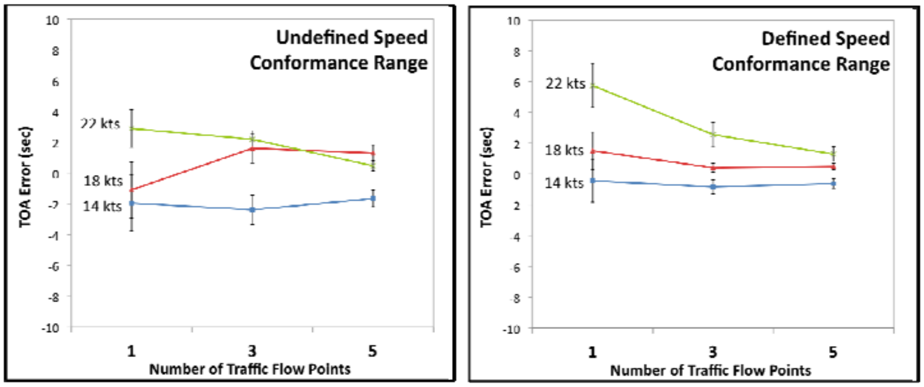


Figure 3. Mean TOA Error as a function of commanded speed and number of traffic flow points for the Undefined and Defined Speed-conformance conditions. Negative TOA errors indicate that the pilot taxied too quickly and therefore arrived early. Positive TOA errors indicate that the pilot taxied too slowly and therefore arrived late. (Error bars are +/- 1 Standard Error).

Percent Dwell Time on PFD Speed

The percentage of the taxi trial during which the pilots' eyes fixated (termed percent dwell time, PDT) on the current speed read-out displayed on the PFD is shown in Figure 4 (left). The Undefined and Defined conditions respectively produced 2.4 and 3.3 times more visual fixation time on the PFD speed display when compared to that of the current-day baseline trial (*M*=7.55 PDT) where pilots taxied with no commanded speed, *F*(2,20)=41.29, *p*<.001. Pilots spent 17.89 and 24.39 percent of the trial looking at the speed display, in the Undefined and Defined conditions respectively. In absolute as well as relative terms, this is a large percentage of the trial to be looking head-down at the speed display when the main duties of the Captain are to navigate and control the aircraft, and maintain awareness and separation from other aircraft taxiing on the airport surface.

Questionnaire Data

On a post-study questionnaire, using a 5-point scale, where 1 = Rarely and 5 = Most of the Time, participants were asked, 'How often did you find yourself focusing on the PFD speed tape when you would have preferred to have been paying attention to the external taxiway environment?' As shown in Figure 4 (right), the results showed a significant effect of Speed-conformance condition, *F*(2,34)=69.19, *p*<.001. Participants reported focusing on the speed tape more than they would have preferred in the Defined Speed-conformance condition (*M*=3.69, *SD*=.79) than in either the Undefined (*M*=3.20, *SD*=.79), *t*(17)=3.35, *p*=.004, or baseline conditions (*M*=1.67, *SD*=.77), *t*(17)=9.71, *p*<.001. Pilots also reported focusing on the speed tape more than they would have preferred in the Undefined than in the baseline condition, *t*(17)=8.59, *p*<.001.

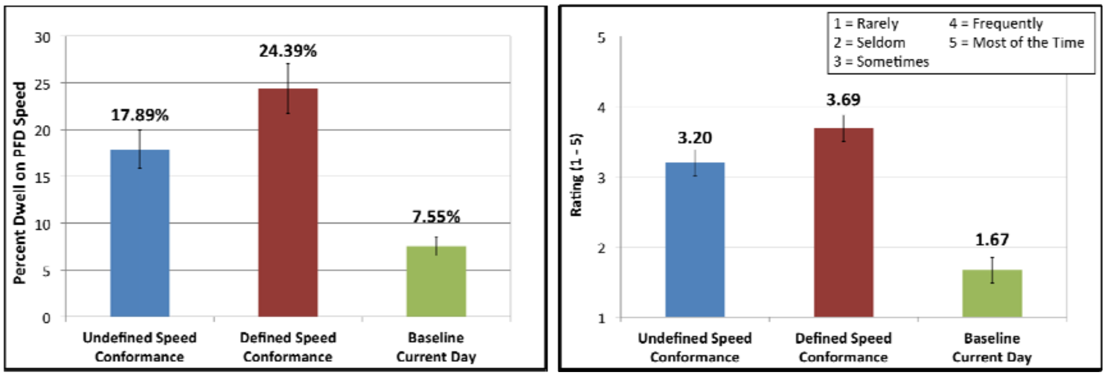


Figure 4. Mean percent dwell time on PFD speed (left). Mean post-trial rating of question: How often did you find yourself focusing on the PFD speed tape when you would have preferred to have been paying attention to the external taxiway environment? (right). Error bars represent +/- 1 standard error.

Participants were asked, ‘Would the demand of having to maintain the required speed conformance range compromise safety in the real world?’ A chi-square goodness-of-fit test revealed that more pilots (n=14) responded that the demand of having to maintain the required speed conformance range in the real world would compromise safety than responded that it would not (n=4), $X^2(1, N=18)=5.56$, $p=.018$.

Participants were also asked, under the simulated future airport concept, in which all aircraft are similarly equipped and conducting time-based taxi operations, ‘Is it a reasonable requirement to stay within +/- 1 kt of a commanded speed?’ A chi-square goodness-of-fit test revealed that more pilots (n=15) responded that staying within +/- 1 kt of a commanded speed was not a reasonable requirement than responded that it was reasonable (n=3), $X^2(1, N=18)=8.00$, $p=.005$. On average, participants reported that a +/- 3.7 kt range would be reasonable. It should be noted that this would likely result in very poor TOA errors – a simple calculation indicates that depending on the speed, each 1 kt error bias over a 12,000 ft taxi results in 20-40 sec error.

Discussion

A previous NextGen taxi simulation study (Foyle et al., 2009) showed that pilots were not able to achieve accurate RTAs when issued speed-based taxi clearances (unless provided with an enabling flight deck algorithm and display). However, in the Foyle et al. study, pilots were not given explicit speed profiles (acceleration/deceleration) or explicit speed-conformance requirements, so as to require aircraft taxi handling that would be comparable to current-day operations. The current simulation was conducted to expand on this result. The present simulation experiment required explicit speed profiles and manipulated speed conformance. Both factors may have caused the poor RTA performance in the previous (Foyle et al.) simulation.

The present simulation demonstrated that pilots were able to taxi their aircraft according to specified speed profiles, resulting in quite good RTA performance. Unfortunately, however, this required the pilot to view the head-down speed display 2.4 to 3.3 times more than in a current-day baseline condition. Pilots overwhelmingly (14 of 18) felt that this would have a negative impact on safety, interfering with the primary taxi tasks to navigate the aircraft and maintain visual separation from other aircraft and obstacles. Thus, although pilots are able to follow a specified speed profile in response to a taxi clearance incorporating a commanded speed, resulting in good RTA performance, it also likely results in unsafe surface operations. As suggested by Foyle et al. (2009), a flight deck display aid may support pilot taxi RTA performance with reasonable and safe workload.

Acknowledgments

This work was funded by the NASA Airspace Systems Program / NextGen Concepts and Technology Development (CTD) Project / Safe and Efficient Surface Operations (SESO) Element. The authors are indebted to Glenn Meyer (Dell Services, Federal Government) for experimental and analysis software support.

References

Cheng, V. H. L., Sharma, V., & Foyle, D. C. (2001). A study of aircraft taxi performance for enhancing airport surface traffic control. *IEEE Transactions on Intelligent Transportation Systems*, *2*(2), 39 - 54.

Cheng, V. H. L., 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.

Foyle, D. C., Hooey, B. L., Kunkle, C. L., Schwirzke, F. J., & Bakowski, D. L. (2009). Piloted Simulation of NextGen Time-based Taxi Clearances and Tailored Departures. *Proceedings of the 2009 IEEE/AIAA Integrated Communications, Navigation and Surveillance Conference (ICNS).* Arlington, VA, May 2009.

Joint Planning and Development Office (2009). Concept of Operations for the Next Generation Air Transport System, v3.0 (October 1, 2009). Retrieved from www.jpdo.gov/library/NextGen_ConOps_v3%200.pdf

Rathinam, S., Montoya, J., & 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.

BETTER-THAN-VISUAL TECHNOLOGIES FOR NEXT GENERATION AIR TRANSPORTATION SYSTEM TERMINAL MANEUVERING AREA OPERATIONS

Lawrence J. Prinzel III, Randall E. Bailey, Kevin J. Shelton, Denise R. Jones,
Lynda J. Kramer, Jarvis J. Arthur III, Steve P. Williams, Bryan E. Barmore,
Kyle E. Ellis, and Sherri A. Rehfeld

NASA Langley Research Center
Crew Systems and Aviation Operations Branch
Hampton, VA

A consortium of industry, academia and government agencies are devising new concepts for future U.S. aviation operations under the Next Generation Air Transportation System (NextGen). Many key capabilities are being identified to enable NextGen, including the concept of Equivalent Visual Operations (EVO) – replicating the capacity and safety of today's visual flight rules (VFR) in all-weather conditions. NASA is striving to develop the technologies and knowledge to enable EVO and to extend EVO towards a "Better-Than-Visual" (BTV) operational concept. The BTV operational concept uses an electronic means to provide sufficient visual references of the external world and other required flight references on flight deck displays that enable VFR-like operational tempos and maintain and improve the safety of VFR while using VFR-like procedures in all-weather conditions. NASA Langley Research Center (LaRC) research on technologies to enable the concept of BTV is described.

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 significantly increased system capacity to that of current operating levels. New capabilities are envisioned for NextGen, including four-dimensional trajectory (4DT)-based operations, performance-based navigation, EVO, super density arrival/departure operations, network-centric operations, and digital data-link communication.

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 might 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. A high-level description of NASA Langley Research Center flight deck interface technology and research issues for these areas are described with references for further reading.

Synthetic and Enhanced Vision Systems

Synthetic and Enhanced Vision System (SEVS) technologies are emerging as standard equipage on today's flight deck. These technologies form the backbone of a BTV operational concept (Bailey, Prinzel, Kramer, and Young, 2011). SEVS generates intuitive visual references for the flight crew/pilot to fly the aircraft as if in visual flight conditions independent of the actual visibility or lighting conditions (see Figure 1). NASA LaRC research aims to extend the present-day SEVS concepts to enable VFR-like operational tempos and maintain and improve the safety of VFR while using VFR-like procedures in all-weather conditions. To meet this potential, research is focused on SEVS technology development and human-in-the-loop performance to enable a 'visual' approach, landing, roll-out, and surface operations down to 300 ft actual Runway Visibility Range. This BTV operational concept suggests that the minimum aviation system performance standard for BTV technologies should be that as defined by human performance in the same operation using windows during today's VFR operations. Significant research is required to quantify this hypothesized performance standard, and more importantly, to determine if it is indeed an appropriate and sufficient standard for BTV. SEVS work includes the development of fusion methods for synthetic and enhanced vision systems; feature extraction by use of real-time imaging sensors; on-board navigational, sensor, and database integrity monitoring; and appropriate display methods for Head-Up Displays (HUD) and Head-Worn Displays (HWD), primary flight and navigation displays, and electronic flight bags.

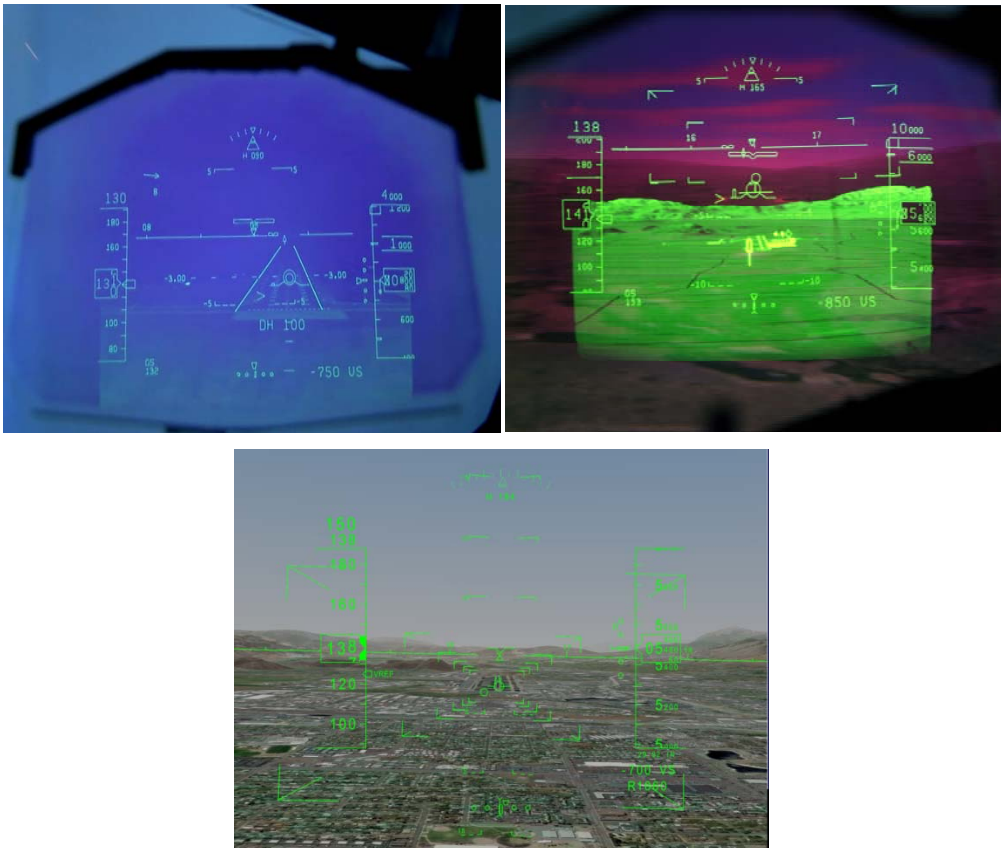


Figure 1. Enhanced Vision HUD (Upper Left), Synthetic Vision HUD (Upper Right), Synthetic Primary Flight Display

Flight Deck Interval Management

Flight Deck Interval Management (FIM) leverages advancements in communications, surveillance, and navigation (CNS) to enable flight crews to precisely space their aircraft relative to another aircraft in the terminal maneuvering area. The goal is to improve airport throughput and reduce delays. Under FIM, the air traffic controller instructs the participating aircraft to achieve an assigned inter-arrival spacing interval at the runway threshold, relative to another aircraft, using on-board automation. The flight crew then takes responsibility to actively fly the FIM operation but the Air Navigation Services Provider (ANSP) retains the responsibility for aircraft separation. NASA LaRC research has demonstrated the efficacy of the concept and system-wide and algorithm effects (Barmore, 2009). Research has recently been completed showing the synergistic potential of combining FIM and SEVS technologies (Figure 2), broaching the concept of BTV. Simulation testing showed the ability of flight crews to self-separate, wherein the pilot/flight crew accepted responsibility for separation from the designated ‘paired’ aircraft, and maintained an “equivalent visual contact” through the use of ADS-B In and SEVS technologies (Figure 3). Spacing intervals during self-separation approaches followed VFR-like operational profiles while maintaining a very high degree of flight precision, stabilized approach procedures, and excellent traffic/situation awareness. Further, ego- and exo-centric display concepts for terrain, traffic, and airport surface conditions kept the flight crew ‘ahead’ of the operation and allowed them to easily manage the arrival, through landing, roll-out, and turn-off with acceptable workload and sufficient spare attention/workload capacity to easily react to non-normal events that were intentionally staged in the experiment.

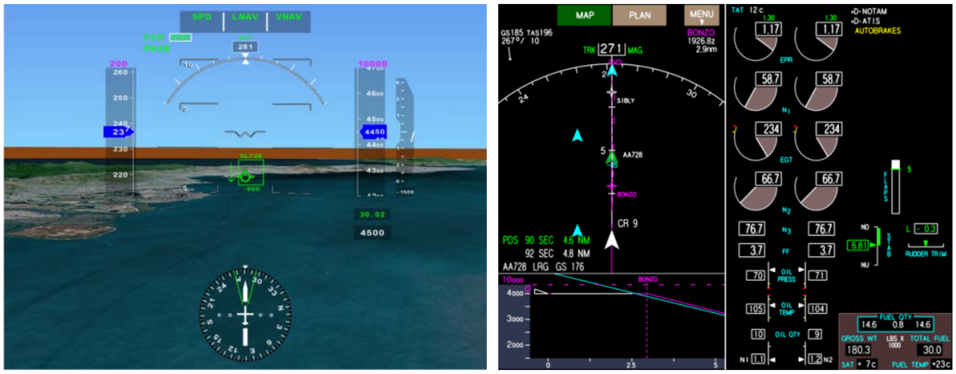
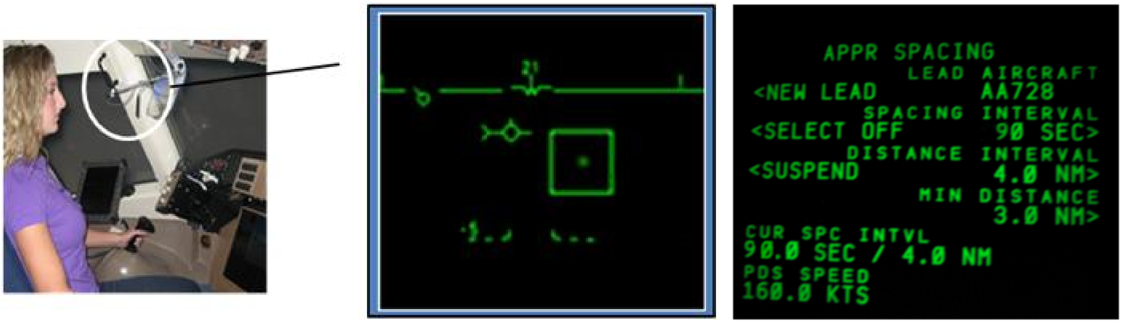


Figure 2. Flight Deck Interval Management Primary Flight (left) and Navigation Display (right)



APPR SPACING
LEAD AIRCRAFT
<NEW LEAD AA728
SPACING INTERVAL
<SELECT OFF 90 SEC>
DISTANCE INTERVAL
<SUSPEND 4.0 NM>
MIN DISTANCE
3.0 NM>
CUR SPC INTVL
90.0 SEC / 4.0 NM
PDS SPEED
160.0 KTS

Figure 3. HUD Forward-Looking Infrared of Paired Traffic and Flight Management Computer Page

Performance-Based Navigation

Performance-Based Navigation (PBN) is the umbrella term for navigation procedures being proposed to support NextGen operations which will enable aircraft to fly precisely desired flight paths leading to reduced delays, emissions, noise, and fuel costs. NASA LaRC flight deck display research has focused on the intuitive display of 4D trajectory-based operations through the use of synthetic vision pathway displays (Kramer et al, 2004; Prinzel et al, 2004) and flight path guidance and symbology (Kramer et al, 2003). This work also includes advanced decision support tools, notably Mission Rehearsal Tools (Figure 4; Arthur et al, 2006), which allow the pilot/flight crew to visualize the operation, preview procedures, and more importantly, preview the ANSP-proposed operation, the weather, and traffic using a user-friendly interface to evaluate and assess the proposed procedure before acceptance, or perform 'what-if' analyses of the proposed or alternate plans.

Surface Flight Deck Displays

Previous research from Taxiway Navigation and Situation Awareness (T-NASA) 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 (e.g., see Foyle et al, 1996; McCann et al, 1998). The use of the HUD was central to this work to promote 'eyes-out' operations, ensuring that the pilot in control used the available visual cues for tactical path control and traffic/airport awareness, augmented by conformal HUD symbology. Recent research suggests that the use of HWDs which provide unlimited field-of-regard and integrated synthetic and enhanced vision with HUD-like symbology might offer additional benefits (Figure 5; Arthur, Prinzel, et al., 2006). Research is being conducted to address the operational confounds of symbology or imagery occlusion and/or attention capture when using a HWD during this 'augmented reality' environment.

Follow normal procedure except: At 9235' MSL (2700' AFL), clean up on schedule, climb at Ref+80 select MCT. Fly published ground track to 10,000' MSL passing 219G, but no later than F204K, turn right direct JESIE. Continue climb to assigned altitude.

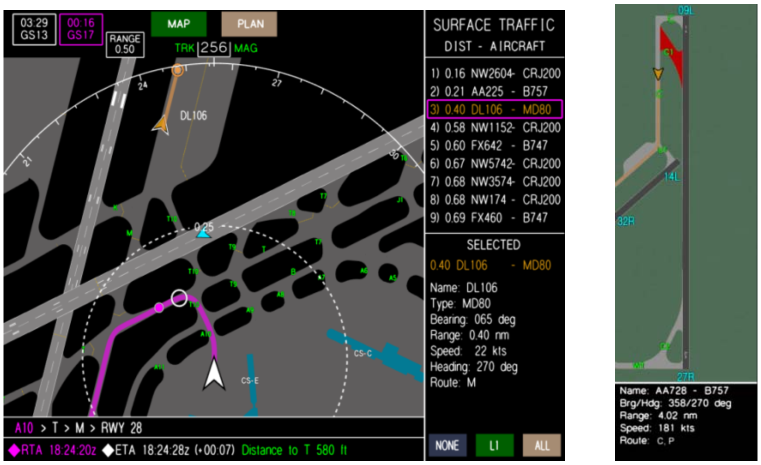
Figure 4. Electronic Flight Bag On-Board Graphical PBN Planning Tools

The surface flight deck displays are not limited to head-up and head-mounted displays but extend to flight deck (cockpit) displays of traffic information (CDTI) for surface operations (Figures 6). Research has demonstrated that surface map displays can significantly enhance situation awareness and NASA LaRC concepts have focused on intuitive graphical display, including traffic, ownship path, other traffic status and intent, and airport status information emerging from Flight Information Services-Broadcast capabilities. This work becomes critically important as emerging NextGen concepts consider trajectory-based operations on the surface, such as 4DT surface guidance (see Cheng et al, 2004; Rathinam, Montoya, and Jung, 2008). The explicit display of intent information for surface routing was found to significantly enhance pilot awareness – critical when considering that intersecting taxiways and runways create potential collision opportunities and in limited visibility conditions, the flight crew may not be aware of which aircraft is first through an intersection or what their planned routing involves. Other innovations include decision support and interface tools to improve surface operations, such as text-to-speech and speech-to-text data-link interfaces and graphical display of turn-by-turn progressive taxi instructions, take-off and roll-out guidance, and runway exit turn-off braking guidance.

Figure 5. Surface HUD / HWD Displays

Conflict Detection and Alerting

Research to develop data and design guidelines is actively being conducted to enable a comprehensive layer of indications, cautions, and warnings for safe TMA operations (Figure 7). This work initially started with runway incursion prevention where T-NASA surface display concepts were enhanced with active monitoring including predictive runway collision alerting and, if necessary, audible and visual alerting for deviations from the assigned taxi route and unauthorized crossing of a hold line (e.g., Jones et al, 2001). This work has been expanded for tactical and strategic surface operations awareness and Conflict Detection and Resolution (CD&R) functionality for NextGen operations, including taxiway and 4DT surface operational concepts (Green 2006, Jones 2002 and 2005, Jones, et. al., 2001, Jones and Prinzel, 2006; 2011; Jones, Prinzel, et al., 2010). This work includes the criticality of surveillance performance and operational scenario interactions emerging in NextGen. Monte Carlo and human-in-the-loop testing are being conducted in a complementary fashion to identify the desired/required operational CD&R functionality, including definitions of acceptable missed detection and nuisance alerting for NextGen.



SURFACE TRAFFIC
DIST - AIRCRAFT

1) 0.16 NW2604- CRJ200
2) 0.21 AA225 - B757
3) 0.40 DL106 - MD80
4) 0.58 NW1152- CRJ200
5) 0.60 FX642 - B747
6) 0.67 NW5742- CRJ200
7) 0.68 NW3574- CRJ200
8) 0.68 NW174 - CRJ200
9) 0.69 FX460 - B747

SELECTED
0.40 DL106 - MD80

Name: DL106
Type: MD80
Bearing: 065 deg
Range: 0.40 nm
Speed: 22 kts
Heading: 270 deg
Route: M

NONE LI ALL

◆RTA 18:24:20z ◆ETA 18:24:28z (+00:07) Distance to T 580 ft

Name: AA728 - B757
Brg/Hdg: 358/270 deg
Range: 4.02 nm
Speed: 181 kts
Route: C, P

Figure 6. Flight Deck 4DT Surface Map Display and Runway Inset Mode

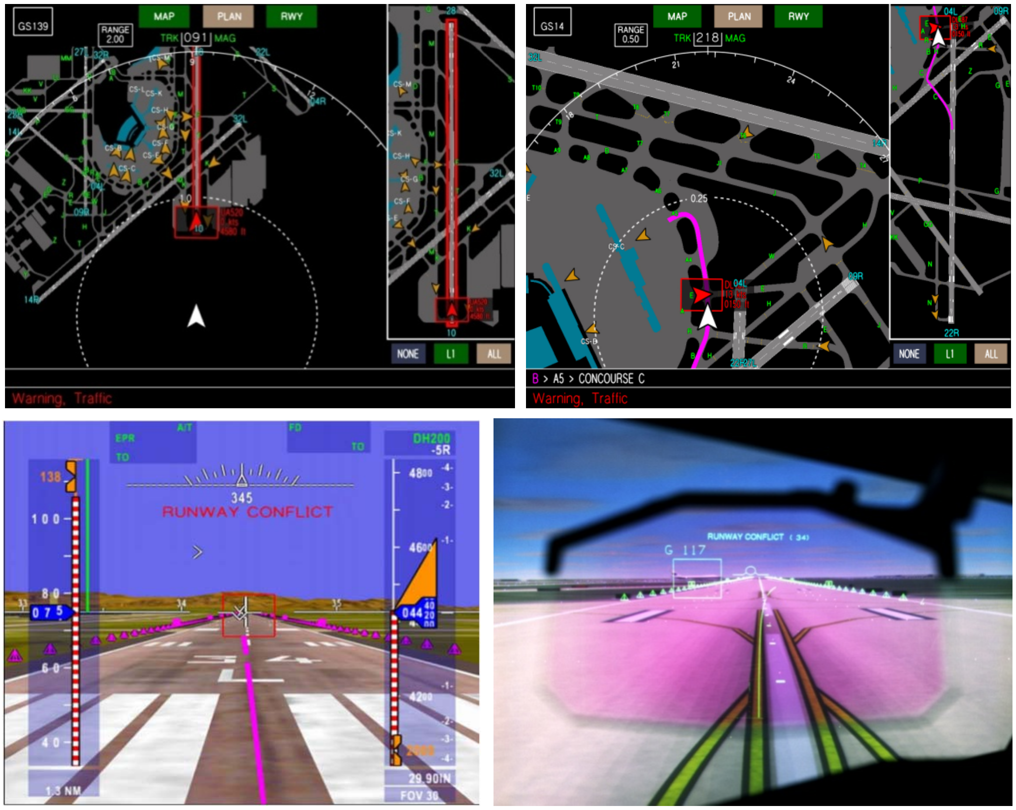


Figure 7. Conflict Detection and Alerting Display Examples

Data Communications

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 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. NASA LaRC research has focused on the issues of data communications and prescriptions to enable NextGen operations (Figure 8; Prinzel, Shelton, et al., 2010). Research is being conducted to essentially identify (and retain) the best features of the present-day radio-based 'party-line' environment, while identifying and introducing the best features of a future data-link communications environment to assist in building NextGen.


ATC messages

18:28:22z KORD Ground
NASA 557, Taxi to Rwy 28
via A10 T M, RTA 3:38

18:29:19z KORD Ground
MSG: WILCO

CONTROLLER PILOT DATALINK (CPDLC/ATC)
Flight: NA557
Date: 08/26/10
Time: 18:04:56 UTC

ATC MESSAGE
TO:
FROM:
MSG:

Figure 8. Data Communication Display Examples

References

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.

Bailey, R.E., Prinzel, L.J., Kramer, L.J., & Young, S.D.: *Concept of Operations for Integrated Intelligent Flight Deck Displays and Decision Support Technologies.* NASA Langley Research Center, NASA/TM-2011-000000, Hampton, VA. Feb 2011.

Barmore, B. (2009). Airborne precision spacing: A trajectory-based approach to improve terminal area operations.

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., 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.

Green, David F., 2006, *Runway Safety Monitor Algorithm for Single and Crossing Runway Incursion Detection and Alerting*, NASA/CR-2006-214275.

Jones, Denise R., 2002, *Runway Incursion Prevention System Simulation Evaluation*, Proceedings of the AIAA/IEEE 21st Digital Avionics Systems Conference.

Jones, Denise R., 2005, *Runway Incursion Prevention System Testing at the Wallops Flight Facility*, Proceedings of the SPIE Defense & Security Symposium.

Jones, Denise R., Quach, C. C., & Young, S. D., 2001, *Runway Incursion Prevention System – Demonstration and Testing at the Dallas/Fort Worth International Airport*, Proceedings of the 20th Digital Avionics Systems Conference.

Jones, Denise R., & Prinzel, Lawrence J., III, 2006, *Runway Incursion Prevention for General Aviation Operations*, Proceedings of the 25th Digital Avionics Systems Conference.

Kramer, L. J., Prinzel III, L. J., Bailey, R. E., & Arthur III, J. J.: *Synthetic Vision Enhances Situation Awareness and RNP Capabilities for Terrain-Challenged Approaches*. AIAA's 3rd Annual Aviation Technology, Integration, and Operations (ATIO) Technical Forum, AIAA 2003-6814, Nov. 2003.

Kramer, L.J., Prinzel, L.J., Arthur, J.J., & Bailey, R.E.: *Pathway Design Effects on Synthetic Vision Head-Up Displays*. Proceedings of SPIE, Enhanced and Synthetic Vision 2004, Editor: Jacques G. Verly, Volume 5424 paper 12 Apr. 2004.

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.

Prinzel III, L.J., Arthur III, J.J., Kramer, L.J., & Bailey, R.E.: *Pathway Concepts Experiment for Head-Down Synthetic Vision Displays*. Proceedings of SPIE, Enhanced and Synthetic Vision 2004, Editor: Jacques G. Verly, Volume 5424 paper 12 April 2004

Prinzel, (submitted, 2011). *Flight deck-based delegated separation: Evaluation of a merging and spacing system with synthetic and enhanced vision technology*. Digital Avionics Conference, 30.

Rathinam, S., Montoya, J., & 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.

Acknowledgements

The authors would like to acknowledge the significant contributions of Dr. R. Michael Norman of the Boeing Commercial Aircraft in this work. The support of NASA's Aviation Safety Program and Dr. Steve Young, Principal Investigator, is also gratefully appreciated. Finally, the authors would like to acknowledge the contributions of many NASA Langley technicians, programmers, and engineers who support/have supported these research projects. Notable among them are Sharon Otero, Glover Barker, Jerry Karwac, Wei Anderson, Thomas Feigh, Chris Harrison, Victoria Chung, Brian Baxley, Sean Kenny, Brent Weathered, and Lisa Rippy.

HUMAN FACTORS ISSUES OF NAVIGATION REFERENCE SYSTEM WAYPOINTS

Shawn Pruchnicki & Bonny Christopher
San Jose State University Research Foundation at NASA Ames Research Center
Moffett Field, California

Barbara K. Burian
NASA Ames Research Center
Moffett Field, California

As part of the Next Generation Air Transportation System initiative the Navigation Reference System waypoint grid was developed to realize additional benefits of area navigation. Despite industry and government involvement in the original design of the grid, it has been met by operators and air traffic controllers with limited enthusiasm. The FAA is sponsoring research to identify human factors issues that might explain this lack of usage and the development of mitigations or recommendations for those issues discovered. In this paper, we will discuss our initial examination of the Navigation Reference System and review potential recommendations to several areas for improvement with specific focus on changes to waypoint nomenclature.

The Next Generation Air Transportation System (NextGen) was initiated, in part, in response to a predicted two to three fold increase in air traffic by the year 2025 as compared to 2003 levels in the United States (Joint Program and Development Office [JPDO], 2007) (Federal Aviation Administration [FAA], 2009). However, the continued reliance on ground-based navigational aids in an environment with increasing air traffic density will limit the achievement of NextGen goals. Current ground based navigational aids are often placed near population centers leaving large geographical areas uncovered. This variable density in conjunction with traffic saturation in metropolitan areas sometimes forces a singular flow of traffic into merge points that are often responsible for system wide delays (Boetig & Timmerman, 2003), especially in the United States northeast corridor. Modern day operators with area navigation (RNAV) and satellite-based navigation abilities (i.e., Global Positioning Systems (GPS)) can now navigate directly to any point in space desired (FAA, 2006) and as such have increased flexibility regarding navigation decision making. Consequently, to fully gain the benefits offered with RNAV operations, the national airspace system as a whole must be designed to accommodate requests for more efficient direct routing. To meet this need, the FAA High Altitude Redesign (HAR) team developed the Navigational Reference System (NRS).

The Navigational Reference System

The NRS is a grid of approximately1600 RNAV waypoints that cover the continental United States and are defined through the intersection of lines of latitude and longitude (See Figure 1). To ensure a user friendly system, the development of NRS waypoint nomenclature was guided by the following objectives (Boetig & Timmerman, 2003; Hannigan, 2009).

- Be easy to communicate
- Have a low potential for error
- Be consistent with principles that guide naming of navigational fixes
- Be intuitive as to the general location of the fix (i.e., provide "geographic" awareness)
- Incur only minimal changes to ground automation (i.e., database changes only)
- Support implementation across the United States
- Be easier to use than fixes delineated by full latitude and longitude coordinates

Additional considerations were that NRS waypoint names should be no more difficult than current waypoints to enter into flight management system (FMS) computers and flight planning software. Also, the NRS should utilize the currently underused RNAV capabilities of many aircraft in high altitude airspace, and the grid should be of sufficient density to support tactical use without significantly adding mileage to an aircraft's route.

Traditionally named waypoints are composed of five letters which are meant to be pronounceable (e.g. CURLY). Frequently the name of these waypoints is randomly assigned with the exception of those associated with geographic or other local features such as the BEARZ waypoint near the city of Chicago referencing the Chicago Bears football team. In contrast, NRS waypoints consist of both letters and numbers and have a distinctive naming pattern in which geographical information is embedded in their name (described below). Because they include both letters and numbers, NRS waypoints are not pronounceable as a single word but rather require the pronunciation of each character separately (e.g. KD54K is pronounced as kilo-delta-five-four-kilo).

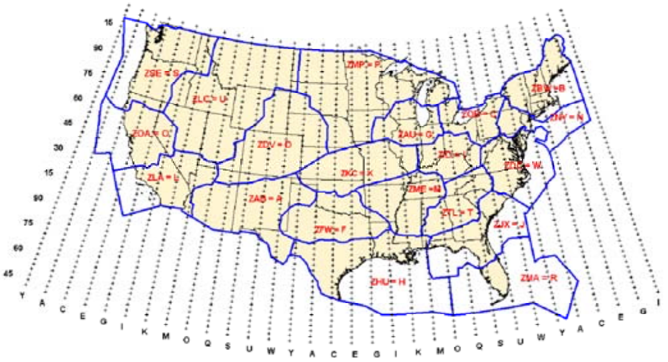


Figure 1. Current distribution of 1600 NRS waypoints and ARTCC regions (Borowski et al. 2004).

NRS waypoint names are composed of two letters followed by two numbers, followed by a single letter (See Figure 2). The first and second characters of NRS waypoints are the FIR identifier for the United States ("K") and the FIR subdivision, or ARTCC center in which the waypoint is located (e.g. "D" for Denver ARTCC). The third and fourth characters are a number group representing the latitude of the waypoint. These numbers begin at the equator with *00* and advances north and south from *01* to *90* and correspond to every 10 minutes of latitude and repeating every 15°. The final character in the NRS waypoint is a letter representing the line of longitude for which the waypoint is located. This identifier starts at the prime meridian moving west to east and uses the letters *A* to *Z* while repeating every 26°. To date, the current density of the NRS grid is one waypoint spaced every 30 minutes of latitude and every 2° of longitude. Possible future expansion will space one waypoint every 10 minutes of latitude and 1° of longitude. This nomenclature system was intended to provide information to users about each waypoint's geographic location: first within the United States, then within which ARTCC airspace, and then narrowed down even further to a specific line of latitude and longitude (See Figure 2).

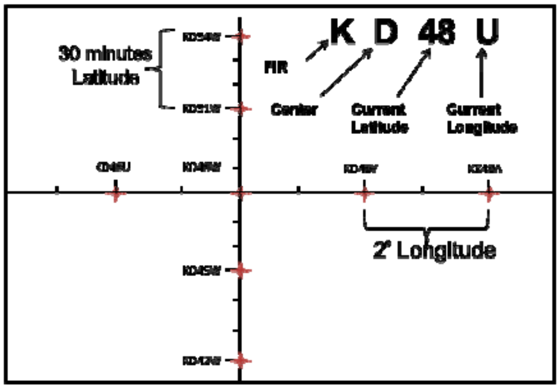


Figure 2. NRS waypoint grid structure and nomenclature.

To ensure the proposed nomenclature would offer the advantages already discussed, prior to deployment, the MITRE Center for Advanced Aviation System Development (CAASD) conducted a series of studies in which the use of the NRS grid was compared to traditional longitude and latitude coordinates. The results of this research found that the naming convention was rated as "easy to use" and "acceptable" by both controllers and pilots alike when compared to full longitude and latitude coordinates (Boetig, Domino, & Olmos, 2004; Borowski, Wendling, &

Mills, 2004; Domino, Ball, Helleberg, Mills, & Rowe, 2003; Domino, Boetig, & Olmos, 2004). Although it is recognized that the creation of any new navigation system is expected to produce a period of adjustment for all users, we have found an apparent industry wide reluctance to utilize the NRS grid despite the pre-deployment finding of grid and waypoint acceptability. Our research has revealed several human factors issues which may help explain why NRS waypoints have been underutilized.

NRS Human Factors Considerations

We evaluated the human factors issues of NRS waypoints from the multiple perspectives of those responsible for their development and current users including pilots, flight planners, dispatchers, FMS database managers, air traffic controllers, and air traffic control managers and supervisors. In addition to extensive interviews and site visits, we also completed an exhaustive review of the literature and incident and accident database searchers. For a complete description of study methodology, see Burian, Pruchnicki, and Christopher (2010).

As in the studies conducted prior to NRS grid implementation (e.g., Domino, et al., 2003), we found that with sufficient information or training, those interviewed understood the intent and structure of both NRS waypoint names and grid structure. Pilots and controllers did not view NRS waypoints any differently than traditionally named waypoints when seeing them on a flight plan. Pilots did not believe NRS waypoints contributed to any particular CRM issues on the flight deck or required any changes to pilot flying and pilot monitoring roles and responsibilities. Dispatchers stated that NRS waypoints provide greater flexibility in route planning, especially in the western portion of the US where fewer ground based navigation aids exist and were enthusiastic about using them. Despite these positive aspects, there are several issues discovered that might explain their limited use.

One item we frequently heard from both controllers and pilots was that the lack of ability to overlay the NRS grid structure over their respective radar and navigation displays greatly reduced the usability of NRS waypoints in their daily operations. This limitation is only problematic for those NRS waypoints that are not currently part of the route of flight being flown. That is, NRS waypoints that are programmed into either the controller's or the pilot's computer systems are displayed as they are part of the entered route of flight. Although this constraint provided little to no ramifications when flight planning for pilots (i.e. strategic use), both groups reported that their ability to use the NRS waypoints in a tactical fashion once a flight was underway was essentially nil. Examples of tactical use include short term deviations around small areas of intense weather or the creation of parallel traffic flows. It became clear to us that any future attempts to increase system wide NRS utilization must be accompanied by an improvement in display capabilities for pilots and controllers alike.

Pilots had additional challenges related to their use of the FMS that restricted practical NRS functionality such as the limited amount of memory available in many of the FMSs in aircraft currently being flown. Rapid expansion in RNAV procedures and corresponding RNAV waypoint development has significantly limited the amount of memory capacity available for the addition of NRS waypoints. To illustrate this limitation, one US air carrier we visited produced a map of the United States where large geographical sections of NRS waypoints had to be removed from their FMS databases due to memory space limitations. Essentially, they were forced to choose which parts of the country they felt they were most likely to utilize NRS waypoints and those areas where they were not. Because of this, they are not only losing the routing flexibility that NRS provides but also the additional burden that is placed on their pilots in not knowing which waypoints were in the database and which are not and for aircraft schedulers who must know in which parts of the country specific aircraft can be allowed to fly.

In summary, throughout our interviews with current users, we discovered that the NRS grid meets some of the expectations that the system was designed to offer. Most groups reported that to some degree they liked the NRS concept even if they had problems with the way it was currently implemented and several individuals stated that they use NRS waypoints during route planning. We found it remarkable that when examining these issues with all groups, there was significant commonality with respect to the operational challenges they faced when trying to utilize NRS waypoints. Issues related to waypoint naming convention (discussed below), the absence of NRS waypoints presented on displays, and charting issues permeated our data. Additional concerns such as FMS database restrictions and En Route Automation Modernization (ERAM) considerations were mentioned by pilots and air traffic controllers, respectively.

NRS Waypoint Nomenclature Considerations

We discovered that several of the issues already mentioned, and others that are covered in more detail in Burian, Pruchnicki, and Christopher (2010), pertained to waypoint nomenclature (i.e., approach to naming waypoints: KD54U). Members from both the flight deck and ATC communities reported that they found the NRS waypoint nomenclature problematic in its current form and contributed to difficulties in using NRS waypoints in their day-to-day operations.

One NRS waypoint communication issue that was hypothesized prior to data collection was that frequency congestion would be aggravated due to the increased time it takes to verbalize a NRS waypoint as compared to traditionally named RNAV waypoints (Borowski, et al., 2004). A named RNAV waypoint is typically a pronounceable one-, two-, or three-syllable word, however each character in a NRS waypoint name generally must be verbalized separately using the phonetic alphabet and numbers; with the exception that the two numerals denoting the latitude line can be phrased as two separate numbers or one (e.g., “54” can be spoken as “five- four” or as the single number “fifty-four”). Through our interviews and searches of Aviation Safety Reporting System (ASRS), airline Aviation Safety Action Program (ASAP), and Air Traffic Quality Assurance (ATQA) incident reports we failed to find any reports of concern over the time it takes to verbalize NRS waypoints over the radio. It is possible, however, that this may become a concern in the future if NRS waypoint tactical usage increases while still using voice communications (prior to data-link). Nonetheless, we did identify some communication concerns with regard to NRS waypoint nomenclature.

Consistent with the controllers in one of the MITRE CAASD pre-deployment studies (Domino, et al., 2003), our pilots and controllers alike felt that the inclusion of the letter “K” in front of each waypoint was cumbersome and unnecessary. This is especially true since NRS waypoints have not been adopted outside of the United States as originally expected. As discussed earlier, the second letter in NRS waypoint name are the single letter identifiers for the ARTCC in which the waypoint is located. It was intended that providing the ARTCC identifier as part of the waypoint name would help provide some degree of “geographical knowledge” to pilots and controllers, not only about the location of the waypoint but its relationship to the aircraft’s route of flight. Our interviews with dispatchers, flight planners and controllers suggest that this nomenclature does in fact provide some degree of geographical knowledge to these populations of users. However, dispatchers and flight planners at some of the air carriers we visited still exhibited some difficulty in finding specific NRS waypoints on en-route charts despite knowing in which Center’s airspace the waypoint was located and despite their having a good understanding of the grid structure (These difficulties went beyond issues in chart readability). Interviews with pilots confirmed our suspicions that ARTCC identifiers are not commonly known and provided little to no geographical awareness. Pilots also suggested that because ARTCC boundaries are irregularly shaped and are generally unknown to flight crewmembers, including an ARTCC identifier as part of an NRS waypoint name is of little utility. (Center airspace boundaries are indicated on en-route charts but they are not very conspicuous and flight crews typically depend upon electronic navigation displays, which do not show air space boundaries, rather than on paper charts during flight). Furthermore, the amount of airspace assigned to each ARTCC is quite large. Pilots we interviewed stated that even if they knew the ARTCC single letter identifiers, additional specificity would be required to assist them in actually locating a specific waypoint within that Center’s boundaries.

The two numbers and single letter that signify latitude and longitude lines in NRS waypoint names should, in theory, provided this necessary specificity but many we interviewed found them to be of little help. One individual summed up particularly well the concerns expressed by many we spoke to:

> “The grid system, while generally understandable with a key diagram in hand, is not intuitive. It requires learning a new coordinate system that conflicts with an existing one. The pseudo-latitude is problematic to my 44 years of flying. The alpha (longitude) key at the bottom of the NRS [diagram] also seems counter-intuitive; it "increases" (alphabetically) in an easterly direction while actual longitude decreases... Most confusing though, I believe, may be the "latitude" number that is not the actual latitude. I understand the system's goal is greater precision, but believe it increases the potential for error and increased workload.”

Furthermore, in a few ASAP reports we discovered that occasional transposition of characters within a waypoint occurred and that the similarity of waypoint names in a route could cause confusion and lead to data entry errors (e.g., KG78K-KP90G-KP09A).

Cognitive Limitations

When humans are presented with information that will be immediately used, this information is held in working memory. It is well understood that there are significant limitations to working memory capacity which can actually decrease during times of stress (Baddeley, 1987). Research has shown that on average, when not under stress, working memory capacity is seven, plus or minus two, "items" or "pieces" of data (7 ± 2; i.e., five to nine items; Miller, 1956). An item or "piece" of data might be a single "thing", such as one digit in a person's phone number, or it might actually be several "things" that together carry a single unit of meaning, such as several letters that together make up a person's first name. Some information held in a person's working memory that is full to capacity will have to drop out to make room for new information that must be remembered.

Working memory limitations have important significance with regard to the design of NRS waypoint nomenclature. A traditional RNAV waypoint name such as "AZELL" is one item or piece of data to hold in working memory because it spells a single pronounceable word. Although the word itself may be meaningless, because it forms a pronounceable "word," it comprises a single unit of information. NRS waypoints, on the other hand, do not "chunk" together to form a single unit of information. The waypoint KD54U is comprised of three to five units of information. It is comprised of three units if: a) the initial "K" is ignored because all NRS waypoints begin with "K" so one does not need to commit it to memory, and b) the numerals signifying latitude are treated as a single number, thus: Delta – fifty-four – Uniform. It comprises five units of information when each character is remembered and the numerals are treated as two separate numbers, thus: Kilo – Delta – Five – Four – Uniform. Therefore, when considering verbal communication and the possible reliance on working memory until the information can be written down, entered into a FMS, or typed on a DSR keyboard, one NRS waypoint alone can come very close to filling human working memory capacity. Remembering two NRS waypoints in a spoken clearance could easily exceed this capacity.

When examining normal human working memory capacity and limitations, it is important to consider the environmental or operational context in which the requirement to hold information in working memory, until it can be acted upon occurs. That is, a 7 ± 2 working memory capacity may be more applicable to the environment in which it was discovered, the laboratory, rather than to other environments such as busy flight decks or air traffic control work stations, which are full of multiple concurrent tasks and distractions. The association found between errors in reading back a clearance, which is often held in working memory until it can be "read back," (Barshi & Healy, 2002; Cardosi, 1993; Prinzo, Hendrix & Hendrix, 2006), has led to the recommendation that air traffic controllers include no more than three items of information when issuing a clearance (e.g., altitude, heading, new ATC frequency). This appreciation for the possible normal reduction of working memory capacity in typical aviation operations should be considered when evaluating any new recommended approaches to the naming of NRS waypoints.

Conclusion

Through the course of this study we discovered that although most individuals we spoke to understood and appreciated the intended advantages of the NRS waypoint grid, they felt that a number of issues impeded realization of those advantages. To ensure the greatest utility of the NRS gird, we suggest that the findings in this report be used as a starting point and that individuals representing all sectors of the NRS waypoint user community be involved in developing potential solutions. In particular, emphasis should be given to the human factors issues associated with NRS waypoint nomenclature and displays which contribute to the most significant limitations in use of the grid by pilots and controllers. A wide variety of solutions should be generated and explored such as changes to NRS waypoint nomenclature, changes to depiction of NRS waypoints on charts and displays, NRS waypoint applications in electronic flight bags, and the feasibility of retrofits or upgrades to FMS and DSR databases and displays, among others. The solutions that are proposed must be evaluated against proposed NextGen airspace changes (e.g., dynamic sector boundaries, generic airspace at high altitudes, etc.), and all potential solutions must be tested and validated, prior to adoption and implementation.

Acknowledgements

The authors would like to express their appreciation to Steve Casner, Ben Berman, and Mary Connors who provided thoughtful and helpful comments on an earlier draft of this paper. This work was sponsored by the FAA and was funded through the FAA Human Factors Research and Engineering Office**.**

References

Baddeley, A. D. (1987). *Working Memory.* Oxford, England: Oxford University Press.

Barshi, I., & Healy, A. F. (2002). The effects of mental representation on performance in a navigation task. *Memory & Cognition, 30*(8), 1189-1203.

Boetig, R.C., Domino, D.A., & Olmos, B.O. (2004). *Airline pilot usability assessment of the navigation reference system* (MTR 04W0000011). McLean, VA: MITRE CAASD, The MITRE Corporation.

Boetig, R.C., & Timmerman, J. (2003). *The HAR navigation reference system* (MP 03W0000152). McLean, VA: MITRE CAASD, The MITRE Corporation.

Borowski, M., Wendling, V.S., & Mills, S.H. (2004). *HAR NRS usability study final report* (MP 04W0000163). McLean, VA: MITRE CAASD, The MITRE Corporation.

Burian, B. K., Pruchnicki, S., & Christopher, B., (2010). *Human factors evaluation of the implementation of the navigation reference system (NRS) Phase 1 report.* FAA Office of Human Factors Research and Engineering NextGen Task,09-AJP61FGI-0101.

Cardosi, K. (1993). *An analysis of en route controller-pilot voice communications.* DOT Report no. DOT/FAA/RD-93/11. Washington, DC: U.S. Department of Transportation, Federal Aviation Administration.

Domino, D.A., Ball, C.G., Helleberg, J.R., Mills, S.H., & Rowe, D.W. (2003). *Navigation reference system (NRS) naming convention usability study preliminary findings.* McLean, VA: MITRE CAASD, The MITRE Corporation.

Domino, D.A., Boetig, R.C., & Olmos, B.O. (2004). *Corporate pilot usability assessment of the navigation reference system* (MTR 04W0000033). McLean, VA: MITRE CAASD, The MITRE Corporation.

Federal Aviation Administration. (2006). *Roadmap for Performance-Based Navigation: Evolution for Area navigation (RNAV) and Required Navigation Performance (RNP) Capabilities 2006-2025.* Washington, DC: FAA.

Federal Aviation Administration. (October 2009). *Play Book.* Air Traffic Control System Command Center, National Severe Weather Playbook. Effective October 22, 2009 to December 16, 2009.

Hannigan, M. (2009). *Navigational reference system (NRS) and performance based routing (PBR).* Presentation to the CNS Task Force.

Joint Planning and Development Office, (2007). *Concept of operations for the next generation air transportation system* (Version 2.0).

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 63, 81-97.

Prinzo, O. V., Hendrix, A. M., Hendrix, R. (2006). *The outcome of ATC message complexity on pilot readback performance.* DOT Technical Report no.DOT/FAA/AM-06/25. Washington, DC: U.S. Department of Transportation, Federal Aviation Administration.

HUMAN PERFORMANCE TRAINING:
SUCCESSES AND FAILURES IN CIVIL AVIATION

Alan E. Diehl, Ph.D., ATP, CPE
Retired NTSB, FAA & USAF Air Safety Investigator
Albuquerque, New Mexico

For over twenty years, two similar types of training have attempted to reduce aviator caused accidents. Crew Resource Management was widely adopted by U.S. airlines, and has generally been credited with helping to dramatically reduce their accident rates. Interestingly, CRM was embraced with little scientific evidence that it could actually reduce operational errors. In contrast, Aeronautical Decision Making, aimed at general aviation users, underwent a series of double-blind experiments before being adopted. And, although some users suggested ADM training significantly reduced their accidents, it was never fully implemented. This may explain the limited improvement in general aviation accident rates, when compared with the airline rates for the last two decades.

"Human error" has historically been associated with approximately three quarters of all aviation accidents, and various studies have indicated aviator decision making failures were associated with the majority of fatal civilian crashes, (e.g., Diehl, 1992). Two major training efforts were launched in the 1980s to deal with these problems. These similar programs, Crew Resource Management and Aeronautical Decision Making, were proffered as cost-effective and quick methods of reducing such accidents.

Crew Resource Management

While serving as a National Transportation Safety Board investigator, I drafted the first recommendation calling for the implementation of Crew Resource Management training by this nation's airlines. This occurred after a 1978 United Airlines DC-8 crash. The crew detected a landing gear unsafe light, and entered a holding pattern, only to run out of fuel 66 minutes later over the suburbs of Portland, Oregon. The captain, fixating on the gear light problem, had ignored repeated queries from the other two crew members about their fuel status, (NTSB, 1979).

Per the NTSB recommendation, United instituted CRM training in 1981. And within several years most U.S. major carriers followed suit. Although, the Federal Aviation Administration did not formally require CRM for all FAR Part 121 carriers until 1995.

NTSB investigators must carefully justify proposed recommendations. So I examined the limited evidence that CRM training would work. I reviewed the research conducted earlier under the direction of Dr. John Lauber at the NASA Ames Research Center, including the simulator studies done by Dr. H. P. Ruffell-Smith. In addition, contemporary research on the recognition of "subtle incapacitation" done at United Airlines as well as Line Oriented Flight Training being conducted by Northwest Airlines, looked promising.

As an adjunct professor of management, I also recognized the similarity of CRM to another type of training used in industrial settings. Total Quality Management empowered workers to speak-up and actively participate in decision making about products and production processes. TQM had been successfully utilized for decades and was credited with helping the successful renaissance in Japanese industry after World War II. Ironically, it was the brainchild of an American, Dr. Edwards Deming.

CRM-related errors also seemed to be factors in several well-known crashes. These accidents included the 1972 Eastern Airlines L-1011 crash into the Everglades and the tragic collision of the KLM and Pan-American B-747s on the runway at Tenerife in 1977. Such catastrophes suggested that issues like loss of situational awareness, communication breakdowns, as well as individual and collective judgment errors needed to be addressed.

Thus, adopting CRM training seemed logical, even without any statistical data or controlled experiments to "prove" it worked. In any case, I felt a new type of training to manage resources on the flight deck was an idea whose time had come, and the NTSB members, as well as United's management and the union apparently agreed.

The FAA authorized an18-month evaluation of the prototype CRM program. The Agency even assisted United to defer some of the costs of the program's development, and the two-day training sessions for its flight crews. The FAA did this by granting the airline an "exemption" allowing its captains to take line checks, instead of going to simulator training every six months. Thus, captains only received simulator training once a year. I was aware of this arrangement, because I had transferred from the NTSB to become the FAA headquarters program scientist for human performance, in 1981.

However, some people legitimately questioned whether this type of training *really* worked. The term "psycho-babble" was used by some FAA bureaucrats. But I retorted that CRM training provided a host of pragmatic techniques to reduce the probability and the consequences of errors -- what others would later call "threat and error management."

Other airlines soon approached the FAA requesting exemptions for their proposed CRM programs, but not all of these programs were well designed. This was a legitimate concern for the Agency. So, in 1983, I hired Dr. Richard Jensen to examine cutting-edge CRM training programs in use around the globe, and to show the FAA how best to certify such programs and airmen enrolled therein.

I was also delighted to hear that NASA was also developing instruments like the Cockpit Management Attitude Questionnaire to evaluate the views of airmen about CRM. However, such measures could not answer the question of whether this training reduced the number of errors during "normal" much less emergency operations, (Wiener, Kanki & Helmreich, 1993).

But perhaps the best evidence that CRM actually worked came from a couple of accidents experienced by United Airlines crews after being exposed to the concepts for several years. In August 1989 a United DC-10's center engine exploded in-flight, severing hydraulic lines and eliminating all aerodynamic controls. The only way to control the jet was through differential power of the two remaining engines, (NTSB, 1990a).

The DC-10 crew led by Captain Al Haynes allowed another off-duty captain handle the thrust levers, and they were able to successful crash land at the Sioux City airport. This widely reported feat of airmanship saved the lives of the majority of the passengers aboard the ill-fated craft. Captain Haynes has repeatedly stated the use of CRM was essential in their success that fateful afternoon, (Haynes, 1991).

Several months earlier another United crew faced an almost as demanding situation. Their 747 had departed Honolulu, but as they passed through 22,000 feet, a forward cargo door failed, ripping away part of the aircraft's fuselage. The explosive decompression ejected nine passengers and baggage, and damaged both starboard engines, along with the wing's leading edge devices, (NTSB, 1990b).

Using various CRM techniques they managed to safely return to Hawaii without the further loss of life. For instance, the second officer soon recognized they were far too heavy for the rapidly up-coming two-engine-out emergency landing. But, he just could not get the bewildered captain's attention, so he quickly decided to dump fuel himself. I appreciated the irony of this situation, because in my 1978 United DC-8 crash, the second officer could not get the captain's attention about their *low* fuel state. Certainly, CRM techniques played a role in saving this 747 and its passengers.

Obviously, trying to make precise statistical statements about extremely infrequent events is always a problem. It has been noted elsewhere that because airline accidents are so infrequent it is difficult to establish the precise role that CRM has played in the reduction in these accidents, (Helmreich, Merritt & Wilhelm, 1999).

Furthermore, many things besides the use of CRM have changed in the aviation system in recent decades. These changes included: the advent of highly automated airliners, more effective air traffic management procedures, installation of newer types of cockpit warning devices, all of which probably helped enhance airline safety.

Incidentally, the USAF Military Airlift Command began using CRM in 1985. This change was ordered by their commander and took place in just three-months. While military transports faced different flying conditions, often with less experienced crews, it was interesting to examine the statistical changes that occurred after this program began.

I computed there was a 51% reduction in mishaps during the five-year period after the program's implementation, when compared with the previous five-year period. This was statistically significant at the .05 level. And unlike the protracted airline implementation period, little else changed for this command during that decade-long time span. Thus, this military evidence suggested CRM was a factor in the improved airline safety, (Diehl, 1992).

Aeronautical Decision Making

The other major type of human performance training focused on general aviation pilots. It was originally called "judgment training" and the initial research was led by Dr. Richard Jensen under FAA contract in the mid-1970s. A major conclusion of his landmark study was that 52% of U.S. fatal general aviation accidents were caused by pilot judgment errors. Jensen also recommended that prototype training materials be developed and their effectiveness be measured, (Jensen & Benel, 1977).

The "judgment training" label continued through the development of the early manuals and other materials. This work was done at Embry Riddle Aeronautical University, under the leadership of Dr. Jerry Berlin. I was soon asked by FAA officials to assume the technical direction of the program as well as the evaluation of the prototype materials. The FAA wanted to insure such training was effective before proceeding further.

Judgment training materials taught many of the same concepts used by airline CRM programs, but the information and case studies were drawn from general aviation situations and single pilot operations. Subject pilots learned about the effects that several types of hazardous attitudes have on performance, as well as concepts such as stress management, risk management, and attention management.

And they were also given various “rules and tools” to help improve their decision-making. For example, the “I’M SAFE” mnemonic reminded subjects that before flying they needed to insure they were free of the following conditions: Illness, Medications, Stress, Alcohol, Fatigue, and to ensure they had Eaten properly.

A major challenge in evaluating this training was whether or not pilots would use the concepts in “normal” situations. Determining if those pilots who had received judgment training performed differently than control group subjects was a requirement. It was decided to measure the decision-making of all pilots during short cross-country flights when they did not know their judgment was being carefully observed.

After receiving the prototype judgment training materials, Embry Riddle subject pilots made 17% fewer decisional errors than student pilots in the control group who had not received this training. These results, while preliminary, suggested pilot judgment could be improved through training, (Berlin, et. al., 1982).

Various organizations around the world became interested in the topic. For instance, Gary Livack and his employer the General Aviation Manufacturers Association offered to help develop cutting-edge audiovisual materials along with improved training manuals.

Dr. Georgette Bush of Transport Canada led the effort there, and conducted two additional studies. One study examined private pilots and another involved pilots training for their commercial licenses. Both studies used the new and improved training manuals, and the commercial students also were exposed to more sophisticated audiovisual materials. The two double-blind studies in Canada indicated private pilots receiving judgment training averaged 9% fewer errors, while commercial pilots (who got a more comprehensive program) averaged 40% fewer errors, when compared to those pilots assigned to the respective control groups, (Bush and Diehl, 1983).

An additional study of private pilot judgment was undertaken for the Australian government by Dr. Ross Telfer. While another study, sponsored by the USAF and directed by Dr. Tom Connelly, used instrument-rated ROTC cadets. Here pilots faced simulated emergencies in GAT-1 instrument training devices.

In all five experiments, pilots receiving judgment training outperformed their contemporaries at statistically significant levels. The measured improvements averaged from a low of 8% fewer errors up to 46% fewer errors in the case of the instrument rated pilots study. This wide variation in the amounts of experimental group improvement was attributed to: the differences in the quality of the training manuals, the sophistication of other training media, and the emphasis placed on this training by instructors, (Diehl, 1992).

The FAA and the Aircraft Owners and Pilots Association wondered what would be the effect of making judgment training available, but not mandatory. So I made another study of private pilots at ten fixed base operators in the U.S. This time the experimental subjects were simply given the training manuals without any encouragement or instructions on how to use them.

The FAA was also evaluating new sectional chart designs at the time, which provided an opportunity to observe the "normal" behavior of both experimental and control group subjects. Incidentally, the observers posed as cartographers, while unobtrusively recording the in flight judgment performance of subjects.

The results of this well-controlled double-blind study produced a 10% reduction in errors for the experiment pilots who were given the manuals. While statistically significant, the results also suggested the training would be much more effective if the subjects believed they were going to be evaluated on the materials during actual FAA flight tests, (Diehl, 1992).

And it was soon concluded the "judgment" was not the best label for such training. This was because many student pilots were older, highly successful, professional people who did not think their youngish instructors could improve their judgment. Hence, I decided to rebrand the training as Aeronautical Decision Making. And a series of ADM training manuals was developed for various categories of pilots, such as Students and Private Pilots, Instrument Pilots, Air Taxi Pilots, Helicopter Pilots, etc.

Shortly after these training manuals were completed in 1987, I transferred from FAA to the USAF, and several other key personnel associated ADM development also went on to other assignments. The result was this training was never fully implemented by the FAA. In fact, it was not until 1997 that it became "theoretically" mandatory training.

Interestingly, some people in the helicopter community immediately embraced this training. Bell Helicopters Inc. and Petroleum Helicopters Inc. vigorously pursued ADM training for the pilots flying their equipment. These two organizations soon reported dramatic drops in their respective accidents rates. Bell saw a 48% drop in their U.S. Jetrangers accidents, while PHI experienced a 54% drop in their accidents, (Fox, 1991; Adams & Diehl, 1988).

Unfortunately, unlike Bell and PHI, the broader general aviation community never thoroughly applied ADM concepts. Ironically, a 1999 study by former FAA inspector, Doug Hawley, confirmed this fact. He determined that Certified Flight Instructors spend very little time teaching ADM. In fact, of the instructors he surveyed, 33% had never heard of ADM, only 13% knew it was actually mandatory, and a mere 3% could explain how to obtain the FAA materials needed to teach these concepts, (Hawley, 1999).

Conclusions

The widely applied CRM training programs appear to have helped dramatically reduce airline accidents in recent decades. Furthermore, the failure to give general aviation pilots ADM training may well have contributed to the relatively paltry reduction in their accidents.

The U.S. general aviation accident data stands in sharp contrast with the progress made by this nation's airlines over the past 20 years. In fact, the latest data available from the NTSB (for 1990 to 2009) indicates the accident rate for all fatal FAR Part 121 airlines decreased by 78%, while the rate for all fatal general aviation accidents declined by only 15 %, (NTSB, 2010). While annual fatal accident rates, especially for airlines are often volatile, other metrics tell a similar story.

Concerns about drawing firm conclusions on the safety of such different types of flying are certainly relevant. However, the large differences in the *changes* in their respective accident rates during this twenty-year period suggests further study of the roles played by CRM and ADM training is warranted.

References

Adams, R.J. & Diehl, A.E. (1988). The Human Factor in Helicopter Safety. *Vertiflight,* July/August, 34, No. 4.

Berlin, J.I., Gruber, J.M., Holmes, C.W., Jensen, P.K., Lau, J.R., Mills, J.W. & O'Kane J.M. (1982). Pilot Judgment Training and Evaluation. DOT/FAA Report CT-82-56, Vol I, II & III, Washington, DC.

Diehl, A.E. (1992). The Effectiveness of Civil and Military Cockpit Management Training Programs. *Flight Safety Foundation, Proceedings of the 45th International Air Safety Seminar*, Long Beach, CA.

Fox, R.G. (1991). Identifying the Hazard Is Only the Beginning. *Hazard Prevention*, 27, No. 3.

Jensen, R.S. & Benel, R.A. (1977). Judgment Evaluation and Instruction in Civil Pilot Training. DOT/FAA Report RD-78-24, Washington, DC.

Hawley, D. (1999). The Implementation of Aeronautical Decision-Making in General Aviation Training Environments, Unpublished M.A.S. Thesis, Embry Riddle Aeronautical University, Albuquerque, NM.

Haynes, A. (1991). The Crash of United Flight 232, Presentation to the NASA Dryden Flight Research Facility, May 24, Edwards, AFB, CA.

Helmreich, R.L., Merritt. A.C. & Wilhelm, J.A. (1999). The Evolution of Crew Resource Management Training in Commercial Aviation. *International Journal of Aviation Psychology*, 9(1), 19-32.

NTSB (1979). Aircraft Accident Report, United Airlines, DC-8, Report No AAR-79-7, Washington, DC.

NTSB (1990a). Aircraft Accident Report, United Airlines, DC-10, Report No AAR-90-6, Washington, DC.

NTSB (1990b). Aircraft Accident Report, United Airlines, B-747, Report No AAR-90-1, Washington, DC.

NTSB (2010). Annual Review of Aircraft Accident Data: Accidents, Fatalities, and Rates, 1990 - 2009, 14 CFR 121, (table 5), and Accidents, Fatalities, and Rates, 1990 - 2009, U.S. General Aviation (table 10), Washington, DC.

Weiner, E.L., Kanki, B.G. & Helmreich, R.L. (1993). *Cockpit Resource Management,* Academic Press, San Diego, CA.

DEVELOPMENT OF A CRM SKILLS MEASUREMENT METHOD INCLUDING THREAT AND ERROR MANAGEMENT CONCEPT

Tomoko Iijima, Hiroka Tsuda and Fumio Noda
Japan Aerospace Exploration Agency,
Tokyo, Japan

The Japan Aerospace Exploration Agency (JAXA) has developed a Crew Resource Management (CRM) Skills measurement method that includes a Threat and Error Management (TEM) concept and identifies a crew's level of CRM Skills by the way in which they manage human errors and threats. To validate the method, a CRM Skills measurement experiment was carried out by four raters using Line Oriented Flight Training (LOFT) scenarios. To increase inter-rater reliability, the raters collated their results to develop "True Scores". The experiment identified factors contributing to individual scoring differences between the raters and provided information for improving the CRM Skills rating sheet, inter-rater reliability training and LOFT scenarios.

CRM training is considered to be one of the most effective means of reducing human errors or minimizing their effects. Demands for greater operational safety and economy are will require more effective and efficient training and to achieve this, CRM Skills measurement will be necessary to evaluate objectively those skills that have been learned and to identify those that are inadequate. To allow CRM Skills training to be incorporated into pilot training and line operations, JAXA has developed behavioral markers by which CRM Skills may be identified (Iijima *et al.*, 2003) and a CRM Skills measurement method to assess the effectiveness of CRM training and to continue its improvement (Noda *et al.*, 2005, Tsuda *et al.*, 2006).

CRM Skills measurement relies upon the subjective scoring of crew behaviors observed in LOFT scenarios by "raters", who are typically pilots assigned to an airline's training department. However, since the scores are subjective, there is an issue of variability between different raters' scores for the same observed behavior. For example, Tsuda *et al.* found individual differences between nine raters who participated in CRM Skills measurement experiments. The main factor contributing to these individual differences was found to be differing rater viewpoints when observing crew behaviors. This indicates that inter-rater reliability training is necessary to standardize raters' viewpoints and their rating criteria.

The sixth annex of the International Civil Aviation Organization (ICAO) treaty now requires as an international standard that flight crew training must include human abilities and limitations, including Threat and Error Management (TEM), and assess competency in these areas. To meet this requirement, it is considered that a CRM Skills measurement method that includes a TEM concept is necessary.

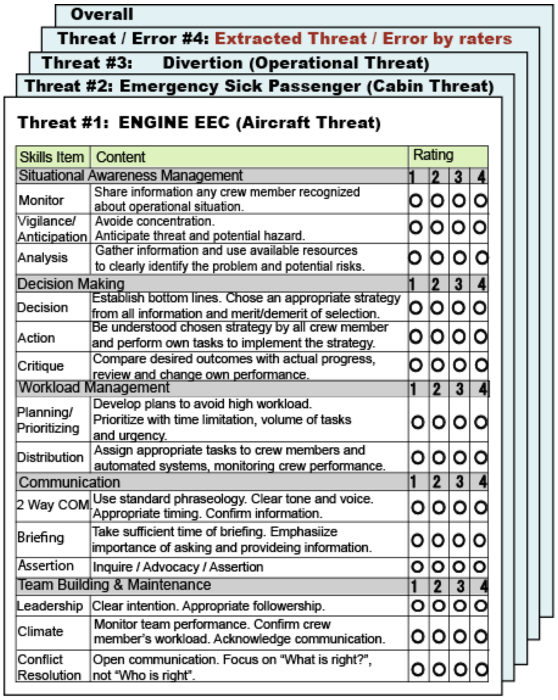
We therefore propose a new CRM Skills rating technique in which raters measure CRM Skills for each threat included in LOFT scenarios. The measurement adopts a TEM concept. To reduce individual scoring differences, raters are directed to evaluate only crew behavior when managing or mismanaging threat/errors, and to standardize their evaluations they hold discussions to derive a "True Score" result for each CRM skill item (Baker *et al.*, 1999). This paper describes the proposed CRM Skills measurement method and the verification of its validity.

Design of CRM Skills Measurement Method

To develop a CRM Skills measurement method, it is necessary to develop a CRM Skills rating sheet, effective LOFT scenarios that exercise CRM Skills for TEM, and to address inter-rater reliability training.

CRM Skills Rating Sheets

The proposed CRM Skills measuring method is a subjective scoring technique in which raters evaluate a crew's CRM Skills from their threat/error management behaviors and for each skill assign a numerical score on a four-point scale: 1 = "Ineffective", 2= "Adequate", 3= "Effective" and 4 = "Highly Effective". Scores are recorded on a CRM Skills rating sheet along with written narrative comments. The purposes of the CRM Skills rating sheet are to reduce score differences due to differences in individual rater viewpoints when observing threat management behaviors, and to conduct measurements appropriate to the TEM concept. To compare measuring CRM Skills on a per threat basis with measuring skills on a per flight phase basis, we developed two corresponding CRM Skills rating sheets shown in Figures 1 and 2.



Overall

Threat / Error #4: Extracted Threat / Error by raters

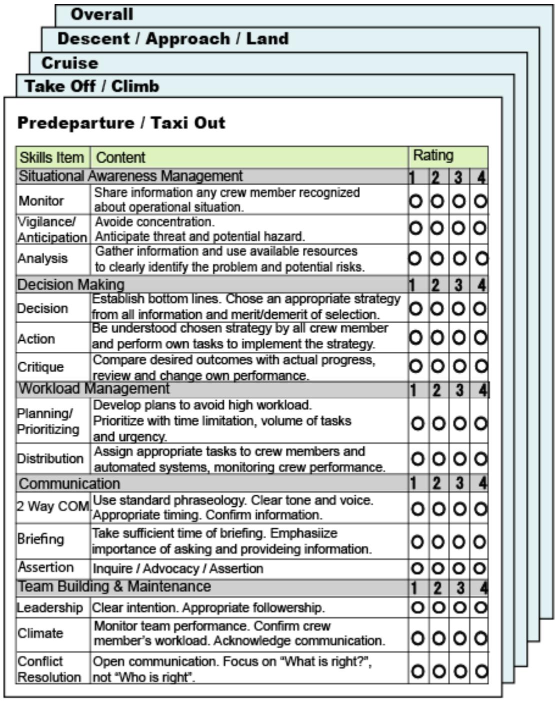
Threat #3: Divertion (Operational Threat)

Threat #2: Emergency Sick Passenger (Cabin Threat)

Threat #1: ENGINE EEC (Aircraft Threat)

Skills Item	Content	Rating			
Situational Awareness Management		1	2	3	4
Monitor	Share information any crew member recognized about operational situation.	○	○	○	○
Vigilance/ Anticipation	Avoide concentration. Anticipate threat and potential hazard.	○	○	○	○
Analysis	Gather information and use available resources to clearly identify the problem and potential risks.	○	○	○	○
Decision Making		1	2	3	4
Decision	Establish bottom lines. Chose an appropriate strategy from all information and merit/demerit of selection.	○	○	○	○
Action	Be understood chosen strategy by all crew member and perform own tasks to implement the strategy.	○	○	○	○
Critique	Compare desired outcomes with actual progress, review and change own performance.	○	○	○	○
Workload Management		1	2	3	4
Planning/ Prioritizing	Develop plans to avoid high workload. Prioritize with time limitation, volume of tasks and urgency.	○	○	○	○
Distribution	Assign appropriate tasks to crew members and automated systems, monitoring crew performance.	○	○	○	○
Communication		1	2	3	4
2 Way COM	Use standard phraseology. Clear tone and voice. Appropriate timing. Confirm information.	○	○	○	○
Briefing	Take sufficient time of briefing. Emphasiize importance of asking and provideing information.	○	○	○	○
Assertion	Inquire / Advocacy / Assertion	○	○	○	○
Team Building & Maintenance		1	2	3	4
Leadership	Clear intention. Appropriate followership.	○	○	○	○
Climate	Monitor team performance. Confirm crew member's workload. Acknowledge communication.	○	○	○	○
Conflict Resolution	Open communication. Focus on "What is right?", not "Who is right".	○	○	○	○

Fig. 1 *Per Threat CRM Skills Rating Sheet*



Overall

Descent / Approach / Land

Cruise

Take Off / Climb

Predeparture / Taxi Out

Skills Item	Content	Rating			
Situational Awareness Management		1	2	3	4
Monitor	Share information any crew member recognized about operational situation.	○	○	○	○
Vigilance/ Anticipation	Avoide concentration. Anticipate threat and potential hazard.	○	○	○	○
Analysis	Gather information and use available resources to clearly identify the problem and potential risks.	○	○	○	○
Decision Making		1	2	3	4
Decision	Establish bottom lines. Chose an appropriate strategy from all information and merit/demerit of selection.	○	○	○	○
Action	Be understood chosen strategy by all crew member and perform own tasks to implement the strategy.	○	○	○	○
Critique	Compare desired outcomes with actual progress, review and change own performance.	○	○	○	○
Workload Management		1	2	3	4
Planning/ Prioritizing	Develop plans to avoid high workload. Prioritize with time limitation, volume of tasks and urgency.	○	○	○	○
Distribution	Assign appropriate tasks to crew members and automated systems, monitoring crew performance.	○	○	○	○
Communication		1	2	3	4
2 Way COM	Use standard phraseology. Clear tone and voice. Appropriate timing. Confirm information.	○	○	○	○
Briefing	Take sufficient time of briefing. Emphasiize importance of asking and provideing information.	○	○	○	○
Assertion	Inquire / Advocacy / Assertion	○	○	○	○
Team Building & Maintenance		1	2	3	4
Leadership	Clear intention. Appropriate followership.	○	○	○	○
Climate	Monitor team performance. Confirm crew member's workload. Acknowledge communication.	○	○	○	○
Conflict Resolution	Open communication. Focus on "What is right?", not "Who is right".	○	○	○	○

Fig. 2 *Per Flight Phase CRM Skills Rating Sheet*

Selection of Raters

We selected four raters, identified here as A, B, C and D, for the CRM Skills measuring experiment. Each rater was an experienced captain (average flying time: 9,125 hours, average pilot in command time: 3,025 hours) who had worked in a CRM training-related department of an airline. Rater B had experience as a LOFT instructor and rater D had experience as a check airman. Rater A had aircrew experience of the aircraft type in one of the LOFT scenarios (scenario 3) mentioned below. All the raters learned the proposed CRM Skills behavioral markers, the scoring procedure and the scenario contents before the experiments.

Simulated LOFT Scenarios

Three simulated LOFT scenarios to measure CRM Skills were selected from existing recordings of LOFT exercises. Figure 3 shows the threat codes (Klinect *et al.*, 2001, e.g. Aircraft Threat) included in each scenario. Scenario 3 includes many kinds of threat types while the others have fewer. Scenarios 1 and 2 include three Aircraft threats. In scenario 2, Aircraft threats appear continuously during the Takeoff/Climb phase. In scenario 1, on the other hand, although Aircraft threats appear continuously during the Descent / Approach / Land phase, an Arrival threat is inserted between Aircraft threats in this phase.

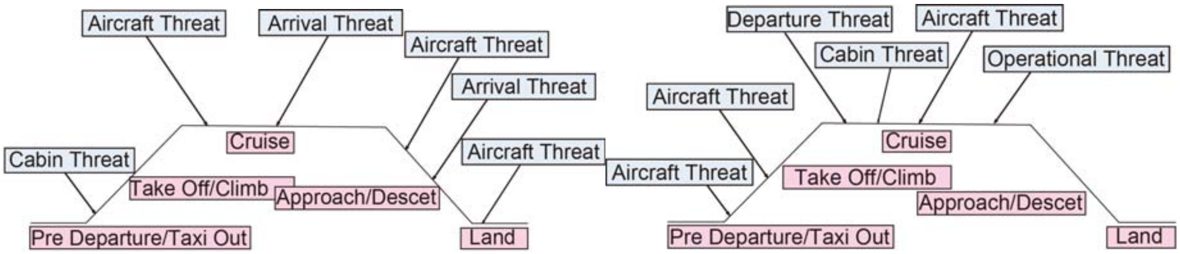
Experimental procedure

The experiment was carried out in three steps:

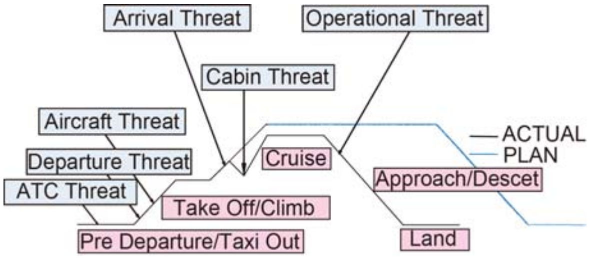
(1) The raters watched video recordings of the three simulated LOFT sessions and completed both types of CRM Skills rating sheets (per flight phase and per threat). The raters were also asked to make notes as appropriate on a CRM Skills observation sheet (Noda *et al.*, 2005, Tsuda *et al.*, 2006) while watching the recordings.

(2) After step (1), we examined the ratings and selected the scenario which had the least differences between the four raters' scores. Based on the rating scores of this scenario, the four raters then discussed those CRM Skills items which they had scored differently. This discussion was a trial to allow the raters to compare their scoring rationales and to decide a "True Score" for each Skill item on which they all agreed.

(3) After step (2), the raters again watched the recording of the scenario with the greatest rating differences, and again completed the two types of rating sheets to verify the validity of the True Score discussions.



a. *Scenario 1* *b. Scenario 2*



c. Scenario 3

Figure 3 *Scenario Structure of simulated LOFT sessions*

Results

Validity of the Proposed CRM Skills Rating Sheet for Each Threat

From step (2) of the experiment procedure, the difference between raters' scores was found to be greatest for scenario 2, and so scenario 2 was scored again after the raters had discussed their rationales. Figure 4 shows average values of standard deviation (SD) of scores of all CRM Skills items for each rater and case. Cases 1, 2, 3 and 4 in Fig. 4 indicate respectively scenarios 1, 2, 3 and the results of the second rating of scenario 2.

It is apparent from the figure that all average SD values are lower for the per threat CRM Skills rating sheet than the per flight phase rating sheet. The unpaired *t* test showed that this difference was statistically significant for cases 1 ($P=0.037<0.05$) and 3 ($P<0.000$), but not for cases 2 ($P=0.098>0.05$) and 4 ($P=0.114>0.05$).

Raters' comments indicate that the main advantages and disadvantages of each type of CRM Skills rating sheet are as follows:

(1) Per flight phase rating

- When the same CRM Skill is observed several times in the same flight phase, the scores for each instance of the Skill could potentially cancel out (nullify) each other. For example, if a rater observes both effective and ineffective behaviors for the "Monitoring" skills during the same flight phase, it is possible that his score will be the average (balance) of these behaviors, and the effective and ineffective behaviors might not appear in the final analysis.
- The "per flight phase" approach allows an overall evaluation of a crew's skills. On the other hand, the "per threat" approach is limited to evaluating crew behaviors when managing (or mismanaging) threats and errors and does not give an overall evaluation.

(2) Per threat rating

- It is possible to rate in detail.
- Timing of evaluation is sometimes difficult because some threats are persistent. For example, in the case of a passenger being taken ill, the rater might be confused as to precisely when to evaluate the crew's management behavior (when the threat first appears, or at some point later) because the crew may continue to address the threat at a later point in time.
- With the per threat rating sheet, it can be difficult to score CRM skills that a crew exercises or fails to exercise because the skill may not be related to any threat on the sheet.

Validity of Discussions to Introduce "True Score"

In step (2) of the experiment procedure, the raters discussed their scoring of cases 1 and 3 in order to derive a "True Score" for each CRM skill item. After this discussion, the raters again observed and rated scenario 2, and the result was analyzed as case 4. The differences between cases 2 and 4 are therefore due to the discussion between the raters and the raters becoming familiar with the scenario, since both cases used the same LOFT scenario.

As is apparent from Fig. 4, the values of variance for case 4 are greater than for case 2 for both types of rating sheet. However, the paired *t* test reveals that these differences are not statistically significant for either per threat rating (P=0.399>0.05) or per flight phase rating (P=0.382>0.05). It is possible, though, that there are concrete differences between cases 2 and 4 that cannot be identified from only the average value of SD. To explain the changes from case 2 to case 4, Figure 5 shows the proportions of each score for these cases. It is clear from the figure that case 2 has the greater proportion of "3 point" and "blank" scores for each item. These results indicate that raters' scores tend to be biased towards middle values. In case 4, the proportion of "3 point" scores decreases, but the proportions of "2 point" and "1 point" scores increase, indicating that scoring tendencies change from a "Central tendency" (Baker *et al.*, 1999) to a more varied evaluation.

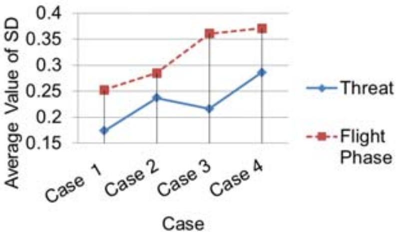
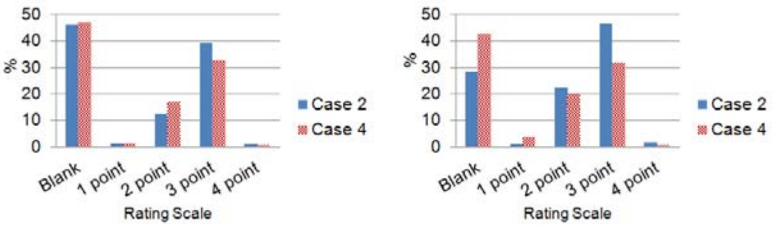


Fig. 4 *Average SD for each case*



a. *Per Threat Rating Sheet* b. *Per Flight Phase Rating Sheet*

Fig. 5 *Percent of Rating for Case 2 vs. Case 4(twice of scenario 2)*

Effect of Included Threat Codes in Simulated LOFT Scenarios

To investigate whether the variance of ratings is affected by the threat types (threat codes) included in the scenarios, Table 2 shows the correspondence between these two quantities. As shown in the table, the average values for the four raters are greatest for scenario 3 (case 3), which includes equal numbers of all threat types. However, we cannot say that it is generally true that scenarios that have equal numbers of each threat code will give greater average scores, because we did not compare scores with another scenario using the same crew who carried out scenario 3.

It is considered that the timing of insertion of a threat into a scenario is important as well as the type of threat. Both scenarios 1 and 2 include three Aircraft threats. While Aircraft threats appear continuously in scenario 2 as shown in Fig. 3, scenario 1 has another type of threat inserted between two Aircraft threats, so Aircraft threats do not appear continuously. The average rating values were lower and variance values were greater in scenario 2, in which the Aircraft threats were presented continuously.

Table 2 *Threat Code (Threat Type) vs. Average Value of Rating and SD*

Threat Code / Case	Aircraft	Departure/ Arrival	Cabin	Operational	ATC	Average Value of Rating		SD
Case 1	3 2 1			0 0		Phase[*1] 3.	17	0.467
						Threat[*2] 3.	12	0.432
Case 2	3 1 1			1 0		Phase 2.	69	0.540
						Threat 2.	71	0.544
Case 3	1 2 1			1 1		Phase	3.39 0.	490
						Threat	3.26 0.	441
Case 4	3 1 1			1 0		Phase 2.	52	0.652
						Threat 2.	63	0.570

Phase: the per flight phase rating sheet, Threat: the per threat rating sheet

Discussion

The final objective of the CRM Skills measurement method is to answer two questions: What CRM Skills does a crew lack corresponding to each type of threat and error, and what training will the crew require to adequately manage threats and errors. For developing such a CRM Skills measurement method, attention needs to be focused on three issues: The CRM Skills rating sheet as a tractable tool for CRM Skills data acquisition; inter-rater reliability training to ensure that reliable data are obtained; and the content of LOFT scenarios. We discuss these issues using the results obtained from the experiment.

Validity of the Proposed Per Threat CRM Skills Rating Sheet

Judging from the results of average values of SD of rating for cases 1 (scenario 1) and 3 (scenario 3), the proposed per threat CRM Skills rating sheet assisted the raters in having a consistent viewpoint by assessing crew behaviors when managing or mismanaging threats / errors. However, individual differences between raters still remained in the results of cases 2 and 4 (scenario 2). Rater comments, mentioned in the Results section above, indicate that the variances of rating for the per flight phase rating sheet were greater than those for the per threat rating sheet because the scoring decisions were different. If raters observed both “effective behavior” and “ineffective behavior” during the same flight phase, one rater might score an average between these behaviors while another rater might score based only one of the behaviors. On the other hand, one rater’s comments gave insight as to why individual rater differences existed even for the per threat rating sheet: the timing of when to evaluate a crew’s threat management behavior is difficult because some threats persist for some time after appearing. Another reason identified is that each rater may focus on different CRM Skills items; e.g. Rater A might score 2 points for the Assertion skill for a given threat, while Rater B might not focus on Assertion and score blank for this skill but instead score 2 points for Leadership for the same threat. This is related to the issue of whether or not it is necessary to standardize the CRM Skills items on which raters should focus.

This study’s analysis was based on the average value and variance of ratings. However, it is considered that analysis should not only use these numerical values but should take into account the raters’ individual judgments of the relative importance of each CRM Skills item. The validity of the CRM Skills rating sheet will then be verified and inter-rater reliability training will be conducted by considering scores weighted by raters’ judgment of importance. For example, if most raters assign a low score for a “Threat X”, even if they focus on different CRM Skills items, the analyst should be feed back that the crew’s management of “Threat X” is very weak, but might then indicate those CRM Skills on which each rater focused as points for improvement e.g. “Assertion”, “Leadership”. A CRM Skills rating sheet that uses a “weighted value” of ratings is likely to be discussed in future.

Inter-Rater Reliability Training

As the result of the discussions to introduce “True Scores”, raters’ scoring tendencies changed from making “ambiguous ratings” to “clear ratings” based on definite judgments; for example, before the discussions some raters hesitated before finally scoring 3 points for an item, but after the discussions their scoring changed from this central tendency to scoring clearly 2 points or 4 points. This is obvious from Fig. 5 and Table 3. As is apparent from Table 3, the proportions of 3-point scores by raters A and B were markedly lower in case 4 (re-rating scenario 2 after discussion). Rater D, who had check airman experience, commented that although it was difficult in case 2 to assign scores of below 3 points (when rating scenario 2 before discussions), in case 4 he was able to assign scores of below 3 points not from the viewpoint of pass or fail, but considering the need for retraining.

This change of rater D’s scoring tendency is revealed from the increased proportion of 1-point scores in Table 3. Rater A’s “central” scoring tendency, by which he tended to score average values, was also improved by the discussion as mentioned in the Results section. Rater A commented although he scored 3 points even for crew behavior which he could not observe in case 2, his rating method changed clearly in case 4 in that crew behaviors which he was not able to observe were scored blank. These findings show that the discussions between raters to introduce “True Scores” contribute to their changing interpretations of the rating scale and avoiding the “central tendency”.

The duration of the discussions, only three hours, was too short to achieve standardization of the raters. Although it was insufficient to achieve totally consistent scoring by the raters, however, some viewpoints such as the interpretation of the rating scale and examples of crew behaviors corresponding to each grade (1-point, 2-point, 3-point and 4-point) could be standardized. The trial discussions therefore helped to familiarize raters with the scoring method, and it is supposed that actual inter-rater reliability training will be conducted in future by repeating the scoring and discussion between raters. It is considered that by such training each rater will understand the rating errors that are easy to commit and be familiarized with the scoring method, and then agreement on the interpretation of the rating scale and CRM Skills items will be performed through an iterative process of scoring and discussions. However, we could not draw any conclusions as to the number of iterations that will be required.

Table 3 Change of Rating from Case 2 to Case 4 *(per flight phase rating sheet)*

		Rater A	Rater B	Rater C	Rater D
Average	2^{*1}	2.77 2.	61 2.	67	2.67
	4^{*2}	2.71	2.42	2.53	2.42
SD	2	0.505 0	.495 0	.586	0.595
	4	0.579	0.620	0.567	0.807
Number of 1 point	2	2 0 0			0
	4	2	1	1	5
Number of 2 point	2 8		15	14	13
	4	6	17	13	9
Number of 3 point	2	43 23	20		18
	4	26	12	18	16
Number of 4 point	2	0 0 2			2
	4	0	1	0	1
Number of Blank	2 3		18	20	23
	4	22	25	24	25

2: Case 2 (rated in scenario 2 before the discussion), 4: Case 4 (rated in scenario 2 after the discussion)

Development of LOFT Scenarios for Threat and Error Management Training

As mentioned in the Results section, it is possible that the variance of rating is affected by the type and timing of threats that appear in scenarios. Variance of rating was greatest for scenario 2, in which a technical event (an Aircraft threat) appeared continuously. Scenario 2 appears to have been aimed at operating procedures training, and it is possible that execution of CRM Skills was hardly observed for this scenario since crew behavior in technical events requires technical skills to execute a prescribed procedure rather than CRM skills in general. Some raters evaluated the fact that captain made decisions by himself without communication with other crew members, because some SOPs does not require much discussion between crew members since the procedures are clearly specified , while other raters evaluated the crew's behavior based on only the captain's "Leadership". It is considered that such issues contributed to the higher variance of rating in scenario 2 than other scenarios.

If the purpose of a LOFT scenario is not procedures training but exercising CRM Skills for TEM, it is necessary to carefully investigate which types of threat are appropriate to be included in the scenario and their timing. A scenario which generates a large variance and low average value of rating between raters is considered inappropriate as a CRM Skills training scenario.

Conclusion

A CRM Skills Measurement Method which includes a Threat and Error Management concept was proposed, and its validity was verified by a CRM Skills measurement experiment.

The results of the experiment showed that the proposed per threat CRM Skills Rating sheet assisted raters to have consistent viewpoint by assessing crew behavior when managing or mismanaging threats / errors, but individual differences between raters still remained. Additionally, discussion between the raters to introduce "True Scores" prompted them to clarify their rationale for scoring. Analysis of the results indicates that factors contributing to individual scoring differences include not only the contents of the CRM Skills Rating sheets and the inter-rater reliability training method, but also the threat types included in simulated LOFT scenarios and the timing of their appearance.

References

Baker D. P., Mulqueen C. and Dismukes R. K. (1999). Training pilot instructors to assess CRM: The utility of frame-of-reference (FOR) training, International Aviation Training Symposium.

Iijima T., Noda F., Sudo K., Muraoka K. and Funabiki K. (2003). Development of CRM skills behavioral markers, TR-1465, National Aerospace Laboratory Report.

Klinect, J. R., Wilhelm, J. A. and Helmrech, R. L. (2001). LOSA Error Code Book 9.0, Proc. of 1st LOSA week.

Noda F., Iijima T., Tsuda H., Sudo K., Yamamori H. and Kobayashi H. (2005). Assessment of CRM Skill Indicators, 44th Aircraft Symposium, No. 3G11.

Tsuda H., Iijima T. and Noda F. (2006). Development of CRM Skills Measuring Method, ICAS2006.

MISHAP REDUCTION TRAINING FOR C-130J CREWS

Robert Nullmeyer
711th Human Performance wing
Mesa AZ

Alan Spiker
Anacapa Sciences
Santa Barbara CA

Greg Deen
Lockheed Martin Simulation, Training, and Support
Little Rock AFB AR

The C-130J is an advanced automation aircraft flown with a smaller crew than its tactical airlift predecessors. It is seeing increased action in theater. The Air Mobility Command sponsored a multi-prong project to improve C-130J aircrew training for operations in high threat environments: (1) analyze C-130J and related platform (C-130 E/H, C-17) mishap reports, (2) survey C-130J crew real world experiences regarding threats to safety, and (3) observe C-130J crews as they plan and execute a challenging, tactically relevant simulator scenario. The Air Force C-130J safety record is enviable—no crew-caused Class A mishaps and low rates across all mishap categories relative to other airlift platforms. Human factors frequently cited in Class B and C reports included checklist errors, distraction, task prioritization, and decision making. Incident and mishap reports both frequently mention problems arising from events external to the crew. In the real-world experience survey, several crews described events that closely paralleled events in the simulator scenario. During the challenging simulator scenario, crews generally accomplished the mission but quality of performance varied considerably. Lower performing crews often struggled with risk management during planning and mission evaluation during execution. They also tended to under-utilize their loadmasters and had difficulty choosing the most appropriate levels of automation during mission execution. Recommendations included an increased emphasis on threat and error management during training and addressing specific skills at particular points in the syllabus.

A longstanding and central role of human error in aviation mishaps is well documented. For example, Helmreich and Fouchee (1993) reported that, from 1959 to 1989, flight crew actions were causal in more than 70% of worldwide air carrier accidents involving aircraft damaged beyond repair. More recently, researchers are reporting diminishing proportions of air carrier mishaps world-wide being attributed to operator error. Baker, Qaing, Rebok, and Li (2008) reported a statistically significant drop in air carrier mishaps involving human error, from 42% in the 1980s to 25% in 1998-2002. Dismukes, Berman, and Loukopoulos (2007) reported that rates of "crew caused" accidents dropped by 50% between 1978-1990 and 1991-2001. While mishaps attributed to crew error may be declining, they still represent a large portion of the total count, so the nature of that error remains an obvious issue. The National Transportation Safety Board (NTSB) reported that 68% of crew-caused accidents from 1989-1990 involved tactical decisions, especially the failure to execute a go-around given an unstable approach, and 84% involved inadequate monitoring/ challenging (NTSB, 1994). Dismukes and his colleagues analyzed more recent (1991-2001) NTSB reports and found 19 US air carrier accidents where crew error played a central role. Descriptions were provided for each of these events, and accident statistics were compared with those from the earlier NTSB analysis. They found that crews faced the same basic challenges in more recent years--74% of recent crew-related accident reports cited tactical decisions and 68% cited monitoring/challenging errors.

By the late 1990s, it became increasingly recognized that errors were only part of the safety picture and that a broader understanding of all threats to safety was essential to develop effective responses, and several airlines moved toward a threat and error management (TEM) model of safety interventions (Helmreich, Klinect, & Wilhelm, 2001). TEM is an organizational response to safety that includes analyses of operational data from line oriented safety audits, flight data recorders, and crew self reports of hazardous situations. The goal is to understand and manage threats to safe operations. From these data sources, several solutions may emerge, including changes in operating procedures or equipment and tailored training interventions. This TEM approach in commercial aviation training is recognized as a "best practice" by the International Civil Aviation Organization, the International Air Transport Association, the National Air Transport Association, and the U.S. Federal Aviation Administration.

In contrast to trends in airline safety, a recent Air Force review of accidents through 2009 revealed that percentages attributed to human factors remained relatively constant at about 70% (Heupel, Gardetto, Dopslaf, Hughes, Williams, & Johnson, 2010). A recent review of mishap rates across *all* Unites States military services revealed little systematic improvement in mishap rates over the past 20 years (Unites States Coast Guard Safety Center, 2009). O'Connor, Hahn, & Nullmeyer (2010) recently reviewed military Crew Resource Management (CRM) training across both the United States armed services and other allied countries, and found no military training programs today that pursue TEM as aggressively as has been the case in commercial aviation.

Air Mobility Command sponsored a series of efforts to better understand threats to safety for C-130J crews and how these crews responded to those threats: (1) review relevant safety data; (2) survey mission qualified crews to solicit accounts of "threats" to flight safety, some type of crew error, and/or a successful mitigation strategy or technique; and (3) observe mission qualified crews perform a tactically demanding capstone simulator scenario during annual refresher training. The overall objective was to provide actionable recommendations for updating C-130J training to better reflect TEM concepts. The focus in this paper is on the first and third of these elements.

Results

Accident and Incident Trends.

The Accident Investigation Board (AIB) homepage (http://usaf.aib.law.af.mil) summarizes AIB reports from Class A accidents (fatality or permanent disability, loss of aircraft, or more than $2 million damage). The Air Force Safety Center (AFSC) home page (http://afsafety.af.mil/) provides considerable summary mishap statistical information, including hours flown and mishap frequencies, by aircraft type and year. Flying hours per year are essential for translating frequencies into mishap rates. AFSC also maintains an electronic database of accident reports, as well as high accident potential and hazardous air traffic self-reports. Data from all of these sources were used to review Air Force airlift mishap trends.

C-130 Class A mishap rates across all models fell consistently from 1970 through 2000, with .73 mishaps per 100,000 flying hours in the 1970's, .56 in the 1980's, and .26 in the 1990's. The rate increased slightly from 2000-2009, to .35 per 100,000 flying hours, and that pattern of slight increases in Class A mishap rates was common throughout the Department of Defense. In response to rising mishap rates, all Services began a series of safety initiatives in 2005. C-130 class A mishap rates dropped from 2.9 in 2001-2005 to 1.6 from 2006-2010, suggesting that the safety initiatives may be proving successful. C-130J flight operations began in the late 1990's. The C-130J safety record compared to earlier C-130 models of is enviable – one Class A ground mishap in 2008. There have been approximately 150 C-130 Class B mishaps in the past decade. Six of these involved a J model C-130, and of these, only one involved crew factors. Two conclusions were drawn from these trends. First, the drop across the past two decades that has been documented in air carrier human factors mishaps was not as pronounced in Air Force accident data. Second, the C-130J community has had an enviable safety record to date.

We broadened the scope of mishap analyses to include C-130E and H Class A mishaps given the shared tactical airlift mission, and to include C-17 mishaps given similarities in cockpit display technology. Mishaps occurred from 1999 through 2009. In both of these other aircraft, numerous mishaps involved human error. There were 13 C-130 E and H Class A mishaps, of which eleven had notable crew factors. Of 22 C-17 Class A mishaps, nine involved notable crew factors. The top human factors in C-17, legacy C-130, and C-130 J mishaps are shown in Table 1. For the C-130J, the human factors came from one Class B mishap and eight Class C mishaps.

The factors in Table 1 come from reports where human factors analyses were based on the Department of Defense Human Factors Analysis and Classification System, or DoD HFACS. DoD HFACS is well documented in multiple places, including the Navy and Coast Guard safety center web sites. Several patterns emerged from our analysis. Inattention was frequently cited in both aircraft with highly automated cockpits (C-17 and C-130J) but not legacy C-130 mishaps. A similar pattern emerged with the highly related factors of cognitive task oversaturation and task misprioritization. Both refer to challenges with managing multiple tasks simultaneously. Risk assessment and decision making are also highly related. HFACS distinguishes between risk assessment and decisions during planning from those behaviors during mission execution. All three platforms cite this skill area during mission execution. Violating stated rules shows up in all three lists. Procedural guidance/publications appears in C-17 and C-130 accidents. A common theme when citing this factor is the difficulty keeping crews current given rapidly changing tactics, techniques, and procedures. While this was not among the top factors in C-130J accident reports, it has emerged as a common factor across the entire Air Force (Air force Safety Center, 2009).

Table 1. The Most Frequently Cited Mishap Human Factors in Three USAF Aircraft Types

C-17 (Class A)	C-130 E/H (Class A)	C-130J (Class B & C)
Risk assessment during operations	**Procedural guidance/publications**	**Distraction**
Procedural guidance/publications	**Risk assessment during operations**	**Task misprioritization**
Inattention	**Mission planning**	**Decision making during operations**
Cognitive task oversaturation	**Planned inappropriate operations**	**Checklist error**
Channelized attention	**Violation - lack of discipline**	**Violation - lack of discipline**
Fatigue physiological mental	**Channelized attention**	**Inattention**
Planned inappropriate operations	**Local training**	**Complacency**
Overcontrol/undercontrol	**Necessary action delayed**	**Misperception of conditions**
Violation - lack of discipline	**Miscommunication**	**Breakdown in visual scan**

Air Force crews are increasingly documenting near misses in flight or on the ground through Hazardous Air Traffic Reports or HATRs. Over 700 HATRs have been submitted involving C-130s of all types from 1999 through 2009. Of these, 20 involved C-130J aircraft. HATRs tend to be short (about 2 pages) summaries of what transpired. Air Traffic Control errors were cited in half of these HATRS, and errors by other pilots were cited in most others. C-130J crew errors were contributing factors in only two of these incidents, further supporting the need to address externally generated threats to safety, rather than focus solely on C-130J in-cockpit error.

Crew Performance in a Simulator-Based Tactical Scenario.

Participants: Twenty mission qualified C-130J crews were observed as they planned a realistic, tactically complex mission; executed that mission in a full crew, high fidelity weapon system trainer; and debriefed the mission. A C-130J crew consists of an aircraft commander (AC), a co-pilot (CP), and a loadmaster (LM).

The Scenario: Annual refresher training for C-130J crews builds up to an exercise where mission qualified crews plan a challenging tactical airlift mission and then execute it in a C-130J Weapon System Trainer, a full crew, high fidelity simulator. An Afghanistan scenario was used in this study. It was specifically designed to interject a number of "threats" or problems to the crews, beginning in planning and continuing throughout the mission. For example, in planning, a Block 5.4 aircraft was specified. This has weight, fuel, and takeoff performance implications that should be considered during planning. Ground threats enroute and around the landing zone were briefed, and the landing zone runway sloped steeply on one end. During execution, challenges included fairly stringent altitude restrictions in the initial climb-out, a marked shift in wind direction affecting an already demanding approach to the landing zone, selected equipment failure, lack of unloading equipment where cargo was being delivered, a cargo fire, and loss of an engine that necessitated a divert decision. Many similar challenges were mentioned by crews in our real-world TEM survey.

Data Collection. Crew performance was independently rated by two observers who used structured data collection forms that were tailored to the scenario being observed. All ratings were based on five-point behaviorally anchored scales, ranging from 1= poor to 5 = exceptional. One observer rated crew CRM processes during three phases of the exercise (planning, execution, and debrief). A second observer rated crew performance. CRM ratings addressed the six Air Force Instruction 11-290 content areas of mission evaluation, task management, situation awareness, crew coordination, communication, and risk management/decision making. Crew performance during planning was rated as seven separately graded items contingency considerations, takeoff and landing data (TOLD), and decision quality, among others. During execution, 14 items were rated including use of automation, checklists, time control, aircraft handling, response to emergencies. The two observers rated their respective areas independently, and did not compare scores or try to reach consensus.

Data Analysis. An initial analysis of the C-130J simulator refresher study data addressed the relationships between ratings that were assigned by the observer for crew process and the observer for mission performance. Subsequent analyses addressed (1) how well the 20 crews handled the "threats" that were administered to them during the session, (2) observer comments that characterize the process behaviors of the most and least successful crews, and (3) the influence of demographic background on CRM behaviors and mission performance.

Overall Ratings. One major question to be answered is the degree to which CRM crew processes predicted mission performance. Figure 1 provides a graphic of the mission execution data from the 20 crews in which the x-axis corresponds to the average CRM process rating received with the y-axis indicating mission performance. Since some crews have identical ratings for CRM process and mission performance, we have indicated multiple crew presence by making the dot bigger with a circle around it. Also, each crew in that position is identified by a crew number above and to the right of the dot. A sizeable positive correlation was observed between crew process and mission performance ($r = .60$, $t = 3.17$, $df = 18$, $p < .01$). In keeping with this relationship, the majority of the crew plots should be located in quadrants I and III. Indeed, from Figure 1 we can see that only one crew, Crew #12, is found in Quadrant II, which is indicative of higher mission performance (average = 4) than would be predicted based on their CRM process rating (average = 2). No crews are found in Quadrant IV (lower performance, higher process). Moreover, two other crews can be said to have a "mismatch" between CRM process and mission performance, as defined by a difference in rating greater than 1 unit. These are Crew #7 (process = 3, performance = 1) and Crew #10 (process = 4.5, performance = 3). The remaining 17 crews are in line with a consistency in ratings between CRM process and mission performance. In fact, 11 of our 20 crews have identical ratings for CRM and performance, either (2, 2) (Crew #4, 11, 13, 14, 20), (3, 3) (Crew #3, 19), (4, 4) (Crew #6, 8, 15), or (5, 5) (Crew #17). For the most part, then, crews that have better CRM process behaviors had better mission performance.

Characteristics of "Strong" and "Weak" Crews. We can use this consistency between CRM and performance to unequivocally identify the strongest and weakest refresher crews during this scenario. Specifically, we labeled the crews whose process-performance plots fall high in Quadrant I as our strongest crews. Looking at Figure 1, we can see by this definition that our four "strong" crews are: #6, 8, 15, and 17. Similarly, we looked to crew plots in Quadrant IV for our "weak" crews. Here, we had five such crews: #4, 11, 13, 14, and 20. In subsequent analyses, we examined the observer protocols for the strongest crews to identify best practice CRM behaviors and corresponding mission performance elements that are representative of what the most successful crews do. In like fashion, we extracted typical crew interaction breakdowns, errors, and associated problems with our weakest crews to generate a list of "avoid these" behaviors. Several highlight of that analysis follow:

- Higher quality mission planning, both in terms of time spent and activities performed, resulted in superior performance. Strong crews tended to plan for the most likely contingencies, particularly with regard to go-arounds at the LZ. Indeed, we observed that thorough LZ study was a key performance-determinant in the study.
- Without exception, the most successful crews in the simulator study treated and utilized the LM as an integral part of the crew throughout the mission. Indeed, we found that effective utilization of the LM was a notable predictor of mission performance in its own right. Thus, it is quite clear that the C-130J, despite its advanced automation, is a true 3-person cockpit (at least during parts of the mission) rather than a 2- or 2.5-person cockpit.
- The ability of crews to program and re-program take-off and landing data (TOLD) during the mission was a major determinant of success. In particular, crews who accepted the TOLD data given them in planning without verification suffered high workload later as the threats began to mount up. Moreover, failure to check data and verify calculations caused some serious problems during LZ approach and landing, resulting in a less successful mission. Strong crews actively calculated and verified as much TOLD data as possible while on the ground.
- Effective set-up and use of automation was a major characteristic of successful crews and its absence was typically a reliable predictor of poor performance. From our analysis of the simulator study data, it became clear that automation use is a continuum in which either extreme – over-reliance on automation to the exclusion of pilotage skills or under-utilization such that the pilot spends the majority of his/her time hand flying – will result in degraded performance. Indeed, several crews where the aircraft commander rarely used the autopilot were extremely task-saturated, resulting in poor mission performance.

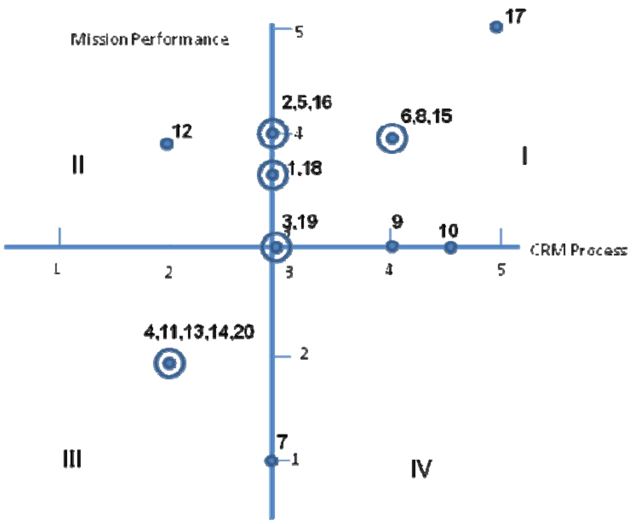


Figure 1. Crew plots of average crew process and mission performance ratings.

- Fortunately, most crew members took the mission seriously, not just as a simulator session, and did their best throughout. However, there were several crews who treated the mission as a typical CONUS operation despite the clear difficulties described during planning. This attitude, coupled with a lack of clear leadership and a closed or partly-closed environment for crew member inputs, creates a bad mix for conducting a mission and, not surprisingly, crews with such an attitude scored at the bottom on mission performance.

Crew Demographics and Mission Performance. Average C-130 flying hours for both C-130J and legacy C-130 are shown in Table 2 for members of both stronger and weaker crews. Comparing the top two rows, we see that there are some fairly sizeable differences between the two groups in average experience. However, because the within-group variances are also fairly large, we elected to perform independent t-tests to determine if any of these differences reached statistical significance. The bottom row depicts the probability of a significant difference given a t-test with seven degrees of freedom. As can be seen, the only significant strong vs weak crew difference is for the LM legacy C-130 hours. Specifically, crews whose LM's had a large number of legacy C-130 flying hours were more likely to part of a crew that was classified "weak" rather than "strong." Crews where the LM had extensive legacy C-130 experience – where the LM was viewed as a backend crewmember only – were at a disadvantage.

Table 2. Average Flight Experience for "Strong" and "Weak" Study Crews.

Crew Classification	**Average AC C-130J Hours**	**Average AC Other C-130 Hours**	**Average CP C-130J Hours**	**Average CP Other C-130 Hours**	**Average LM C-130J Hours**	**Average LM Other C-130 Hours**	**Average Total J Hours**
Strong 1138		1556	968	745	1326	417	3431
Weak 930		910	896	1410	1660	1890	3486
p-value (based on indep t-test)	.63 .	46	.86	.43	.30	**.03**	.93

Implications for Training

The C-130 Hercules aircraft has been in production for over 50 years, serves untold missions, and is utilized by numerous services around the world. The advancements in aircraft technology were minor when compared to the introduction of the "J" version. The "classic" C-130 typically required a crew of four in the cockpit, and, for the USAF, two LMs in the cargo compartment. Appropriate flying speeds, fuel endurance, route headings and weight and balance of cargo were calculated manually using charts and slide-rule technology. The navigators used sextants for overwater route navigation into the ‘80s, slowly replaced by INS and GPS systems. The C-130J reduced the flight crew to two pilots and one loadmaster by taking advantage of advances in computers and avionics.

The military mission, however, became more complex. It only stood to reason that the skill-sets taught to the legacy aircrews needed to change, and perhaps since a flight engineer and navigator are no longer in the cockpit does not mean their activities are no longer required. On the contrary, the tasks of the past exist; those and many more are now within the computers on board. The LMs no longer crank a bicycle chain to unlock cargo or physically remove a connecting link to fully lower the aft ramp; remote electric switches now perform that task. Training aircrews to understand and interface with the computers that operation the aircraft systems will be the major challenges to convert legacy crews to the automated operating system. This report confirmed that the pilot’s prowess with the automation is a key requirement for training. Workload management, even with the aid of automated systems, can still overwhelm a 2-pilot crew. It would seem advisable, then, to encourage the 3-person cockpit concept during refresher training, particularly for crews where extensive legacy C-130 experience is present. When considering the emerging objective of TEM, adaptability will become a key skill for the crews. Tactical airlift, by its nature, includes short-notice changes to the planned mission.

This report clearly indicates the need for increased training in the understanding and use of automation to ensure pilots are fluent in the operating system of the flight management computers, and the 3 crewmembers perform with naturalistic synergy. Loadmasters are no longer just “cargo-guys in the back” but now must play a key role in supporting cockpit activities, particularly in times of high taskloads and changes in mission requirements. While the C-130J aircraft is a significant advancement in aircraft technology, it demands equally significant changes in CRM skills. These changes call for increased mission performance training, as a crew, in realistic simulator scenarios. To address this, the C-130J training system for the USAF will introduce a thoroughly revised training curricula beginning FY 2012. This new plan will incorporate a new part-task trainer that emulates the automated operations of the flight deck. These training aids will concentrate on automation-management skills that are currently taught during simulator lessons. Loadmaster fuselage trainers will also have these training aids to allow the loadmaster students opportunity to acquire proficiency in the cockpit tasks needed to support the pilots prior to their joining the pilots in the flight simulators. This shift in training will allow a reduction of simulator events within the training program, and allow the simulator missions to focus more on CRM and mission-accomplishment skills by the entire crew, for which the full crew, full-motion simulator is better suited.

References

Baker, S.P., Qiang, Y., Rebok, G.W., and Li, G. (2008). Pilot error in air carrier mishaps: longitudinal trends among 558 reports, 1982-2002. *Aviation, Space, and Environmental Medicine*, 79(1), 2-6.

Dismukes, R. K., Berman, B. A., & Loukopoulos, L. D. (2007). *The limits of expertise: Rethinking pilot error and the causes of airline accidents*. Aldershot, UK: Ashgate Publishing Co.

Helmreich, R.L. & Foushee, H.C. (1993). Why crew resource management? Empirical and theoretical bases of human factors training in aviation. In E.L. Wiener, B.G. Kanki, and R.L. Helmreich (Eds.), *Cockpit resource management*, San Diego, CA: Academic Press.

Helmreich, R.L., Klinect, J.R., & Wilhelm, J.A. (2001). System safety and threat and error management: the line operations safety audit (LOSA). In Proceedings of the Eleventh International Symposium on Aviation Psychology (pp. 1-6). Columbus OH: The Ohio State University.

Heupel, K., Gardetto, P., Dopslaf, E., Hughes, T., Wiliams, S. & Johnson, B. (2010). USAF aviation safety: FY2009 in review. Presented at the 2010 annual meeting of the Aerospace Medical Association, Phoenix AZ.

Loukopoulos, L.D., Dismukes, R.K., & Barshi, I. (2001). Cockpit interruptions and distractions: a line observation study. In *Proceedings of the Eleventh International Symposium on Aviation Psychology*. Columbus OH: The Ohio State University.

National Transportation Safety Board [NTSB] (1994). *A Review of Flightcrew-involved Major Accidents of US crriers, 1978 through 1990.* (Report No. PB94-917001, NTSB/SS-94/01.

U.S. Coast Guard (2008). *FY07aviation safety report,* available at http://www.uscg.mil/safety.

THE CONTINUING CHALLENGE OF AVIATION SAFETY IN AFRICA.

Paper to be presented at the 16th International Symposium on Aviation Psychology, Dayton, OH, May 2011

Dr. Maxine Lubner, Dr. Stephen Braccio, Mr. Michael Bartron, Mr. Al Logie.

This study follows an epidemiological approach to examine possible predictors of and current interventions for safety in aviation transportation in two regions with widely different safety records: New York (NY) representing several regions in the US, and Tanzania (Tz), representing several regions in Africa. For most transportation modes, NY has one of the best and Tz has among the worst safety records. This paper identifies some of the similarities and differences between the two regions in order to find ways to improve the safety record in Tz and to ensure that safety continues to improve in NY.

Several US, African, and international public and private entities offer that safety is a serious and growing problem, with injuries accounting for approximately 1 in 8 deaths among males and 1 in 14 deaths among females worldwide (MacKenzie, 2000). In addition, they agree that with today's market globalization, to promote the economy and quality of life of one's own region, other struggling regions must also be enhanced. Local, national, and global connectivity is required for efficient commerce. Connectivity, in turn, requires ongoing security and safety. By improving transportation safety we may also find cost-effective ways to improve both the NY and Tz regions' economies and living standards.

NY is facing new economic, technological and safety challenges, such as those related to insufficient capacity in all modes of transportation. In NY, and across the US, congestion threatens safety, such as by runway incursions, and hampers economic growth by for example, increased delays or emissions pollution. "The great challenge is that of stimulating capacity growth through increased system efficiency, as infrastructure growth will likely be constrained" (Schubert, 2003).

This paper examined some predictors of safety in a preliminary manner, with the assumption that each region's problems and successes can inform the other. By describing factors such as international and government structures, safety culture, training and retention of the workforce, and statistical reports of accident data, we identify several safety predictors. We suggest that patterns of predictors may emerge that will solve the puzzle of why some regions continue to experience disproportionally high accident rates. Some predictors are common to all regions and modes, such as attempts to introduce and maintain a safety management system, a safety sub-culture, and implement advanced training. While technology improvements are necessary, they may not be sufficient to ensure transportation safety. Funding and government support remain challenges in both regions, although most officials and researchers agree that funding alone is insufficient to address all safety issues. Other predictors vary by context, modifiability and cost-effectiveness. Similarities in some of the accident rates can be found.

Our theoretical approach includes examination of the study variables from an epidemiological perspective, where three major levels of variables are examined (see Table 1). In addition, we suggest that competing hypotheses, particularly within the 'environment' level of our model, may serve to explain some of the observed differences in accident rates by world region.

We suggest that the persistence of high accident rates in East Africa in comparison to the US and specifically New York, is primarily related to the environmental level variable of economics (i.e. relative wealth) of each region. As a result of limited funding, the East African governments are constrained in terms of development of their infrastructure, regulations, policies, training and safety sub-cultures, safety management systems, and availability of technology and equipment.

If availability of funding is the issue, and we consider that New York is more economically advantaged than Tanzania, then:

In the more advantaged region, we assume that industry-leading technology and equipment would be more readily available and well-funded. Although we would expect to see accidents resulting from all three predictors in Table 1, in the more economically advantaged regions we expect a larger percentage of accidents resulting from the organizational and individual variables (i.e. policies and human factors issues). Additional resources for technologies and equipment may then only have relatively little impact on improving aviation safety in these advantaged regions.

In the disadvantaged region, technology and hardware improvements would have a relatively larger impact on aviation safety, as the more fundamental infrastructure issues may be present.

Table 1. Theoretical Model: Predictors and Outcomes for NY and Tz

Predictors and Interventions	Outcomes
Environment: Geography, climate, regulations, policies, economy, conflicts, general culture, energy, transportation infrastructure	Accidents, incidents, (violations, hazardous events)
Organization: Professional organizations, businesses, local transportation and safety sub-cultures, safety management system, technology, hardware, facilities	
Individual: Training types and levels, facilities, equipment, attitudes, licensing	

Methods

Findings related to aviation safety were drawn from a review of literature; secondary analyses of publicly available data; and from personal communications and internal reports, or reports specially generated for this study from government officials and researchers, mostly in Tanzania or South Africa (see Appendix A.) Because most information collected on Tz was not publicly available, the results present more information on Tz than NY. Next, we defined a testable model on best practices related to safety culture, government policies, and availability of technology. Our findings also suggested ways to design future data collection to test our theoretical questions about the persistent, high, aviation accident rate in Tz.

Results

The results are presented by comparing Tz and/or the five East African Community (EAC) countries with NY and/or the USA in terms of their overall economic status to determine their relative advantaged vs. disadvantaged status; their aviation accident rates to indicate the extent of the problems with safety; and examine the regions' environmental level variables' differences in terms of their infrastructure, policies and oversight; attempts to implement a safety culture; and their current initiatives to upgrade and improve their equipment and technologies, using the particular example of NextGen implementation for satellite based air traffic control and navigation systems.

Economic as indicators of advantaged vs. disadvantaged status

NY and USA are economically far more advantaged than Tz, as evidenced by the GDP of the US, which in 2010 was estimated to be $14.62 trillion and for NY, - $981 billion. In Tz the GPD was estimated at $US 22.43 billion. In 2010, the income per capita for the US was $47,400, for NY, $50,205 (2009), and in Tz it was $1,500 (www.cia.gov, www.cenus.gov, www.bea.ov, 2011).

Tz is the largest of the Eastern African countries, and has established a strong history of political stability. Agriculture and the services industries are the largest sectors of the economy. Tz has the most (6) World Heritage sites in sub- Saharan Africa, yet lags behind other countries in the region such as Kenya and South Africa in the development of the tourism sector.

However, the country has an aggressive plan to address deficiencies, which include full liberalization of the air transport industry (The Citizen, 2010). The government would raise the funds on its own and in concert with private investors. Restrictions will be removed on routes, capacity, code-sharing agreements, and tariffs, while strengthened government regulations and policy will ensure that operations meet international safety standards. While these plans are positive, Tanzania has had difficulty with other sectors' development and regulation. Bradford (2009) notes that the telecommunications sector showed uneven regulatory activity throughout its growth period, with a lack of resources that led to "under researched and under theorized" elements within regulatory governance.

Aviation accident rates indicate the extent of safety problems in NY and TZ

According to the International Air Transport Association (IATA), North America's hull loss (of Western-built jet aircraft) accident rate was 0.10 per million flights. In contrast, Africa had the worst rate in the world, 7.41. This rate was lower than the 2009 rate of 9.94, but the improvement is not considered sufficient by the global community. There were four hull losses with African carriers in 2010. African carriers are 2% of global traffic, but 23% of global western-built jet hull losses.

The 2010 global accident rate was 0.61, or one accident for every 1.6 million flights. This is a significant improvement of the 0.71 rate recorded in 2009 (one accident for 1.4 million flights). The 2010 rate was the lowest in aviation history, below the 2006 rate of 0.65. In the decade, 2001-10, the accident rate has been cut 42%. (www.iata.org, 2011).

The total number of Tanzanian air accidents, while small in absolute terms, represents a rate that highlights the need for safety improvements (see Table 2). Rates for 2007-9 are not yet published, but some details regarding the occurrences were provided by TCAA (2011). For example, in 2009 and 2010, most accidents involved Cessna aircraft. The pilot's ages ranged from 64 - 22, mean=30 years, while their flight hours ranged from 35-20,000, median of 1,700 hours (TCAA, 2009).

Table 2. Accident/Incident Performance in Tanzania

Occurrences	1995/96	2004/05	2005/06	2007	2009	2010
Departures by local & foreign airlines	38,796	79,727	78,213			
Passengers by local & foreign operators 784,635		2,031,359	2,302,105			
Accidents to local and foreign operators	6	5	3	3	0	1
Incidents to local and foreign operators	25	14	10	1	15	6
Fatal accidents	1	1	1	3	0	
Fatalities	1	8	5			
Accident rate per 100,000 departures –local	17.46	5.97	4.41			
Accidents/100,000 departures -local and foreign	15.47	6.27	3.83			
Incident rate per 100,000 departures –local	69.84	19.40	11.78			
Incidents/100,000 departures -local and foreign	64.44	17.56	14.06			
Fatal accidents/100,000 departures – local	2.91	1.49	1.44			
Fatal accidents/100,000 departures-local & foreign	2.58	1.25	1.28			
Fatality rate per 100,000 passengers –local	0.19	0	0.34			
Fatalities/100,000 passengers –local & foreign	0.12	0.39	0.21			

(TCAA, 2007, 2009, 2011; Personal Communication, 2006)

Although the accident rate has declined steadily over the eleven year period 1995-2006, the fatality rate for local operators remains about three times higher than that for foreign operators. Cumulatively, over this period, local operators were responsible for 2.12 and foreign operators for 0.72 fatal accidents per 100,000 departures. For incidents, local operators account for almost eight times the rate for the same period (12.64 for local vs. 1.64 for foreign operators) (TCAA, 2007).

By comparison, in 2005, US all transportation fatalities totaled 45,650, with highway accounting for most of the deaths, at 43,443 but 616 from aviation transportation (www.ntsb.gov, 2008). The aviation accident rate for all accidents in the USA in 2009 per 100,000 hours for Part 121 scheduled carriers was 0.149; for non-scheduled carriers, the rate was 0.753; for Part 135 carriers, commuters it was 0.685; and for on demand carriers 1.63; and for general aviation it was 7.20. A total of 534 people died from civil aviation accidents in 2009 (www.ntsb.gov, 2011).

Size of the aviation industries in the two regions

Not only is the accident rate in Tz higher than that of NY, its industry is also relatively smaller, especially when taking into account the population differences of the two regions as shown in Tables 3 and 4 below. The total number of pilots registered in Tz in 2008 was approximately 875 (Personal Communication, 2006; TCAA, 2009) and of those, there were approximately 600 pilots with a commercial certificate (see Table 3).

Table 3. Number of aviation professionals

Number of aviation professionals	Tanzania (2006/8)	US (2008/9)	NY (2009)
Population	37,445,392	307,006,550	19,541,453
Total Pilots	**997 (875 in 2008)**	**613,746**	**16,906**
Student Pilots	45	80,989	2,837
Private Pilots	301	222,596	7,382
Commercial Pilots	362 **(600 in 2008)**	124,746	3,645
Airline Transport Pilot	180	146,838	2,951
Total Non-Pilot Airmen	**30**	**678,181**	
Mechanics 326,276			
Air Navigation Services Engineers	21		
Air traffic Controllers	79	26,200	

(Personal Communication, 2006; TCAA, 2011; FAA, 2011)

There are relatively few domestically registered aircraft in absolute numbers or per capita, in Tz, in its neighboring EAC countries, and even in South Africa, which is the regional economic and transportation leader in comparison with those in NY and the USA (see Table 4).

Table 4. Domestically registered A/C by population & per capita for regions (2008)

Country	Population (millions)	Domestically registered aircraft	A/C Per 100,000 Capita
Tanzania 30		67	0.22

Kenya 30		175	0.58
Uganda 22		20	0.09
Rwanda 8.3		4	0.05
Burundi 6.4		4	0.6
EAC - Totals	100	275	0.28
South Africa	44	350	0.8
US Air Carrier	305.7	8,225	2.7
US General Aviation	305.7	224,352	73.4
USA – Totals	305.7	232,577	76.08

(Personal Communication, 2006, TCAA 2011, FAA 2011)

Other comparisons indicate that the Tz government does not fully fund the aviation industry and its infrastructure, which together with its absolute lack of equipment and technology, might contribute to Tz's higher accident rates.

Government funding for aviation in the United States consists of a combination of local initiatives and federal expenditures. The US FAA budget ($000) was $17,066,062 in 2009 (actual), $16,082,731 in 2010 (enacted) and $16,468,000 (requested) in 2011. In the 2011 requested budget, 'Safety' comprises ("000") $8,687,258 or 53%. The 2011 budget request provides a total of $1,143 million in technological support for 'NextGen'(www.dot.gov, 2011).

According to the TCAA, air travel in Tz is more expensive than in other countries in that region. The total Revenue for TCAA in 2006 was US$8,494 million, less total Operating Expenditure US$7,505 million. The Operating Expenditures were organized by administrative rather than program category, such as staff costs, repairs and maintenance, financial expenses (www.tcaa.go.tz, 2011). In the Tanzania Finance Minister, budget speech for 2010/11, the government allocated US$ 1,096.6 million to the infrastructure sector in FY-2010/11 budget, an increase of 12.7 % compared to FY-2008/09. The air industry was projected to grow by 7.9 % in 2009/10, but according to service providers, oppressive regulations, high taxes, levies, and poorly maintained runways and facilities prevented even greater growth (Toroka, 2010).

Capacity increases over the past decade as reported by the Official Airline Guide (OAG) (www.oag.com, 2011) reported in terms of the number of available seats worldwide, has increased 40%. The Americas and Europe showed moderate increases in capacity, with greater increases for Africa (11%), Asia (12%) and the Middle East (13%).

Similarly, for cargo, Africa is showing a strong growth rate: According to IATA, in 2010 Africa showed a much larger growth rate of cargo or freight traffic (28.5%) than North America and/or Europe (23.3%). Surprisingly, the Asia/Pacific growth at25.6% is less than Africa's (IATA, 2010). In absolute terms, however, the differences in size of the cargo industry are also large. In Tz approximately 1600 tons of cargo was transported in 2006 (www.nbs.go.tz, 2008), while in NY approximately 284,000 short tons of freight was transported in 2008 (www.bts.gov, 2008).

According to TCAA (2011), international aircraft movements increased by 5.9% from 23,593 in 2004/05 to 24,996 in 2005/06. This was a result of increased weekly frequencies by foreign airlines, from 67 to 78. The number of international passengers handled, increased from 914,446 passengers handled in 2004/05 to 1,021,822 in 2005/06. This was a result of operators using larger aircraft and increased frequencies.

Domestic traffic recorded an increase in aircraft movements of 5.6%, from the total of 123,420 movements in 2004/05 to 130,435 movements in 2005/06 and a 12.9% increase in passenger traffic, from 1,107,352 to 1,250,563 passengers in the same period.

Overall, aircraft movements increased by 5.7%, while passenger traffic increased by 12.4% between 2004/05 and 2005/06 (TCAA, 2011). Aircraft movements are expected to increase from 178,551 in 2008/09 to 192,620 in 2009/2010 (Toroka, 2010).

In the US, there were 35,143,152 foreign carrier enplanements and 733,836,574 total enplanements in 2008. In NY State there were 7,069,353 foreign carrier enplanements and 44,453,732 total enplanements in 2008. In the New York City metropolitan region, there were approximately 45 million total enplanements in 2008 and 42 million in 2009 (www.bts.gov, 2011).

New York has the dubious honor of being the leader in congestion related problems for the country. Almost 25% of all air traffic delays can be traced to this region (www.panynj.gov, 2011). Although there are battles over policy and budgets, all parties agree that the new technologies will help and, importantly, that safety could be compromised if the congestion issues are not resolved. Safety issues related to congestion in the US include those related to aging infrastructure, airlines' financial stress and an increase in runway incursions (www.ata.org; www.faa.gov; www.panynj.gov, 2011).

In many parts of Africa, insufficient infrastructure development functions to reduce capacity for air travel, hampering regional economic development. To address these issues, the Ministry of Infrastructure Development developed 'Tanzania Vision 2025', which has goals similar to FAA's NextGen. The purpose of both countries' plans is to transform the current air transportation system, using, for example, satellite based navigation to replace radar to meet future air transportation needs. These initiatives are designed to improve safety, efficiency, environmental issues, integrate national defense, homeland security and address the economic needs of the global civil aviation industry (Personal Communication, 2006; TCAA 2011).

Thus, infrastructure development, whether congested or unfilled, must be addressed in order to promote economic development in both regions. Because of the impacts on policy, strategic planning and operations, infrastructure is considered as an environmental level safety risk factor for both regions. In Tz, however, the additional burden of economic obstacles to also improve technology and equipment would add to the safety problems and increase the overall accident rate in comparison to that of NY.

Implementing a safety culture.

There appear to be differences between Tz, and NY and USA in terms of their infrastructure, policies and oversight. Both the US and Tz conform to ICAO's plans to implement a safety culture, but may differ in terms of what technology exists and what policy steps are being taken. The 2010 ICAO meeting for the EAC highlighted 1) countries or states should establish and maintain effective and sustainable safety overflight systems and establish regional agencies for safety oversight and accident investigation, 2) aviation safety culture of African aviation service providers should be enhanced, and 3) a time frame for addressing deficiencies was to be established.

Tanzania, as with many African countries, faces economic pressures that might potentially derail any attempts to regionalize safety efforts. For instance, in 2005, Air Tanzania's operating license was suspended by the TCAA. TCAA's concerns centered on "compromised flight safety" issues that included improper aircraft inspections and shortage of qualified technicians and pilots. There was a real possibility that the airline would not recover and would cease operations (The Mercury, 2008).

In 2009 the Tanzania Minister for Infrastructure Development bemoaned the lack of qualified personnel in the aviation sector. Because this is a regional issue, the Minister implored the ICAO to support collaboration efforts to improve aviation safety in East Africa. Tanzanian officials pointed to a lack of policies, regulatory structure and the ability to retain qualified personnel as the main factors depressing safety statistics (Africa News, 2009). Bradford (2001) noted that the chronic shortage of qualified personal cannot be filled effectively with the use of expatriates.

Collaboration among the East African Community (EAC) is not always a given. A 2008 conference held in Arusha, Tz, showed that prior efforts to establish collective air traffic safety standards had not been fully ratified by any member nations except Uganda (BBC Monitoring Africa, 2008). Adequate funding is lacking in the EAC's push to improve air safety. In northern regions and in South Africa, air traffic control systems are long in place. Not coincidentally, many of these countries also have a history of regulation and enforcement. However, in much of the African interior, pilots are flying over vast areas of uncontrolled airspace. With large percentages of the population of these countries engaged in daily struggle to provide for their most basic needs, it is difficult for their governments to commit to costly technology projects such as satellite based air navigations systems. As recently as October 2010, the EAC's Civil Aviation Safety and Security Oversight Agency (CASSOA) addressed the funding issue. CASSOA is required to address deficiencies identified by ICAO but admits that neither they as an organization nor the individual member states, including Tanzania, possess the financial or technical abilities to address the deficiencies. ICAO agrees that a major hurdle to overcome is the lack of qualified personnel who typically leave for more lucrative positions upon completion of their training.

In the United States, safety management has a long history dating back to the 1940's. Lu, Wetmore, and Przetak (2006) noted that the FAA has advocated the use of System Safety protocols for the last 20 years for accident prevention and enhancement of safety management. The FAA's current approach to the implementation of a nationwide Safety Management System (SMS) is to develop a single set of rules for all branches of the industry rather than addressing each stakeholder's concerns on an individual basis. The FAA notes that most elements of inherent in an SMS are already employed within the industry (www.faa.gov, 2011; Flight International, 2008).

In summary, we demonstrate that there are persistent differences between NY and Tz in terms of their economies and their aviation accident rates, industry size and safety culture implementation capabilities. We recommend that further research be conducted to more directly test the competing hypotheses derived from our theoretical model. We would collect data on causes of aviation accidents and incidents corresponding to the three levels of variables in our model in each region. Testing the model's competing hypotheses about the contributions of infrastructure and technology to safety in economically different regions would direct formation of cost-effective solution(s). Finally, verified prediction(s) and solution(s) could address the relationship between environmental, organizational and individual level variables to aviation safety in both economically disadvantaged and advantaged regions of the world.

REFERENCES

1. Air Transport Association,www.ata.org, retrieved August 2008
2. BBC Monitoring Africa, (2008). Uganda: East African regional body raps members over integration laws.
3. Bradford (2001) Indigenizing Organizational Change: Localization in Tanzania and Malawi; An African Centric Alliance. *Journal of Managerial Psychology*. Volume 16:1
4. Elsayed-Elkhouly, S. and Buda, R. (1997) A Cross-Cultural Comparison of Value Systems of Egyptians, Americans, Africans, and Arab Executives. *International Journal of Commerce & Management.* Volume 7
5. Ernest, S. (2010) Tanzania aviation plan will be global envy and (2009) Tanzania; Aviation operations threatened. in Africa News: The Citizen.
6. Lubner, M., Barton, M., Logie, A., Chen, G. (2008) "Preliminary examination of multimodal transportation safety, in Tanzania and New York, USA." *International Annual Roads Federation Conference*, Tanzania
7. Lu, C., Wetmore, M. and Przetak, R. (2006) Another Approach To Enhance Airline Safety: Using Management Safety Tools. *Journal of Air Transportation*. 11:2
8. MacKenzie, E. J. (2000). Epidemiology of injuries: current trends and future challenges. *Epidemiol Rev. 2000;22(1):112-9*, Retrieved December 12, 2007
9. Official Airline Guide (OAG), www.oag.com, retrieved February 28, 2011
10. Ott, J. (2001) Rising African safety culture paves way for new projects. *The East African (Nairobi*)
11. International Aviation Transport Association (IATA), www.iata.org, February 25, 2011
12. International Civil Aviation Organization (ICAO) www.icao.org, February 25, 2011
13. Port Authority of New York and New Jersey, www.panynj.gov, November 20, 2008, February 25, 2011
14. Personal Communication (2006) Kamhabwa, J. and Alloo, M. – see Appendix A
15. Toroka, E. (2010) Tanzania: stifling regulations, taxes play havoc with aviation business. *Business Times, Tanzania*, www.businesstimes.co.tz, retrieved February 27, 2011
16. Tanzania Civil Aviation Authority (TCAA) (2011), www.tcaa.go.tz, retrieved February 27, 2011
17. Tanzania National Bureau of Statistics (2011), www.nbs.go.tz, retrieved June 18, 2008
18. US Federal Aviation Administration, www.faa.gov, retrieved February 23, 2011
19. US Bureau of Labor Statistics, www.bts.gov, retrieved December 23, 2008, 2011
20. US Census, www.census.gov, quickfacts.census.gov, retrieved February 23, 2011
21. US Central Intelligence Agency, www.cia.gov, retrieved February 23, 2011
22. US Bureau of Economic Analysis, www.bea.gov, retrieved February 23, 2011
23. van Gorp, A. and Maitland, C. (2009) Regulatory Innovations In Tanzania: The Role Of Administrative Capabilities and Regulatory Governance. *The Journal of Policy, Regulation and Strategy for Telecommunications, Information and Media.* 11: 1
24. World Bank. Tanzania receives additional funding for local government support project. M2 Presswire. Coventry, retrieved June 18, 2008

APPENDIX A

1. Eng. Omar Abdullah Chambo, Permanent Secretary, Ministry of Infrastructure Development
2. Margaret T. Munyagi (Mrs.), Director General, Tanzania Civil Aviation Authority (TCAA)
3. Eng. John Njawa, Director, Safety Regulations, Acting Chief Inspector of Accidents
4. Ephraem C M Mrema, Chief Executive, Tanzania National Roads Agency (TANROADS)
5. Eng. Hagai Bishanga, Tanzania Technology Transfer Center
6. Eng. Light Choboya, Tanzania Technology Transfer Center
7. Dr. Estomihi Masaoe, Head of Transportation and Geotechnical Engineering Department
8. College of Engineering and Technology, University of Dar Es Salaam
9. Robert Mwesigwa, Technical Coordinator, Civil Aviation Safety & Security Oversight Agency
10. Dr. Bridget Ssamula, CSIR, South Africa
11. Mr. Julius Kamhabwa, Chief, Civil Aviation Security Inspector, Tanzania
12. Dr. Mumtazahussein Alloo, Director, Air Navigation Services, TCAA
13. Mr. Prosper Tesha, General Director, Tanzania Airports Authority
14. ICAO – Safety and investigator officials
15. US DOT: librarian researchers, policy dept. interviews
16. Dr. Bill Bramble, US NTSB

OCCURRENCE HUMAN FACTORS ANALYSIS MODEL (OHFAM)

1. Luo Min 2. Rong Mei 3.Li Jing
China Academy of Civil Aviation Science and Technology

Beijing 100028，China

In order to enhance the classification, analysis and utilization of safety information, China civil aviation develops a more suitable model named “Occurrence Human Factors Analysis Model(OHFAM)” ,which based on the “Human Factors Analysis and Classification System” (HFACS) and the actual operation conditions and characteristics of China civil aviation. This model consists of five layers and especially adds the layer of “Government Supervision” which reflects the deficiency of regulatory authorities in Safety Supervision. On the basis of safety information analysis and extensive research, OHFAM considers safety culture and operational characteristics of China civil aviation, and offers five layers in details from flight, maintenance and air traffic control.

In this paper, the basic elements and functions of OHFAM are introduced. And then we use this model to analyze the incidents of the last five years (2006-2010) of China, and sum up the main contributing factors.

OHFAM’s Origin and Characteristics

Since 1940, International Aviation indicated 75% of the accidents were due to one or more "human error", and then we got the point gradually that “to error is human”, it means that we should not only concern about human error, but more in-depth study on the organization and management factors behind the human error.

The most well-known human factors analysis model proposed by James Reason in 1990 is “Swiss cheese model”. According to this model, the failures of all levels interact and lead to disastrous consequences; the failures are the "holes” of system at different levels 1)[1]. Based on Swiss cheese model, Wiegmann and Shappell1)[2]1)[3] developed the Human Factors Analysis and Classification System（HFACS）,this model helps to define the “holes of cheese model” to promote the availability in accident investigation and information analysis.

However, the existing human factors analysis model cannot fully meet China's actual needs. On one hand, the classification of factors and items in China is slight different, on the other hand, the existing model does not involve the top management of civil aviation authorities - the government factors. Therefore, after the combination of HFACS and human factors of previous research results 1)[4] , we develop a more suitable model named “Occurrence Human Factors Analysis Model (OHFAM)” for China civil aviation human factors analysis. It has five layers including “Unsafe Behavior”, “Preconditions for Unsafe Behavior”, “Department Management”, “Organizational Influence” and “Government Supervision”. The model clarifies its sub-categories of factors (as shown in Figure 1) and gives various items as the expression of each factor. The OHFAM’s checklist gives targeted items to differ from flight, maintenance and ATC in “Unsafe Behavior” and “Preconditions for Unsafe Behavior”, expanding the model’s availability.

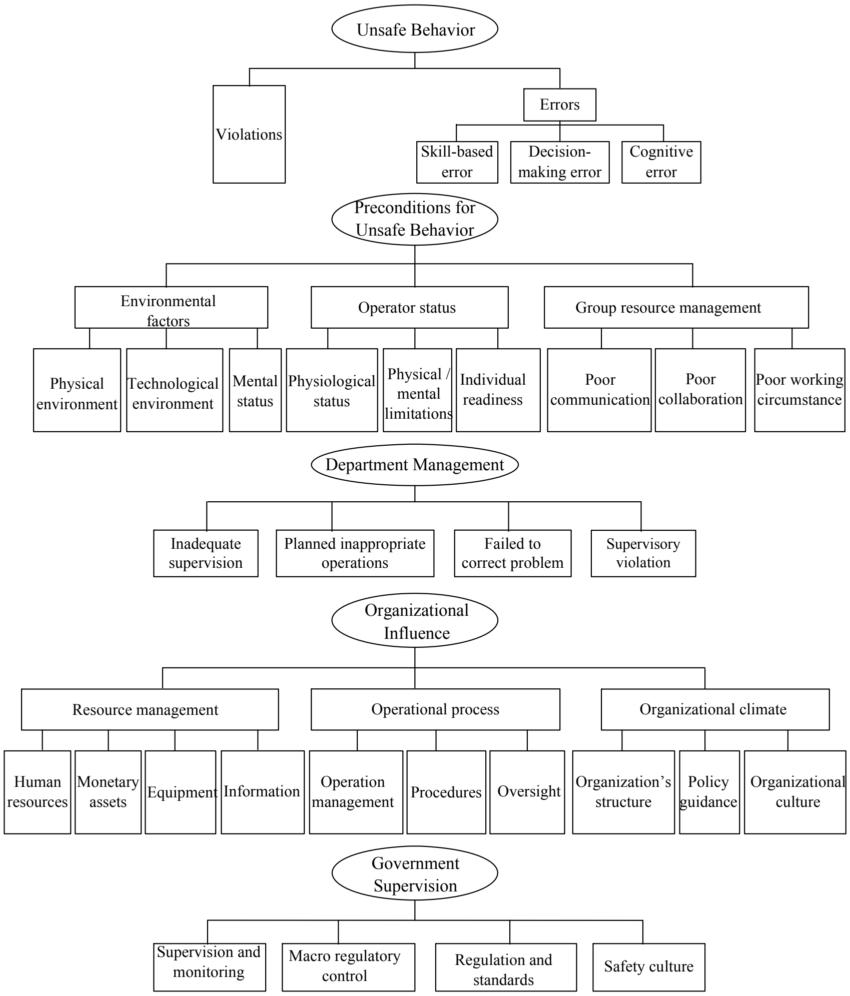


Figure 1 OHFAM Framework

Unsafe Behavior

"Unsafe behavior" refers to the frontline operator' behavior that violates the objective laws of safe operation during the work process, and lead to occurrence directly. In OHFAM, "unsafe behavior" can be classified into two categories: violations and errors. Violations represent the willful disregard for the rules and regulations that ensure the safety of flight. Errors represent the mental or physical activities of individuals that fail to achieve their intended outcome. It can be classified into three categories: skill-based error, decision-making error and cognitive error.

Preconditions for Unsafe Behavior

"Precondition for unsafe behavior" is the adverse objective and subjective conditions which cause the unsafe behavior. In OHFAM, it can be classified into three categories: "environmental factors", "operator status" and "group resource management". "Environmental factors" refers to the physical and technological environment which reduces the operator's performance or lead to unsafe behavior. "Operator status" is the personal condition which may reduce the operator's performance, including "mental status", "physiological status", "physical/mental limitations", and "individual readiness". Besides, "Group resource management" refers to the poor communication or collaboration among the operator and their group which may lead to the occurrence.

Department Management

"Department Management" refers to the supervisor at the middle management does not effectively control or monitor the status of front-line operators and their operating environment, In OHFAM, it can be classified into four categories: "inadequate supervision", "planned inappropriate operations", "failed to correct problem" and "supervisory violation"."Inadequate supervision" refers to the middle managers failed to fully perform its regulatory functions. "Planned inappropriate operations" refer to the department managers do not consider completely when they make the operational plans or the plans are unreasonable. "Failed to correct problem" refers to those instances when the deficiencies of individuals, equipments ,trainings or other related safety areas are "known" to the middle manager, yet are allowed to continue unabated."Supervisory violation" refers to those instances when existing rules and regulations are willfully disregarded by middle manager.

Organizational Influence

"Organization influence" refers to the top management's inappropriate decision. In OHFAM, it can be classified into three categories: "resource management", "operational process" and "organizational climate". "Resource Management" encompasses the decision of corporate-level regarding the allocation and maintenance of organizational assets such as human resources, monetary assets, and equipment and information. "Operational process" refers to the processes that govern the daily activities, establish and maintain of standardized operation procedures, and keep balances between the workforce and management. It can be reflected from the operation management, procedures, and oversight. "Organizational climate" refers to a broad class of organizational variables that influence worker performance. It can be reflected from the organization's structure, policy guidance and organizational culture.

Government Supervision

"Government Supervision" reflects the deficiency of regulatory authorities in safety supervision. In OHFAM, it can be classified into four categories: "supervision and monitoring", "macro regulatory control", "regulation and standards" and "safety culture". "Supervision and monitoring" refers to civil aviation regulatory authorities supervise and evaluate their enterprises, and expose the related illegal and violation activities. "Macro regulatory control" refers to civil aviation regulatory authorities' plan of infrastructure development and the pace or the direction of industry to achieve a goal of a safe, stable and sustainable development. "Regulation and standards" is the rational induction and systematic summary of the safety law, working experience and practice of aviation operation. "Safety culture" encompasses scientific safety management concept, the right values and the strict criterion.

The Analysis of Crew Caused Incidents Based On OHFAM

We use the OHFAM to analyze the 76 crew caused incidents[1] of the last five years from 2006 to 2010 in China, and determine what kind of the factors and items happened frequent (as shown in Table 1).

Table1

Analysis of the Last 5 Years Crew Caused Incidents Based on OHFAM

Layer(Factors)			fn	Items	fn
Unsafe Behavior	Violation 42			a) Violate flight operation procedure	22
				b) Violate ATC instruction	7
	Error	Skill-based error	124	a) Delayed manipulation	13
				b) Poor attention distribution	13
				c) Delayed deviation amendment	12
		Decision-making error	36	a) Wrong decision of go around	21
				b) Improper takeoff\ landing decision	9
		Cognitive error	37	a) Misjudge distance/altitude/airspeed	12
				b) Loss of location consciousness	5
Precondition for Unsafe Behavior	Environmental factors		99	a) Wind	19
				b) Rain\Snow	17
				c) Poor visibility	15
	Operator status		122	a) Unfamiliar with airport	21
				b) Poor practical experience	18
				c) Poor theoretical knowledge	16
				d) Fluke mind	10

---

[1] Data from the China Civil Aviation Safety Information Reporting Database.

			e) Fatigue 3	
	Group resource management	117	a) Lack of explicit distribution b) Insufficient cross-check c) Fail to conduct adequate brief d) Lack of proper authority gradient	18 13 12 9
Department Management	Inadequate supervision	4	Fail to implement supervision	3
	Planned inappropriate operations	20	Crewmember mismatch	12
	Failed to correct problem	1	Fail to report unsafe information	1
Organization Influence	Resource management	112	a) Insufficient CRM training b) Lack of theoretical training c) Lack of manipulation training	34 17 17
	Operational process	14	a) Lack of risk management b) Lack of supervision	4 4
	Organizational climate	17	a) Poor safety culture b) Insufficient group consciousness	10 3
Government Supervision	Regulation and standards	3	Incomplete regulation	3
	Supervision and monitoring	4	Absent supervision	3

Note： fn is an frequency which is the numbers of this factor or item per 100 crew caused incidents.

According to the frequency of the factors and items described above, it can be summarized in the following significant issues.

1) Poor flight techniques. Analysis results show that the most common unsafe behavior is skill-based error, which mainly reflects on “Delayed manipulation”, “Poor attention distribution” and “Delayed deviation amendment”. These errors indicate the pilot’s flight techniques are not proficient.

2) Lack of teamwork. Analysis results show that the “Group resource management” of “precondition for unsafe behavior” is the most important factors contributed to the incidents. It mainly reflects on “Lack of explicit distribution”, “Insufficient cross-check” and “Fail to conduct adequate brief”. These errors indicate there are not effective

communications between crew members.

3) Crewmembers mismatch. Result about the group resource management also reflects the crew manning or scheduling is inappropriate, especially on the authority gradient. Because most companies have adopted the automatic scheduling to arrange the qualified pilots in a short time, but it cannot take into account the pilots' technical characteristics, operating custom, personality characteristics and other factors. The mismatch is always an issue.

4) Crew fatigue needs more attention. Crew fatigue is one of the significant problems of the past five years. Airlines in China faced larger transportation pressure year by year, but were forced to operate at full capacity. In this situation, the serious phenomenon is the shortfall of pilots especially the qualified caption, finally resulted in a heavy workload and the reduced rest time, and then induced fatigue.

Conclusion

Based on the HFACS, this paper establishes the OHFAM considering features of China civil aviation. In summary, the model can provide a systematic approach for accident investigation, which is conducive to guide the information gathering and information analysis; The model considers the deep factors on the government layer, which is conducive to find that the government regulatory problems; The model verify the theory that the accident was caused by the interactive multi-level factors, which is conducive to provide the safety recommendations and measures.

References

[1] Reason, J. (1990). Human Error. Cambridge: Cambridge University Press.

[2] Douglas A.Wiegmann, Scott A.Shappell. (2000). The Human Factors Analysis and Classification System- HFACS (DOT/FA A/A M-00/7). Washington, DC: Office of Aviation Medicine.

[3] Douglas A.Wiegmann, Scott A.Shappell(2003). A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System. ISBN-13: 978-0754618737. Ashgate Publishing.

[4] China Civil Aviation Human Factors Workgroup (2003). Civil Aviation Human Factors Training Manual. Beijing: China Civil Aviation Press.

A PRELIMINARY ANALYSIS OF AERONAUTICAL SERVICES IN AIR NAVIGATION ACTIVITY

Cabral, Lisia Maria Espinola da Silva Pacheco
Costa, Igor
Leal, Luciana da Costa
Madureira, Dulce Maria Saint Pastous
Mendes, Stael Prata
Moreira, Edméia
Reuter, Cristina Juliet de Jesus
Leão, Silvyanne di Paula Santos
Tavares, Allan
INFRAERO - Brazilian Airport Infrastructure Enterprise, Brazil

DECEA - Air Space Control Department regulates Air Navigation activity in Brazil and authorizes military and civilian organizations to provide aeronautical services. This article presents a preliminary analysis of brazilian Air Navigation activity involving the following aeronautical services provided by a civilian organization: Air Traffic Control, Management and Telecommunication; Aeronautical Meteorology; and Aeronautical Information. This analysis is based on civilian psychologists´ professional practice and has the main purpose to subsidize Human Factors preventive approaches and studies to contribute to more proactive and predictive safety interventions; improvements on organizational safety culture, workers´ performance and interfaces (BRASIL, 2010a; ABERGO, 2010).

1. Air Navigation Activity in Brazil

DECEA - Air Space Control Department is a military institution subordinated to COMAER - Aeronautical Command, subordinated to Defense Ministry, responsible to regulate and inspect brazilian Air Navigation activity, based on ICAO - International Civil Aviation Organizational publishing about the following specialized aeronautical services: Air Traffic Control, Management and Telecommunication; Aeronautical Meteorology and Cartography; Aeronautical Information; Flight Inspection; Search and Rescue. Therefore, it is also the brazilian aeronautical authority which homologates and authorizes SISCEAB - Brazilian Air Space Control System following systemic links to provide these services: 4 military CINDACTA - Air Space Control and Defense Centers; one military SRPV - Flight Protection Regional Service; 79 military Detachments; several civilian EPTA - Telecommunications and Air Traffic Services Stations; and the civilian INFRAERO - Brazilian Airport Infrastructure Enterprise. Initially, a summary of the main aeronautical services in Air Navigation activity will be presented, emphasizing the ones provided by INFRAERO, which will be the focus of this analysis, as follows: Air Traffic Control, Management and Telecommunication; Aeronautical Meteorology; and Aeronautical Information (BRASIL, 2010a; ABERGO, 2010).

1.1. Air Traffic Control, Management and Telecommunication Service

Air Traffic Services consist of standard phraseology communication between ATC - Air Traffic Controller and Air Traffic Center or between ATC and aircraft pilots, aiming at achieving safety flight purpose. The complexity level of air traffic scenery determines the type of service that shall be offered by the following operational segments:

1.1.1. Aeronautical Telecommunications Radio Stations. They provide appropriate flight information services to aircraft pilots about the existence of other aircrafts and obstacles in the same air space. There are more than 90 stations installed into brazilian aerodromes, of which 72 are INFRAERO stations operated by PNA-OEA - Air Navigation Professionals-Aeronautical Station Operators, responsible to transmit messages using AFTN - Aeronautical Fixed Telecommunications Network, of which fifty are AFIS - Aerodrome Flight Information Services.

1.1.2. TWR – Aerodrome Control Towers. They provide Aerodrome Control Services of aircraft maneuver, take-off, landing and overflying, for collision avoidance with other aircrafts, obstacles and vehicles if moving in the aerodrome runway. TWR jurisdiction area embraces aerodrome´s air traffic circuit and maneuver. INFRAERO has 22 TWR, operated by ATC named PTA - Air Traffic Professionals.

1.1.3. APP - Approach Centers. They provide Approach Control Services of aircraft take-off and landing, for its appropriate separation of other aircrafts or obstacles. TMA - Terminal Maneuvering Area and CTR - Control Zone are APP jurisdiction space. In Brazil, there are 47 APP, of which 13 are INFRAERO APP, operated by PTA.

1.1.4. ACC – Area Control Center. INFRAERO doesn't provide this service, which is operated by DECEA ATC for monitoring aircrafts during air route, to guarantee their safety separation. FIR - Flight Information Regions consist of ACC jurisdiction area, which embraces several TMA and air routes. There are 5 ACC settled in Brazil, linked by a communication structure involving: 380 DECEA SMA - Aeronautical Mobile Service, which are radio communication stations; DECEA SFA - Aeronautical Fixed Service for telecommunication among different Air

Traffic Centers by phone networks; and DATACOM AFTN for communication of planning, aircraft landing / take-off, arriving / retarding flights, engine failures monitoring and logistic purposes among aircrafts and air companies.

1.1.5. Air Navigation Groups and Units. They provide complementary air traffic control services´ support to give information flight. INFRAERO has 69 Air Navigation Groups and 51 Air Navigation Units.

1.2. Aeronautical Meteorology Service

INFRAERO MEG - Meteorologist and PMET - Meteorologist Technician are responsible for this service. So, they publish noticed and predicted meteorological information involving visualization, treatment and diffusion, associated to DECEA REDEMET and OPMET networks' coordination. This service is composed by:

1.2.1. CNMA - Aeronautical Meteorogical National Center. It is settled at CINDACTA I, situated at Brasília City (Federal District) and operated by 2 meteorological information bases: OPMET, responsible to make routine national and international meteorological reports (METAR, TAF, SPECI, SIGMET); and REDEMET (BRASIL, 2010b), which uses AFTN to publish meteorological information in order to integrate meteorological stations known as REM - Meteorological Stations Networks. There are 3 types of REM: EMS - Surface Meteorological Stations / Classes I, II and III, operated at aerodromes to collect meteorological information about landing runway conditions and its codification and transmission for meteorological data basis services, of which 68 are INFRAERO EMS, using INFORMET network; EMA - Altitude Meteorological Stations, which are equipped with a hydrogen gas balloon attached to a sounding lead with sensors and GPS - Global Position System, to be thrown at atmosphere, aiming at collecting, codifying and transmitting information for data basis of Aeronautical Meteorology Vigilance System of the World Meteorology Organization, and of which 5 are INFRAERO EMA; and ERM - Meteorological Radar Stations, operated by DECEA CMV - Vigilance Meteorological Centers to complement meteorology vigilance in adverse conditions for air operations of high Air Traffic density areas.

1.2.2. CMV - Vigilance Meteorological Centers. There are 4 DECEA CMV in Brazil, localized at ACC and responsible for FIR meteorological conditions vigilance that may affect air operations. INFRAERO has no CMV.

1.2.3. CMA - Aerodrome Meteorological Centers. Their purpose is to support air operations and air traffic services operated at aerodromes, and to diffuse meteorological information and weather forecast predicted by other centers to the crews and the flight operator dispatchers. INFRAERO has 68 CMA linked to EMS.

1.2.4. CMM - Military Meteorological Centers. They are situated at COMAER air bases operated by DECEA to support military aviation in restricted places. INFRAERO, as a civilian organization, has no CMM.

1.3. AIS - Aeronautical Information Service

It consists of collecting, generating, processing and publishing necessary information for planning and execution of a safety flight. INFRAERO has 66 AIS Rooms operated by PNA-TIA - Air Navigation Professionals-Aeronautical Information Technicians, based on DECEA AIS publications, as follows:

1.3.1. IAIP - Integrated Aeronautical Information Publication. Documents based on ICAO standardized publications: AIP - Aeronautical Information Publication, with aeronautical permanent information and long duration modifications´ registers; AIP Supplement, which publishes AIP temporary and permanent modifications and organizes its changes; NOTAM - Notice to Airmen, which complements AIP, ROTAER - Air Routes Auxiliar Manual, AIP Supplement and Aeronautical Charters, containing information to establish or modify any aeronautical installation, service, procedure or danger for operators uncharged of flight operations; BIP - Previous Flight Information Bulletin, prepared by AIS operators or emitted by NOTAM data basis, to attend pilots planning flights; AIC - Aeronautical Information Circular, with explanation, advisement, administration or technical information.

1.3.2. ROTAER - Air Routes Auxiliar Manual. It consists of DECEA brazilian publications created to help pilots to plan their flights and to navigate the national territory.

2. INFRAERO Aeronautical Services in Air Navigation Activity

2.1. Purpose

EPTA and INFRAERO provide brazilian aeronautical civilian services in Air Navigation activity, homologated by DECEA. The ones provided by INFRAERO are: Air Traffic Control, Management and Telecommunication; Aeronautical Meteorology and Information. The main purpose of this article is to make a preliminary analysis of INFRAERO services developed by organic operators, based on its psychologists´ practice contribution, aiming at subsidizing Human Factors approaches and studies to implement proactive and predictive safety interventions; and improvements on organizational safety culture, workers´ performance and interfaces.

2.2. INFRAERO main characteristics

INFRAERO is a civilian mixed economy and an indirect public administration enterprise guided by brazilian CLT - Labor Rules Consolidation and, in which concerns Air Navigation activity, also by DECEA regulations

(BRASIL, 2009b; BRASIL, 2008c). Its mission consists of: "providing Airport and Air Navigation structure and services, contributing to the national integration and the sustainable development of the country" (BRASIL, 2010c). It has 67 airports, of which 28 are international and the others are domestic ones. It has 28.000 employees, of which 13.293 are organic and the others are outsourced staff. It has 1.585 professionals working on aeronautical services: 576 PTA; 38 MEG; 332 PMET; 432 PNA-OEA; 198 PNA-TIA; 19 PNA-ESP - Air Navigation Professionals-Specialists (BRASIL, 2011). INFRAERO has: 72 Aeronautical Telecommunications Radio Stations operated by PNA-OEA, of which 50 are AFIS; 22 TWR; 13 APP; 69 Air Navigation Groups; 51 Air Navigation Technical Units; 68 EMS and CMA; 5 EMA; and 66 AIS Rooms operated by PNA-TIA.

2.3. Problematic

Every DECEA regulation is based on ICAO publications (ICAO, 1998; ICAO, 2002; ICAO, 2000; ICAO, 2005; ICAO, 2006), which, from the 90 decade on, started to establish Human Factors and Ergonomics requisites with the purpose of: minimizing human errors as the main contributor aspect of aeronautical accidents; and defining instruments and methods to identify, monitor and control systemic threats that may motivate them. ICAO Human Factors and Ergonomics specific documents admit human error as a normal condition, but introduce instruments, techniques and methods to understand, address and learn about possible systemic threats which may lead to it, aiming at reducing human contributions to aeronautical accidents (BRASIL, 2010d).

Human Factors is: “The group of sciences which studies all elements which contribute to men interactive relations, in a certain context, surrounded by several systems, determined by their dynamics, efficiency and efficacy. It concerns the optimization of human well being and system global performance for adapting work sets to human characteristics, abilities and limitations, for an efficient, effective and safe performance” (BRASIL, 2010d).

In order to attend ICAO requirements (ICAO, 1998; ICAO, 2002; ICAO, 2000; ICAO, 2005; ICAO, 2006), brazilian aeronautical authorities – DECEA, ANAC - National Civil Aviation Agency and CENIPA - Aeronautical Accident Prevention and Investigation Center – do their best to bring up to date their own regulations, based on ICAO Human Factors and Ergonomics requisites, in which there are relevant psychological aspects to be issued.

Psychologists´ participation on aeronautical accidents and incidents prevention and investigation in Brazil is an old practice, but on air traffic control incidents prevention and investigation is a recent initiative, requiring, besides their professional qualification (BRASIL,2008b), also the definition of interdisciplinary procedures to improve operators performance on aeronautical services in Air Navigation activity. Therefore, since 2008, INFRAERO psychologists have been overcoming their place in this field (BRASIL, 2010f), but, still, much has to be done to reduce the distance between work prescriptions and workers real practice (GUÉRIN, 2006), improve operators work rules and procedures (DÉJOURS, 2008), monitor human errors and organizational threats (REASON, 1990 and 1997), make possible predictive and proactive safety interventions, reduce air traffic incidents.

2.4. Human Factors Indicators - Psychological Aspects

This article is based on INFRAERO psychologists´ practice of the last 3 years on Human Factors related to aeronautical services in Air Navigation activity. Since 2008, DECEA regulations (BRASIL, 2009b; BRASIL, 2008c) started to require psychologists’ participation on air traffic incidents prevention and investigation, what was accomplished by INFRAERO, which had, initially, only 2 psychologists in Air Navigation sectors for this job. In 2010, the need for more psychologists became evident in this area, a regulation was elaborated (BRASIL, 2010f) and more 8 psychologists were acquired, one for each region of Brazil.

According to ICAO, there are 3 types of safety interventions in aviation: predictive, proactive and reactive (ICAO, 2006). Reactive safety intervention after aeronautical occurrences happens to be more common. On the other hand, predictive and proactive safety interventions, before aeronautical occurrences, require efforts to intensify Human Factors approaches and studies to identify Human Factors indicators to be continuously managed.

CENIPA recent publications’ upgrade (BRASIL, 2008) emitted a combination of safety regulations related, not only, to Material Factors contribution, but also to Human Factors contribution for aeronautical accidents investigation analysis, introducing a new classification consisted of 3 aspects: psychological, physician and operational (BRASIL, 2008d). Before 2008, there were only the first 2 Human Factors’ aspects – psychological and physician, and the operational aspect was analyzed separately, as an Operational Factor, similar to Material Factors.

Each day more, modern advanced technological development requires an interdisciplinary performance among professionals of different branches, so that they may apply Human Factors concepts, interacting with their scientific, technical and specialized knowledge for a better understanding of operators´ real work (GUÉRIN, 2006) and organizational environments (REASON, 1990 and 1997). INFRAERO CADOC - Aeronautical Occurrences Data Basis is a System developed to register INFRAERO aeronautical services information in Air Navigation activity all over the country, not including Human Factors indicators related to Psychological Aspects of operators´ performance yet. This represents a blank to be fulfilled, in CADOC or other appropriate Data Basis System, considering predictive and proactive measures to be profitable before air traffic incidents and occurrences´ consolidation. INFRAERO psychologists´ latest practice may contribute to understand and define them, as follows.

2.4.1. Air traffic control incident investigation. Since 2008, DECEA introduced a regulation (BRASIL, 2008c) to oblige psychologists´ investigation of air traffic control incident, by defining standardized procedures to analyze Human Factors psychological aspects concerning individual, psychosocial and organizational variables. Every time an air traffic incident occurs at INFRAERO, psychologists ought to investigate these aspects contribution possibility in interface with operational investigators. The investigation process result consists of filling out RICEA - Air Control Incident Report, with both psychological and operational aspects, each of them considered as a parcel of Human Factors contribution to the incident analyzed, making possible integrated *feedback* and actions for necessary improvements on safety. This consists of a reactive safety intervention, in accomplishment to DECEA regulation (BRASIL, 2008c), and needs to be monitored by a Human Factors Indicators Data Basis System, in complement to CADOC, for a permanent follow-up of aeronautical occurrences causes´ statistics indicating what should be improved and helping to adopt a more predictive and proactive safety intervention in the future.

2.4.2. TRM. Since 2005, in accomplishment to ICAO publication (ICAO, 1998; ICAO, 2005), DECEA started to implement TRM Training (BRASIL, 2005a; BRASIL, 2009b and 2010e). In 2008, INFRAERO, in accomplishment to DECEA regulation (BRASIL, 2005a; BRASIL, 2009b and 2010e), introduced DECEA Facilitator TRM Training, as a support to prepare INFRAERO own facilitators (psychologists and operators) to implement internal TRM, which main purpose is: introduce team techniques to improve team behavior abilities at work related to leadership, decision making, situational awareness, communication, stress management and team interaction (BRASIL, 2005a). It was a positive experience, but, still, much has to be done to develop methodologies to identify operational restrictions that need to be improved with TRM and taken to operational practice routine, as a continuous process to assure TRM effectiveness. This requires TRM understanding as a global Program and represents a challenge to be reached by psychologists and operators´ interdisciplinary intervention focused to: TRM initial consciousness as part of organizational culture; operators’ performance follow-up after training for periodic feedback and improvements; INFRAERO basis TRM instruction as a civilian reference.

2.4.3. Operators health inspection. Pilots, flight attendants, ATC and PNA-OEA are, annually, required to upgrade their CCF - Heath Certification Ability at CEMAL - Aerospatial Medicine Center, which regulates aeronautical professionals health inspection (BRASIL, 2003). CEMAL is a military institution, subordinated to DIRSA - Aeronautical Health Directory, subordinated to COMAER, subordinated to Defense Ministry. INFRAERO PTA (ATC) and PNA-OEA have do be submitted to CEMAL health inspection but, as a civilian public administration enterprise, operators, regulated by CLT, are also obliged to make periodical health exams with an internal labor physician staff, in accomplishment to MTE - Labor Ministry regulations (BRASIL, 1978). This points out to a double procedure of operators’ health evaluation process, based on 2 different sources of regulations – Defense Ministry and Labor Ministry – indicating the possibility of divergent health restrictions diagnosis emitted by physicians from both institutions, which may take operators, temporarily, out of work. Still, if the operator stays out of work for health reasons for more than 15 days, he must be submitted to INSS - National Social Security Institute medical expertise, regulated by MPS - Social Security Ministry (BRASIL, 2004), which represents a third institution, based on a diverse regulation. INSS medical expertise often doesn’t know enough about operators´ job attributions on aeronautical services in Air Navigation activity for emitting a well based health report. Therefore, misunderstandings resulted from over-prescriptions, concerning operators´ health conditions, may happen because of an apparent inappropriate conduction of their health destination and recovery, what may end up deviating them, definitely, from work operation, without perspectives of return. In 2010, INFRAERO psychologists tried to search for a solution to this situation by promoting an interdisciplinary meeting among physician representatives of the different institutions here referred – CEMAL, INFRAERO and INSS, which pointed out to the possibility of realizing a Mixed Health Council composed by CEMAL and INSS physicians, inside CEMAL, with the main purpose of deciding together about operators health restrictions instead of sending them to INSS medical expertise evaluation, outside CEMAL. Mixed Health Council is a procedure that has been conducted for pilots and flight attendants health restrictions´ cases for years, but not for INFRAERO PTA (ATC) and PNA-OEA (BRASIL, 1967; BRASIL, 1968), because of the existence of three different institutions regulations – Defense Ministry, MTE and MPS, required to be upgraded, which represents a strong obstacle for this necessary change. This consists of a Human Factors indicator which requires, not only, well based integrated interventions on Human Factors, but, also, political determination to make that change. Besides, there is another Human Factors indicator about health subject that just started to be studied by INFRAREO psychologists, involving operators´ absence from work causes and operators´ out of work for illness causes, which needs to be monitored.

2.4.4. Audits, inspections and other safety procedures. Air Navigation activity variability requires a minimum acceptable level risk management. Therefore, ADSO - Operational Safety Audit / VSO - Operational Safety

Inspection are instruments for monitoring SGSO - Operational Safety Management System (ICAO, 2005; BRASIL, 2008a) in accomplishment to Operational Safety Program. Before SGSO, DECEA and INFRAERO implemented a Quality Program, requiring Quality Audits, in accomplishment to DECEA regulations (BRASIL, 2009b). There are some differences between DECEA and INFRAERO Safety Operational Program and Quality Program implementation. One involves the fact that DECEA developed a unique ADSO / VSO for both Operational Safety and Quality Programs, using a common instrument and application form resulting in only one report, which facilitates this practice; and INFRAERO developed a different and complementary ADSO / VSO for both Programs, representing a duplicity on application forms and reports related to safety monitoring procedures, which may bring some difficulties to both processes. Another difference concerns that DECEA ADSO / VSO, has psychologists' participation in the inspectors' group, which may represent a misunderstanding on the way of conducting this process; and INFRAERO psychologists don't participate of ADSO / VSO, because of the belief that inspection function may create a barrier for this kind of professionals to operators' real needs, which must be understood and supported by them. INFRAERO psychologists use to make periodic visits to operational sets in order to observe operators´ work performance, make interviews and look for uncomfortable situations to be improved. Operators also visit INFRAERO psychologists to talk about work situations and relationships, so that latent conditions (REASON, 1990 and 1997) may, gradually and periodically, be visualized and issued. Until 2010, there were no INFRAERO standardized procedures for this psychological attribution, but, in 2011, there were created the following Programs to be implemented: Operators Psychological Follow-up Program, using psychological instruments involving psychological tests, questionnaires and interviews; and Psychoactive Substances Safety Program, aiming at increasing consciousness about chemistry substances abuse. Before standardization, psychologists had more freedom to make observations and researches; after standardization, they got more precision information to be compared, monitored and improved. A balance between subjective observation without standardization and objective information with standardization must be reached, so that appropriate proactive and predictive safety interventions for each situation can be implemented. Studies to validate the referred Programs and instruments consist of relevant Human Factors indicators to be monitored by a needed Human Factors Data Basis System.

2.4.5. Operators acquirement and transfer process. INFRAERO public service exams for staff acquirement are implemented regionally and the participants compete for job positions according to their classification. So, each may be placed for work at any city belonging to the region he once applied the exams, not always coincident with the place he lives with his family. In case of backlists, the participant may accept the opportunity to place a job position out of the initial region he took exams, in order not to wait too long to be called, by signing a contract determining the possibility of his transfer to any operational set all over the country, according to the enterprise services needs, which may happen when, for instance, there are changes on the airport operation affecting the operational set of aeronautical services and reducing job shifts, scales and staff. In this case, is hard to preserve the appropriate operation function and find someone interested to be transferred. On the other hand, there are many cases of transfer resulting from the operator will, which happens because of his option for a job position at an operational set situated out of the region he originally made exams, often into a very small city, without a minimum social-cultural structure for living, and far away from where he was born or has family, which leads him to ask for transfer to another city where he can get a better living, closer to his family. In this case, it turns out to be a transfer process not that simple, because getting a job position at a better operational set at a developed city depends on another operator's will to go to another operational set. This process frequently last too much time, leading the operator either to overcome the situation and adapt himself to it, giving up the transfer process; or to loose motivation, exposing himself to develop illnesses, possibly reflecting on delays, absences and firings, prejudicing his stability at work, which takes too long for him to be consciousness of the heath symptoms, justifying, in time, medical help (GUÉRIN, 2006); besides, the delay of the transfer process may, gradually, contribute to deteriorate, even more, his health condition. INFRAERO has a proper regulation (BRASIL, 2005b) to prescribe the transfer process, but doesn´t have solutions for the problems described: the operator's transfer before his health consummation by not satisfying his expectation for the transfer; and the transfer process emerged from the enterprise initiative with a negative impact over the operator. In both cases, he ends up asking for psychologists' help, who also don't have solutions, trying to sensitize managers about them. This brings on the need for studies related to: public service exams rules changes involving operators allocation; transfer process regulation upgrade based on the difficulties here exposed; precise information about possible transfers and firings causes aiming at defining appropriate methodologies, parameters, criteria and encouragements to conduct proactively this process.

2.4.6. Instructors training and ability. To be an INFRAERO operator, he must realize public service exams followed by a specific operational course at ICEA - Space Air Control Institute, situated at São José dos Campos City, SP, during 4 to 8 months, depending on the function, as part of the selective process for admission. After this course, according to the participant' classification, he chooses a job operational set to work and realizes health exams by psychologists and physicians, but this step should be done before ICEA Course, aiming at reducing INFRAERO cost investment in case of participants reproof, as the Course is paid by it. The last selection step corresponds to On-the-Job-Operational-Training, taught and supervised by internal instructors of the operational set chosen by the operator to work. There is no instructor standardized profile for this training, the instructor is selected according to his local operational time and experience. Besides, he doesn't always have a specific course ability to prepare him as an instructor and doesn't perceive financially for this practice, which represents a different treatment compared to other INFRAERO organic instructors. It is also relevant to consider the lack of standardization related to contents and methodologies for all instructors to conduct this training, which may lead the new operator to possible concepts and practices' errors, hard to be corrected later. These problems require interdisciplinary studies (operators, psychologists and instructors), focused to this training teaching process, aiming at: determining a specific standardized Program in terms of content, methodology and comprehension levels, to each INFRAERO aeronautical service training; developing an Instructor Ability Course to prepare instructors for this function.

2.4.7. Air traffic control per operator. According to DECEA, operators workload is related to the maximum number of simultaneous controlled aircrafts by one operator and is evaluated according to the sum of times dispended with: communication, and transmission/reception; coordination manual activities (strips fulfillment); planning and air traffic distribution activities. The time duration of diary operational service shifts may vary from six to twelve hours of continuous work, according to: simple and accumulated operational positions workload; operational service function time; and operational set characteristics. Operators must do their specific tasks on a turn function system in every shift service aiming at: standardizing procedures and distributing workload equally; and maintaining operators in good technical conditions to perform any task of his specialty. Besides, the computation of operational staff of ATC sectors per operational position is calculated based on: shifts services positions X number of hours per shift X 30 days = staff number per month and per shift / prescribed monthly workload = operational staff per shift; where operational staff of operational sector is the sum of operational staff of all shifts (BRASIL, 2007). DECEA inspection towards this prescription not always follows the velocity of air traffic aeronautical service variability and dynamics in Air Navigation activity. SGTC - Management TWR System is used by INFRAERO TRW ATC to monitor high levels of air traffic control with the possibility to manage operation time per operator, representing a follow-up instrument in complement of DECEA regulation (BRASIL, 2007). But the use of this facility is not always prioritized in operational sets because of operational immediate demands of fast air traffic responses by operators, which consists a gap in terms of controlling the distribution of operators work on operational positions for monitoring their performance profitability. Workload per operator in operational air traffic control performance requires be better studied with psychologists' participation to be continuously monitoring.

3. Conclusion

INFRAERO aeronautical services in Air Navigation activity points out to some relevant Human Factors Indicators here analyzed: air traffic control incident investigation is a recent attribution of INFRAERO psychologists to analyze psychological aspects in complement to operational aspects contributing to their occurrence, pointing out to the need of aeronautical occurrences causes´ statistics; TRM has to be a Program consolidated as a continuous practice on the organizational culture, requiring a follow-up methodology to monitor its results after Training and also an internal instruction; operators health inspection needs CEMAL Mixed Health Council for INFRAERO PTA (ATC) and PNA-OEA health restrictions, and studies related to operators´ absence from work causes and operators´ out of work for illness causes; audits, inspections and other safety procedures consider that it isn´t appropriate psychologists participation on inspections and audits, but it´s necessary to validate INFRAERO Operators Psychological Follow Program, Psychoactive Substances Safety Program and respective instruments; operators acquirement and transfer process points out to the need of studies to reduce operators transfers and firings; instructors training and ability requires a standardized instruction Program, an instructor profile and an Instructor Ability Course; and air traffic control per operator needs to be prioritized to monitor operational performance and workload. The analysis was based on INFRAERO psychologists' experiences and emphasis was done to the need of a Human Factors Indicators Data Basis System as an instrument to continuously monitor Human Factors issues.

This analysis could comprehend other Human Factors discussions, but were selected the ones considered more relevant to evoke reflections with scientific basis about safety approaches and studies demands to understand: the purpose of ICAO (ICAO, 1998; ICAO, 2002; ICAO, 2000; ICAO, 2005; ICAO, 2006) and DECEA (BRASIL, 2009b; BRASIL, 2008c) prescriptions as a reference to be accomplished, but not as a barrier for necessary improvements on real work (GUÉRIN, 2006), considering its variability and dynamics compared to required tasks; the understand operators work rules and procedures (DÉJOURS, 2008) as an aid to subsidize operators performance, not making them static enough to hide system faults, human errors and organizational threats (REASON, 1990 and 1997), which must become evident to be treated; the importance to keep opened communication channels among operational sets and crews, and upgrade knowledge to transform operators performance in more safe, comfortable and effective. This may guide the implementation of more predictive and proactive safety interventions to INFRAERO aeronautical services in Air Navigation activity, contributing to improvements on organizational safety culture, workers´ performance and interfaces (BRASIL, 2010a; ABERGO, 2010).

4. Acknowledgments

Thanks to everyone – psychologists, operators, managers – who made possible this preliminary analysis of INFRAERO aeronautical services in Air Navigation activity, trying to understand how Human Factors indicators approaches and studies may contribute to improve operational sets and crews to a more predictive and proactive safety culture and interventions with psychologists´ participation.

References

ABERGO. Brazilian Ergonomics Association (2010). ABERGO Congress: Preliminary analysis of Human Factors psychological aspects in brazilian Air Navigation Activity. In http:// www.abergo.org.br/abergo2010.
BRASIL. DECEA (2010a). In http://www.decea.gov.br.
___. DECEA (2003). ICA – Aeronautical Command Instruction 160-1: Health inspection regulation instruction. RJ.
___. CENIPA (2008a). NSCA – Aeronautical Command System Regulation 3-3: Safety operational management. Brasília/DF.
___. CENIPA - (2008b). NSCA 3-10: SIPAER Human Resources ability and education. Brasília/DF.
___. DECEA (2008c). MCA – Aeronautical Command Manual 63-7: Psychological aspects on air traffic incidents investigation. RJ.
___. CENIPA (2008d). NSCA 3-1: Vocabulary concepts, expressions and abreviations for SIPAER use. Brasília/DF.
___. DECEA (2005a). ICA 37-288: TRM Training minimum curriculum focusing error management - ASE-001. RJ.
___. DECEA (2007). ICA 100-30: ATC personnel planning. RJ.
___. DECEA (2009a). ICA 63-15: Health inspection for ATCO and OEA aeronautical medical certification. RJ.
___. DECEA (2009b, 2010e). ICA 63-16: Aeronautical Accident Safety Program. Rio de Janeiro/RJ.
___. Aeronautical Ministry (1967). Law-decret n^{o}. 158: Health Mixed Council. Brasília/DF.
___. MSPS (1968). Service Instruction n^{o}. SSS-502,4, Section I – BS/INPS 60: Health Mixed Council. INPS, Brasília/DF, 23 de fevereiro de 1968.
___. MPS (2004). Lei n^{o}. 10.876: Medical carrier on social security. Brasília/ DF.
___. DECEA (2010b). In: http://www.redemet.aer.mil.br.
___. INFRAERO (2010c). In http://infraero.gov.br.
___. INFRAERO (2005b). NI – Internal Regulation 18.07/B (APE): Staff transfer and moving. Brasília/DF.
___. INFRAERO (2010f). MP – Procedures' Manual 16.08 (NAE – Air Navigation): Psychological Human Factors. Brasília, DF, 16.09.2010.
___. MTE (1978). NR - Regulation 7: Health Occupational Medical Control Program. Portaria GM n.º 3.214. Brasília/DF.
___. ITA - Aeronautical Technological Institute (2010d). Text: Psychologist and operational safety – The psychologist performance on safety. Master degree professional course on aviation safety and proceeded air navigation ability, Discipline AS-733/Psychology on aviation. São José dos Campos/SP.
___. Union Official Diary (2011). Page 10, ISSN 1677-7050, Section 2, N^{o}. 43, wednesday, 02.03.2011: INFRAERO, Administration Directory, Staff Board Notification. In http://www.in.gov.br/autenticidade.html/codigo 00022011030200010.
-Déjours, C. (2008). Work technology and organization: valuation of work put to test of real work, a critique of valuation basis. Ed. Edgard Blucher Ltda., São Paulo/SP.
-Guérin, F. et al. (2006). Understanding work to transform it: a practice of ergonomics. Vanzolini Foundation, Ed. Edgard Blucher Ltda., São Paulo/SP.
-ICAO (2005). Doc. 9859/AN 460: SMM - Safety Management System. Montréal, Canada. In http://www.icao.int.
___. (2002). Doc 9806/AN 763: Human Factors guidelines for safety audits manual. Montréal, Canada. In http://www.icao.int.
___. (2000). Doc 9758/AN 966: Human Factors guidelines for air traffic management systems. Montréal, Canada. In http://www.icao.int.
___. (1998). Doc 9683/AN 950: Human Factors training manual. Montréal, Canada. In http://www.icao.int.
Reason, J (1990). Human error. New York: Cambridge University Press.
___. (1997). Managing the risks on organizational accidents. Aldershot (England): Ashgate Publishing Limited.

HOW THOROUGHLY DO PROPOSED NEXTGEN MID-TERM OPERATIONAL IMPROVEMENTS ADDRESS EXISTING THREATS?

Jon Holbrook
San Jose State University
Moffett Field, CA

Antonio Puentes
San Jose State University
Moffett Field, CA

Nicole Stasio
San Jose State University
Moffett Field, CA

Kimberly Jobe
San Jose State University
Moffett Field, CA

Lori McDonnell
San Jose State University
Moffett Field, CA

Bettina Beard
NASA Ames Research Center
Moffett Field, CA

The goals of the Federal Aviation Administration's (FAA) Next Generation Air Transportation System (NextGen) include improved safety, increased capacity, increased efficiency, and reduced environmental impact. The FAA has developed 46 mid-term Operational Improvements (OIs) to facilitate initial realization of these benefits in the 2015 – 2018 timeframe. These OIs describe changes in technologies, policies and procedures from current-day air and ground operations designed to mitigate safety, capacity, efficiency, and environmental issues. The main goal of this project was to investigate how thoroughly threats to safety present in today's operations are addressed by the OIs. These threats, without mitigation, could remain threats in the mid-term, potentially compromising the intended NextGen safety benefits. To address this concern, we extracted threats to safety from 200 Aviation Safety Reporting System incident reports filed by tower air traffic controllers over a five-year period. We then evaluated whether these threats are addressed by the mid-term OIs.

The Next Generation Air Transportation System (NextGen) is a modernization effort of the National Airspace System (NAS) to increase efficiency, capacity and safety. The FAA outlined their vision in a set of Operational Improvements (OIs) that describe new capabilities and procedural changes intended to meet future traffic demands, avoid sources of delay, reduce emissions, fuel burn and noise, and improve safety. These OIs are intended to be implemented over the 2015 to 2018 time period, representing the Next-Gen "mid-term" vision for the NAS. Successful implementation of the proposed mid-term OIs is fraught with challenges, ranging from identifying top candidates for improvement to implementing those improvements in a way that successfully addresses the intended problems without introducing new and potentially worse problems (Sheridan, 2006). Early identification of the applicability of proposed improvements to existing problems supports investment decisions by providing criteria for prioritizing OIs based on the specific operational challenges they address.

The primary goal of this paper was to identify to what extent current safety concerns are addressed by planned improvements to the NAS. Despite an excellent safety record in the current-day NAS, some operational practices and actions or inactions by pilots and controllers represent threats to safety. It is unclear whether the proposed mid-term OIs address the range of safety threats present in current operations. These threats, without mitigation, could remain threats in the mid-term, potentially compromising the intended NextGen safety benefits. A set of Aviation Safety Reporting System (ASRS) incident reports submitted by tower air traffic controllers was analyzed to determine how well the NextGen OIs might be expected to mitigate specific safety issues in air traffic control operations. Although reports submitted to the ASRS are voluntary and therefore do not necessarily offer a representative or comprehensive picture of operations, these reports do provide valuable information about many of the types of operational challenges that controllers and pilots face and the circumstances that contribute to those challenges.

Method

This study examined a sample of 200 ASRS incident reports filed by tower air traffic controllers involving air carrier (part 121) operations over a five-year period (between Jan 2005 and June 2010). Tower positions included Ground, Local, Flight Data/Clearance Delivery, Handoff/Assist, and Supervisor/Controller In Charge. The set of analyzed reports was restricted to the 35 Operational Evolution Partnership (OEP) airports, which serve as hubs for airline operations in major metropolitan areas. A team of five human factors researchers (a subset of the authors) coded the reports based on two sets of criteria described below. Researchers coded different subsets of the reports for each set of coding criteria, such that each report was read by at least two researchers. In addition, 4% of the reports were coded by all five researchers and discussed as a group to ensure consistency across codings.

Coding for Relevant Operational Improvements

For each report, incidents were defined based on operational roles/positions implicated as contributing to the events, thus each report could include several incidents (e.g., a *pilot* landed on an unassigned runway, and the *local controller* was distracted and unaware until alerted by another aircraft in position for departure on that runway [ACN 879171]). For each incident, mid-term OIs were identified that, if implemented appropriately, could have prevented or mitigated the consequences of the incident. It was possible to assign zero, one, or multiple OIs to a given incident. Coding was based on comparisons of incident narratives against OI descriptions provided in the FAA NAS Enterprise Architecture (2009). Coding also included a justification statement for why the OI was relevant or, for the incidents for which no OI was identified, a statement describing the specific unaddressed operational challenge.

Coding for Causal Categories

The Department of Defense (DoD) Human Factors Analysis and Classification System (HFACS) was used to classify the reports according to causal factors (O'Connor, 2008; Wiegmann & Shappell, 1997). Determinations of HFACS code applicability were based on comparisons of incident narratives against HFACS code descriptions (DoD HFACS, 2005). As with the OI coding, it was possible to assign zero, one, or multiple HFACS codes to an incident.

Results

Two hundred twenty five incidents were identified in the 200 ASRS reports. Caution should be used in interpreting the absolute or relative numbers of incidents reported, because the frequency of unreported events is not known. However, the reported incidents sample from a wide range of tower operations and roles (Table 1).

Table 1
Number of Incidents by Position and Phase of Flight

Incident Position	# of Incidents	Phase of Flight	# of Incidents
Ground	17	Descent	24
Local	163	Final approach	42
Supervisor	11	Landing	41
TMC	2	Taxi in	35
TRACON	6	Taxi out	23
Pilot	26	Takeoff	80
		Departure	28
		Go around/Missed approach	15

Note. Multiple phases of flight may be associated with one incident.

Thirty-four of the 35 OEP airports were represented in the incident database. No incidents that met the search criteria were available for DCA via the ASRS online search tool. The median number of analyzed incidents per airport was 5, with a range from 1 to 23.

Two of the 46 mid-term operational improvements were excluded from analysis a priori. These OIs (109303 and 109304) describe enhancements to the Aviation Safety Information Analysis and Sharing (ASIAS) system. ASIAS is an existing mechanism for integrating, analyzing and sharing aviation safety data and information, including ASRS reports. Because every analyzed incident was reported via ASRS, the enhanced ASIAS system described in the OIs would be relevant to every incident and therefore not diagnostic. Of the remaining 44 IOs, 27 were used in incident coding at least once, for a total of 304 instances of relevant OI identification across all incidents. The remaining 17 OIs were not identified as relevant for any incidents. We also identified 68 instances in which no OI was relevant to the incident. A depiction of the relative frequencies of OIs used in the incident coding can be seen in Figure 1.

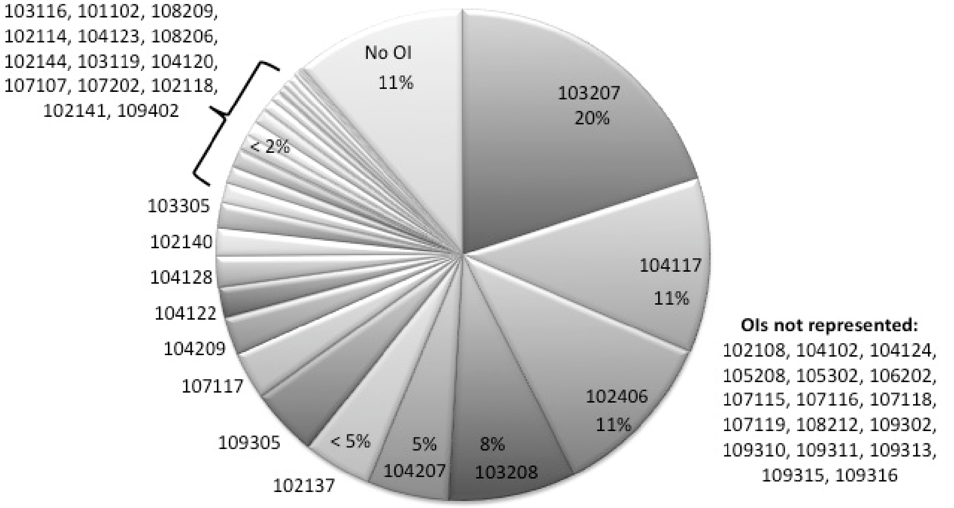


Figure 1. The relative frequencies of OIs identified (and not identified) in incident coding.

Twenty-two of 23 HFACS codes were identified as relevant across the ASRS incidents, for a total of 469 HFACS codings. All incidents were associated with at least one HFACS code, with a range from 1 to 6 HFACS codes per incident (mean = 2.1).

For each incident, we determined which of the previously identified relevant OIs were associated with each identified HFACS code. A given OI might be associated with one or more HFACS codes for that incident. From this analysis, a correlation matrix was constructed based on the OI and HFACS coding to identify codes that tended to co-occur. Statistically significant correlations among pairs of OIs and HFACS codes are shown in Table 2. Due to the increased likelihood of Type I errors in a large correlation matrix, the criterion for statistical significance was set at p < .001.

Table 2
Correlations among OIs and HFACS Codes

OIs/HFACS Codes	χ^2 (1, N=225)	ϕ Coefficient
OI 103207 : OI 103208	24.90	.333
OI 103207 : OI 102406	16.83	.274
OI 103208 : OI 102406	28.09	.353
OI 109305 : Technological Envir.	14.60	.255
OI 109305 : Violation Based on Risk Assessment	48.78	.467
Technical Envir. : Violation Based on Risk Assessment	13.69	.247
OI 103208 : 104207	11.93	.230

Note. All reported correlations were significant at p < .001.

Analysis of the relevant OI to HFACS category associations revealed that, overall, 80% of the instances of HFACS coding were associated with relevant OIs. Figure 2 shows the proportion of incidents associated with relevant OIs partitioned by HFACS categories. Further examination revealed a fairly consistent and broad pattern of OI coverage across HFACS categories, with extreme values (i.e., more than 33% of incidents with no relevant OI) only for categories comprising a single observation (i.e., Supervisory Violations and Perceptual Factors). This finding indicates that, overall, the mid-term OIs could help to mitigate or prevent the breadth of incidents explored in this study, however some incidents were identified for almost all HFACS categories that were not addressed by the OIs.

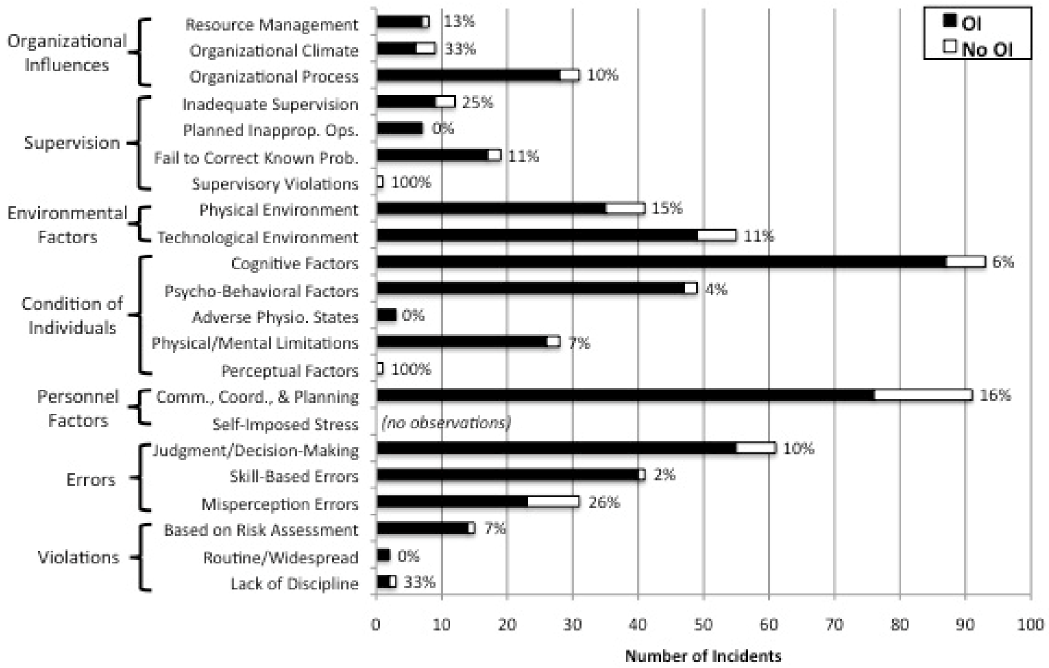


Figure 2. The number of incidents associated with each HFACS category, along with the percentage of incidents that did and did not have associated OIs. Percentages shown in the figure are for incidents without associated OIs.

Discussion

In this report, we considered a set of recent operational incidents from major US commercial airports and identified NextGen mid-term Operational Improvements that could mitigate or prevent those incidents (see Figure 1). Two noteworthy aspects of this analysis were the relatively small number of OIs that comprised the majority of the coded instances (e.g., four OIs account for 50% of codings), and the relatively large number of OIs (17) not represented in the coding at all. The four most represented OIs were 103207 (Improved Runway Safety Situational Awareness for Controllers), 104117 (Improved Management of Arrival/Surface/Departure Flow Operations), 102406 (Provide Full Surface Situation Information), and 103208 (Improved Runway Safety Situational Awareness for Pilots). These OIs are heavily focused on runway operations, providing enhanced alerting for runway incursions and identification of aircraft position on the surface, as well as sequencing and scheduling based on wake and aircraft performance characteristics. The frequent occurrence of these OIs is not surprising given the heavy bias toward incidents involving the local controller position (see Table 1). Similarly, all of these OIs support various levels of situation awareness (SA), from perception (e.g., digital display of the airport environment in 102406) to comprehension (e.g., runway status alerting capabilities in 103207 and 103208) to projection (e.g., support for planning arrival, departure and surface operations in 104117; see Endsley, 1995, for a thorough discussion of levels of SA). Loss of SA is implicated under Cognitive Factors, the most frequently assigned HFACS code in this data set, so it is again reasonable to expect that SA-related OIs would be prominently featured.

Seventeen OIs were not represented at all in the current study. Further analysis indicated that these OIs tend to fall into at least one of three categories: 1) OIs unrelated to tower operations (e.g., 102108, Oceanic In-trail Climb and Descent); 2) OIs unrelated to safety (e.g., 109316, Increased Use of Alternative Aviation Fuels); or 3) OIs targeted to flight deck enhancements (e.g., 107115, Low Visibility/Ceiling Takeoff Operations; recall that the incident reports used in this analysis were filed by tower controllers, not pilots). Future ASRS analyses involving pilot-reported incidents and incidents from other phases of flight (e.g., en route) could address many of the OIs not represented in the current study. However the OIs unrelated to safety would likely remain a blind spot in any analysis of ASRS reports.

The current analyses indicated that, overall, the mid-term OIs address approximately 80% of the HFACS codings. Omitting the two HFACS codes for which there was only a single incident, this value climbs to 88%. However, three HFACS codes showed a substantially lower level of OI relevance (see Figure 2): Organizational Climate, Inadequate Supervision, and Misperception Errors. In the cases of Organizational Climate and Inadequate

Supervision, these types of incidents are somewhat beyond the scope of the OIs, which focus heavily on technological advances and automation to address problems in the NAS. However, these problems are very likely to persist in the NAS, and should be addressed through other FAA initiatives and training procedures. Misperception errors involve limitations in human sensation and perception resulting in misjudging of distances and rates of change, as well as susceptibility to sensory illusions. This class of errors is largely insensitive to mitigation by training or experience, but could be mitigated with appropriate application of automation. Addressing these types of errors is within the scope of NextGen. Future research should focus on identifying specific human research-based requirements for NextGen automation to effectively mitigate misperception errors.

Several statistically significant correlations among OIs and between OIs and HFACS codes were identified (see Table 2). Further examination of these relationships may be useful in understanding the degree and nature of overlap among the OIs, which may have implications for how the OIs should be implemented. The following OIs, 102406 (Provide Full Surface Situation Information), 103207 (Improved Runway Safety Situational Awareness for Controllers) and 103208 (Improve Runway Safety Situational Awareness for Pilots) share the common theme of enhancing surface information and improving SA for controllers and pilots. Implementation of these OIs has the potential of preventing runway incursions or other loss-of-SA incidents that have the potential for catastrophic consequences. Surface display maps should give controllers better information about the exact location of all surface traffic and provide alerts for unauthorized movement, potentially preventing incidents like the following:

> When Air Carrier X came to a stop, he was in the Intersection of Runway 10/28 and then turned onto Runway 28 towards the Air Carrier Y jet who had just read back the take off clearance. I forcefully said to Air Carrier X to go straight down the Runway (33L), don't turn on an active Runway 28 at least 3 times (ACN 701804).

Further inspection of these OIs and relevant incidents provides insights about how the OIs might work together to enhance SA for controllers and flight crew. OI 102406 contains the framework of a surveillance system by positioning sensors around the airfield and providing the infrastructure to communicate vehicle data to controller information systems. OIs 103207 and 103208 rely upon the surveillance system to populate specific controller and aircraft displays to further enhance SA. Both of these OIs integrate surface situation information, visually provide the users with a picture of ground traffic position and provide alerts when aircraft are at risk of runway incursions.

OI 109305 (Improved Safety for NextGen Evolution) was positively correlated with the HFACS codes Technological Environment and Violation Based on Risk Assessment. This OI is intended to mitigate safety risks associated with increased automation in NextGen by providing enhanced methods to optimize human-automation interaction, monitoring system safety performance to accelerate the detection of unrecognized safety risks, and providing advanced training concepts which will maintain levels of proficiency for humans to conduct safe operations in place of degraded or failed automation. One major challenge for automation implementation involves procedures for addressing false alerts. In the current operational environment, controllers are instructed to comply with automated alerts regardless of what they visibly perceive. However, several reported instances of false alerts placed controllers in the position of choosing between following procedures (i.e., comply with the automated alert) or executing the "safer" course of action, based on their expertise (e.g., electing not to issue a go-around after an obviously false AMASS alert [ACN 876688]). This illustrates an instance of a Violation Based on Risk Assessment in response to a Technological Environment factor. As automation increases with NextGen implementation, OI 109305 may become increasingly relevant. However, this OI represents a post-implementation means of addressing human-automation interaction issues that might be better addressed through early application of human-automation interaction principles and research in the development of NextGen automation. The methods and concepts that serve as the critical elements of this OI should be applied to automation design across NextGen to decrease the likelihood or severity of human-automation interaction problems in the first place.

OI 103208 (Improve Runway Safety Situational Awareness for Pilots) is intended to improve runway safety operations by providing pilots with improved awareness of their location on the airport surface and by providing runway incursion alerting capabilities. OI 104207 (Enhanced Surface Traffic Operations) proposes that data communication will be the principle means of communication between aircraft and controllers for clearances, amendments, and requests during the mid-term. Anticipated benefits arising from these capabilities include improved efficiency, reduced frequency congestion, and enhanced safety due to avoided readback/hearback errors. Results from the present study indicate a positive correlation between these two OIs, which is not surprising given that a miscommunication or misperception on either the part of the pilot or controller might result in disorientation on the airport surface, leading to a runway incursion, rejected takeoff, cancelled clearance, or worse.

Both of these concepts address how technology might be used to improve safety with respect to airport surface movements. The following narrative illustrates how these two OIs might work together to mitigate these types of errors:

> LCL W CTLR was issuing taxiing instructions [to a B757] to turn *away from* RWY 9L and go S, not N on TXWY N. It appeared that the ACFT was not following instructions. I immediately cancelled TKOF CLRNC to [another ACFT on RWY 9L] and advised him to exit the RWY. The AMASS alerted… After review of the voice recordings, the B757 PLT crossed the RWY without ATC AUTH (ACN 750426).

A recent examination of runway incursions found that pilot deviations accounted for 57% of the total number of runway incursions in the United States (Rankin, 2008), while another reported that pilot deviations accounted for 60% of incursions and that ATC operational errors accounted for another 20% during a 4 year period (Young & Jones, 2001), suggesting that roughly 80% of these errors might have been avoided if the technology suggested in these two OIs were available. Data communication would virtually eliminate readback/hearback errors, and if taxi route information were integrated graphically into flight deck moving map displays, this would enable pilots to visualize the taxi instruction and routing constraints, helping to improve pilots' SA. Integrated displays coupled with the ground and flight deck runway incursion alerting capabilities described in these concepts should result in more robust taxi route coordination and location detection by pilots and controllers, reducing the likelihood of runway incursions.

In 68 instances (11% of all cases), incidents were identified for which no relevant OI was assigned. Commonalities among these instances have not been systematically explored, but this task is planned for future investigation. Although OI relevance across the range of incidents addressed in this study was found to be quite high, virtually all HFACS categories included some incidents for which no relevant OI was identified. Systematic exploration within each HFACS category of differences between incidents with and without identified relevant OIs is also planned. These analyses could provide insights into additional possible NextGen improvements.

Acknowledgements

This work was supported by DOT/FAA – NASA interagency agreement DTFAWA-09-X-80020 and by the NASA Airspace Systems Program SAIE Project.

References

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors 37* (1), 32–64.

Federal Aviation Administration (2009). National Airspace System NAS Enterprise Architecture (NASEA) 3.0, January, Washington, D.C.

O'Connor, P. (2008). HFACS with an additional layer of granularity: Validity and utility in accident analysis. *Aviation, Space, and Environmental Medicine, 79* (6), 599-606.

Rankin, W. B. (2008). Runway incursions: An industry examination of FAA initiatives and objectives. *International Journal of Applied Aviation Studies, 8* (2), 225-240.

Sheridan, T. B. (2006). A review of human-automation interaction failures and lessons learned. Cambridge, MA: Volpe National Transportation Systems Center.

U.S. Department of Defense (2005). Department of Defense Human Factors Analysis and Classification System: A mishap investigation and data analysis tool. Retrieved from http://www.uscg.mil/safety/docs/ergo_hfacs/hfacs.pdf.

Wiegmann, D. A., & Shappell, S. A. (2001). Human error analysis of commercial aviation accidents: application of the Human Factors Analysis and Classification System (HFACS). *Aviation, Space, and Environmental Medicine, 72,* 1006–1016.

Young, S. D., & Jones, D. R. (2001). Runway incursion prevention: A technology solution. *Proceedings of the Flight Safety Foundation 54th Annual International Air Safety Seminar, 54*, 1-22. Athens, Greece: Flight Safety Foundation.

HUMAN-CENTERED DESIGN TO SUPPORT FLEXIBILITY AND ADAPTABILITY IN AIRPORT SURFACE MANAGEMENT

Alicia Borgman Fernandes
Philip J. Smith
The Ohio State University
Columbus, OH

Airport surface delays can impact airport surface congestion, operational costs, environmental emissions, and passenger satisfaction. We report on structured interviews and observational studies at two US airports with different approaches to managing surface delays. Each approach requires human judgment to set and adapt control parameters to manage departure flows. We contrast these approaches in terms of: 1) distribution of roles and responsibilities; 2) human judgments required; 3) enforcement of flight operator compliance; and 4) tools for coordination and decision support. Guidance is provided for designing and implementing human-centered surface management programs based on an analysis of these approaches.

This paper discusses programs at two major US airports for metering departures on the airport surface. Both reduce surface congestion by constraining the rate at which flight crews request access to the active movement surface and are only two examples of processes developed locally to manage departures on the surface. The discussion is based on structured interviews and observations at the airports during summer 2010. This is part of a larger project aimed at integrating airport surface and airspace constraints in departure management. Note that the focus of this discussion is on approaches to managing departures on the surface and not arrivals, although interactions between arrivals and departures also can contribute to surface congestion.

At the John F. Kennedy International Airport (JFK), the Airport Operator and Ramp Control Tower (RCT) Operators make up a Surface Group that started the metering program in spring 2010. The Surface Group adapted a previously developed winter operations slot allocation program for daily use. The metering program assigns aircraft to 15-minute time windows according to the expected departure rate during the afternoon departure push. Flight crews are not to contact the Air Traffic Control Tower (ATCT) for permission to taxi to the departure runway more than 5 minutes before their time window. Time window allocation and coordination of RCT intent to use their time windows is facilitated by software tools.

The ATCT at the Newark Liberty International Airport (EWR) has used a metering program for several years. ATCT controllers assign a specific metering time to each departure. The flight crew is not to contact the Ground Controller before that time. Times are provided to flight crews via radio, which the RCTs monitor. We visited the ATCT, Airport Operations Center, and one RCT at JFK, and the ATCT and one RCT at EWR.

Both metering programs were developed by local airport stakeholders and rely on a small number of people to meet the challenges of surface and airspace constraints. We observed and interviewed a number of these people to identify some of the judgments they make about capacity, demand, and traffic flow. These judgments help manage traffic on the airport surface in response to changing conditions in the airspace and on the ground. Each program has been locally successful and provides insights into features that facilitate success and contribute to challenges.

The paper is organized as follows: We discuss how human judgments are influenced by system designs in terms of the distribution of roles and responsibilities in the system. We conclude by outlining design considerations for providing flexibility and adaptability in surface departure management programs.

Impacts of System Design on Human Judgments

The way in which roles and responsibilities are distributed in a system has strong implications for the judgments and decisions people in the system are called on to make. This and other aspects of system design limits the strategies available to people for identifying constraints in the world and adapting to meet those constraints. In this section we discuss the distribution of roles and responsibilities in the metering programs at JFK and EWR, judgments and decisions made by people in different roles, and some of the tools available to support them.

Distribution of Roles and Responsibilities

Both metering programs have a central authority that monitors the status of traffic on the airport surface and determines how the metering program should be adapted accordingly. We have identified six key responsibilities in managing both metering programs, but the distribution of the responsibilities is different in the two programs. Table 1 provides an overview of the assignment of these responsibilities at the two airports. Note that JFK has created new roles in developing their metering program and has distributed the responsibilities more widely than EWR.

Table 1, *Distribution of Judgment Responsibilities in the Metering Programs at JFK and EWR.*

Judgment Responsibility	JFK	EWR
Determine whether to meter	Surface Group and RCTs	ATCT
Estimate departure capacity	Metering Desk	ATCT
Allocate capacity to flight operators	Metering Desk	ATCT
Allocate capacity to individual flights	RCTs	ATCT and RCTs
Enforce compliance	Surface Group and RCTs	ATCT
Adjust program to changing conditions	Metering Desk and RCTs	ATCT

Determine whether to meter. The Surface Group at JFK decided to meter departures every day. The Surface Group meets weekly to discuss issues and refine the program. Metering is likely to remain in daily use as long as Surface Group members agree to participate. In contrast, at EWR the ATCT decides each day whether to meter departures. According to an RCT Manager the decision to meter "depends on the ATIS [weather conditions] and the runway configuration". It also depends on traffic flow. One ATCT Manager said, "If it'll take 45 minutes for them to get out of here but they're moving, I won't put times on them. Now if they're taking 45 minutes and they move two feet, I'll put times on them." Note that the ATCT does not have specific software tools to support this process.

Estimate departure capacity. Metering depends on an estimate of the airport departure capacity over some planning horizon. The nature of this estimate differs between the two metering programs. At JFK, departure capacity is defined in terms of the number of aircraft allowed to request entry to the active movement surface during each 15-minute time window. A Metering Desk tactically manages the program and estimates the departure rate. They have tools that aid this estimate such as a historical database suggesting an average rate based on weather conditions and airport configuration. However, they use judgment to adjust the rate to current conditions.

At EWR departure capacity is defined based on the length of time flight crews are told to wait before contacting Ground Control to request entry to the active movement surface. According to RCT personnel at EWR, the length of assigned delays reflects the amount of time that the ATCT expects is required to reduce the departure queue to a manageable length. They said that typically this is 45-60 minutes when metering is in effect. It is not clear that an estimate of departure capacity is ever explicitly made in the context of metering, although an implicit estimate of departure capacity seems necessary to provide useful delay times.

Allocate capacity to flight operators. Each metering program has a different approach to allocating estimated departure capacity to flight operators. At EWR, the ATCT determines the delay to assign to flights. Flight crews call the ATCT to state that they are ready to push back and taxi to the runway for departure. The ATCT tells each flight crew what time to contact Ground Control. The order of metering times is approximately First-Come, First-Served (FCFS) according to the order in which flight crews call. To support this task the ATCT has a control position, Flow Control, which takes the initial radio contact from flight crews and provides metering times to them.

At JFK, the Metering Desk determines how many departures to allocate to each flight operator for each time window. A "Slot Calculator" tool supports this task. They assign individual flights to time windows according to the "Slot Calculator" recommendation and a Ration By Schedule philosophy (Wambsganss, 2001; Smith, Geddes & Beatty, 2008). The software automatically communicates these assignments to the RCT responsible for each flight.

Allocate capacity to individual flights. RCTs at both airports have some control over the order in which flights taxi out for departure, as well as locations of aircraft waiting for their assigned time windows. This is explicitly part of the metering program design at JFK, but is less explicit at EWR. At JFK, each RCT determines which flight should use each time window. RCTs may request to swap pairs of flights assigned to different time windows in order to better align the time windows to their business needs. Similarly, if a departure is not expected to be ready in time to use its time window, the RCT requests a later time for that flight. The software supporting the metering program at JFK has a group chat window that is the chief tool the RCTs use to request changes in time window assignments (e.g., “pls swap 4017 with 246”). The Metering Desk is responsible for reviewing the requests. If they determine that the swap can be made they make the change. If the swap cannot be made, the Metering Desk types that they cannot make the swap, sometimes adding a reason (e.g., the two flights are assigned to different runways). The software also supports other ways to request changes in time window assignments. For example, the RCT can drag a flight from one time window to another to request a change. This action automatically changes the color and location of the call sign for that flight on the Metering Desk display. The Metering Desk then determines whether the change can be made and changes the color of the call sign accordingly.

At EWR, the influence of RCTs over the allocation of departure capacity to individual flights is more implicit. That is, RCTs can control the order in which their flight crews contact Flow Control to receive a metering time, thus influencing the order of their flights’ metering times under the approximate FCFS system. However, swapping two flights’ metering times is “not really done in practice,” according to an RCT Manager, although exceptions do occur. For example, “if a departure had to come back to the gate and we really wanted to get him out we would call the Tower to ask if we can arrange a swap.” Note that this process of arranging for a flight to be treated as a priority is similar to that at most airports, where such arrangements occur on a case-by-case basis.

Enforce compliance. The metering programs are enforced differently. When the Flow Controller at EWR assigns a metering time to a flight, he or she writes that time on the flight progress strip before handing off the strip to the Ground Controller. When the flight crew contacts the Ground Controller to request entry to the active movement surface, the Ground Controller checks the metering time written on the flight strip. “The FAA [Ground Controller] doesn’t want to hear from you before” the time written on the strip, said an RCT Manager.

Because participation at JFK is voluntary, RCTs are responsible for ensuring that flight crews do not contact the ATCT early to request entry to the active movement surface. However, sometimes a flight crew requests to taxi early and the ATCT may grant this request. When this happens it is typical for other RCTs to note the early departure in the chat window. For example, during our observation a flight departed roughly thirty minutes before its time window. Immediately, an RCT wrote a chat message noting the occurrence. The offending RCT Operator then receives a “slap on the wrist” at the next Surface Group meeting in an effort to eliminate such behavior. However, because participation is voluntary, if one or more RCTs perceive that participation puts them at a competitive disadvantage the process can break down.

Adjust program to changing conditions. It might be necessary to adjust metering program parameters such as due to changes in weather conditions or aircraft maintenance problems that impact surface traffic. This task requires expertise in judging the departure capacity as well as how changes in the departure capacity are likely to impact traffic on the active movement surface. It also requires judging how actions taken to modify the time window allocations are likely to impact surface traffic. The ATCT at EWR monitors departure traffic and modifies metering times if necessary. An RCT Manager said, “Sometimes the FAA [ATCT] will get behind and they’ll add 10 minutes to the taxi times. … They do adjust taxi times up if the queue gets too short.” The ATCT does not have tools to explicitly support them in this task and therefore must rely on ATCT Managers’ and Controllers’ expertise.

The Metering Desk adjusts the program at JFK when necessary. When they identify conditions that are likely to cause a change in the departure rate they determine how to adjust the time window allocation. Although they have the departure queue as a reservoir to absorb incorrect judgments, they do not receive immediate feedback on their actions because of the time required for the length of the departure queue to adjust to changes. For example, during our observation an emergency landing closed the departure runway for nine minutes. As soon as the runway closed (and before they knew how long it would remain closed), the Metering Desk needed to modify the time windows such that the runway queue would not grow too big if the runway remained closed for an extended period. They also needed to ensure that the queue would not run dry if the runway did not remain closed very long. They discussed the situation as they monitored traffic on the surface display and decided to move all departures back 15 minutes, starting with the time window 45 minutes after the runway was closed. One of the individuals involved in the decision reported the following reasoning for selecting that time window: “I know they’re probably not ready, and it’s far enough into the future that if they are ready and we do have slots available we can move them up.” The surface display is a key tool allowing the Metering Desk to view the status of traffic on the airport surface.

RCTs at JFK decide how to use time windows to best meet their organizational goals. They may use any of their time windows for nearly any of their departures. The Metering Desk must approve any changes to the original allocation. The RCTs have a software tool showing the current allocation, and most have towers allowing them to see at least their ramp areas. They also have strategies for identifying high priority flights and monitoring the metering program. For example, the RCT observed at JFK has its own metering desk for managing that RCT's time windows and communicating with the Airport Operator Metering Desk. One employee staffing this desk said they monitor "times, changes in times, and which flights have [time windows]. … good [time windows] are within ten or 15 minutes" of the scheduled departure time, and, "I ask what's going on when it's more than thirty minutes."

RCTs at both airports choose what time each flight should push back relative to its time window based on gate availability, the metering time, and current delay metrics for the flight according to the Department of Transportation (DOT). "If the [time window] is more than about 70 minutes" after the flight's scheduled departure time, individuals at one RCT at JFK delay boarding if possible. However, if another flight needs the gate, the RCT boards the aircraft and holds it in the ramp area until its time window arrives. If the aircraft is to remain at the gate the RCT must also decide at what time passengers should board the aircraft and at what time the aircraft should push back in order to make the gate available to other aircraft.

Design Considerations for Surface Departure Metering Programs

Design guidelines for surface departure metering programs can be derived from many sources. There are several general distributed work system design guidelines (e.g., Bowers, Salas, & Jentsch, 2006; Hinds & Kiesler, 2002; Smith, McCoy, & Orasanu, 2001; Smith, Spencer, & Billings, 2007). Successes and challenges in existing metering programs provide additional considerations. Our study of the programs discussed here leads us to posit several such design considerations, including: program status visibility, enforcement, flexibility, physical constraints, perception of equity, reservoirs for absorbing variability, information management, and technology support.

Program status visibility. The Metering Desk at JFK has a surface display showing the locations of all aircraft on the airport surface. In addition, they listen to the Ground Control radio frequency. Similarly, the location of the ATCT at EWR provides a view of all aircraft on the active movement surface. This visibility of the airport surface enables those responsible for monitoring and adapting the metering programs to view the number of aircraft currently in the departure queue. They also can detect patterns in surface traffic and take steps to avoid congestion. Without such visibility, the Metering Desk would be "blind," according to one member of the JFK Surface Group. Airport surface displays seem to be very useful tools in guiding departure management decision-making. These displays also can include user-customized alerts that can help with attention management issues associated with monitoring surface traffic and other competing tasks (Spencer, et al., 2005).

Enforcement. Our observations, interviews, and additional discussions with participants in these metering programs lead us to believe that enforcement by the ATCT as demonstrated at EWR is important to ensure success. The ATCT is the only true authority in allowing or denying access to the active movement surface.

Flexibility. A key goal of Collaborative Air Traffic Management is to "accommodate flight operator preferences to the maximum extent possible" (FAA, 2010). Typically this is implemented as procedures for flight operators to swap departure time windows as at JFK. However, RCT Managers at EWR reported that even without such processes the metering program there usually provides them sufficient flexibility to ensure that their highest priority flights are accommodated. When necessary, they contact the ATCT by phone to make requests to expedite specific flights. This works adequately for cases where a single high priority flight is involved, but may be problematic if an RCT Manager would like to expedite several high priority flights. A useful feature for departure management programs seems to be mechanisms for RCTs to express a set of high priority flights to the ATCT, and for the ATCT to expedite those departures without giving that flight operator a competitive advantage.

The Federal Aviation Administration (FAA) reports a goal of flight operator flexibility (FAA, 2010). ATCT flexibility also can be important. Tools to support flight operators in communicating their priorities can make it easier for ATC to accommodate them. An ATCT Manager at JFK said that under the metering program, "There aren't as many aircraft out there so they [ATCT controllers] have more flexibility to move them around." Reducing the number of flights on the active movement surface (and hence in the departure queue) improves the ability of ATCT controllers to manage the departure sequence. This can provide them with more flexibility to deal with dynamic departure fix constraints. In addition, the reduced number of taxiing flights increases ATCT flexibility by decreasing the time required to change the runway configuration at JFK. During two days of observations the change from one to two departure runways was accomplished in 7 and 9 minutes respectively. Metering Desk, ATCT, and Airport Operator personnel said that 30 to 40 minutes was typical before the metering program.

Physical constraints. If an airport faces congestion on the active movement surface, a metering program can manage the flow of departures to the active movement surface and help avoid gridlock. At both airports discussed here, personnel credit the metering program with decreasing surface congestion. However, it must be noted that those flights whose entry to the active movement surface is delayed by metering must be held somewhere. The physical geography of the airport may cause congestion to be moved from one place to another (such as from the departure queue to an ad hoc holding area as at JFK). Even if physical space is limited, ad hoc holding areas can have advantages. For example, aircraft in holding areas are, for the most part, out of the way of other aircraft that are actively taxiing to the runway for departure. In addition, aircraft in the holding area can have one or both engines off, reducing fuel burn. This reduces emissions and potentially can decrease flight operator operating costs. Mechanisms for creating and using holding areas, whether ad hoc or permanent, can help to ease surface congestion.

Perception of equity. Most ATCTs manage departures according to an approximate FCFS process because it is perceived to provide equitable access to all flight operators. The ATCT sequences flights for departure according to the order in which flight crews call to request access to the active movement surface, with minor exceptions to increase runway throughput. ATCT personnel at both airports stated that ensuring that the flight operators perceive equitable treatment is important. A modified version of Ration By Schedule (Wambsganss, 2001; Smith, Geddes & Beatty, 2008) has largely been accepted by the aviation community as a useful surrogate to FCFS for other air traffic management programs (such as ground delay and airspace flow programs). This approach offers a potential alternative to allow better airport surface management while still providing equitable treatment to flight operators.

Reservoirs for absorbing variability. At both airports, the departure queue acts as a reservoir to absorb variability in the departure rate and uncertainty in departure capacity estimates. That is, if the Metering Desk at JFK or the ATCT at EWR overestimates the departure capacity and allows more aircraft onto the active movement surface than necessary, the departure queue will increase in length. Those responsible for managing the program then can take action to decrease the length of the departure queue. In addition, RCTs at both airports provide an additional reservoir with departures that are ready and waiting for their time windows to arrive. If the Metering Desk at JFK or the ATCT at EWR underestimates the departure capacity and allows fewer departures onto the active movement surface than necessary, there are departures that can quickly be allowed to enter the active movement surface before their time windows to quickly feed flights to the departure queue. Such reservoirs seem to be an important feature of departure management programs.

Information management. Information requirements should be identified in the design process and supported in technologies and processes built into the system. One major benefit of the metering program at JFK is that RCTs have better information about the time at which each flight is likely to actually take off. The metering program allows RCTs to make a more informed decision as to whether flights should be delayed at the gate or whether the flight should be delayed with the passengers on board and the aircraft pushed back from the gate. In addition, the ability to swap flights with different time windows allows RCTs to prioritize departures so that higher priority flights can depart sooner than they would in a typical FCFS system. Metering at EWR improves ATCT information management. According to an RCT Manager, metering can “help the Ground Controller… [so] they know what to expect… [by] managing who’s coming out.” The ATCT sees an advantage in using the Flow Control position, whether or not they are metering: “We may not issue [metering] times, but we’ll have them [flight crews] monitor the frequency. It helps manage the Ground frequency because the aircraft aren’t all calling at once.” Program designs should include processes for communicating information to those that need it, when they need it.

Technology support. Understanding the tasks people perform in existing collaborative departure management programs helps in developing technology to support them. For example, surface displays showing the locations of aircraft on the surface seem to be a key technology, particularly if the departure management program is managed by personnel who are not located in the ATCT. In addition, people working in such systems today perform some low level tasks that can be automated, such as assigning individual flights to time windows at JFK (e.g., see Brinton, Lent, & Provan, 2010). They perform these tasks to achieve higher level goals, such as increasing or decreasing the length of the departure queue. It would be possible for a person to set a target queue size and have software develop a plan to achieve that goal. Software might also detect that conditions have shifted and support people in determining the best way to adapt the plan. Technology can help ensure that parties are kept appropriately informed of program status and the activities of others, while also reducing the effort involved in managing priorities and adapting to changing conditions. User-customized alerts can help people manage their attention and workload when they can be faced with several competing tasks (Spencer, et al., 2005).

Conclusion

Departure metering programs can help people to better manage departure demand when it exceeds airport capacity. These programs require humans with sufficient expertise to make predictions about airport departure capacity on a given day and to determine how that capacity should be allocated to flight operators and individual departures. They also require humans to monitor the programs and determine how best to adapt them to changing conditions. The design of such systems strongly influences the judgments people are called on to make. For example, the distribution of roles and responsibilities in the system determines who is required to make each judgment as well as strategies available to them to make and carry out decisions. They need to have tools that support them in these judgments. To complement general design guidelines for distributed work systems, we provide design considerations specific to departure metering programs. The overall theme of these guidelines is that the assignment of roles and responsibilities to different people and decision support technologies need to provide the people involved with *appropriate* individual and shared situation awareness. People also need the ability to communicate their plans to each other and to the decision support tools that they are using to help manage airport surface traffic.

Acknowledgments

The FAA Human Factors Research & Engineering Group coordinated the research requirement and its principal representative acquired, funded, and technically managed execution of the research service. The authors also wish to thank the many personnel who granted interviews for the study and provided feedback on early drafts.

References

Bowers, C., Salas, E., & Jentsch, F. (2006). *Creating High-Tech teams: Practical Guidance on Work Performance and Technology.* Washington, DC: American Psychological Association.

Brinton, C., Lent, S., & Provan, C. (2010). Field test results of Collaborative Departure Queue Management. *Proceedings of the 29th Digital Avionics Systems Conference (DASC).* Salt Lake City, UT. October 2010.

Federal Aviation Administration (2010). *NextGen Solution Set Collaborative Air Traffic Management (CATM).* May 5, 2010. Retrieved from http://www.faa.gov/about/initiatives/nextgen/portfolio/sol_sets/catm/index.cfm.

Hinds, P., & Kiesler, S. (2002). *Distributed Work*, Cambridge, MA: MIT Press.

Smith, P. J., Geddes, N. & Beatty, R. (2008). Human-centered design of decision support systems. In A. Sears & J. Jacko (Eds.), *Handbook of Human-Computer Interaction, 2nd Edition* (pp. 573-602). Mahwah, NJ: Lawrence Erlbaum Associates.

Smith, P. J., McCoy, E., & Orasanu, J. (2001). Distributed cooperative problem-solving in the air traffic management system. In G. Klein & E. Salas (Eds.), *Linking expertise and naturalistic decision making* (pp. 369-384). Mahwah, NJ: Erlbaum.

Smith, P. J., Spencer, A., and Billings, C. (2007). Strategies for designing distributed systems: case studies in the design of an air traffic management system. *Cognition, Technology and Work*, 9: 39-49.

Spencer, A., Smith, P.J., and Billings, C. (2005). Airport resource management and decision aids for airlines. *Proceedings of the 2005 International Symposium on Aviation Psychology.* Oklahoma City.

Wambsganss, M. C. (2001). Collaborative decision making in air traffic management. In L. Bianco, P. Dell'Olmo, & A.R. Odoni (Eds.), *New concepts and methods in air traffic management* (pp. 1-16). Berlin: Springer.

DIGITAL TAXI CLEARANCES IN AIRPORT TRAFFIC CONTROL TOWERS – INTERFACE DESIGN AND RESULTS FROM A HIGH-FIDELITY SIMULATION EXPERIMENT

Todd R. Truitt, Ph.D.
Federal Aviation Administration
Atlantic City International Airport, New Jersey 08405

The experiment examined the effects of Digital taxi (D-Taxi) clearances when either 40% or 75% of the departure aircraft were data link equipped and compared these conditions to voice-only operations. Sixteen Airport Traffic Control Tower controllers used the Tower Operations Digital Data System at the ground control and local control positions to control airport traffic with a 270-degree simulated out-the-window view. The controllers worked at each position in each experimental condition, provided ratings of subjective workload, and responded to questionnaires. D-Taxi clearances for departure aircraft reduced voice communication between pilots and the ground controller when 75% of the aircraft were data link equipped, without increasing workload or heads-down time. On average, the controllers took 1.28 s to construct and issue a D-Taxi clearance, and they preferred D-Taxi to voice because it reduced their communication duties and removed the potential for readback and hearback errors.

The air traffic controllers in Federal Aviation Administration (FAA) Airport Traffic Control Towers (ATCTs) manage the aircraft moving on the airport surface as well as airborne aircraft that are arriving or departing the airport. The ground controller establishes a sequence for departures and provides taxi clearances to maintain efficient use of the runways and a smooth flow of aircraft to and from the ramp. Currently, ground controllers issue all clearances to pilots via radio voice communications. The FAA has identified Data Communications (Data Comm) as an enabling technology for Next Generation Air Transportation System (NextGen) operations to manage aircraft on the airport surface (Joint Planning Development Office, 2007). In particular, the FAA proposes that ATCT controllers use Data Comm for digital-taxi (D-Taxi) clearances. D-Taxi clearances for taxi-out (i.e., departure) operations may include information such as current Automatic Terminal Information System (ATIS) information or Data Link Operational Terminal Information System (D-OTIS) information, detailed taxi route instructions, and "hold short" instructions. Controllers would not use D-Taxi for time-critical events such as clearances to begin taxi movements, to cross runways, or to transfer control. Fundamental to enabling an effective Data Comm system will be the controllers' ability to manage these digital messages, using effective display designs and supporting automation capabilities in a mixed equipage environment. ATCT and airport surface operations may benefit as a result of the interaction between Data Comm and automated decision support tools, including (a) increased safety in the National Airspace System by providing tools for conformance monitoring and increasing situation awareness, (b) improved controller productivity by reducing controller and pilot workload, and (c) increased capacity by enabling advanced operations via an effective user interface that will reduce instances of human error. To implement Data Comm, the FAA must conduct concept and operational research to identify the advantages and limitations of data link communications. The Data Comm research will also generate an initial set of system requirements that all stakeholders can use as a foundation for future development. The experiment presented here examined Data Comm concepts for ATCTs using a high fidelity, human-in-the-loop simulation.

Method

Current ATCT controllers worked in a high-fidelity simulator at the FAA Research, Development, and Human Factors Laboratory to compare two levels of aircraft Data Comm equipage and a voice only, baseline condition. The Data Comm capability allowed the controllers to receive digital requests for taxi clearances and issue D-Taxi clearances for outbound aircraft. The experiment used a 2 (Run number – First vs. Second) x 3 (Condition – Voice Only, 40% Data Comm, 75% Data Comm) repeated measures design.

Participants

Sixteen Certified Professional Controller (CPC) ATCT controllers served as participants (14 male, 2 female). They were on average 41.3 yrs (*SD* = 8.5) of age, and had worked as an ATCT controller for an average of 17.0 yrs (*SD* = 9.6). All of the controllers had normal or corrected-to-normal vision.

Apparatus

The ATCT simulator consisted of three controller workstation displays situated within a 270-degree out-the-window (OTW) view that comprised nine, 73″ high definition televisions. A standard 20″ computer monitor presented Standard Terminal Automation Replacement System (STARS) radar data. Two VarTech Systems, Inc. touchscreen displays enabled the Tower Operations Digital Data System (TODDS). All of the displays rested on one of two height-adjustable tables. We placed a Workload Assessment Keypad (WAK; Stein, 1985) at each controller position and mounted two cameras and two microphones above each controller position. The cameras recorded interactions with the TODDS and the microphones recorded all ambient sound. We mounted a camera in front of the controllers to capture their point of gaze (e.g., TODDS, OTW). Six simulation pilots, each with their own workstation and communications system, communicated with the ground and local control positions.

The TODDS consisted of two touchscreen displays, one for each control position. Each touchscreen had an active display area of 17″ (43.2 cm) wide and 12.75″ (32.4 cm) high with a 1,600 x 1,200-pixel format, and had a viewing angle of 85 degrees. Each touchscreen had an associated Airport Surface Detection Equipment – Model X (ASDE-X) keyboard and a trackball/keypad as additional input devices. The TODDS presented electronic fight data in the form of Flight Data Elements (FDEs), surface surveillance data – including aircraft position and associated data blocks – weather information, and the ability to construct and issue D-Taxi clearances. This integrated design placed all of the most important information on a single display for each controller position to simplify information presentation and to reduce the controllers' need shift their visual attention among multiple displays. Figure 1 presents screenshots from the TODDS ground and local control positions. For more information about the TODDS, see Truitt (2006, 2008).

Figure 1. The TODDS ground (left) and local (right) control positions including FDEs, surface surveillance data, and weather information.

TODDS D-Taxi graphical user interface. For the current experiment, we built upon the existing TODDS D-Taxi graphical user interface and functionality (Truitt, 2008) by providing controllers with a means to modify the taxi clearance. A data link indicator appeared on the left side of the FDE and in the center of the second line of the data block for each departure aircraft that was data link equipped. The data link indicators appeared at both the ground and local control positions, but only the ground controller could issue or cancel a D-Taxi clearance. Aircraft that were data link equipped had an upright triangle that was either open grey (unowned) or filled white (owned). Only one controller position could own an aircraft at any given time (see Figure 2).

Figure 2. Data link capability indicators in the data blocks of owned (left) and unowned (right) aircraft.

When a data link equipped aircraft reached a ramp spot and was ready to taxi, the pilot sent a D-Taxi request via data link to the ground controller and the data link indicator changed to a light blue triangle and pointed to the left. When the ground controller noticed the pilot request, he or she selected the aircraft's data block or FDE to activate the D-Taxi button, highlight the flight data of interest (FDE and data block), and open the readout area in the upper corner above the FDE list. The readout area displayed the specific pilot request in addition to the full set of flight data for the selected aircraft. The ground controller could defer the pilot's request by deselecting the aircraft which caused the data link indicator to appear as an upright, light blue, open triangle. Deselecting the aircraft also closed the readout area and deactivated the D-Taxi button (see Figure 3).

Figure 3. Data link indicator in data blocks for a pilot request (left), and a deferred request (right).

When the ground controller selected the activated D-Taxi button, the TODDS entered the D-Taxi construction mode and an opaque "screen" appeared over the surface surveillance display. The screen dimmed everything except for the aircraft and related elements of interest and prevented the controller from selecting or moving other FDEs or data blocks. In addition to the opaque screen, the TODDS presented a proposed taxi route (indicated by a white line) that included hold short points and a set of D-Taxi construction mode buttons. In the D-Taxi construction mode, the controller could edit the taxi route, add or remove hold short points, send the D-Taxi clearance to the pilot, or cancel the D-Taxi construction and return to the normal mode. To edit the taxi route, the controller placed his or her fingertip on the displayed route and then dragged the route to a different taxiway. Once the controller selected the Send button and transmitted the D-Taxi clearance to the aircraft, the data link indicator in the FDE and data block pointed to the right. Each D-Taxi clearance included the current ATIS information and the taxi route including hold short points. If a data link transmission failed, the data link indicator turned red and the controller could resend the D-Taxi clearance or contact the pilot by voice. If the data link transmission was successful and the pilot accepted the D-Taxi clearance, then the D-Taxi accepted indicator (an unfilled circle) appeared in the aircraft's data block and FDE. If the D-Taxi clearance included a hold short clearance, then the hold short indicator also appeared on the left side of the aircraft's data block (see Figure 4).

Figure 4. Data comm indicators in data blocks for a sent (left), failed (center), and accepted (right) D-Taxi clearance.

The controller then contacted the pilot via voice and instructed the pilot to "resume taxi" to begin the taxi operation. We required the controller to initiate a taxi movement via voice to establish that voice communication was operational and to allow the controller to determine the sequence and timing of departure aircraft. Aircraft with a D-Taxi clearance were subject to taxiway conformance monitoring. The FDE and data block text of a nonconforming aircraft flashed red on both the ground and local control positions until one of the controllers selected the data block or FDE. The data block and FDE continued to display with red text until the aircraft returned to conformance, or until the ground controller canceled the D-Taxi clearance by selecting the D-Taxi cancel button.

Airport traffic scenarios. We used an airport configuration, which was based on Boston, to develop one 40-min base air traffic scenario. Using a 27/33 runway configuration, the scenario had an arrival rate of 36 aircraft/hr and a departure rate of 42 aircraft/hr with arrivals and departures on runways 27, 33L, and 33R. We modified the base scenario by changing the aircraft call signs to create 12 different versions of the same scenario. This reduced the potential effects of traffic demand while preventing the controllers from recognizing identical scenarios. The controllers worked each version of the scenario in a different random order for each group. A scripted data link failure occurred for one aircraft in each scenario.

Procedure

The controllers worked in groups of two. After signing an informed consent statement, they completed a biographical questionnaire, and received training on the airport and procedures. They then completed a touchscreen training protocol and received training on the TODDS before engaging in six practice scenarios (one at each controller position in each experimental condition). The controllers then completed the experimental scenarios where they controlled airport traffic and provided online measures of subjective workload (WAK) every 5 min. The experimental conditions were presented in a counterbalanced order. The controllers completed questionnaires after each scenario and at the end of the experiment.

Results

Airport System Operations

There was a significant effect of Condition on the duration that aircraft waited on the ramp, *F*(2, 30) = 9.81, *p* < .001. Aircraft waited on the ramp longer in the 40% Data Comm condition (*M* = 74.1 s, *SD* = 17.9) and in the 75% Data Comm condition (*M* = 77.6 s, *SD* = 15.7) compared to the Voice Only condition (*M* = 57.9 s, *SD* = 9.1), *HSD*(30) = 11.73, *p* = .005 and *p* < .001, respectively. The mean taxi-out duration generally decreased as Data Comm capabilities increased from the Voice Only (*M* = 1015.9 s, *SD* = 137.4) to 40% Data Comm (*M* = 982.3 s, *SD* = 83.0) to 75% Data Comm (*M* = 963.8 s, *SD* = 132.3), although the main effect of Condition was not significant. It is possible that holding aircraft on the ramp longer as a result of data link delays resulted in less congestion on the taxiways. Although the differences in taxi-out duration were not statistically significant, they may be operationally significant and could offset the additional time that aircraft waited on the ramp in the Data Comm conditions. On average, 29 arrivals and 33 departures occurred in each condition. D-Taxi did not affect the duration of taxi-in operations, or the number or duration of delays.

Communications

There was no statistical difference between the number of radio transmissions made from the ground control position to the pilots in the Voice Only (*M* = 154.2, *SD* = 22.0), 40% Data Comm (*M* = 155.7, *SD* = 18.5), and 75% Data Comm (*M* = 150.6, *SD* = 22.5) conditions, but there was a significant effect of Condition for the duration of radio transmissions, *F*(2, 30) = 18.67, *p* < .001. The controllers made shorter radio transmissions at the ground control position in the 40% Data Comm condition (*M* = 4.2, *SD* = 0.4) and in the 75% Data Comm condition (*M* =

3.9, *SD* = 0.2) compared to in the Voice Only condition (*M* = 4.5, *SD* = 0.4), *HSD*(30) = .25, *p* = .008 and *p* < .001, respectively. The controllers also made shorter radio transmissions in the 75% Data Comm condition compared to in the 40% Data Comm condition, *p* = .021. Compared to the Voice Only condition, the 75% Data Comm condition saved the controllers about 2.3 min/hr on the radio frequency (0.6 s x 150 transmissions = 90 s or 1.5 min per 40 min scenario).

There was a significant effect of Condition on the number of radio transmissions from the pilots to the ground control position, *F*(2, 30) = 0.88, *p* < .001. The pilots made fewer radio transmissions to the ground control position in the 75% Data Comm (*M* = 187.1, *SD* = 23.6) condition compared to the Voice Only (*M* = 219.6, *SD* = 22.9) condition and the 40% Data Comm (*M* = 206.9, *SD* = 20.2) condition, *HSD*(30) = 17.33, *p* < .001 and *p* = .022, respectively. There was no difference between the number of radio transmissions made in the Voice Only and the 40% Data Comm conditions.

There was also a significant effect of Condition on the duration of radio transmissions from the pilots to the ground control position, *F*(2, 30) = 9.25, *p* < .001. Radio transmissions were shorter on average in the 75% Data Comm (*M* = 2.9 s, *SD* = 0.2) condition compared to the Voice Only (*M* = 3.2 s, *SD* = 0.3) condition and the 40% Data Comm (*M* = 3.1 s, *SD* = 0.2) condition, *HSD*(30) = .16, *p* < .001 and *p* = .019, respectively. Although these differences are relatively small, they may become operationally significant over time. The pilots reduced their time on the radio frequency by about 1.5 min/hr when 75% of the aircraft were data link equipped compared to the Voice Only condition (0.3 s x 200 transmissions = 1 min per 40 min scenario). Overall, the 75% Data Comm condition reduced usage of the ground control radio frequency by 3.8 min/hr, or 6.3%, compared to the Voice Only condition.

Workload

Controller workload was unaffected by the use of D-Taxi clearances. Although the controllers' WAK ratings changed between rating intervals, they did not vary between conditions. Either the D-Taxi operations generated as much workload as voice communications or the controllers reallocated the cognitive and physical resources needed for voice communications to other tasks, including the D-Taxi processes.

Point of Gaze

There was a significant main effect of Object on the total duration of looks, *F*(4, 60) = 298.63, *p* < .001. The post hoc analysis showed that the ground controllers looked at the TODDS longer than at any other object, *HSD*(60) = 444.52, *p* < .001, and they looked at the OTW view significantly longer than they looked at the WAK, Local control position, and Miscellaneous objects, all *p* < .001. When the controllers worked at the ground control position, they looked at TODDS for 30 min and 35 s (*M* = 1835.0 s, *SD* = 112.4) and at the OTW view for 8 min and 37 s (*M* = 517.3 s, *SD* = 110.9) of the 40 min scenario. The controllers spent the remaining 48 s by dividing their visual attention between the WAK (*M* = 11.2 s, *SD* = 1.2), Local control position (*M* = 23.6 s, *SD* = 6.1), and Miscellaneous objects (*M* = 13.12 s, *SD* = 2.8). Even though the controllers used the D-Taxi functions of TODDS in the 40% and 75% Data Comm conditions, they spent the same amount of time looking at the TODDS and OTW in the Data Comm conditions compared to the Voice Only condition.

D-Taxi Usability

We recorded the number and types of actions the controllers performed with the TODDS during each scenario. We then calculated an error rate (ER) for each action type by dividing the number of failed actions (F) by the sum of successful actions (S) and failed actions (F), so that ER = F/(S+F). The overall error rate for the ground control position of the TODDS was 3.3% across all scenarios. Overall, the controllers were able to construct and issue a D-Taxi clearance in only 1.28 s on average, including the time consumed by usability errors.

Controller Opinion

The controllers reported that it took little effort to issue either a canned or user-defined D-Taxi clearance. They reported that it was easy to manage flight data. They had a high awareness of current and projected aircraft positions, potential runway incursions, and the overall traffic situation. Although one controller thought the failed Data Comm indicator was too subtle ("I like the red, but it needs to stand out a little more"), other controllers thought the symbol was "very usable." Workload due to controller-pilot communications was moderate as was overall workload, but workload at the local control position was slightly higher than workload at the ground control position. The controllers rated the safety of operations at the ground control position as high, but they thought it was less safe at the local control position. The controllers may have been concerned about safety at the local control position because of the complex, busy traffic scenarios that involved intersecting runways.

The controllers thought that D-Taxi would have a positive effect on their ability to control airport traffic. They indicated that it was easier to detect an aircraft deviation if an aircraft had received a D-Taxi clearance ($M = 2.4$, $SD = 1.1$) compared to an aircraft that received a voice clearance ($M = 4.3$, $SD = 2.3$), $t(15) = 3.58$, $p = .003$. However, the controllers did not perceive a difference in the effort needed to manage an aircraft that deviated from its taxi clearance, given the type of taxi clearance. Some controllers thought that D-Taxi may be useful for taxi-in operations, but they did not elaborate on a particular set of procedures that would be likely to work. Others said that D-Taxi for arrivals would be "more challenging" and could result in blocked runway exits and increased pilot workload. Some controllers stated that they would use D-Taxi as much as possible. Others stated that they would not use D-Taxi during rapidly changing weather conditions, during low visibility, during times that the pilot was unfamiliar with the airport, during times that an aircraft had an Expected Departure Clearance Time, or during emergencies. The controllers said that reduced voice communications was one of the greatest benefits of D-Taxi because it reduced the risk of readback and hearback errors, reduced confusion, and improved situation awareness. They thought that the TODDS, as an overall system including D-Taxi, would have a positive effect on the ability to control airport traffic. Although the controllers said that the TODDS had a learning curve that one must overcome before taking full advantage of the tool, they listed a number of benefits, including the fact that it integrates all of the most important information into a single screen. They thought that the TODDS would work in low visibility conditions and that it would improve safety, reduce workload, and reduce the need for verbal communication and coordination.

For complete details of this experiment, see Truitt and Muldoon (2010).

References

Joint Planning and Development Office. (2007, August). *Next generation air transportation system research and development plan FY 2009-2013*. Retrieved from http://www.jpdo.gov/ library/NextGen_v2.0.pdf

Stein, E. S. (1985). *Air traffic controller workload: An examination of workload probe* (DOT/FAA/CT-TN84/24). Atlantic City International Airport, NJ: Federal Aviation Administration Technical Center.

Truitt, T. R. (2008). *Tower Operations Digital Data System – Concept refinement and description of new features* (DOT/FAA/TC-08/09). Atlantic City International Airport, NJ: U.S. Department of Transportation, Federal Aviation Administration.

Truitt, T. R., & Muldoon, R. (2010). *Data Communications Segment 2 Airport Traffic Control Tower Human-in-the-Loop Simulation* (DOT/FAA/TC-10/05). Atlantic City International Airport, NJ: U.S. Department of Transportation, Federal Aviation Administration.

SITUATION AWARENESS, WORKLOAD, AND PERFORMANCE IN MIDTERM NEXTGEN: EFFECT OF DYNAMIC VARIATIONS IN AIRCRAFT EQUIPAGE LEVELS

Thomas Z. Strybel, Kim-Phuong L. Vu, L. Paige Bacon, Sabrina Silk Billinghurst, Robert Conrad Rorie,
Joshua M. Kraut, Corey Morgan
Center for Human Factors in Advanced Aeronautics Technologies
California State University, Long Beach
Long Beach, CA

Vernol Battiste
San Jose State University Foundation; NASA Ames Research Center
San Jose, CA; Moffett Field, CA

Walter Johnson
NASA Ames Research Center
Moffett Field, CA

NextGen changes in air traffic management promise to bring many benefits to the current airspace system, but these changes must be evaluated for their impact on mid-term air traffic management in which mixed-equipage is certain. We examined mixed equipage environments in which the equipage levels changed over the course of the scenario to reflect changes in sector characteristic over the course of a day or controller's work shift. Six retired ATCs managed mixed-equipage traffic that either began with low levels of NextGen equipped aircraft and increased midway through the scenario or vice-versa. These were compared to a scenario in which the equipage mix was held constant. ATC performance, workload and situation awareness were affected differently by these scenarios.

The Next Generation Air Transportation System (NextGen), is a program being developed under the guidance of several government agencies, including the Federal Aviation Administration (FAA), Department of Transportation, and NASA, under the umbrella of the Joint Planning and Development Office (JPDO, 2007). The goals of NextGen include expanding the capacity of the National Airspace System (NAS) to handle 2-3 times current day traffic to accommodate projected increase in air travel over the next 10-20 years (see e.g., FAA Implementation Plan, March 2010). NextGen tools, technologies, and concepts will be implemented in phases (JPDO, 2007). Phase 1 (2007-2011) included research and development of avionics technologies for enabling NextGen concepts and procedures being planned for subsequent phases. One of these technologies, Data Comm, allows pilot and controllers to communicate using a text-based messaging system in lieu of voice communications.

The use of Data Comm is anticipated to reduce operator workload under high traffic environments by reducing the serial mode of voice transmissions (Kerns, 1991), reducing ambiguity of the intended message (Flathers, 1987), and reduce working memory demands associated with remembering auditory messages (Hinton & Lohr, 1988). In mid-term NextGen, Data Comm in the cockpit can be integrated with the flight management system. With integrated Data Comm, ATC and pilots can negotiate flight plan requests, and agreed-upon changes loaded into both ground and aircraft systems. ATCs and pilots may also benefit from automated conflict resolution which provides recommended speed or trajectory changes to resolve a conflict. Controllers managing Data Comm equipped aircraft (AC) could benefit from other NextGen technologies such as trial planners and conflict probes.

Although the potential benefits of Data Comm, along with changes to the NAS brought about by NextGen procedures would eventually lead to most aircraft being equipped with Data Comm, the adoption of Data Comm by commercial and private carriers is likely to develop gradually over time due to the cost associated with equipping aircraft. Consequently, air traffic controllers in the near- and mid-term will be challenged to manage AC having very different capabilities. In addition, controllers may not always benefit from NextGen tools for conflict detection and resolution because it requires that both ACs in conflict be appropriately equipped with NextGen technologies. Controllers will need to rely on their "manual" air traffic management skills for detecting and resolving conflicts between AC pairs involving one or more unequipped aircraft. A mixed-equipage fleet may also increase workload because ATCs must maintain awareness of the equipage level of all ACs in the sector.

Preliminary investigations of air traffic management in mixed equipage airspace have shown that the percentage of equipped aircraft affects ATC performance and workload. Corker et al. (1999) examined the effects of mixed equipage environments on conflict detection and ATC workload. For equipped AC in this investigation, pilots had some responsibility for maintaining separation from other traffic. The authors showed that the equipage levels of 80 and 100% reduced workload, but that a 20% equipage level increased workload over the level achieved with 100% equipage. Prevot et al. (2005) reported that controllers who participated in mixed equipage airspace simulations were somewhat negative about the impact of mixed equipage on situation awareness and safety. Specifically, controllers reported that it would be slightly more difficult to detect non-conforming aircraft and more difficult to cope with unplanned events in mixed-equipage environments. Willems et al. (2008) showed that for 70% equipage levels, ATCs could manage a 33% increase in traffic over current day levels, but could not handle a 66% increase in traffic. Willems et al. also observed that in a 70% equipage environment, controllers attempted to uplink Data Comm messages to aircraft were not equipped to receive it. Hah et al. (2010) compared sector equipage levels of 0%, 10%, 50%, and 100% in a simulation of an en route sector. They noted that significant contributions of Data Comm required at least 50% equipped aircraft in the sector. Moreover, ATCs reported high percentages of equipped traffic changed the way they managed traffic.

In summary, NextGen changes in air traffic management promise to bring many benefits to the current airspace system, but these changes must be evaluated for their impact on mid-term air traffic management in which mixed-equipage is certain. In the present simulation, we investigated another aspect of mixed equipage traffic airspaces. We examined mixed equipage environments in which the equipage levels change over the course of the scenario, because the equipage mixture will most likely be dynamic or changing over the course of a day or controller's work shift. These may bring about changes in ATC workload, which has been shown to reduce operator situation awareness (SA; Hallbert, 1997). ATCs managed traffic in a simulated en route sector having an average of 50% equipped aircraft, but the mixture in some scenarios began with a lower level of equipped aircraft and increased midway through scenario, or vice versa. Workload and situation awareness were measured with an online probe technique that is a variant of the Situation Present Assessment Method (SPAM; Durso & Dattel, 2004) because SA information is not only limited to the contents of working memory, but also can be distributed across the operators' task environment (Chiappe, Strybel & Vu, submitted).

Method

Participants

Seven retired ATCs participated in the simulation. All were former radar-certified TRACON ATCs with 9-25 years of experience in either Southern California or Bay Area TRACON facilities. One participant had eight years experience in an en route center, and one had participated in previous simulations at NASA Ames Research Center. None reported having real-world experience with the sector being simulated in the present study.

Apparatus and Scenarios

The simulation was run using the Multi Aircraft Control System (MACS), Aeronautical Datalink and Radar Simulator (ADRS) developed by the Airspace Operations Laboratory at NASA Ames Research Center (e.g., Prevot et al., 2006). Participants managed traffic in simulated sector ZID-91, using MACS configured as a Digital System Replacement (DSR) display having advanced tools (integrated Data Comm, conflict alerts and probes, and trial planners) that could be used to manage AC designated as equipped. For the remaining non-equipped AC, participants managed traffic with voice commands. All ATC instructions were executed by a pseudopilot located in an adjacent room. ATC participants determined the status of each AC's equipage by checking the data tag for a diamond located to the left of the call sign indicating that the aircraft was equipped. Online situation awareness and workload probes were presented on a touch-screen workstation located adjacent to the DSR display.

Six experimental trials were run, each lasting approximately 50 minutes, with each scenario containing on average 50% equipped and 50% unequipped ACs. Within some scenarios the percentage of equipped AC changed midway through the scenario. Two scenarios started with 25% equipped AC. Midway through the scenario the percentage increased to 75% and remained roughly constant thereafter. Two scenarios began with 75% equipped AC, and the percentage decreased to 25% at the halfway mark. Finally, two scenarios had a constant mix of 50% equipped AC throughout the scenario. For each equipage mixture, one scenario contained 8 conflicts and the other six conflicts, as shown in Table 1, with half the conflicts occurring between equipped aircraft (and subsequently producing an alert on the DSR.) Of the remaining conflicts, the number of conflicts between one equipped and one unequipped AC, and between two unequipped ACs was counterbalanced. In addition, the number of conflicts in the first and second half of each scenario was equated. Note also in Table 1, that the second half of the scenario contained more AC, which reflects potential increases in traffic that occurs at particular times of day.

Table 1

Traffic Characteristics (Percent Equipped AC, Number Alerted Conflicts) of the Scenarios

Scenario	% Equipped by Half	Total Num. Conflicts	Num. Equipped Conflicts	Num. Un-equipped Conflicts	Ave. Num.AC: 1st Half	Ave. Num.AC: 2nd Half
1 2	5-75	6	3	3	6.6	9.1
2 5	0-50	6	3	3	7.7	10.6
3 7	5-25	6	3	3	7.6	14.0
4 2	5-75	8	4	4	7.6	10.9
5 5	0-50	8	4	4	7.5	12.1
6 7	5-25	8	4	4	7.0	9.0

Procedure

Participants were given one week of training consisting of a briefing on traffic flows, ATC procedures, the MACS DSR interface, conflict alerting and use of the trial planner, and online probe questions. Following the briefing, participants practiced managing traffic using current day procedures, with NextGen tools, and with probe questions (see Kiken et al., 2011). The experimental sessions were run in the week after training over two days. Each participant ran one each of the scenarios described in Table 1, while simultaneously responding to probe questions. In each experimental scenario, 16 probe questions were presented, one every three minutes, beginning four minutes into the scenario. The questions were designed to assess ATC's awareness of conflicts between aircraft (e.g., "In what area of your sector will the next conflict occur?"), information relating to sector status (e.g., "Are there more equipped AC in your sector at this moment?") and workload (7-pt scale similar to the ATWIT). Workload prompts were administered four times at regular intervals. The sequence of events for each probe question was as follows. The ATC received a "Ready for Question?" prompt on the touch screen accompanied by an audio alert. The participant responded by touching a button on the screen when he/she had sufficient excess capacity to answer a question. Once the participant responded affirmatively, a probe question and response alternatives were displayed on the panel and the participant answered the question by pushing one of the response buttons. If the participant did not respond to the "Ready" prompt within two minutes the query was withdrawn.

Results and Discussion

Measures of workload, performance, and situation measures were analyzed with repeated measures ANOVAs. For time-based variables, the analyses were done on log transforms but the results are shown in seconds. Two measures of workload, ATWIT ratings and time to respond to ready prompt, were analyzed with three-factor repeated-measures analyses of variance, with the factors of percent equipped AC, number of conflicts and scenario half. For ATWIT ratings, a significant three-way interaction was obtained ($p < .05$), as shown in Figure 1. For each scenario, the ATWIT ratings increased in the second half, with the greatest increase occurring for the 75%-25% scenario with 6 conflicts, and the 50%-50% scenario with 8 conflicts. Moreover, in the second halves of both these scenarios, the mean rating was very high (M=4.9 for both scenarios), and ATC provided ratings of 6 or higher roughly 50% of the time. From Table 1, these scenarios contained the highest traffic densities. The three-way interaction was marginally significant for ready latency ($p < .06$). As shown in Figure 2, for scenarios containing 6 conflicts, ready latency increased in the second half when the scenario contained 25%-75% equipped AC, but decreased in the second half when the scenario contained 75%-25% equipped AC. For 8-conflict scenarios, the mean ready latency increased in the second half for scenarios containing 50%-50% and 75%-25% equipped AC. Obviously, these ready latencies are inconsistent with ATWIT ratings. We believe this is due to the fact that ATCs did not respond to Ready prompts frequently in the second half of the scenario. For example, in the 50%-50% scenario with 8 conflicts, the ready prompt was ignored (and timed out) on 35% of the workload probe queries. Nevertheless, a marginally significant correlation was obtained between ATWIT rating and Ready latency (r=.20; p= .07). Only ATWIT ratings were significantly correlated with average number AC being worked (r =.21, p<.05), suggesting ATWIT subjective workload is related to ATCs' perception of the number of AC being managed.

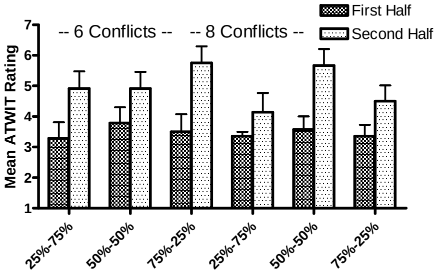


Figure 1. Mean ATWIT Rating as a function of equipage and number of conflicts.

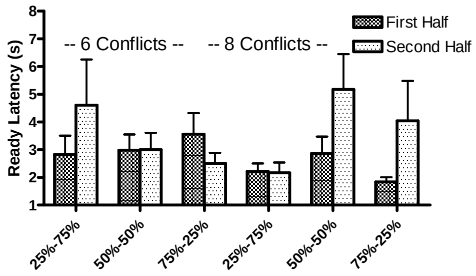


Figure 2. Mean Latency to Ready Prompt as a function of equipage and number of conflicts.

Performance

Based on the extremely high workload in two scenarios, subsequent analyses was limited to those scenarios having roughly the same traffic densities, 25%-75% with 6 and 8 conflicts, 50%-50% with 8 conflicts, and 75%-25% with 6 conflicts. Performance was measured in terms of safety and efficiency. The number of alerted LOS and non-alerted LOS were analyzed for second halves only because no LOS occurred in the first half of any scenario.

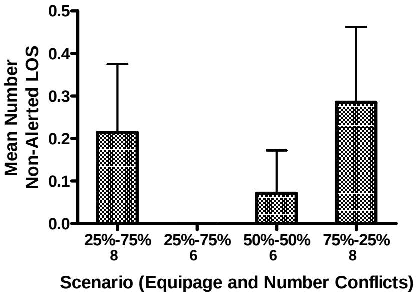


Figure 3. Average non-alerted LOS as a function of scenario for the second half.

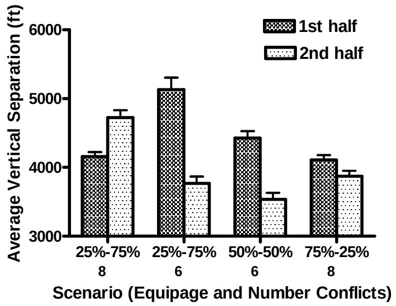


Figure 4. Vertical separation between AC as a function of scenario and scenario half.

For non-alerted LOS, a marginally significant effect of scenario was obtained (*p* < .07). As shown in Figure 3, more LOS occurred in scenarios containing 8 conflicts, and equipage did not seem to affect the number of LOS. For vertical and average lateral separation between AC, repeated measures analyses of variance were performed with the factors being scenario and scenario half. A significant main effect of scenario and a significant interaction between scenario and scenario half (*ps* < .001) were obtained, as shown in Figure 4. Vertical separation was greater on average for the 25%-75% scenarios compared with both 50%-50% and 75%-25 % scenarios. For the 25%-75% scenario with 8 conflicts, vertical separation increased in the second half. In all other scenarios, average vertical separation decreased. A main effect of scenario and scenario half was obtained for average lateral separation, shown in Figure 5 (*p* < .001). Lateral separation was lower for the 25%-75% equipage scenarios compared with the remaining scenarios. For each scenario, lateral separation increased in the second half, and the interaction with scenario was not significant. Therefore, when the scenario began with 25% equipped AC and this increased to 75% AC, vertical separation was greater and lateral separation smaller than when the scenario contained 50% equipage throughout or began with 75% equipage and this decreased to 25% equipage.

Sector efficiency was measured with average handoff delay and average time working an AC. For handoff delay, only scenario half was significant (*p* < .002); handoff delay increased in the second half of the scenario, presumably because of the increase in traffic. For average time spent working AC, significant main effects of scenario and scenario half were obtained (*p*s < .002). On average, ATCs spent the least amount of time working on AC in the 50%-50% scenario compared with the remaining scenarios. This may not be due to the number of conflicts, however, because the longest average time was for the 25%-75% scenario with 6 conflicts.

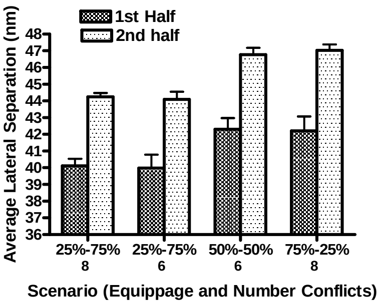


Figure 5. Average lateral separation (nm) as a function of scenario and scenario half.

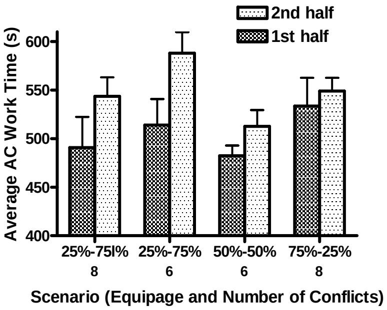


Figure 6. Average time spent working AC as a function of scenario and scenario half.

Situation Awareness

Situation awareness was measured with the accuracy and latency of responses to online probe queries relating to conflicts or sector status. Probe latencies were analyzed separately for conflict and sector status queries. For sector status queries, a significant interaction between scenario and scenario half was obtained, as shown in Figure 7. For most scenarios, probe latencies were either unaffected by scenario half, or they increased minimally. Probe latencies for the 50%-50% scenarios were nearly identical, meaning that awareness of sector status did not change when the percent of equipped AC remained constant. On the other hand, for 75%-25% equipage, sector probe latency increased in the second half by 3 s on average, suggesting that awareness for this information was lowered by the decreasing percentage of equipped AC. The effect of scenario on conflict probe latency was marginally significant ($p < .10$), but the effect of scenario half was nonsignificant.

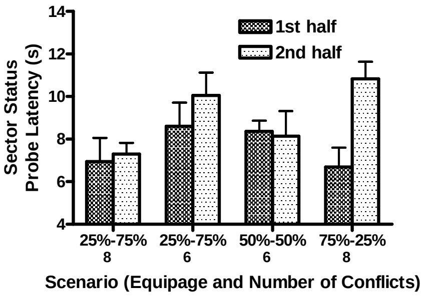


Figure 7. Average probe latency to sector status queries as a function of scenario and scenario half.

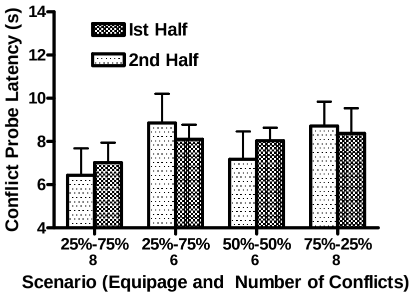


Figure 8. Average probe latency to questions on conflicts as a function of scenario and scenario half.

Conflict probe latencies, shown in Figure 8, were lowest at 25%-75% with 8 conflicts, and slightly higher for the remaining scenarios. To check the validity of our probe latencies, we computed correlations between sector probe latency, conflict probe latency, and performance metrics. Sector probe latencies were significantly correlated with handoff delay ($r= .29$; $p<.05$. Conflict probe latencies and number LOS was marginally correlated, $r = -.26$; $p < .07$. The correlation between conflict probe latencies and average time spent working AC was significant $r=-.35$; $p<.01$.

Conclusion

Our preliminary investigation showed that changes in equipage within a scenario affected ATC performance, workload, and situation awareness, but the effects depended on the specific measure. By limiting our analysis to scenarios with roughly the same traffic densities, we determined that when the percentage of equipped AC was initially low and then increased during the scenario, vertical separation was greater and lateral separation smaller than when the percentage of equipped AC was constant or began high and then decreased during the scenario. Combined, these changes in vertical and lateral separation suggest that ATCs changed their strategies for managing traffic based on changes in equipage. The number of LOS was affected only by the number of conflicts, and not equipage levels. LOS occurred only in the second half of each scenario. ATC perceived workload (ATWIT) increased in the second half of the scenario regardless of equipage levels, because of to the increase in traffic during the second half. Situation awareness of sector status information was lower (probe latencies higher) in the second half, and this change was greatest when equipage began at 75% and decreased to 25%. Conflict probe latencies were not dependent on scenario half, because the number of conflicts in each scenario half was equal. In summary, these results indicate that changes in mixed equipage traffic within a scenario should be investigated further to determine the extent to which ATCs change traffic management strategies in response to equipage level.

Acknowledgements

This simulation was partially supported by NASA cooperative agreement NNA06CN30A: *NRA: Metrics for Situation Awareness, Workload, and Performance in Separation Assurance Systems* (Walter Johnson, TM). Preparation of this paper was supported by NASA cooperative agreement NNX09AU66A, *Group 5 University Research Center: Center for Human Factors in Advanced Aeronautics Technologies* (Brenda Collins, TM).

References

Chiappe, D. L., Strybel, T. Z., Vu, K.-P. L. (2010). Review of Situation Awareness Theories from the Perspective of Cognitive Science. Submitted for publication.

Corker, K., Fleming, K., & Lane, J. (1999). Measuring controller reactions to free flight in a complex transition sector. *Journal of Air Traffic Control, 4,* 9–16.

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

Flathers, G.W. (1987). Development of an air ground data exchange concept: Flight deck perspective (NASA Contractor Rep. No. 4074). Hampton, VA: NASA Langley Research Center.

Hallbert, Bruce P. (1997) Situation awareness and operator performance: results from simulator-based studies. *Proceedings of the IEEE sixth annual Human Factors Meeting*, Orlando, Florida.

Hinton, D.A. & Lohr, G.W. (1988). Simulator investigation of digital data link ATC communications in a single-pilot operations (NASA Tech. Paper 2837). Hampton, VA: NASA Langley Research Center.

Joint Planning and Development Office. (2007). *Concept of operations for the Next Generation Air Transportation System, Version 2.0.*Washington, DC: Author.

Kerns, K. (1991). Data-link communications between controllers and pilots: A review and synthesis of the simulation literature. *The International Journal of Aviation Psychology, 1* (3), 181-204.

Kiken, A., Rorie, R.C., Bacon, L. P., Billinghurst, S., Kraut, J.M., Strybel, T. Z., Vu, K.-P. L., & Battiste, V. (in press). Effect of ATC training with NextGen tools and online situation awareness and workload probes on operator performance. *HCII 2011, Lecture Notes in Computer Science.*

Prevot, T., Smith, N., Palmer, E., Mercer, J. Lee, P., Homola, J. & Callantine, T. (2006). The Airspace Operations Laboratory (AOL) at NASA Ames Research Center. *AIAA Modeling and Simulation Technologies Conference and Exhibit* (AIAA 2006-6112). Keystone, Colorado.

Salmon, P., Stanton, N., Walker, G. & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics, 37* (2), 225-238.

Willems, B., & Hah, S. (2008). *Future En Route Workstation Study (FEWS II): Part 1 – Automation Integration Research* (DOT/FAA/TC-08/14,II). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Willems, B., Hah, S., & Schulz, K. (2010). *En route data communications: Experimental human factors evaluation* (DOT/FAA/TC-10/06). Atlantic City International Airport, NJ: FAA William J. Hughes Technical Center.

IMPROVE CRM AND ADVOCATE “ONE BALLOT VETO”

1. Shi Ya-jie 2. Li Jing

China Academy of Civil Aviation Science and Technology

Beijing 100028，China

Accident investigations showed that most accidents are caused by the flight crew. But in many cases, if the captain can listen to first officer or the observer suggestion, such as "go-around ", for safety reasons at the key phase of final approach, many flying accidents/incidents can be avoided. Unfortunately, the captain sometimes would still chose to continue the final approach due to over-confidence, then leading to disastrous consequences in the end. CRM(Crew Resource Management) endeavored to reduced the crew human errors and improve the safety level all the while, but there is no effective way to eliminate this kinds of problems at yet. In this paper, this kind of problems is deeply analyzed from some areas, such as the crew attitude, cultural characteristics, operation standards and regulations, etc. And recommended CAA to improve CRM by advocating the ‘One Ballot Veto’ system(—at the key phase of take off and approach, once someone argue against the current operation for the safety reason, the captain must hear the suggestions and take the operation which are more benefit to the safety at once), so that reduce these problems.

The worst disaster in airline history occurred in Tenerife on March 27th, 1977 when a captain of a KLM 747 insisted on commencing a takeoff without clearance in a heavy fog even with the knowledge that a Pan Am 747 taxiing down the runway and had not yet reported clear of the runway. The first officer on KLM 747 had some doubts on whether the take-off clearance had been received and tried to convey his concerns to the captain, but the captain ignored his advice and pushed the power up without any hesitation which caused two Boeing 747 jumbo jets collided on the runway and 583 people were killed[1]. Another instance was a Air China 767 crashed in Pusan, South Korean on 15 April, 2002. During the final approach, the F/O’s advisory on immediate go around was not disregarded by the captain and hence the best opportunity to avoid flying into the terrain was missed[2]. Similar situation are very common among different incidents and accidents occurred in the world.

What caused the captain rejecting safety recommendation from the other crew member and missed the last opportunity to break accident chain and eventually avoid the disaster? Why shouldn’t we take some measures to stop such tragedy? On the basis of analysis on some problems existed in CRM, a new concept-“one ballot veto” is proposed as well as some solutions are discussed in this paper.

CRM Status

The target of CRM is to achieve the highest level of efficiency and safety by making use of all available resources effectively. Since the conception of CRM was put forward in 1979, it has been developed and perfected all the way along to its fifth stage-“Threat and Error Management (TEM)”. Statistics shows it plays an important role in enhance flight safety and efficiency[3,4].

However, the following table (see Table 1) indicates that there are total amount of 323 fatal accidents happened world widely in last ten years together with a death toll of 8646, besides no remarkable decrease in the number of the accidents. Moreover, many investigation reports demonstrate more than 70% of the accidents are caused by human factors i.e. fatigue, decision mistake or ignoring automatic system warning, etc[4]. Therefore, to improve CRM, some problems need to be studied further.

Table 1 *Statistical Summary on Worldwide Fatal Accidents 1998-2007*

Year	2007 200	6	2005	2004	2003	2002	2001	2000	1999	1998	**T**otal
Number of accidents	26 27		35	28 25		37	28 36		42	39	323
Death toll	750 888	1	059	429 679	1	101	768 108	2	671	1219 8	646

Note：1)Data in the above table is from ASN；2) Aircraft refer to multi-engine commercial airplanes with 14-seating or above.

Among them, the problems related to CRM in the two accidents mentioned before is a kind of severe and unsolved one. One common feature involved in this type of accidents (or incidents) is that, during the critical point in approach and landing (or take-off), to deal with the abnormal situation appeared, one of the crew members (or ATC controller, GPWS system) propose some safety advice or alert (e.g. request for attention to altitude, go-around or take-off abortion etc), which are unfortunately discarded by the captain, who due to CRM defect keep on the wrong side of aircraft control or safety operation, missed the last chance to survive the mistake chain and consequently lead the disaster.

According to Boeing company[5] (see Figure 1), most of the accidents are occurred at take-off and approaching to land. The number of the accidents happened during these two phases amounts 51% of the total, in contrast to the duration of the phases is only 6% of the total flight time. It is easily found that any CRM problem occurred during takeoff and landing are likely threaten flight safety seriously and resulted in fatal disaster. On the basis of accident chain theory, any interruption in the chain connection could prevent the accident from happening. As far as the accident or incident discussed in this paper, the final decision made by the captain is the last minute to interrupt the chain, but if the captain still persist in his fault even after the other member suggesting to go-around or abort take-off or warning triggered by onboard instruments, it is impossible to avoid the accident. As a result, how to solve such kind of problem by means of develop CRM is of great importance to strengthen global air transportation safety. In fact, it has been noticed in aviation industry already and it is the major factor to push forward the adoption of CRM. In order to communication in cockpit, a lot of books have been published as well as some principles including "double check" and "being most conservative". Through analysis, "one ballot veto" principle was promoted by us and illustrated as the follows:

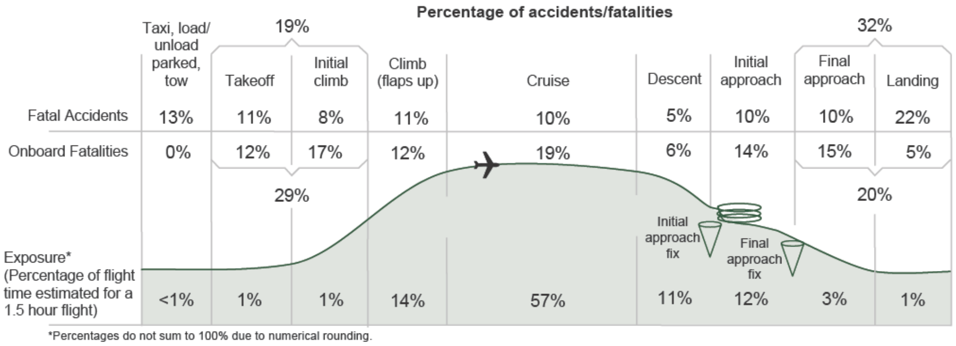


Fig.1 fatal accidents and onboard fatalities by phase of flight (1997-2006)

Analysis on Causes to These Problems

Due to some deficiency of CRM, the flight crew is unable to prevent the accident at critical moment by interrupting the accident chain, which is usually divided into two types:

1）Captain doesn’t accept safety recommendations from F/O or warning system, for examples of the two accidents mentioned in the first part.

2）For some reason, crew member doesn’t express any fear or objection to captain’s wrong decision or action.

Most of the cases belong to type 1, since type 2 is an extreme situation rarely occurred nevertheless more serious and hard to deal with. The following factors are found to be the causes of the problem.

Personal Attitude of Crew Members

The study showed that one of the main causes to CRM problem is the inappropriate personal attitude of the crew.

1）Inappropriate personal attitude of captain:

- Autarchy;
- Actuation;
- Overconfidence;
- Show-off.

Being an absolute leader in the cockpit, the captain is likely to regard any safety advice by any other crew member as a challenge to his authority or interruption to his order so that respond in a negative and autarchical behavior. Some captains are so overconfident that they don’t think any mistake he or she could make as well as any disaster could happen to them. Besides, someone may take a risk to show his power and boldness [6]. It is often to find these improper attitudes in the cockpit with two captains. More and more rule violation cases are found in the cockpit due to improper competition or struggle led by overconfidence. Moreover, as an operator, some captains intend to control everything and take immediate response to any situation which often resulted in a rush action. Usually, the best way to handle emergency situation is to go around instead of trying to land, unfortunately, go around is regarded as “being incapable to perform” among lots of the pilots.

2）Inappropriate attitude of F/O:

- In awe of authority;
- In fear of make mistake;
- Irresponsible;
- Usual assumption.

On the contrary to the captain’s attitude, other crew members are possibly passive and take no action to the captain’s dangerous operation at critical moment. Being a F/O, he or she is probably hesitant to give any negative comments when flying with an experienced and capable captain. Even if they point out the captain’s errors, they would like to use some mild words. For example, being a trainee, the F/O would not tell a captain the final approach speed is 15knots slower than required, on the other hand, he or she may say that “the speed seems a little bit slow” or in even more polite way by reminding the captain of “wind shear may be encountered”, which are dangerous rather than ineffective since it may screen the severity. To take over control of the aircraft from the captain is supposed even more difficult, simply because the F/O have to worry about losing his job.

Training on CRM

How to improve communication in the cockpit is covered by CRM training syllabus, but up to now, no effective training course is set up targeted to help the captain abandon his or her stubbornness at the urgent moment. Why haven't these problems been worked out effectively? It is found that the main reason is that the captain in the cockpit is over-authorized and at an ultimate leadership. Of course, the existing power unbalanced reality is the exact reason for putting forward the concept of CRM, but it is a pity that no effective solution is achieved. How does the crew make a safety beneficial decision at crucial point? When is the proper time for the F/O to extend his worries on safety issue or how to put forward his argument against the captain? When shall the F/O take the control actively so as to ensure safety? All of these questions are required to be considered in CRM training courses.

Regulations and Standards

Actually, there is specific instruction on when to go around or abort taking off in flight rules or standards currently in force. In accordance with REASON theory, the operator of the aircraft is the last safety defense line. However, it is impossible for a human being to make no mistake. For a variety of reason, the operator may ignore some regulations helpful for safety and take a risky action. Additionally, the current regulations and standards on going around and take-off abortion are focused on technical point of view as well as there is no particular explanation on how to ensure its implementation and what kind of rights could support the F/O.

Cultural Features

It the social culture to lead the F/O are easily obey the authority and seldom resist, moreover, the control column is in the hand of the captain[7]. This phenomenon is very common, especially in Asian area, just because the social or national culture is in favor to the person with superiority and power. Meanwhile, the Asian people are more easily to accept the social estate and power inequality in their daily life which eventually facilitate autarchical habit built by the powerful one, which make it hardly possible to overcome the authority limitation for the junior and object the instruction from the senior even if in case of emergency. It is not so bad in western countries.

Advocate "One Ballot Veto" and Improve CRM

Creating strong CRM is the basis for safety assurance. Scientific and reasonable CRM is very helpful to maintain the control of the aircraft in bad condition, which is the first priority to ensure safety. On the basis of the above analysis, the principle of "one ballot veto" is advocated so as to improve CRM, prevent those evitable accidents from happening and enhance aviation safety in China.

Definition of "One Ballot Veto"

The principle of "one ballot veto" refers that the captain should accept advice and take immediate action once any member of the crew (including flight crew or ATC controllers) shows uncertainty or objection to continue current operation (i.e. take-off, approach and landing) and provides recommendations favorable to flight control and safety (i.e. abort take-off, maintain altitude or go around) according to relative rules and standards. It is initiated basing on the guideline of "safety first" always implemented in China civil aviation. The main purpose for the initiative is to drive high attention paid on any safety advice or disagreement during flight and choose the safest plan at some critical phases. Two circumstances are listed, firstly, during the final approach, in case of any condition be harmful to control or safety, one of the crew member suggests to give up approach (go around) i.e. oppose to take on a risk of approaching to a final landing, the captain must go around. Secondly, before take-off, if some crew member has any doubt on safety, the captain must abort take-off and check it.

Function of “One Ballot Veto”

The principle of “one ballot veto” may contribute to improve CRM in several aspects;

At critical time, it is very helpful for the crew to make a safety beneficial decision, take an action accordingly and interrupt the accident chain so that the accident would not happen, as it is demonstrated by case 3 and case 4 in appendix. Suppose the aircraft is at low level or some emergency situation occurred, it may make the situation worse if the crew’s attentions are distracted by arguing which method is more economical or simple, or the captain sticks on his own opinion deliberately for showing his flying skill. It will be improved if the principle of “one ballot veto” is implemented in a proper way and safety will be enhanced.

The principle will contribute to train the crew members to get into the habit of good communication, which in return protect the captain against being dogmatic and stubborn for a variety of reason as well as relieve the F/O’s tension to “say no”. Besides, it will assist the crew members better understand their own roles and duties and dedicate to flight safety.

It will be of great support to carry out the guideline of “Safety First” and establish more active cockpit atmosphere and reporting culture.

Conclusions and Suggestions

Through the above analysis, it is concluded that the principle of “one ballot veto” does play an active role in solving the problem related to CRM discussed in the paper. It is believed that nobody will query or disagree the captain’s decision and action without any reason. Once some other voice is heard, it must be for the purpose of flight safety assurance instead of challenging the captain’s authority. As far as it is concerned, the captain has dual responsibilities, one for controlling the aircraft, another for collecting team members’ opinions and making a final decision. Therefore, rather than destroy the captain’s authority and deprive his decision making right, advocating “one ballot veto” will certainly improve CRM, build up favorable safety culture and encourage the captain to accept more reliable measures suggested and avoid accident at critical fight phases. Meanwhile, it should be clarified that the principle does not mean anyone could take a measure immediately without the captain’s approval once he or she propose some safety advice, otherwise, it will go to another extreme.

In order to implement the principle of “one ballot veto” correctly, improve CRM and enhance flight safety, it is strongly recommended to do the followings:

１）Further research should be conducted to distinguish those factors harmful to successful CRM and find different solutions. The current flight rules and standards should be revised so as to specify the conditions applicable to the principle as well as corresponding implementation directory and authority necessary should be made.

２）Strengthen CRM training by increasing more cases study to assist the pilots realizing the severity of the problems and educate them to be able to prevent the accident by taking the advantage of the principle, especially in case of emergency. It is very important to help them to recognize “once the approaching is not successful, going around quickly may be the last chance for evading accident. During approach to landing, if you are not sure of current situation and confident with the safety of continuing to land, you’d better not to take any adventure and go around immediately since it is probably the last opportunity to replace the accident report by the reporting of “it is the worst approach I have ever done”. On the other hand, CRM training course should cover the responsibility and action taken by the other crew members once they observe any decision threatening safety by the captain. It is necessary to require pilots understanding the accidents involving human factors discussed herein and set up guides about when and how to take the control from the captain so as to prevent accident.

3)Relative reporting system should be established to encourage pilots provide written reports if they find any decision obviously endanger safe operation. Penalty regulations should be published for punishing those captains clinging on their action threatening safety.

References

[1] Aircraft Accident Report: Tenerife, Canary Islands, Mar. 3, 1977, National Transportation Safety Board, Washington, D.C., USA.

[2] Compilation of China Civil Aviation Accidents (7), General Administration of Civil Aviation of China, Beijing, 2007

[3] Tony Kern, Controlling Pilot Error(12): Culture, Environment, and CRM, New York：McGraw-Hill, 2002, ISBN: 0-07-138638-6

[4] Luo Xiaoli, Cockpit Resource Management, Southwest Jiaotong University Press, Chengdu，2002

[5] Statistical Summary of Commercial Jet Airplane Accidents: worldwide Operations 1959-2006, 2008-2-26, http://aviation-safety.net/statistics

[6] Daryl R. Smith, Controlling Pilot Error(4): Controlled Flight into Terrain, New York：McGraw-Hill, 2001, ISBN: 0-07-137411-6

[7] Tony Kern, Controlling Pilot Error(10): Approach and landing, New York：McGraw-Hill, 2001, ISBN: 0-07-137362-4

EMERGENCY AT 35’000 FT.:
HOW COCKPIT AND CABIN CREWS LEAD EACH OTHER TO SAFETY

Bienefeld, Nadine
ETH Zurich
&
Grote, Gudela
ETH Zurich

Many aircraft accidents have illustrated the catastrophic consequences of ineffective leadership. However, the optimal form of leadership during emergencies on board is not yet fully explored, particularly not with regards to its influence on decision making. Several authors have studied decision making errors in the cockpit, but to our knowledge so far, nobody has considered the role of the cabin crew, who in these stressful and challenging circumstances have to closely collaborate with pilots despite obvious differences in their training and culture. This study investigates the influence of collective leadership on the quality of decision making by observing 84 cockpit and cabin crews (N=504) live during a simulated emergency. Results indicate that collective leadership strongly correlates with the quality of the decision and crew performance. To conclude, we discuss the implications of those results for decision making in aviation and recommend changes in the design and content of CRM training.

Introduction

The importance of leadership in effective teamwork is acknowledged without controversy (see Yukl, 2006 for an overview) and leadership is even more relevant where evidently it matters most: in the face of life-threatening hazards and stressful situations as encountered during an emergency on board an aircraft (Baran & Scott, 2010; Yammarino, Mumford, Connelly, & Dionne, 2010).

Recent studies have shown that collective leadership, defined as an ongoing reciprocal interactions process among all team members regardless of their formal organisational rank or authority has positive effects on team performance. Baran & Scott (2010) for instance, have concluded from their observations of fire fighting teams, that collective leadership was most effective in dynamic, stressful and dangerous situations because one single hierarchical leader could not attend to all the required leadership tasks by him- or herself at once. Similar conclusions were found in medical action and anaesthesia teams where both formally assigned leaders and informal leaders fulfilled leadership tasks and thereby increased team performance (e.g. Klein, Ziegert, Knight, & Xiao, 2006; Künzle, Zala-Mezö, Wacker, Kolbe, & Grote, 2009; Xiao, Seagull, Mackenzie, & Klein, 2004; Yun, Faraj, & Henry, 2005).

Decision making is described as one of the most important leadership tasks and decision quality is often used as a direct outcome measure of the leadership process (Vroom & Jago, 1974; Yukl, 2006).

In the past two decades, we have learnt much about decision making under stress and the potential for human error (e.g. Klein, Orasanu, Calderwood, & Zsambok, 1993; Klein, 1993, 1997; Flin, Salas, Strub & Martin, 1997). Studies in the domain of aviation have revealed interesting results with regards to traps for errors in decision making particularly in emergency situations where pilots have to make vital decisions which determine the fate of everybody on board (e.g. Orasanu, Dismukes, & Fischer, 1993; Orasanu, 1994; Orasanu & Serfaty, 1996; Orasanu & Fischer, 1997; Orasanu & Lynne, 1998).

In this paper we will investigate decision making during a simulated emergency on board and the way in which leadership influences the quality of the decision as well as the overall crew performance.

Based on what we know about the positive effects of collective leadership described above, we recorded leadership behaviour in both formal leaders, i.e. captains and informal leaders, i.e. cabin crew members.

To our knowledge cabin crew members have not yet been included in studies on decision making or leadership despite the fact that they play a crucial role in an emergency and have to closely collaborate with pilots taking part in the decision making process. Several aircraft accidents have tragically illustrated what can go wrong when this collaboration fails. For instance in 1983, 23 people on board Air Canada flight 797 were killed after smoke and fire in the cabin became uncontrollable after a tardy and ineffective decision making process involving both pilots and cabin crew members. Due to incomplete information and misunderstandings, pilots had underestimated the risk posed by the smoke in the cabin whereupon they delayed the decision and lost vital time. Had the captain or in fact any other crew member on this eventful flight fulfilled some of the most basic leadership tasks such as maintaining an overview of the situation, supervising or correcting the actions of the other crew members, the decision effectiveness and hence the overall outcome could have been very different (TSBC, 1983).

For this study we have created a similar scenario in which cockpit and cabin crews, due to differing information and closed cockpit door, had to all contribute in order to reach the correct decision (Vroom & Jago, 1974). In that way we were able to observe 504 cockpit and cabin crew members live in a simulator with regards to their leadership behaviour, whereby we followed the functional team leadership theory (Zaccaro, Rittman, & Marks 2001) and classified leadership according to Yukl's leadership taxonomy (2006).

Hypotheses

We derive our assumption from the theory described above and postulate the following hypotheses:

H1: In crews with correct decision making, the overall crew performance is significantly higher than in crews with faulty decision making.

H2: In crews with correct decision making, the overall amount of leadership is significantly higher than in crews with faulty decision making.

H3: Only in crews with correct decision making does collective leadership, i.e. leadership demonstrated by formal leaders (captains = CMDs) as well as leadership shown by informal leaders (cabin crew members = CCMs) significantly predict crew performance.

Methods

Participants

A total of 504 cockpit and cabin crew members (84 crews) of a medium sized European airline voluntarily participated in this study, whereby participants had been chosen randomly depending on their flight schedule.

Flight experience of captains varied from nine to 37 years (M = 26.18, SD = 5.19) and from one year to a maximum of 31 years for first officers (M = 13.08, SD = 6.68). Cabin crew members had between 11 months and 37 years (M = 19.44, SD = 7.99) of professional experience. The age of the participants followed a normal distribution with an average of 50 years for captains, 39 years for first officers and 36 years for cabin crew members. There was a significant sub-group specific gender bias (cockpit crew 100% male and cabin crew predominantly female with 79% against 21% male). Anonymity of all participants was guaranteed.

Procedure

The observations took place during a simulated emergency exercise in the A320 cabin flight simulator which was part of the annual safety training day at the respective airline.

Each crew performed a standardized flight during which they had to deal with an onboard emergency of a critical nature while three trained observers recorded frequencies of various leadership behaviours as described below.

Directly after completion of the simulated flight two independent subject matter experts and safety instructors completed a team performance evaluation form and subjects were asked to fill in a questionnaire to collect team process variables, control variables and biographical information.

Following this all participants received a training oriented debriefing.

Apparatus

The Airbus A320 cabin flight simulator is a special form of high fidelity simulator equipped with a two-man cockpit and a fully furnished passenger cabin seating up to 20 passengers, thus creating a realistic environment for cross-disciplinary mixed team training.

Using hydraulic mechanisms various airplane movements can be simulated and the training of emergency situations is enabled through different manipulations such as system malfunctions, alarms, smoke etc.

To further increase face validity the cabin flight simulator is equipped with the original intercommunication system, cabin signs and emergency equipment. Realistic audio simulation and background noises facilitate further immersion into the experience.

Scenario

For the purpose of this study a specific 20 minutes long standardized scenario was created in which pilots, due to physical separation (closed cockpit door), had to base their decision making on the information they received from the cabin crew. The situation began with a normal course of flight which then developed into a critical emergency situation. While pilots had no indication of a problem in the cockpit throughout the entire exercise, smoke started to develop in the cabin, gradually intensifying. The scenario was developed by the first author together with two subject matter experts and validated by one training captain and seven safety instructors.

Face validity of the scenario was rated as high by 86% of all participants, indicating that they felt the scenario was realistic, that they acted accordingly and that they believed that a real crew would act in the same way.

Leadership

Leadership behaviour was coded by means of a leadership taxonomy based on Yukl's categorical system of 14 managerial position duties and responsibilities (Yukl, 2006) and recorded by means of event sampling in real time, using TrackVivo ©, a data sampling software (SmarTrack, 2009).

Observers and interrater agreement. The first author and three undergraduate psychology majors with current or past background as cockpit or cabin crew served as on-site observers.

To check the accuracy of coding 10% of data (10 crews) were recorded on video and double coded. Cohen's kappa for the different leadership codes ranged from $\kappa = .82$ (consulting others) to $\kappa = 1.0$ (delegate sth. to sb.), indicating excellent interrater reliability.

Performance measure

Performance was assessed by two trained safety instructors and subject matter experts.

For this purpose, a checklist based and time sensitive and weighted performance rating system was developed and validated using Delphi Technique (Clayton, 1997) this rating system was validated by 5 safety experts (experience > 10 years) over the course of three discussion rounds.

Decision making

The correctness of the decision was assessed by a safety instructor and subject matter expert and consisted of one dichotomous variable (Did the pilots reach the correct decision within the predefined time frame? Yes or No). If the correct decision was reached but the critical point in time had passed, the decision was rated as wrong because according to the given scenario, a successful landing would not have been possible anymore.

Orasanu & Lynne (1998) point out that in real life scenarios, there is often no clear standard of "correctness" and that the "best" decision may not be well defined. This is why we chose a scenario which was very simple in this respect. If the information (location, colour, density, development, smell of smoke) was correctly passed on to the pilots, the correct decision could be reached by means of recognition-primed-decision making (RPDM) (Klein, 1993). All subsequent actions were fully under control of the pilots and could be carried out in accordance with the corresponding emergency procedures. If the information received from the cabin was incomplete however, there was significant potential for misunderstanding.

Results

First the data was tested for potential influencing factors such as job experience, experience in a formal leadership position, age and gender. None of them showed significant effects on the statistical models.

All data collected from first officers was excluded from the analyses because interactions between cockpit and cabin crews took place between captains and cabin crew members exclusively.

To compare the leadership and overall performance in crews whose decision making was correct (N=63) with the performance in crews whose decision making was faulty (N=21), we computed two independent *t*-tests which revealed the following results:

On average, the overall performance in crews with correct decision making was higher ($M = 4.68$, $SE = .14$) than in crews with faulty decision making ($M = 3.71$, $SE = .37$). This difference was significant $t\,(82) = 2.88$, $p < .01$.

With regards to leadership we also found significant differences between the groups ($t\,(82) = 2.14$, $p = .03$) whereby crews who reached the correct decision demonstrated more leadership in total ($M = 5.86$, $SE = .53$) than crews who made a faulty decision ($M = 4.60$, $SE = .28$).

For the purpose of identifying the influence of leadership on performance in dependency of the quality of decision making, we split the data by the variable 'decision making' (correct vs. faulty) and computed a hierarchical regression model. As demonstrated in table 1, leadership of captains, entered as a first factor, was a significant predictor for performance in crews who made the correct decision ($B = -1.78$; $SEB = .92$, $\beta = -.22$, $p = .043$) but not in crews with faulty decision making ($B = -2.13$; $SEB = 2.29$, $\beta = -.29$, ns.). Similarly, leadership demonstrated by cabin crew members, entered as a second factor, significantly predicted performance ($B = 4.34$; $SEB = 1.62$, $\beta = .29$, $p = .009$), but again only in crews who reached the correct decision. However, this significant second factor effect reduced the influence of captains' leadership, making it statistically insignificant ($B = -1.55$; $SEB = 0.88$, $\beta = -.19$, $p = .08$). Contrary to our expectation, the interaction effects between predictors one and two were statistically insignificant.

Table 1.

Hierarchical regression model for the effect of leadership on crew performance in dependency of the quality of decision making (decision correct vs. decision wrong)

Decision correct
Crew performance

predictors	B	SEB	β	t	Sig.
Step 1					
(Constant)	21.11	5.12		4.12	.000
Leadership CMD	-1.78	0.92	-.22	-1.93	.043*
Step 2					
(Constant)	-2.44	10.09		-0.24	.810
Leadership CMD	-1.55	0.89	-.19	-1.75	.084
Leadership CCM	4.34	1.62	.29	2.67	.009**
(Constant)	-3.62	18.06		-0.20	.842
Leadership CMD	-1.33	3.00	-.16	-0.44	.659
Leadership CCM	4.58	3.52	.31	1.30	.196
LS CMD * LS CCM	-0.05	0.60	-.03	-.08	.937
N					84

Decision incorrect
Crew performance

predictors	B	SEB	β	t	Sig.
step 1					
(Constant)	11.17	14.23		0.78	.453
Leadership CMD	-2.13	2.30	-.30	-0.93	.376
step 2					
(Constant)	-12.34	16.04		-0.77	.464
Leadership CMD	-3.32	1.99	-.46	-1.66	.135
Leadership CCM	6.55	2.99	.61	2.19	.060
(Constant)	13.51	37.48		0.36	.729
Leadership CMD	-7.97	6.40	-1.11	-1.25	.253
Leadership CCM	.50	8.46	.05	0.06	.955
LS CMD*LS CCM	1.05	1.37	.99	0.77	.468
N					84

Note: CMD = Commander, CCM = Cabin Crew Member

Discussion

This study offers some new evidence to strengthen the notion that leadership plays a crucial role in the decision making process (Vroom & Jago, 1974; Yukl, 2006). Our results demonstrate that significantly more leadership was displayed during the decision making process in crews who reached the correct decision. Furthermore, leadership was a significant predictor for crew performance, but only in crews who had reached the appropriate decision. This effect was insignificant in crews with erroneous decision making.

The more interesting result however is that not only formal leadership (demonstrated by captains) correlated strongly with performance in crews with good decision making, but informal leadership (demonstrated by cabin crew members) correlated even more strongly with crew performance.

These findings contribute to the ongoing research on the effectiveness of collective leadership and go in line with what others (e.g. Baran & Scott, 2010; Klein et al., 2006; Künzle et al., Xiao et al., 2004) have observed in similarly structured teams, implicating a change in the traditional leadership paradigm where leadership is seen as centralized within one single hierarchical leader (see Yukl, 2006 for an overview).

The call for proactive participation and support in the leadership process by informal leaders however will not be adequately answered if crew members lack the necessary knowledge, skills and attitudes. Should the active participation of so called ‘followers’ in the decision making and leadership process continue to demonstrate positive effects on performance, the implications for training would be great in that every crew member would have to be trained to effectively execute some of the necessary leadership tasks. Before that is the case though, more research on the subject is needed and potential negative effects (e.g. competing leaders creating chaos and diffusion of responsibility) need to be considered.

What we suggest instead is that cockpit and cabin crews should each act as a system of redundancy within and between themselves in correspondence with the credo of ‘Crew Resource Management’ (CRM) in that all available resources of the crew should be used for the purpose of a safe and efficient flight operation (e.g. Helmreich, et al., 1999).

Specifically we propose that the importance of leadership and decision making be addressed in CRM-training involving both cockpit and cabin crews. All crew members should be given the opportunity to train the skills and behaviours that are needed for effective interdisciplinary collaboration in an emergency by having to interact with each other during task-oriented, practical training sessions using realistic scenarios.

Limitations

The sample was drawn from only one airline and in only one industry. Although this controls for the influence of contextual issues, it raises the issue that the results may not generalize to other organizational contexts. Given the structural similarity of teams working in other high risk environments such as medicine, policing or fire fighting, we would argue however, that these findings are transferable and may have important implications also for their training.

Acknowledgements

We gratefully acknowledge the financial support by the ETH Research Grant for independent researchers and kindly thank Swiss International Air Lines Ltd. and Swiss Aviation Training Ltd. for their cooperation and support.

References

Baer, M., & Frese, M. (2003). Innovation is not enough: climates for initiative and psychological safety, process innovations, and firm performance. *Journal of Organizational Behavior, 24*(1), 45-68.

Baran, B. E., & Scott, C. W. (2010). Organizing Ambiguity: A Grounded Theory of Leadership and Sensemaking Within Dangerous Contexts. *Military Psychology, 22*(1 supp 1), 42 - 69.

Bass, B. M. (1990). *Bass and Stogdill's Handbook of Leadership* (3rd ed.). New York: Free Press.

Bureau of Transportation Statistics (2010). Transportation accidents and deaths per billion passenger kilometres in the U.S., from http://www.bts.gov/

Campbell, D. J., Hannah, S. T., & Matthews, M. D. (2010). Leadership in Military and Other Dangerous Contexts: Introduction to the Special Topic Issue. *Military Psychology, 22*(1 supp 1), 1 - 14.

Chen, G., & Bliese, P. D. (2002). The Role of Different Levels of Leadership in Predicting Self- and Collective Efficacy: Evidence for Discontinuity. *Journal of Applied Psychology, 87*(3), 549-556.

Clayton, M. J. (1997). Delphi: a technique to harness expert opinion for critical decision-making tasks in education. *Educational Psychology, 17*(4), 373 - 386.

Cohen, M.S. (1993). The naturalistic basis of decision biases. In G. Klein, J . Orasanu, R. Calderwood & C. Zsambok (Eds.), Decision making in action: Models and methods (pp. 51-99). Norwood, NJ: Ablex.

Cohen, M.S., Freeman, J.T. & Thompson, B. (1998). Critical thinking skills in tactical decision making: A model and a training strategy. J.A. Cannon-Bowers & E. Salas (eds.), *Making decisions under stress* (pp. 155-189). Washington, DC: APA.

Edmondson, A. C. (2003). Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *Journal of Management Studies, 40*, 1419-1452.

Flin R, Salas E, Strub M, Martin L. Decision Making Under Stress. Aldershot: Ashgate, 1997.

Flin, R. M., L. Goeters, K. Hörmann, H.J. Amalberti, R. Valot, C. Nijhuis, H. (1998). Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills. In D. H. a. H. C. Muir (Ed.), Contemporary Issues in Human Factors and Aviation Safety Aldershot: Ashgate.

Ginnett, R. C. (1993). Crews as groups: Their formation and their leadership. In E. L. Wiener, B. G. Kanki & R. L. Helmreich (Eds.), *Cockpit resource mangement* (pp. 71-98). San Diego, CA: Academic Press.

Gladstein, D. L., & Reilly, N. P. (1985). Group Decision Making under Threat: The Tycoon Game. *The Academy of Management Journal, 28*(3), 613-627.

Hackman, J. R. (1993). Team, leaders, and organizations: New directions for crew-oriented flight training. In E. L. Wiener, B. G Kanki & R. L. Helmreich (Eds.), *Cockpit resource mangement* (pp. 47-69). San Diego, CA: Academic Press.

Hannah, S. T., Uhl-Bien, M., Avolio, B. J., & Cavarretta, F. L. (2009). A framework for examining leadership in extreme contexts. *The Leadership Quarterly, 20*(6), 897-919.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica, 47*(1), 153-161.

Helmreich, R. L. (2000). On Error Management: Lessons from Aviation. *BMJ: British Medical Journal, 320*(7237), 781-785.

Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The Evolution of Crew Resource Management Training in Commercial Aviation. *The International Journal of Aviation Psychology, 9*(1), 19 - 32.

International Air Transport Association (2009). 2009 Traffic results.

Isenberg, D. J. (1981). Some effects of time-pressure on vertical structure and decision-making accuracy in small groups. *Organizational Behavior and Human Performance, 27*(1), 119-134.

Jones, D. G., & Endsley, M. R. (1996). Sources of situation awareness errors in aviation. *Aviation, space, and environmental medicine, 67*(6), 507-512.

Klein, G. (1993). A recognition primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. Zsambok (Eds.), Decision making in action:

Models and methods (pp. 138-147). Norwood, NJ: Ablex.

Klein, G. (1997). The current status of the naturalistic decision making framework. In R. Flin, E. Salas, M. Strub & L. Martin (Eds.) *Decision Making Under Stress: Emerging Themes and Applications.* Aldershot, UK: Ashgate.

Klein, G., Orasanu, J., Calderwood, R. & Zsambok, C. (Eds.) (1993). *Decision making in action: Models and methods.* Norwood, NJ: Ablex.

Klein, K. J., Ziegert, J. C., Knight, A. P., & Xiao, Y. (2006). Dynamic delegation: Shared, hierarchical, and deindividualized leadership in extreme action teams. *Administrative Science Quarterly, 51*, 590–621.

Künzle, B., Zala-Mezö, E., Wacker, J., Kolbe, M., & Grote, G. (2009). Leadership in anaesthesia teams: The most effective leadership is shared.

Nembhard, I. M., & Edmondson, A. C. (2006). Making it safe: the effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *Journal of Organizational Behavior, 27*(7), 941-966.

Orasanu, J., Dismukes, K. & Fischer, U. (1993). Decision errors in the cockpit. *In Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, 1,* 363-367. Santa Monica, CA: Human Factors and Ergonomics Society.

Orasanu, J. (1994). Shared problem models and flight crew performance. *Aviation psychology in practice.* Aldershot, United Kingdom.

Orasanu, J. & Serfaty, D.(1996). Distributed Team Decision Making: Understanding the Whole as well as the Parts, *Human Factors and Ergonomics Society Annual Meeting Proceedings, Cognitive Engineering and Decision Making (4),* 158-161.

Orasanu, J. & Fischer, U. (1997). Finding decisions in natural environments: The view from the cockpit. In C. Zsambok & G. Klein (Eds.) *Naturalistic Decision Making.* (pp. 343-357). Hillsdale, NJ: Lawrence Erlbaum Associates.

Orasanu, J. & Lynne, M. (1998). Errors in Aviation Decision Making: A Factor in Accidents and Incidents. *HESSD,* 100-107.

Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It Is Not How Much You Have but How You Use It: Toward a Rational Use of Simulation to Support Aviation Training. *The International Journal of Aviation Psychology, 8*(3), 197 - 208.

Scott, C. W., & Trethewey, A. (2008). Organizational Discourse and the Appraisal of Occupational Hazards: Interpretive Repertoires, Heedful Interrelating, and Identity at Work. *Journal of Applied Communication Research, 36*(3), 298 - 317.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004). Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality. *Risk Analysis, 24*(2), 311-322.

Small, S. D., Wuerz, R. C., Simon, R., Shapiro, N., Conn, A., & Setnik, G. (1999). Demonstration of High-fidelity Simulation Team Training for Emergency Medicine. *Academic Emergency Medicine, 6*(4), 312-323.

SmarTrack (2009). TrackVivo behavioural observation system for live recording of concurrent events in complex behavioural processes. , from http://www.smartrack.co.uk/

Transportation Safety Board of Canada (TSBC) (1983). Accident investigation report of Air Canada 797.

Tschan, F., Semmer, N. K., Gautschi, D., Hunziker, P., Spychiger, M., & Marsch, S. U. (2006). Leading to recovery: Group performance and coordinative activities in medical emergency driven groups. *Human Performance, 19*, 277-304.

Vroom, V. H. & Jago, A. G. (1974). Decision making as a social process: normative and descriptive models of leader behavior. Decision Sciences, 5, 743-769.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science, 16*(4), 409-421.

Xiao, Y., Seagull, F. J., Mackenzie, C. F., & Klein, K. (2004). Adaptive leadership in trauma resuscitation teams: A grounded theory approach to video analysis. *Cognition, Technology and Work, 6*, 158-164.

Yammarino, F. J., Mumford, M. D., Connelly, M. S., & Dionne, S. D. (2010). Leadership and Team Dynamics for Dangerous Military Contexts. *Military Psychology, 22*(1 supp 1), 15 - 41.

Yukl, G. (2006). *Leadership in organizations* (5 ed.). Upper Saddle River, NJ: Prentice Hall.

Yun, S., Faraj, S., & Henry, P. S. J. (2005). Contingent leadership and effectiveness of trauma resuscitation teams. *Journal of Applied Psychology, 90*(6), 1288-1296.

Zaccaro, S. J., & Marks, M. A. (1999). The roles of leaders in high-performance teams. In E. S. Associates (Ed.), *Supporting work team effectiveness: Best management practices for fostering high performance* (pp. 95-125). San Francisco: Jossey-Bass.

Zaccaro, S. J., Rittman, A. L., & Marks, M. A. (2001). Team leadership. *The Leadership Quaterly, 12*, 451-483.

BEST CAPTAINS – A SURVEY OF CHARACTERISTICS AND SKILLS OF AIRLINE CAPTAIN EXCELLENCE

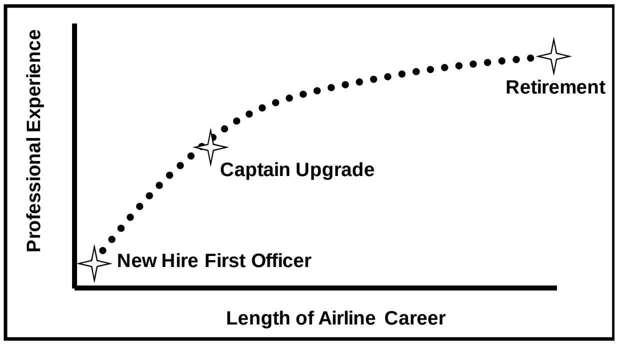
Steve Swauger
Southwest Airlines Pilots' Association
Chandler, Arizona, USA

One training development challenge in the air carrier industry today is the lack of sharing of best practices between pilots who hold the same seat position. Part of the difficulty is identifying traits, techniques, and practices of the best captains. To identify top-performing captains, this survey asked first officers to identify the best captains that they had flown with and to identify the characteristics that earned them that distinction. What emerged was a consistent description of the ideal captain – one who is technically competent, psychologically confident, and who promotes good CRM values. The best captains were then resurveyed to collect and share their best practices on a variety of traits, skills, and values.

Introduction

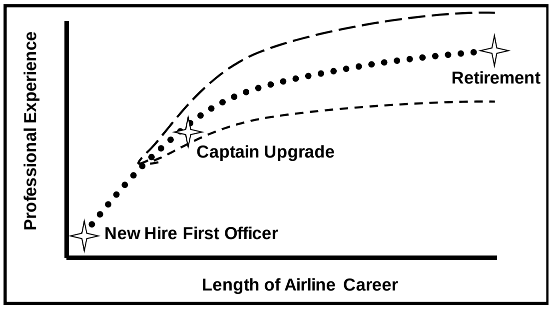
Much has been studied about the early development of pilot skills. Less attention has been devoted toward mature pilots who have arrived at the pinnacle of their career. Having spent many years striving to reach the top, what exactly happens during the many years of flying *after* they reach this milestone? Consider the following graph of a nominal air carrier pilot's career.

Graph 1
Average pilot – career professional experience growth



We reasonably assume that the average pilot demonstrates positive growth of professional knowledge and experience as he or she encounters learning events. As depicted by the dotted line, the growth is steep during the early years when learning events are plentiful and new, but begins to decrease as the pilot's experience grows – limited solely by exposure to new information or events. As pilots near retirement age, the depth of professional knowledge and experience reaches their personal peak.

Graph 2
Range of professional experience growth



On an individual scale, the growth of professional knowledge and experience varies between pilots. While the airline has strong control over the training and performance of new-hire first officers and upgraded captains, personal professional growth beyond this point is dependent largely on individual skill and motivation. Pilots who lack the personal learning skills or drive to continue their professional growth effectively stagnate (as depicted by the lower narrow-dashed line). Captains depicted by this line are not substandard performers. The depth of their knowledge and experience just stops growing. While their growth continues because of their daily exposure to more events and conditions, their assimilation and integration of this new information is often inefficient – resulting in very slow growth . There are two possible reasons why this occurs. First, many of us display a basic human tendency to reduce our level of effort and ambition following achievement of our goals. Having achieved the highest rung on their career ladder, there are no higher steps to achieve. The energy and focus that these pilots used to reach this career goal is often redirected toward other personal ventures, such as hobbies or side businesses. Second, there is very little outside stimulus to grow professionally as an aviator after reaching the top. FAA certification requirements focus on checking performance standards – accurately termed *performance minimums.* As the old adage goes, "if you aim for the minimum, you'll hit it every time." Those minimums do not appreciably change over a pilot's career. Thus, pilots who choose to coast or stagnate for the many remaining years of their careers are empowered to do exactly that.

On the upper extreme of graph 2, some pilots continue a steep growth and learning curve, continuing to assimilate and integrate new knowledge to build their aviation wisdom (shown by the upper wide-dashed line). Their growth only tapers off as they reach the practical limits of exposure to new learning events and conditions. These pilots continue to seek and achieve personal growth even after reaching the captain seat.

What we see is a fairly wide range of potential growth throughout an average airline career, a range that widens after reaching the captain's seat. Clearly, something very significant is happening between captains displaying high-end performance and low-end performance. This is the motivating question behind this study – what are the traits and characteristics demonstrated by the best captains.

Method

Objective

The objective of this study was to identify the characteristics, motivations, and techniques held by the captains who demonstrate high grow in professional experience and knowledge throughout their careers.

Survey Group

This study was conducted at a large passenger airline operating in the United States. All pilots were airline transport pilot (ATP) certified and type-rated in the Boeing 737. All pilots were qualified on all company aircraft. They were assigned to 7 crew bases for scheduling purposes. While pilots voluntarily move between crew bases for upgrade opportunities and personal reasons, the norm is for them to remain at a single crew base – often for their entire careers. So, captains within a crew base would typically fly with first officers from that crew base. Most pilots bid for a month-long flight schedule consisting of pairings lasting one to four days, with a three-day pairing being the statistical mode. Initial flight schedules are awarded in seniority order with a single captain and a single first officer assigned to all pairings for the month. While many pilots shuffle individual pairings to fulfill personal preferences, the same two pilots normally remain together for at least 2-4 days.

The first challenge of this study was identifying the best captains. Fellow captains, supervisory check airmen, and trained observers were deemed unsuitable because of their lack of opportunity to observe <u>all</u> the available captains across many flights. The most obvious survey group was the population of first officers. This was because first officers had the most opportunities to work with the most captains for multi-day pairings. At the time of the survey, the subject airline was flying about 3000 flights per day. In an average year, a first officer flies at least one multi-day pairing (6-20 flights) with 20-50 different captains – almost all from his or her crew base. Additionally, all first officers had served as captains, aircraft commanders, or instructors in their previous flying jobs, so their ability to identify desirable traits was well-established. Thus, a core assumption of this study is that first officers, individually and as a group, have the ability to identify the high-performance captains and that they can identify the character traits that set these captains apart.

First Officer Survey

The entire population of over 2500 first officers were given the opportunity to voluntarily and anonymously complete a single-page survey. The pilots were incentivized to receive a positive-space travel pass (essentially free tickets that they could use or give to anyone). The passes were awarded randomly to the survey respondents. Note: The surveys had a code number that identified the respondent's relative crew base number – only used to identify the drawing winners. The survey asked the first officers for two things. First, "list the names/employee number/domicile of 1, 2 or 3 Captains that you consider exemplify the finest *<u>line operational</u>* skills and practices at [the airline]. Please consider the following areas: flight management, flying skills, situational awareness, open communications, team building, CRM environment, conflict resolution, decision making, instructional ability, and mentoring." This listing of characteristics was included to focus the ratings on traits and characteristics instead of personal popularity. Second, they were asked to "identify some of the specific outstanding qualities that these Pilots demonstrate. What makes them rise to the top of your list."

The first officers were asked to complete some basic demographic data of their relative seniority within their crew base. This identified how long they had been a first officer and by extrapolation, how many different captains they had flown with over the years.

Results

Data Validation

The data was extracted in two ways. First, list by name and number-of-votes was tabulated for the best captains at each crew base. Second, a list of characteristics and attributes was collected across the entire population. Naturally, the most credible first officers were the ones who had flown with the most captains within their crew base. Demographic data was analyzed to ensure that newer first officers (ones who had flown with fewer captains) did not skew the totals away from selections made by more-experienced first officers (ones who had flown with the most captains). While no statistical tests were conducted to ensure consistency, the data appeared to be quite even across first officer groups. The only trait that seemed to show a difference was instructional ability, with newer first officers listing instructional ability much more often than experienced first officers. This is probably because the best captains found fewer opportunities to teach job skills and knowledge to experienced first officers than to new-hires. With that one exception, all other characteristics seemed to be universally distributed throughout the crew bases and between first officer experience levels. The consistency of the data supports the assertion that first officers, regardless of crew base, value the same traits and characteristics in the best captains.

Best Captains Lists

Over 380 captains (out of about 3000) received one or more votes (with the vast majority receiving only one vote). In almost every crew base, one or two names got multiple votes – in one case as high as six. The newest crew base did not have anyone with three or more votes. This was understandable since the crew base was recently formed with pilots drawn from the other 6 bases and insufficient time had elapsed to develop a consensus of who their best captains were. At the remaining 6 bases, any captain receiving 3 or more votes was selected for further survey in the second half of this study.

Best Captains Characteristics

Recall that the first officers were also asked to "identify some of the specific outstanding qualities that these Pilots demonstrate." They were asked to "consider the following areas: flight management, flying skills, situational awareness, open communications, team building, CRM environment, conflict resolution, decision making, instructional ability, and mentoring." By percentage, the reported characteristics broke out as follows:

Table 1

Reported characteristics of the best captains

CRM environment	13.8%	First tier
Personality *	10.7%	
Technical Expertise *	10.3%	
Instructional ability	9.1%	Second tier
Flying skills	8.8%	
Easy-going nature *	8.3%	
Professional	8.2%	
Team building	7.1%	Third tier
Open communications	5.8%	
Company asset *	4.6%	
Flight management	3.1%	
Sense of humor *	3.0%	

Mentoring	2.7%	Fourth tier
Situational awareness	2.1%	
Decision making	1.5%	
Experience *	1.0%	
Conflict resolution	0.0%	

The first officers were not limited by how many or few characteristics they could chose to list. Some provided simple listings, while others provided narratives. Specific characteristics were extracted in these cases.

Listed Characteristics

As stated before, each first officer survey sheet listed nine example characteristics (shown in Table 1- entries displayed with normal font). No effort was made to define these terms, but among professional aviators, there is general consensus in the meanings. Promoting a good CRM environment was the top characteristic with 13.8% of the votes and conflict resolution, didn't receive any votes.

Unsolicited Characteristics and Traits

Of significance is that the first officers added six additional characteristics (shown in italics font and with an asterisk in Table 1). Two of these, personality and technical expertise, ranked 2nd and 3rd overall. Easy-going nature, company asset, and sense of humor ranked in the middle.

Follow-on Best Captains survey

The identified best captains were then asked to complete a multi-page questionnaire. The questionnaire asked them to provide narrative perspectives to 8 trait-areas that the first officers seemed to value most highly: first impressions, open communications, team building, instructing/mentoring, personality, professionalism, and being deliberate/predictable. The captains were also asked to comment on how they would handle three situational line-flying problems. Finally, they were offered a chance to comment on an open-ended essay about their perceived role and goals as a captain. The results were compiled and shared with the entire pilot group through a series of articles in the union safety publication.

Discussion

Best Captains

Clearly, this experimental design did not yield a comprehensive list of best captains, nor was it intended to. Many extremely-talented newer captains were undercounted because they had flown with fewer first officers than high-timers. The survey tended to favor longer-serving captains who had flown with many first officers within a single crew base. Additionally, there was a concern that training check airmen would be over-counted because of their higher exposure to new-hire first officers. While several check airmen did make the list, their overall numbers were close to their relative population percentage. Of significance is that the best captains seemed to be cut from the same cloth. Regardless of crew base or relative seniority, they each seemed to hold the same values and approach their role as a captain in similar ways. In the series of articles that shared their actual questionnaire narratives, it was clear that, while there were differences in personality and technique, all the best captains held the same core values and professional growth ethic. These articles served as a way to spread the perspectives and practices of the best captains throughout the pilot group. These articles were very well received.

Characteristics:

Recall that the first officers were offered a list of 9 characteristics to guide their best captain selection. It is extremely significant that they generated 6 additional characteristics and that two of these, personality and technical expertise, rated in the top three – both first tier characteristics. That their collective group consciousness generated these categories and that they rated them so highly speaks to the high value that first officers placed on these traits.

Also significant, many of the buzz-word characteristics that we write about in the industry, like situational awareness, decision making, flight management, and mentoring ranked surprisingly low. Arguably, many of these do comprise components of the higher rated traits. So, perhaps these results only indicate the first officers' desire to identify more-descriptive traits over generalized traits. As general trend conclusions, technical expertise was valued over experience, flight management, situational awareness, and decision making; CRM environment was valued over open communications, team building, and conflict resolution; and personality was valued highly in all measures.

The Ideal Captain:

We should note that since the overwhelming majority of airline operations are routine, the results of this study paint the picture of the ideal, every day, normal-operations captain. As a single sentence summary of the results, first officers valued captains who were technically competent, psychologically confident and promote an effective CRM environment. What emerges is an image of a captain who creates and promotes a good work environment, knows the job, and is relaxed and comfortable in his or her own skin.

It is safe to extrapolate these results into emergency environments as well. What more can we ask of captains faced with severe emergencies than to calmly and confidently direct the efforts of the flight team to accurately handle the emergency?

Additionally, this study was completed at a single airline with a long and unique cultural history. While results may vary with other airlines, it is reasonable to assume that the characteristics and techniques discovered here can apply elsewhere.

Finally, while we cannot assume that we identified all the traits that typify high-end captains from Graph 2, we can safely conclude that this study did accurately identify a consistent subset that was valued by first officers – and certainly offer good role model traits to promote throughout the airline.

Best Captain Value Sharing:

The purpose of this study was to determine those characteristics that distinguish the high professional growth captains and to share them throughout the pilot group. To date, seven articles have been published in the pilots' union safety newspaper. Following is a selection of direct comments from the first officers. They further paint a picture of the ideal best captain.

professional competence: outstanding aviators who knew the procedures and the aircraft.

- "Knows how to fly by the book with excellence"
- "Phenomenal stick skills – passengers think they're riding a magic carpet"
- "Outstanding cockpit management"
- "Tremendous assets to the airline"
- "Excellent mission management"
- "Deliberate in actions and pace"
- "Knows the book cold"
- "Super situational awareness – flies a smooth airplane and keeps the passengers comfortable"
- "Professional approach to skilled piloting"
- "PhD in Boxology (Note: The onboard flight management computer is called "the Box".)"
- "The most standardized captain I've ever flown with"
- "The consummate professional pilot"
- "Takes the time to project cause-and-effect in everything"
- "Stays ahead of a developing situation and builds plans and contingencies"
- "Smoothest hands on the tiller and yoke of anyone I have flown with"

personality: easy-going and relaxed personality.

- "Calm experience" (two words that say so much)
- "Puts customer, crew, and others ahead of his own agenda"
- "Natural leader"
- "Laid back atmosphere"
- "Courteous and considerate"
- "Very proactive and an outstanding leader in the jet"

- “Super professional”
- “Egos don’t get in the way of their flying”
- “Grace under pressure”
- “Extremely personable and friendly – they enjoy getting to know you”
- “Fun, fun, fun to fly with”
- “Omits personal bias”
- “Relaxed demeanor is soothing in a stressful situation”
- “Always willing to give a helping hand”
- “No matter what phase of flight, he can make you laugh”
- “Patient and allows everyone to get the job done right”
- “I’d fly every trip for the rest of my career with any one of these captains”
- “They showed great personal integrity”
- “Attitude and professionalism was contagious”
- “They know their stuff, but show humility”
- “Excellent balance of professional knowledge, even-keeled personality, and airmanship”
- “Too much fun! Got banned from flying together by the chief pilot”
- “Not afraid to make tough decisions”

team building: made team building a high priority from the first moment everyone met them.

- “Exemplifies the golden rule”
- “His briefing at the beginning of the pairing is the best I’ve heard”
- “Excellent and open communicator”
- “Makes the pairing fun as well as a learning experience”
- “Goes the extra mile with not only first officers, but flight attendants the whole team”
- “Takes input with respect and grace, even if it’s wrong”
- “Promotes the environment of teamwork the moment you meet him”
- “Takes time to communicate – LISTEN!”
- “Not in a rush mentality”
- “Effectively integrates all inputs to make good decisions”
- “Looks after the whole crew – offers to buy food between flights”
- “Flying the aircraft is a team effort and inputs are truly desired”
- “You really want to do the best for him because you don’t want to let him down”
- “Exudes humble professionalism”
- “They want input, but these people are outstanding airmen and I rarely observe any errors”
- “Makes sure the entire crew is in-the-loop”
- “Keeps me informed of his intentions”
- “Creates a cockpit atmosphere that allows any first officer to do his best”
- “Care about the success of each individual and the company”
- “Treats other SWA employees with dignity and respect – buys McDonalds for the rampers!”
- “Stayed with me in the hospital all night while I was waiting for appendectomy”
- “No one does a better job creating a relaxed environment”
- “Gets to know all the SWA people he can (thus, more aware of whole operation)”
- “They are rapport builders”
- “Always gives a thorough briefing to the incoming crew”

EXAMINING THE VALIDITY OF TRADITIONAL RISKY FLIGHT BEHAVIORAL MEASURES ACROSS A VARIETY OF RISKY FLYING ACTIVITIES

Mr. Justin Drinkwater
Dr. Brett Molesworth
The University of New South Wales
Sydney, Australia

Predicting pilots' willingness to engage in a variety of risky activities has implications for the selection and training of pilots (Drinkwater & Molesworth, 2010). In addition to traditional predictors of safety such as flight experience and age, a variety of measures have been employed that examine pilots' attitudes and risk perceptions (Hunter, 2002). However, in order to test their predictive validity, they are often paired with a single behavioral measure, nominally a simulated flight with a stable level of risk, potentially limiting their ability to predict pilots' risk management behavior accurately. Therefore, the aim of the present study was to examine the stability of these predictors across a variety of risky flying activities. The results revealed risk perception to be the only reliable predictor of pilots' risk management behavior, suggesting that the traditional measures of risky flight behavior may require revision to ensure their efficacy.

Predicting pilots' willingness to engage in a variety of risky activities has implications for the selection and training of pilots (Drinkwater & Molesworth, 2010). In the literature, attitude (Cox & Cox, 1991; Lund, & Rundmo, 2009; Hunter, 2005; Molesworth & Chang, 2009), age (Vroom & Pahl, 1971; Retting et al., 1999), risk perception (Hunter, 2006), and risk tolerance (Hunter, 2002) are samples of the variety of factors that have been linked to risky behavior.

It is thought that risk perceptions, attitudes, and experience influence risk-taking behavior in aviation (Molesworth & Chang, 2009). Indeed, previous research has identified both risk perception (Hunter, 2006; Molesworth & Chang, 2009) and attitude (attitude towards safe flight operation - Hunter, 2005; and attitude towards low altitude flight - Molesworth & Chang, 2009) as predictors of pilots' risky flight behavior. It is generally thought that attitudes affect behavior in the intuitive, or 'positive' direction, that is, a more conservative attitude would lead to a more conservative behavior, and a less conservative attitude, to a less conservative (or riskier) behavior. This expectation arises primarily from the Theory of Planned Behavior (see for example, Ajzen, 1991; Ajzen & Fishbein, 2005; Armitage & Conner, 2001).

In a similar fashion, in line with hypotheses previously put forward in the risk perception literature (see Brewer, Weinstein, Cuite, & Herrington Jr., 2004), risk perception is expected to affect behavior in line with its strength and direction. That is, if one perceives a risk to their safety (or other valuable asset), then it is expected that the individual will modify their behavior to minimise risk.

An underlying assumption with the research in this area, particularly within the general aviation research is that predictors of risky flight behavior are stable across a variety of risky situations. However, this assumption remains untested. Therefore, the aim of the present research was to examine whether the predictors of risky flight behavior are stable across a range of flights from low to high risk. This involved examining the relationship between attitudes, risk perceptions and experiential variables (i.e., age, flight experience, and recent flight time) of Australian General Aviation (GA) pilots and their self-reported behavior. Specifically, the following research was designed to answer the following two questions

Research Questions

1. What is the relationship between the attitudes, risk perception, experiential factors and the behavior of pilots in the general aviation?

2. Are known predictors of risky flight behavior in general aviation contextually sensitive?

Experimental design

The study consisted of a single session, in which participants completed a battery of pen-and-paper surveys. The study was designed to examine the relationship between attitudes towards aviation safety (using Hunter's ASAS, see Hunter, 1995), risk perceptions of aviation-centric situations (using Hunter's Risk Perception Scales 1 and 2, see Hunter, 2005), experiential variables (e.g., age, and flying experience in terms of flight hours), self-reported risk ratings and self-reported risky flight behavior. Self-reported behavior and the reported risk ratings were gathered for three hypothetical flight scenarios. The three scenarios as described below were selected from a total of nine by seven Subject Matter Experts - SMEs (i.e., senior instructors in general aviation). Specifically, SMEs were asked to rate the level of risk in each scenario, with the anchor points being 0 – no risk and 100 – high risk, with the likelihood of death being high. Based on mean ratings, three scenarios were selected at three points on the risk rating, namely low, medium and high. However, prior to determining these scenarios it was important to ensure SME rated the scenarios similarly. The results of inter-rater agreement analysis utilising an intra-class correlation coefficient illustrated good agreement between the raters with respect to the ratings given to the scenarios, R(6) = .875.

The scenarios used in this experiment were the ''Hunter Valley'' Scenario, the "Ferry to LAME" scenario, and the "Moruya" Scenario.

The first scenario, rated as the lowest risk by the expert group was the "Hunter Valley" scenario. In this scenario, pilots were told that they had planned a flight from Camden to Cessnock (approximately 1 hour flight time) in fine weather, with no time pressures. There were appropriate back-up plans in case the weather turned bad or the aircraft was not performing suitably.

The second scenario presented, which was rated as medium risk by the expert group was the "Ferry to LAME" scenario, in which pilots were being asked to ferry their friend's aircraft to an airport approximately one hour away in order that maintenance be performed. The aircraft was very near to its allowable flight time before maintenance was required by regulations, there was poor weather fast approaching the destination airport, and the en-route weather was such that the cloud base was at 1,000 feet above ground level. Therefore, the pilot was under pressure to 'beat' the weather and fly lower than normal in order to get the aircraft to its maintenance appointment.

The final scenario, rated as high risk by the expert group was the "Moruya" scenario. In this scenario participants were asked to fly an aircraft with critically low fuel (23 minutes of fuel remaining, including reserves) to search for a skydiver that had landed away from the normal landing zone for the local skydivers. Whilst not required of the pilots explicitly, search operations may necessitate low flight which would be an additional hazard.

The three distinct levels of risk were used in the experiment for two reasons. The first was to examine pilots' different behavioral responses to the three risk scenarios, while the second was to examine if predictors of these behavioral responses varied depending on the level of risk in each scenario (i.e., low, medium and high-risk). The design also allowed the scenarios presented to pilots to be multi-dimensional with regards to the number and type of risks present. This is as opposed to utilising a uni-dimensional design, in which a single risk is present as is the norm in the literature (see Goh & Wiegmann, 2001; Molesworth, Wiggins, & O'Hare, 2006; O'Hare & Smitheram, 1995). The desired advantage of utilising scenarios that feature multi-dimensional risk is that this more closely resembles the operational environment as hazards are rarely found in isolation. It is hoped that the findings of a study that is grounded in more realistic assumptions and scenarios may be more representative of the situation with regards to the predictors of risk management behavior in the industry.

The flight scenarios were designed such that pilots had the ability to choose and report their actions, rather than only rating the relative risk of the scenario on a scale of 1 to 100, as in Hunter's risk perception scales. In this way, the experiment was designed to be more like a flight simulation than a risk-rating survey as pilots were asked to make the same decisions as they would be forced to make in a simulation (or real aircraft), but they are not asked to undertake the physical act of operating the aircraft.

Since the main aim of the study was to determine if pilots would undertake such flights, the dependent variable under examination was pilots decision to 'Go' (choosing to undertake the flight) or 'No-Go' (choosing not to undertaken the flight). This is opposed to the options available for a flight simulator based experiment in which continuous variables such as total time flown, altitude flown, or speeds reached could be utilised.

Method

Participants

Thirty-eight participants were recruited from flying schools and institutions located in the greater Sydney Basin, NSW Australia. The mean age of the participants was 27.03 years (SD = 14.90), the mean flight experience was 599.16 hours (SD = 2102.67) and the mean total of hours in the past 90 days was 28.56 (SD = 32.95 hours).

Materials and Stimuli

The material consisted of: a demographics questionnaire, ASAS questionnaire (Hunter, 1995), Risk Perception 1 and 2 questionnaires (Hunter, 2002), followed by the three flight scenarios. The completed surveys were collected from the participants directly. All data gathered was entered into SPSS v. 13 for Macintosh. The material and all stimuli were approved in advance from the University of New South Wales Ethics Panel.

Procedure

Participants were informed about the research through two methods: a personal brief given by the researcher and/or an advertisement poster placed on trainee organisations' notice boards. Participants interested in completing the survey required for the research arranged a mutually suitable time in which to undertake the work. Participants were asked to complete the five questionnaires in the following order, the demographic survey, Hunter's Risk perception scale 1 and 2, Hunter's ASAS, and finally the flight scenarios. At the conclusion of the study participants were thanked for their time.

Results

Data Analysis

The main objective of the experiment was to determine the relationship between pilot demographic measures, attitudes, risk perceptions, and the self-reported behavior of pilots. In order to investigate this, a series of Spearman's ρ correlations were employed. This test was used as the self-report behavioral data was nominal data. With alpha set at .05, a series of correlational analyses revealed only four relationships, all within the medium risk scenario (the Ferry to LAME scenario). Specifically, a positive relationship was found between the decision to go and the risk perception factors of Delayed Risk $r(38) = 0.36$, $p = .03$, General Flight Risk $r(38) = 0.33$, $p = .04$, High Risk $r(38) = 0.41$, $p = .01$, and Altitude Risk $r(38) = 0.41$, $p = .01$. No other relationships were noted, (largest $r = -.29$, $p = .08$). The positive direction of these relationships indicate that the pilots that chose to go (coded as one in the analysis) displayed a lower level of risk perception on these factors than did those pilots that elected not to fly (coded as two; see Table 1.).

Two pilots chose not to complete the Hunter Valley scenario (low risk flight), while thirty-five pilots chose not to complete the Ferry to LAME flight (medium risk) and the Moruya flight (high risk).

Table 1.

Correlations Between Behavior and Risk Perception, Attitude & Experiential Variables

Scenario (Expert Risk Level)	Age	Total Hours	Recent Hours	Delayed Risk	Nominal Risk	Immediate High Risk	General Flight Risk	High Risk	Altitude Risk	Driving Risk	Everyday Risk	Self Confidence	Risk Orientation	Safety Orientation
Hunter (Low)	.08	-.17	-.24	.11	.22	-.15	-.08	-.02	.06	-.09	-.05	-.12	.23	-.10
Ferry (Medium)	.04	-.11	-.23	.36*	.24	.28	.33*	.41*	.44**	.07	.14	-.04	-.07	-.10
Moruya (High)	.09	-.18	-.11	.17	-.07	-.05	-.03	.00	.01	.13	.18	-.29	.09	.04

* $p < 0.05$ ** $p < 0.01$

Discussion

The present study was designed to answer two main questions. The first was to examine the predictors of risky flight behavior in general aviation and the second was to examine the contextual sensitivity of these predictors. The results revealed that both attitudinal and experiential factors as measured in the present study were not related to the decision to fly. In terms of the former, this finding is relatively unique and differs from the majority of academic literature in the area of attitude and behavior (Ajzen, 1991; Albarracín, Fishbein, Johnson, & Muellerleile, 2001; Crano & Prislin, 2006; Deery, 1999; Glasman & Albarracín, 2006; Smith & Terry, 2003) in that most studies have found that attitude affects behavior to a greater extent than was evident in the current experiment.

There are many possible causes for this finding. The General Aviation industry in Australia exhibits what is arguably a healthy safety culture, in which safety is a cultural norm. Safety-based publications are distributed to all pilots, and safety management systems are utilised extensively by commercial (and private) operators. In contrast to this, other studies are often undertaken within systems similar to the road environment, where little or no testing, checking, or other safety-related activities like training are undertaken by authorities. An additional difference that may contribute is the self-selection bias caused by the cost and relatively rigorous academic, time and physical requirements of flight training, when compared to the relative economy and ease of driving, or other similarly ubiquitous activities.

In terms of the latter – experiential factors, the results revealed neither age nor experience appeared to be related to the decision to undertake (or not to undertake) any of the flight scenarios. An inference from the finding is that older pilots or more experienced pilots performed no better than their younger counterparts with regards to conservative behavior. This finding is dissimilar to those of the road safety arena, where increasing age has been linked to more conservative behavior (Deery, 1999). As above, there are many possible reasons for this finding, and indeed, all factors mentioned above are valid hypotheses as to the reasons why the current finding differed from those in the literature. With particular reference to the relationship between age and risk-taking, it may be the case in Australian GA that a self-selecting bias is present such that younger candidates are more risk averse than is the population wide norm. Therefore, it is possible that there is less of an initial difference in risk perception between younger and older pilots than there is between younger and older people in the populations, leading to the lack of correlation as found in the current research. This hypothesis however, has not been tested in the current research.

The finding that both total flight hours and recent flight hours are not related to behavior is echoed in the recent aviation literature, with studies in the Australian aviation environment finding that these two experiential factors appear unrelated to behavior (Molesworth & Chang, 2009).

It was found however, that the risk perception factors Delayed Risk, General Flight Risk, High Risk, and finally Altitude Risk were all related to behavior. That is, participants that exhibited a higher level of these risk perception factors were more likely to choose not to fly. This conforms to the accuracy hypothesis of risk perception (Brewer et al., 2004), in which the perception of a risk will lead to compensatory behavior by an individual as an attempt to reduce the amount of risk encountered and therefore perceived. This is also consistent with Hunter's (2006) findings, in which those with higher perceptions of risk were less likely to have experienced hazardous situations in comparison to those that rated the risks as lower.

The scenarios in the current research utilised multi-dimensional risk factors, such that there were at least two risk factors present for each scenario. It appears from the current results that the perception of risks in uni-dimensional risk scenarios (as in Hunter's Risk Perception scales) is related (albeit relatively weakly) to the perception of multi-dimensional risk. That is, risk perception of relatively simple situations seems to be related to risk perception in more complex (with regards to the hazards present) situations.

In relation to the second question, the results did reveal that the predictors of risky flight behavior are contextually sensitive. Moreover, in the present study only the medium risk scenario produced any statistically significant correlations. This is an important finding as it highlights the potential limitations of relying on the traditional predictors of risky behaviour to explain the diversity in all risky behavior. Considering high-risk situations present the greatest level of risk to safety, future research should investigate if there are more appropriate indicators or predictors to explain this often undesired behavior. This finding also points towards a possible limitation in previous research that has employed only a single scenario to measure the attitude/risk perception-behavior relationship. Conversely, the results of the present study may have been influenced by the experiment design, namely a pen and paper exercise opposed to a simulated flight or actual flight. This is another area for future research.

Conclusion

From the current study, it appears that attitude, age and experience do not significantly affect the risk management of pilots in Australian GA. Making sound risk management decisions, given the context, appears to be most dependent upon risk perception, not experiential or attitudinal factors. It was also identified that predictors of risky flight behavior are contextually sensitive. However, these conclusions are based on a pen and paper study and future research should be directed towards replicating these finding in the operational environment or at least within a flight simulator.

Acknowledgments

This paper is based on one part of a PhD dissertation by Mr. Justin Drinkwater that is currently undergoing assessment.

References

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes, 50*, 179-211.

Ajzen, I., & Fishbein, M. (2005). The Influence of Attitudes on Behavior. In D. Albarracín, B. T. Johnson & M. P. Zanna (Eds.), *The Handbook of Attitudes*. Mahwah, NJ: Erlbaum.

Albarracín, D., Fishbein, M., Johnson, B. T., & Muellerleile, P. A. (2001). Theories of Reasoned Action and Planned Behavior as Models of Condom Use: A Meta-Analysis. *Psychological Bulletin, 127*(1), 142-161.

Armitage, C. J., & Conner, M. (2001). Efficacy of the Theory of Planned Behavior: A Meta-Analytic Review. *British Journal of Social Psychology, 40*, 471-499.

Brewer, N. T., Weinstein, N. D., Cuite, C. L., & Herrington Jr., J. E. (2004). Risk Perceptions and Their Relation to Risk Behavior. *Annals of Behavioral Medicine, 27*(2), 125-130.

Cox, S. J., Cox, T., 1991. The structure of employee attitudes to safety: A European example. *Work & Stress, 5,* 93-106.

Crano, W. D., & Prislin, R. (2006). Attitudes and Persuasion. *Annual Review of Psychology, 57*, 345-374.

Deery, H. A. (1999). Hazard and Risk Perception Among Novice Drivers. *Journal of Safety Research, 30*(4), 225-236.

Drinkwater, J. L., & Molesworth, B. R. C. (2010). Pilot See, Pilot Do: Examining the Predictors of Pilots' Risk Management. *Safety Science, 48*, 1445-1451.

Glasman, L. R., & Albarracín, D. (2006). Forming Attitudes That Predict Behavior: A Meta-Analysis of the Attitude-Behavior Relation. *Psychological Bulletin, 132*(5), 778-882.

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*(4), 359-379.

Hunter, D. R. (1995). *Airman Research Questionnaire: Methodology and Overall Results.* Springfield, VA: Federal Aviation Administration.

Hunter, D. R. (2003). Measuring General Aviation Pilot Judgment Using a Situational Judgment Technique. *The International Journal of Aviation Psychology, 13*(4), 373-386.

Hunter, D. R. (2005). Measurement of Hazardous Attitudes among Pilots. *The International Journal of Aviation Psychology, 15*(1), 23-43.

Hunter, D. R. (2006). Risk Perception Among General Aviation Pilots. *The International Journal of Aviation Psychology, 16*(2), 135-144.

Lund, I., Rundmo, T. 2009. Cross-cultural comparisons of traffic safety, risk perception, attitudes and behaviour. *Safety Science, 47,* 547-553.

Molesworth, B. R. C., & Chang, B. (2009). Predicting Pilots' Risk-Taking Behavior through an Implicit Association Test. *Human Factors, 51*(6), 845-857.

Molesworth, B. R. C., Wiggins, M. W., & O'Hare, D. (2006). Improving Pilots' Risk Assessment Skills in Low-Flying Operations: The Role of Feedback and Experience. *Accident Analysis and Prevention, 38*, 954-960.

O'Hare, D., & Smitheram, T. (1995). "Pressing on" into Deteriorating Conditions: An Application of Behavioral Decision Theory to Pilot Decision Making. *International Journal of Aviation Psychology, 5*(4), 351-370.

Retting, R. A., Ulmer, R. G., Williams, A. F., 1999. Prevalence and characteristics of red light running crashes in the United States. *Accident Analysis and Prevention, 31,* 687-694.

Smith, J. R., & Terry, D. J. (2003). Attitude-Behavior Consistency: The Role of Group Norms, Attitude Accessibility, and Mode of Behavioral Decision-Making. *European Journal of Social Psychology, 33*, 591-608.

Ulleberg, P., & Rundmo, T. (2003). Personality, Attitudes and Risk Perception as Predictors of Risky Driving Behavior among Young Drivers. *Safety Science, 41*, 427-443.

MAHLER TO MACH 1

Capt. John Gadzinski
Four Winds Consulting
Virginia Beach, VA

As a classically trained musician who studied with the Boston Symphony and worked for the Opera Company of Boston, the author was able to directly apply the skills used in the high pressure world of professional music performance to become a Navy fighter pilot with 10 years of service and 321 carrier landings. Later, after flying for a major airline and completing his certificate in Aviation Safety at USC, the author worked with the NTSB on major accident investigations and started to observe the link between aviation human factors, safety analysis, and the similarities between accident investigation skills and the skill needed for professional musicians. This presentation will discuss topics including how data in both music and in safety management systems can be heavily influenced by training and talent, and the striking similarities in the skills used. Other topics include human performance and how music auditions and carrier landings share identical mental modeling, how error management and non normal operations are treated in both fields and how the functioning of a chamber orchestra can serve as a model for cockpit situational awareness and CRM. The presentation shows how classical music and aviation safety relate in unique and fascinating ways.

When the World Changes

The Apollo I fire was a tragedy so unexpected and the consequences so unimaginable that it shook the country to its core. Three national heroes were killed in a matter of seconds, only feet away from a crew helpless to do anything. The exercise itself was no more than a dress rehearsal meant to have about as much risk as an evening in front of the television. The political aftermath was equally brutal. The year was 1968 and in front of the US Senate Apollo astronaut Frank Borman was trying to keep the effort to go to the moon alive.

Amazingly the best and brightest minds of the day couldn't foresee what would have been obvious to any high school physics teacher: placing three men in a metal container, clamped shut and filled with 100% pressurized oxygen while throwing hundreds of electrical switches and packed full of flammable plastic was a sure recipe for disaster. The consequences of just one small spark would be catastrophic. And yet that disaster occurred. When asked why this was allowed to happen, the answer was straightforward and eloquent, Borman replied: "a failure of imagination." Even though safety", he said, "was never intentionally compromised …no one ever imagined [a fire] could have happened on the ground. If anyone would have thought of it the test would have been classified as hazardous…but it wasn't, we just didn't think of it."

You don't need to be in NASA to have decisions about risk affect you. We live in a world defined by risk and the rewards that taking them bring to us. We call it a business approach for preventing unnecessary cost and everyone from your local school to airline pilots try to imagine the best way to keep all of us free from the kinds of consequences we feel unable to afford. Why then do failures occur? In recent years we've had professional pilots crash airplanes, one of the largest and most advanced oil platforms blow up, and see established financial institutions fail and threaten to take the entire country down with them, just to name a few. In each case we ask the same questions as the Senate Panel for Apollo 1; why someone, anyone, didn't see it coming? We ponder this despite the fact that the data available to us has increased exponentially. The problem is obviously not our lack of data.

Analyzing the present flood of information and imagining the risks it represents in our complex world takes as much training and skill as it does the Boston Symphony when they look at the data (musical notes) given to them. The only difference is that the symphony is composed of people highly trained, mentally disciplined, possessing the background knowledge, analysis techniques, and communication skills necessary to pull immense meaning from raw notes and present them in the most effective way possible. The people responsible for safety most likely are not so prepared. For both the musician and the airline pilot the question is fundamentally the same: "what turns information into knowledge and knowledge into relevance?"

This question started to appear when I was part of a lengthy investigation with the NTSB. The data was there in black and white and yet two extremely smart people looked at the same information and came to completely different conclusions. In this case it was the Joint Winter Runway Friction Program conducted in collaboration by NASA, the FAA, and Transport Canada. Coming out of this research the US and Canada developed two different guidelines concerning what the hazards were and how best to effectively deal with them. The Canadian's used a friction index that managed both uncertainty and measurable components. The US relied on pilot reports alone. Both were correct from the standpoint of their respective reasoning, so how could one have contributed to the accident but still be viewed as a valid conclusion contributing to safety? Could a risk lay not in what was answered but in the question?

When looked at in hindsight it became clear that there were two visions of "risk" being used. One was engineering and legal risk, the other was an operational risk. The engineering approach sought to eliminate randomness in favor of the most direct path towards what could be considered a consistent correlation between observations and performance. The operational approach sought to minimize engineering variables through equipment standardization and embrace the possibility of corrupt data (randomness) as a variable that itself had to be identified. The data itself didn't point to one approach over the other. Instead it was the interpretation of the individuals themselves that played such a large part in defining hazards and mitigations. This was not the first time a scenario like this has been played out.

One of the more famous quotes came from the Space Shuttle Challenger accident when a Morton Thiokol manager pointedly told someone to "take their engineers hat off and put their manager's hat on." It's not that people were intending to cut corners on safety; they just couldn't imagine what was around those corners. And even if they did, can a hazard reside not only in a potential event but also in someone's ability to communicate risk effectively?

In the Senate Panel on Apollo I, Frank Borman was saying that at the time the people involved in the process didn't have the skills to see how much meaning lay embedded in what was before them. What was put down on paper as a testing protocol was actually a treasure map for the many interconnected perspectives that a story about risk could possibly contain. It was if the people involved were reading a book called "Moby Dick" purely for guidance on how to look for whales with no expectations that it could be talking about anything else. Borman was an engineer and pilot, but his observations were in fact more a reference to art.

You see what your knowledge tells you and when that knowledge changes, what you see changes. When looking at the dawn it makes just as much sense to think that the sun goes around the Earth as it does the other way around. We only see that it that way because over time we've asked the kind of questions that changed our knowledge in the first place. As for the orbit of the sun, the question that first leads to the theory of planetary motion was: "On what day should we celebrate our religious holidays?" Knowing what day it was and how to accurately predict what planet will appear in the sky then lead to Copernicus and a major re-shuffling of our view of the sky. To the church, the question of risk was a political one. How can you run a religion without having some concrete direction on when to do things? Of course what they got in the end was more of a headache than they bargained for.

Today we look at risk from perspectives we chose and sometimes end up with unintended results just like the church did. They come from different viewpoints on what is important such as from lawyers worried about legal risk, from elected officials worried about political risk, from managers worried about financial risk, and from almost everybody worried about social risk. Will any of those people, the lawyers, accountants, bosses, or politicians be worried about the risk of someone's "failure of imagination?" To them there is scant room for imagination because their world is carefully defined by variables that set the boundaries of what is acceptable and not.

Let us now look at how musicians ask questions and about how not only their perspective changes but how they *expect* and want their perspective to change. In fact, if a musician looks at a series of notes and doesn't find some change in perspective it is deemed downright unprofessional!

Anyone with a child in a school band knows what an 8th grade concert sounds like. Now take that same piece of music and give it to the Boston Symphony and it'll probably sound a lot different. You pay good money to go to the movies where they employ a top orchestra to play a soundtrack, whereas you'd feel much differently if some high school band gave the score their best shot. In my experience, company safety briefs and professional safety conferences are a lot like 8th grade concerts. The people who attend are there because they should be there, whether they get anything out of it is up for debate.

To understand how meaning is extracted out of data and presented in a way that is both science and art we first look at the Navy Pilot out at sea.

Music and the Pilot

The carrier landing and the concert stage have many facets in common. First, they are both very much dependant on the mental state of the performer to bring his skills to bear. Unlike the landing on normal runway, the carrier landing is always performed in front of a knowledgeable and critical audience. Every landing on the carrier is graded, de-briefed in person, and posted for all to see. Every landing is also broadcast over TV through the entire ship as one of the channels available for general viewing. The landing requires an ability to repeat a skill that has been practiced beforehand and to do so under stressful conditions for not only he being watched but there is a very real possibility that if done incorrectly, he may not survive.

While for the musician physical survival is rarely at stake, the mental demands are almost exactly similar. For such an event the preparation is accomplished using a multi-step process. First, a large amount of repetition is employed to create a subconscious relationship between mind and body coordination for the skills required. This is first and foremost a physical learning process. Then the musician will employ what is known as technical rehearsals. This is where a the music is played up to the point where a difficulty arises whereupon that particular part is again broken down into component parts, repeated, and strung together again until it flows. Last is the performance rehearsal. This is similar to a LOFT event in that the entire piece of music is played, mistakes and all, in its entirety and examined later.

Aviation's success lies in the fact that many of the tasks and skills required are broken down and institutionalized in procedures. For the carrier pilot this takes place during the approximately 150 landings made in a dedicated training syllabus before the first carrier landing. The technique is similar to the technical rehearsal in that only a few skills are concentrated on during any particular event. Some skills such as flying the aircraft in reference to a specific angle of attack are similar to musician's scales. Steady deliberate repetition is the order of the day. More complex skills are then added as the landing pattern altitudes are refined. Finally the subtleties of optics, flight physics, engine design, glide slope geometry, and wind are introduced to form a broad picture of the event that goes well beyond mere stick and rudder technique.

In the case of both carrier flying and music, we can see that technical correctness alone is not enough, a broader context must also come into play. For the pilot flying onboard ship, for the safety analyst investigating an accident, and for the musician approaching a piece of music, technical skills are merely one set of tools for interpreting a larger canvas of relationships. For this the musician enters a deep well of human factors and physics.

If we are to consider human factors as anything that affects human performance, then music is the study of performance that affects humans. In both music and aviation, human factors strive to produce an intended result and both use known tendencies of cognitive psychology in the process. We start with the basics of breathing and discomfort. When two jet engines are not operating at near similar speeds the result is a kind of audio interference pattern. Most people would rather not have the oscillating noise drumming against their ears for extended periods of time. Increase the frequency of that same audible pattern and speed and you have what is known in music as a dissident chord, equally tough to listen to, and either a compositional effect, or an indication that something's not right.

All sounds produce pitches that can be both heard and unheard. The unheard sounds are what are known as "overtones" in music and represent harmonic frequencies, much the same as they do in mechanics. There's one catch however: the harmonics aren't integers or perfect fractions. In fact if you take a frequency, say the 440 cycles per second of the concert note "A" and double it, you will indeed get an "A" an octave higher (this was a major discovery of Pythagoras). If, however, you wish to create a pitch that will correspond to an appropriate dissonance with relation to that "A", say a minor 7th, it will require a slightly different pitch than if you were simply to mathematically calculate where such a frequency should populate the spectrum of sound. Thus, if you are a professional concert pianist and you are to play Rachmaninoff's concerto in C you would have your piano specifically tuned for that key. If, however, you want to tune your home piano once and be done with it, you would want to consider Bach's "Well

Tempered Clavier," a set of music specifically designed for evenly tuned pianos that cleverly avoids chord combinations that highlight such a discrepancy.

Fine you say, but Bruce Springsteen doesn't care about this and what has all this to do with aviation anyway? The answer is Crew Resource Management. Before we get to that point let us look at one more aspect of music, phrasing.

No matter how simple or complicated the music, it all has to do one thing and that's to tell a story. Garth Brooks and Igor Stravinsky both have one thing in common: their music has the same basic phrase length of a human sentence. Perceptually we are so used to it we don't even notice, but the affect on the mechanics of timekeeping are profound. The data in audio notation is similar to the data in an Flight Data Recorder plot in that it is all relative to a mathematical notation of time. As with our discussion about pitch, however, knowing when to "break the rules" is the stock and trade of the performer. You see when we speak, we subtly alter our rhythm so as to allow us to breathe, place emphasis, create an effect and so on. This also happens when we dance. The most difficult music for an orchestra to play well is a waltz because it's lilting and slightly hesitating rhythm must be accomplished exactly by all members of an orchestra without specific, detailed direction from the conductor.

A true professional musician normally spends years, usually nine to twelve including summer education, full time college, and graduate or advanced studies before his "mental model" of pitch, phrasing, along with physical technique allows him to truly ply his trade for a living. Then he will most likely do what pilots do; attempt to join a group of likeminded folks in producing some sort of desired output. For the pilot this will mean a multi-crewed passenger airline jet. For the musician it means an orchestra or some form of ensemble. While most of us are familiar with the image of the conductor, there is a large area of music that is played without one. These ensembles are known as "chamber groups" and can range from small orchestras to groups of three. It is here we start to notice how situational awareness and leadership are built and enacted.

Situational awareness is the result of a shared mental model among groups of people. In a chamber group what is important is that such a model is expected to *change* and *so are its leaders.* Because how you perform depends on what your relationship is to a commonly shared awareness of pitch structure (B flat for instance), as well as whether or not you or someone else is controlling the phrasing, your role as leader, supporter, or vitally important facilitator is relative to the current situation. As a result the chamber orchestra becomes skilled at rotating the roles of leadership in quite a fluid fashion. This does not mean there is not a dedicated leader responsible for the overall coordination of the activity, merely that while it is taking place, there is a collective understanding that no single person can define all the requirements for a successful outcome.

From an error management point of view the ramifications are obvious. Error is a fact of human performance both on stage and in the cockpit. The chamber orchestra however is more resilient to unexpected slips in performance because changes in leadership are a desired outcome of the group mentality and the mission training is more robust in that there is a greater appreciation for the roles each person plays and a recognition that how an outcome relates to others is more important than the individual effort.

I want to stress again that this is no casual technique. All symphony orchestras are composed of "sections," the violin section, the trombone section, the bass viola section, and so on. Each section has a designated leader that coordinates how the section should phrase and each section performs according to how their pitch lies in relation to the other pitches, rhythms, or melodies that can be involved at the same time. It is similar to how an airport operator, ATC facility, and flight crew operate; the only difference is that the musicians have been heavily trained to understand each other's roles and how they relate to them.

Compare this to the many examples of CRM breakdown seen in the accident databases. Even when looking at a basic safety management issue such as the un-stabilized approach we see that a fluid transfer of leadership based on commonly understood conditions runs into challenges. The reasons for this are both simple and elusive. The first and most obvious is training. We don't train pilots in the underlying science of human cognition anywhere near to the extent that musicians do, but even if we did, the musician has one advantage we don't, positive feedback. Even without an audience a group of musicians can get satisfaction from an exercise where situational awareness has been proficiently maintained because their recognition of success or failure is immediate. For purposes of safety success or failure is often viewed as how well exposure to risk has been managed. But while exposure to risk may be important to those in the safety office, it rarely has any connection with the satisfaction of accomplishing a successful mission of delivering passengers to their destination. While an on time arrival at a destination is easy to comprehend as a successful outcome, how much exposure to risk was amplified in the process can require some imagination.

And so we find ourselves back to Apollo I and the question "what makes a hazard?" There are some things a musician doesn't need to know that professional risk management does. As the musician needs to know phrasing, the safety professional needs to know rhetoric. Both disciplines deal with communicating relevance. As the musician needs to know the physics of sound, the safety professional needs to know about probability from Gaussian to Bayesian to the concepts of the "Black Swan" and the effects of the highly improbable. Both music and probability teach how to relate information. As the musician needs to know music theory and the differences between Gregorian chant and Wagner's "Tristan chord" so the safety performer needs a thorough understanding of human factors. Both deal with how relationships between events can be viewed in different and unique ways.

A "hazard" is really no more than a story presented to an audience that places in their mind's eye a unique image of their world. Professionally trained and skilled musicians invest heavily in the many ways their world can present itself, how all those facets interrelate, and in how to effectively communicate that composition. Most important, however, is that music taps into our fundamental yearning to witness the economy of precision, that the image we are painting for others is exactly, to the finest subtlety, what we mean it to be.

We finish with the comments of astronaut Frank Borman about a failure of imagination. If we start by looking at our safety information; our reports, our data, our investigations, and believe that they hold the answers to *our* questions than we miss an opportunity. If instead we take that information and ponder what questions we can ask about the data itself, what questions we should ask, and what questions would most likely lead to a change in our perception then we start down the process of changing our view of the world in new and unique ways. If we are surprised by what we find then we can re-assure ourselves that it is a view that won't surprise us in the future.

VISUAL VS. AUDITORY MEMORY IN AN AVIATION TASK: A POTENTIAL PERFORMANCE THEORY ANALYSIS

Gayle Hunt[1]
Stephen Rice[1]
David Trafimow[1]
Jeremy Schwark[1]
Joshua Sandry[1]
Lisa Busche[1]
Kasha Geels[1]
[1]New Mexico State University
Las Cruces, New Mexico

Information can be relayed to pilots by visual presentation or auditory presentation; both methods are frequently used. To date, there is quite a bit of conflicting literature regarding which type of communication is most effective for recalling information. The current study tested memory differences between digits presented visually or audibly in number strings. Results showed that performance in the visual presentation condition was superior to the auditory presentation condition. However, when a PPT analysis was conducted on the data, it was revealed that performance in the visual condition was only superior because participants were more consistent in that condition compared to the auditory condition. In other words, had participants been perfectly consistent in both modalities, there would have been no differences in performance. We conclude that inconsistency, and *not* different strategies, was responsible for the increased performance of visual memory over auditory memory.

The current study focuses on visual and auditory digit memory, an area that is well-studied in the basic cognitive memory literature but one that does not always scale up to the real-world in ways that are useful for applied research. In the basic literature, when comparing visual to auditory presentations for memory recall, it is critical to avoid any confounds that might give one presentation an inherent advantage over another. Thus, if auditory digits are presented serially, visual digits also must be presented serially, and the rate of digit presentation must be controlled (e.g. Jensen, 1971).

However, in the real world, it often is not the case that visual digits can be, or should be, presented serially in such a fashion. For example, a pilot reading digits should have access to the entire string of digits at once glance. An air traffic controller reading aircraft call signs has, or should have, immediate access to the entire string of digits. Cases like these can be imagined in many types of applied aviation settings.

Of course, comparing a visual presentation, where all digits are presented simultaneously, to an auditory presentation, where all digits are presented serially, results in an inherent advantage to the visual presentation if the two presentations are compared for overall memory performance. In the visual presentation, participants can review previous digits whereas in the auditory presentation, backtracking is impossible. In this situation, it seems likely that simultaneous visual presentation promotes superior information processing strategies compared to serial auditory presentation, which in turn provides a reason for observed performance to be better in the former case than in the latter one. Alternatively, it is possible that the two types of presentations allow for equally good information processing strategies but that the information processing strategies promoted by simultaneous visual presentation are used more *consistently* than those promoted by serial auditory presentation. Either way, observed performance in the simultaneous visual processing condition should exceed that in the serial auditory condition.

Because the goodness of strategy hypothesis and the consistency hypothesis make the same prediction with respect to observed performance, it is necessary to find some means to distinguish them from each other. Consequently, we focus on the issue of parsing strategy and consistency in digit-recall memory tasks. But before we examine this issue, it is necessary first to briefly review the literature pertaining to visual versus auditory memory.

Visual Versus Auditory Memory

Much work has been done comparing memory of information presented in a visual versus auditory manner. There are conditions where visual presentations cause better performance than auditory presentations (e.g. Hawkins, 1897; Henmon, 1912; Worchester, 1925) but there also are conditions where the reverse is so (e.g. Binet, 1894; Dornbush, 1968; Koch, 1930).

Day and Beach (1950) reviewed the entire body of research prior to 1950 and concluded that visual memory is generally superior to auditory memory; however, studies also show that visual memory degrades more quickly than auditory memory with age (McGhie, Chapman & Lawson, 1965) and that children produce superior performance with auditory presentations compared to visual presentations (Abbot, 1909). These effects may be differentially affected by a number of factors (Beaman, 2002; Jensen, 1971; Norman, 1966; Saults & Cowan, 2007; Sherman & Turvey, 1969).

Several researchers have proposed variations on an argument relating presentation format to learning histories (e.g., Dornbush, 1968; Johnson & Miles, 2009; Saults & Cowan, 2007). To understand this argument, consider that in normal living, people tend to be exposed to auditory stimuli serially and to visual stimuli all at once. Although there doubtless are exceptions, it seems reasonable to suppose that most people have learned to process auditory stimuli in accordance with serial presentations and visual stimuli in accordance with simultaneous presentations. Consequently, serial presentation of stimuli confers an advantage for auditory versus visual presentation, which provides one explanation for the cases where participants perform better in auditory than visual presentation conditions.

Taking this argument seriously implies the possibility of a subtle confound in experiments designed to test visual versus auditory presentation conditions. That is, if participants encounter stimuli serially, with a resulting advantage in the auditory presentation condition relative to the visual presentation condition, it could be due to a better matching of presentation condition with previous learning history rather than due to the superiority of auditory presentations to visual ones. On the other hand, if the participants in the visual presentation condition encounter the stimuli simultaneously, whereas the participants in the auditory presentation condition encounter the stimuli serially, an obtained effect might be due to the simultaneous-serial difference rather than due to the visual-auditory difference. Either way there is a potential confound.

Given that there exists a confound no matter how the experiment is designed, theoretical and methodological considerations cannot, by themselves, determine the confound with which we should choose to live. Rather, we believe that applied considerations also should figure into the decision. And it is obvious that in most applied settings, visual information is presented simultaneously and auditory information is presented serially, in accordance with our past learning histories. Therefore, we choose to resolve the learning history confound and live with the simultaneous-serial confound—a decision that we emphasize is strongly influenced by the fact that our goal is applied rather than basic.

As we pointed out earlier, because the effects of visual versus auditory presentations on digit-recall are so dependent on a multitude of factors, it is unlikely that any one perspective is going to explain all of the data, at least not in the near term. Consequently, it is not our intention to choose a "winner." However, choosing a winner is not the only possible contribution. If it could be determined (a) whether simultaneous visual presentation causes better performance than serial auditory presentation and (b) whether the effect is due to random or nonrandom factors, there would be important implications for applications. Until recently, there has been no way to parse random and nonrandom factors; however, a recent theory—termed potential performance theory (PPT; Trafimow & Rice, 2008, 2009)—does just that.

Potential Performance Theory

Potential Performance Theory is a mathematical theory of task performance that allows researchers to look beyond observed performance and calculate potential performance in the absence of any random factors. According to this theory, one of the primary factors that must be taken into account is consistency. 'Consistency' is a term that can be characterized in many different ways, depending on the area to which it is being applied (Brunswick, 1952; Kelley & Michela, 1980; Orvis, Cunningham, & Kelley, 1975). PPT uses the term consistency to reference the correlation coefficient calculated between two identical blocks of trials in a study. For example, suppose a person is given a pair of two ten-digit numbers and asked to identify whether the pair of numbers is the same or different. The person repeats this task 100 times, each time being presented with a unique set of numbers. Following this task is a short break and a second block of 100 trials. The second block is identical to the first, so the person is giving an answer to each pair of numbers two times. This allows the researcher to calculate a correlation coefficient across the two blocks. This correlation coefficient is also known as the consistency coefficient, which is essentially an inverse measure of randomness. More randomness necessarily implies a lower consistency coefficient while less randomness implies a higher consistency coefficient.

Potential performance, as defined by PPT, is the score a person could achieve in the absence of randomness. When a person's accuracy is greater than chance, more randomness will cause observed performance to decrease. In the absence of randomness, a person's answers to each block of trials should be identical, producing a consistency coefficient of exactly 1.00 and an observed score equal to the potential score. As randomness increases, this coefficient will decrease and cause the observed score to fall below the potential score. Once observed scores are measured and consistency coefficients are calculated, a researcher can use the formulas given by Trafimow and Rice (2008, 2009) to calculate potential scores for each participant. (For lack of space, those formulas are not presented here.)

By virtue of having both observed scores and the results of PPT computations, it is possible to have the three main variables of importance for each participant. These are the observed score or observed proportion of successes, the potential score, and the consistency. In turn, as the experiment to be presented demonstrates, these variables make it possible to disentangle whether differences between simultaneous visual presentations and serial auditory presentations are due to nonrandom factors, random factors, or both.

Current Study

In the current study, participants viewed a 10-digit string of random numbers visually or they were exposed to the same 10-digit string of numbers auditorially, via headphones. Visual presentation where all digits are presented concurrently provides an inherent advantage over an auditory presentation where all digits are presented serially. Thus, our first hypothesis is that visual presentation of digits will result in superior observed performance compared to auditory presentation of digits. However, what is not definitively known is *why* this particular paradigm advantage results in superior performance. One might assume that this advantage results in a superior type of memory processing strategy (nonrandom factor) which would in turn indicate better potential performance. However, it is possible that this advantage also triggers more consistency (or less randomness). Thus, our second hypothesis is that visual presentation of digits will result in more consistent performance compared to auditory presentation of digits. It is entirely plausible that the visual presentation will result in both superior potential performance *and* superior consistency, and thus we have three additional competing hypotheses: *A*: Visual presentation of digits will result in superior potential performance and higher consistency compared to auditory presentation of digits; *B:* Visual presentation of digits will only result in higher consistency compared to auditory presentation of digits; and *C:* Visual presentation of digits will only result in superior potential performance compared to auditory presentation of digits.

Method

Participants

Twenty (13 females) undergraduate students from a large southwestern university participated in the experiment for partial course credit. The mean age was 20.3 (*SD* = 2.70). All participants were tested for normal or corrected-to-normal vision and color-blindness.

Materials and Stimuli

The experimental display was presented via E-prime 1.0. Participants were exposed to 100 visually-presented trials, and 100 auditorially-presented trials. Each set of 100 trials was divided into 2 blocks of 50 trials each. Of these, 25 trials were matched and 25 trials were unmatched. A matched trial consisted of two consecutive display presentations whereby a string of 10 randomly generated numbers were identical across both displays. An unmatched trial consisted of the same format with numbers that were not identical; one number in the string was randomly changed.

Each trial began with a fixation display that was presented for 500 msecs. Following this, a Number display presented the first 10-digit number, followed by a Mask display that remained for 200 msecs. Next, a second Number display presented a 10-digit number that either was identical to, or different from, the first Number display. Whether or not the trial contained a matched or an unmatched display was randomly determined. Following this, a Choice display was presented whereby participants were asked to decide whether or not the two numbers were identical. Participants pressed the J key if they determined that the numbers were identical and the F key if they were not identical. Since PPT methodology requires 2 identical blocks of trials, participants were exposed to 2 blocks of visual stimuli and 2 blocks of auditory stimuli. These blocks were presented randomly and all trials within each block were presented randomly as well. Each block consisted of 50 trials. At the end of each block, participants were able to take a short break. Instructions were given at the beginning of each block to inform participants whether the trials would be presented visually or audibly.

Procedure

Participants first signed a consent form and were then seated in a comfortable chair facing the experimental display. Viewing distance was controlled by a chin rest at 21 inches. Participants first read on-screen instructions and then were given the opportunity to ask questions. Once they were comfortable with the instructions, participants pressed a button to begin the experiment. The entire experiment took approximately 70 minutes. Upon completion, participants were debriefed and dismissed. The experiment employed a within-participants design, whereby all participants were exposed to both the visual and auditory conditions.

Results

The data were analyzed using PPT theorems (see Trafimow & Rice, 2008, 2009) and a series of three *t* tests were conducted. As can be seen from Figure 1, observed performances in the visual condition exceeded those in the auditory condition (*M* = .84 and *M* = .71), *t*(19) = 3.59, *p* < .001, *d* = 1.65. In addition, consistencies in the visual condition exceeded those in the auditory condition (*M* = .56 and *M* = .26), *t*(19) = 4.63, *p* < .0001, *d* = 2.12. Finally, potential scores did not differ discernibly between the visual and auditory conditions (*M* = 1.00 and *M* = .98), *t*(19) = 0.34, *p* = .37, *d* = 0.16.

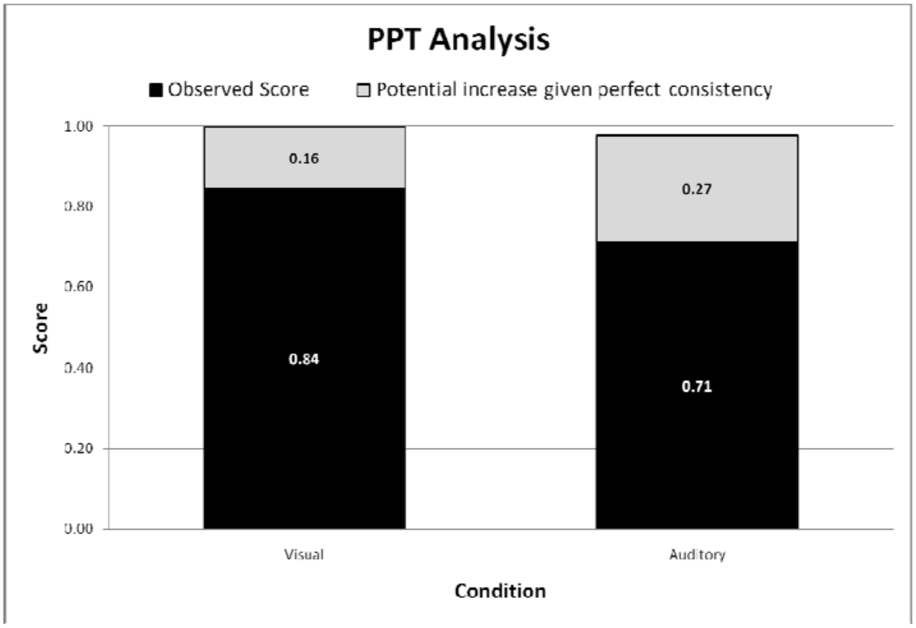


Figure 1. Experimental data.

Discussion

Our first hypothesis was that performance in the visual presentation would exceed that of the auditory presentation. Not surprisingly, this was the case. Clearly, there was an advantage for participants to view all visual digits concurrently. Of course, it could also be the case that the visual presentation would have produced superior performance even if the digits had been presented serially (Jensen, 1971); however, we believe it is not wise to speculate here. Our second hypothesis was that visual presentation of the digits would result in more consistent behaviors. The data clearly show this to be the case. Consistency in the visual presentation exceeded that in the auditory presentation by both a statistically and practically significant amount with an impressive effect size.

We pointed out earlier that PPT implies three classes of explanations for observed effects; these are potential performance, consistency, or both. Because the two types of presentations resulted in differences in observed performances and consistencies, but not in potential performances, it is clear that two of the three possibilities can be ruled out. Put more simply, it is consistency, and not potential performance, that is responsible for the observed effect in the current study. Thus, our third hypothesis (B) appears to be correct, and we can reject its competing hypotheses (A and C).

What does this mean in plain English? Consider again that observed performance is caused by a combination of nonrandom effects and random effects. Suppose that there were no random effects. In that case, the PPT analyses indicate that observed performances in the two stimulus presentation conditions would be equal. In other words, the systematic components of task performance are equally effective in the visual and auditory presentation conditions. The reason for the observed effect, then, is because people are more consistent in the visual presentation condition than in the auditory presentation condition. The stimulus presentation manipulation causes people to behave more randomly in the auditory presentation condition than in the visual presentation condition, thereby reducing observed performance more in the former than in the latter condition. To our knowledge, none of the available literature can account for this.

But the data are even more surprising than that. Consider that mean strategy scores equaled or were very close to unity. This means that practically all of the error in observed performances was due to randomness; there was almost no error due to systematic causes in either condition.

The lack of systematic error also suggests some practical applications. Suppose that one wishes to train people to have better observed performances in the current paradigm, whether with visual or auditory presentations. Because the problem is one of consistency, it would be useful to find out exactly what it is people are doing and train them to do it that all of the time. Possibly, this could be done simply by making people’s strategies explicit, thereby placing them under conscious control.

In conclusion, the data reveal that different levels of potential performance are not responsible for the superior visual memory performance; rather, participants were simply more consistent in the visual condition than in the auditory condition. We believe that designers of equipment to be used by pilots and air traffic controllers should be aware that different modality presentations can differentially affect how consistently operators interact with visual and auditory displays.

Acknowledgements

The authors wish to thank Eric Johnson, Audrey Rosenblatt, David Flemming, Sandra Deming, Eduardo Rubio, Crystal Sandy, Rachael Currier, Crystal Garcia, and Natasha DeVries for their help in collecting data.

References

Abbot, E. E. (1909). Memory consciousness in orthography. *Psychological Monographs, 11*, 127-158.

Beaman, C. P. (2002). Inverting the modality effect in serial recall. *Quarterly Journal of Experimental Psychology, 55A*(2), 371-389.

Binet, A. (1894). *Introduction a la psychologie experimentale*. Paris: Bailliere.

Brunswik, E. (1952). The conceptual framework of psychology. Chicago: University of Chicago Press.

Day, W. F., & Beach, B. R. (1950). *A survey of the research literature comparing the visual and auditory presentation of information*. (AF Technical Report No. 5921) Dayton, Ohio: Wright-Patterson Air Force Base.

Dornbush, R. L. (1968). Input variables in bisensory memory. *Perception and Psychophysics, 4*, 41-44.

Hawkins, C. J. (1897). Experiments on memory types. *Psychological Review, 4*, 289-294.

Henmon, V. A. C. (1912). The relation between mode of presentation and retention. *Psychological Review, 19*, 79-96.

Jensen, A. R. (1971). Individual differences in visual and auditory memory. *Journal of Educational Psychology, 62*(2), 123-131.

Johnson, A. J., & Miles, C. (2009). Single-probe serial position recall: Evidence of modularity for alfactory, visual, and auditory short-term memory. *Quarterly Journal of Experimental Psychology, 62*(2), 267-275.

Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual Review of Psychology, 31*, 457-501.

Koch, H. L. (1930). Some factors affecting the relative efficiency of certain modes of presenting material for memorization. *American Journal of Psychology, 42*, 370-388.

McGhie, A., Chapman, J., & Lawson, J. S. (1965). Changes in immediate memory with age. *British Journal of Psychology, 56*, 69-75.

Norman, D. A. (1966). Acquisition and retention in short-term memory. *Journal of Experimental Psychology, 72*(3), 369-381.

Orvis, B. R., Cunningham, J. D., & Kelley, H. H. (1975). A closer examination of causal inference: The roles of consensus, distinctiveness, and consistency information. *Journal of Personality and Social Psychology, 32*, 605-616.

Saults, J. S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General, 136*(4), 663-684.

Sherman, M. F., & Turvey, M. T. (1969). Modality differences in short-term serial memory as a function of presentation rate. *Journal of Experimental Psychology, 80*(2), 335-338.

Trafimow, D., & Rice, S. (2008). Potential performance theory (PPT): A general theory of task performance applied to morality. *Psychological Review, 115*(2), 447-462.

Trafimow, D., & Rice, S. (2009). Potential Performance Theory (PPT): Describing a methodology for analyzing task performance. *Behavior Research Methods, 41*(2), 359-371.

Worchester, D. A. (1925). Memory by visual and auditory presentation. *Journal of Educational Psychology, 16*, 18-27.

FRACTAL TIME SERIES ANALYSIS OF HUMAN HEARTBEAT INTERVALS FOR PHYSICAL AND MENTAL WORKLOAD

Sheldon M. Russell, Gregory J. Funke, Benjamin A. Knott, Matt Middendorf

Air Force Research Laboratory
Wright-Patterson AFB, OH

As the environments and tasks that teams (in both military and civilian settings) are faced with increase in complexity, standard statistical methods may not fully capture team dynamics and processes. Nonlinear analyses provide alternative, mathematically derived descriptions quantifying the level of complexity and variability inherent in a data set, and may provide a more accurate understanding of dynamic systems. The goal of the present study was to investigate changes in heart interbeat interval associated with task workload using one type of nonlinear analysis, power spectral density analysis. In this study, physical and mental workload were manipulated in separate tasks to explore the contributions of each to interbeat interval variability. Results indicated that spectral analysis can identify large changes in overall workload, but may be insensitive to small or medium changes. However, these conclusions are based on preliminary results; follow-up research is necessary to determine the veracity of these conclusions.

Introduction

Although a relatively new addition to the sciences, the study of nonlinear dynamical systems (formerly known as chaos theory) has yielded interesting findings in a variety of scientific disciplines (an easily accessible general overview of this body of work can be found in Gleick, 1998). These types of systems (referred to from here on as *fractal systems*) are characterized by complex interdependencies in system variables and often unpredictable long-term outcomes. Long term weather patterns are a primary example of this, as the many interdependent factors associated with weather (humidity, solar activity, cloud cover, etc.) make long term predictions of specific weather outcomes very difficult. Although long-term outcomes are unpredictable, studying the behavior of systems over time (via time series analysis) has led interested researchers to conclude that, though specific predictions are not possible, fractal systems will generally behave in structured ways across time and scale, a property known as self-similarity. A variety of mathematical techniques are available to assess the fractal characteristics (including self-similarity) of a system, but the type of interest here involves frequency analysis of the time series output from a system.

Fractal analysis of time series data can involve a variety of frequency analysis techniques, such as Fourier transform, Lomb-periodogram, and wavelet analysis (Malamud & Turcotte, 1999). The goal of frequency analysis is to break the time series into frequency components (sine and cosine base signals) to determine the power spectral density characteristics of the time series. Correlations between these frequency characteristics indicate specific attributes of the system that produced the time series. Specifically, systems that are fractal in nature produce a power spectral density plot that has a power-law distribution (as opposed to, for example, a normal distribution). When the probability of measuring a specific value varies inversely as a function of a power of that same value, it is said to follow a power law.

Many natural systems, including biological processes such as cardiac interbeat intervals (IBIs), have been shown to fit a scaling that has power-law characteristics (see Newman, 2005, for a review). In the area of nonlinear applications to heart rate variability and IBIs, most research has focused on medical applications, examining cardiac risk (Goldberger, 2002), gender differences (Ryan et al., 1994), as well as determining baseline or normal resting heart rate characteristics (Ivanov et al., 1999). For an overview of fractal geometry research in biological systems, including the human heart, see Iannaccone & Khokha (1996).

In addition to research focused on fractal characteristics, other measures of heart rate variability have previously been explored as indices of cognitive workload and stress (see, e.g., O'Donnell & Eggemeier, 1986, for a review). Such research typically compares mean IBI variability across conditions for evidence of differences attributable to experimental manipulations. Overall, decrements in heart rate variability have been observed with increases in mental workload (Kalsbeek, 1971; Kalsbeek & Ettema, 1963). These decrements appear in frequency ranges similar to those reported in the cardiac nonlinear analysis literature.

In the present research we utilized the Fourier transform method, as previous research has demonstrated its accuracy as a method for determining the fractal characteristics of typical heart rate data (McSharry & Malamud, 2005). The following equations were reported in McSharry & Malamud (2005), and describe the Fourier transform of IBI time series data (x_n), where the total duration of the time series, T, is divided into N intervals of size $\delta = T/N$:

$$X_k = \delta \sum_{n=1}^{N} x_n e^{2\pi i n k/N}, \quad k = 1, \ldots, N, \qquad (1)$$

and

$$x_n = \frac{1}{N\delta} \sum_{k=1}^{N} X_k e^{-2\pi i n k}, \quad n = 1, \ldots, N. \qquad (2)$$

The power-spectral density function S_k is then defined as:

$$S_k = \lim_{N \to \infty} \frac{2|X_k|^2}{N\delta}, \quad k = 1, \ldots, N/2. \qquad (3)$$

While in general fractal systems follow power law scaling, correlations between values of S_k for all reported frequencies further specify characteristics of the system that generated the time series. Correlation values for spectral densities typically are reported as $\beta = x$, where x is the value of the exponent in the power equation multiplied by -1. For example, white noise (and other random processes) result in $\beta = 0$, in that there are no correlations between power spectral densities at any frequency value in the time series. Processes with long range correlations (i.e., preceding time series values influence subsequently observed values, a phenomenon termed *persistence*), result in correlations of $\beta = 2$, often called Brownian motion. A very common finding in natural systems is a correlation of $\beta = 1$, often referred to as $1/f$ noise or pink noise. Systems characterized by pink noise are not random, in that frequency values are correlated; however, these systems are not as persistent as systems characterized by Brownian motion. Previous research indicates that, in healthy individuals, a normal, resting cardiac IBI exhibits pink noise characteristics for frequencies below .1 Hz (e.g., McSharry & Malamud, 2005).

In general, deviations from β values of 1 have been shown to be associated with abnormal heart function and health risk. In the most extreme cases, β values of 0 indicate that the heart is fibrillating, and β values higher than 2 indicate states of congestive heart failure. The emerging consensus is that normal, healthy hearts display regular variability in IBI, and too much or too little variability can create significant problems (Iannaccone & Khokha, 1996). The complex nature of heart rate variability has led some to conclude that heart rate dynamics can be termed *multi-fractal*, in that the heart has a variety of different variability characteristics depending on the demands placed on the cardiovascular system (Ivanov, 1999).

Since fractal analysis of IBI has been demonstrated to be diagnostic in clinical settings (e.g., Goldberger, 2002), 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 to the frequency domain, thereby reducing individual differences often associated with physiological measures, which should result in increased sensitivity (Malamud & Turcotte, 1999).

Within the domain of workload and team performance, an initial, exploratory study was conducted by Russell, Funke, Knott, and Knott (2009) to examine the fractal characteristics of IBI during a visual search task. Data for the analysis was drawn from an experiment conducted by Knott, Nelson, McCroskey, and Miller (2007). In that study, task workload was manipulated by varying the number of on screen distracters (8 or 48 objects) in a visual search task. The results of the analysis conducted by Russell et al. (2009) were relatively promising, indicating that it may be possible to observe differences in fractal exponents (β values) by manipulating task demands.

While the results from Russell et al. (2009) were encouraging, the experiment in that study was focused on visual search, rather than workload. The present study expands on the idea that changes in fractal exponents can be indicative of workload by testing several tasks. The present study also explores differences in physical and cognitive workload, with the goal of determining the sensitivity of fractal methods to sources and levels of workload. We initially hypothesized that the fractal exponents associated with primarily cognitive tasks would be differentiable from those associated with physical tasks. We also hypothesized that task difficulty would influence fractal exponents such that more difficult tasks would be associated with increased fractal exponent values. It is worth noting that the results reported herein are preliminary; data collection for this experiment is still ongoing.

Methods

Participants

24 people, 11 men and 13 women, between the ages of 18 and 30 participated in the experiment. Prior to participation, prospective participants were asked to confirm that they met the study's inclusion criteria, i.e., that they did not have abnormal heart conditions (e.g., heart murmurs, pace makers, etc.), that they were not taking any drug that would alter normal circulatory system function (e.g., blood pressure medications, blood thinning medications, etc.), and that they had abstained from caffeine and nicotine for 12 hours prior to the study. Participants who did not meet these requirements were not allowed to participate in the experiment.

Experimental Design

The current experiment utilized a 4 × 2 mixed design. The between-subjects factor was experimental task (mental arithmetic, anagrams, card sort, treadmill). For the purposes of description and subsequent statistical analysis, tasks were also grouped based on their primary sources of demand; mental arithmetic and anagrams required cognitive resources for task performance, and the card sort and treadmill tasks required physical resources. In the two physical task conditions, the within-subjects factor was trial (first, second). For the cognitive tasks, participants also completed 2 experimental trials, but these trials were further differentiated by a manipulation of task difficulty (easy, hard). All experimental trials were 20 minutes in duration. The primary dependent measure examined for this manuscript was the β values of the power spectral density analyses derived from participant heart inter-beat intervals (IBIs).

Cognitive Tasks

Two cognitive tasks (mental arithmetic and anagrams) were selected for this experiment as exemplars of task processes (i.e., working memory, mental computation, and verbal ability) utilized by many people during their normal work activities. In addition, these two types of tasks (i.e. mental calculation and verbal ability) have been used to evaluate the sensitivity and diagnosticity of more traditional measures of heart rate variability (e.g., Nickel & Nachreiner, 2003). The present study also featured a manipulation of task difficulty (easy, hard) between trials to examine the sensitivity of fractal analytic methods to variations in mental load. The order of the presentation (easy-hard or hard-easy) was counterbalanced across participants. A description of each task, including easy and hard conditions, follows below:

Mental arithmetic. In the easy condition, participants were required to calculate the sum of a pair of two digit numbers (e.g., 24 + 48) and input the answer using the keyboard. The hard condition required participants to divide a three digit number by a one digit number (e.g., 210 / 7). All answers in the hard condition were whole numbers. In both conditions, participants were free to respond to the problems at their own pace (i.e., there was no maximum response time to items).

Anagrams. Word lists utilized for this task were drawn from *The Teacher's Word Book of 30,000 Words* (Thorndike and Lorge, 1944). The easy condition included common four letter words; the hard condition word list was drawn from common 6, 7, and 8 letter words. Scrambled words were presented to participants on a computer screen, and participants were required to type in the correct word (e.g., "enes" was presented to the participant, and the correct response was "seen"). Each easy presentation was timed so that participants had 7 seconds to view the scrambled word and input the correct response; in the hard condition, participants had 10 seconds.

Physical Tasks

Two tasks (a card sorting task and walking on a treadmill) were selected to simulate physical exercise as might be encountered by many people during their day-to-day work activities. These conditions were included to examine the sensitivity of fractal analytic methods in discriminating physical and mental sources of task load. As in the cognitive tasks, participants completed two 20-minute trials. However, the experimental task in each trial for the physical tasks was identical. This arrangement was selected to assess the effects of light fatigue on β values and to ensure that the duration of all tasks were comparable. A short description of each physical task is included below:

Card sort. The card sorting task was chosen to generate physical workload levels akin to light office work. Two tables were set up 10 feet apart from each other. Each table contained 8 decks of shuffled playing cards. Participants' task in this condition was to sort a deck of cards (by number) at one table, then move to the second table and sort a second deck of cards. This process (moving between tables and sorting cards) was repeated for the duration of the trial.

Treadmill. The treadmill task was designed to generate moderate physical workload. In the treadmill task, resting heart rate was established using averaged values from the baseline period. The task in this condition was to walk on the treadmill for the duration of the trial. The target heart rate for participants during each trial was 30% above resting baseline, and this goal was achieved by manipulating the speed of the treadmill, leaving the elevation setting at a 0° incline.

Heart IBI Recording

Heart rate information was collected using CleveMed Bio Radio devices sampling at 256 Hz. Custom software was written to detect R-wave peaks from the raw signal. This system required three surface electrodes attached to the participant to detect the electrocardiogram (ECG) signal. Electrodes were attached to the sternum in two places (the manubrium at the top of the sternum and the xiphoid process at the bottom of the sternum), and to the left clavicle (as a ground). The skin below the electrodes was prepared by cleaning each site with alcohol, and then applying NuPrep ECG skin abrasion gel to the site, gently wiping the skin with gauze, and finally applying the electrode pads. Heart rate data was recorded to text files and then imported into Matlab and Microsoft Excel for subsequent analysis.

Procedure

Participants first completed an informed consent document. Next, participants were asked to confirm that they met the study's inclusion criteria. Those who did not meet these requirements were not allowed to participate in the experiment. Participants were then assigned at random to an experimental task condition. The basic procedure for all experimental conditions was to first apply the electrodes and attach the heart rate monitor. Participants were then asked to sit quietly while a 5-minute baseline was recorded. Next, participants performed their assigned task during the first 20-minute experimental trial. Participants were then given a 15 minute rest break. Following the break, participants again completed a five-minute baseline and their experiment task.

Results

Spectral Analysis

All data files were imported into Microsoft Excel and cleaned for erroneous data points due to artifacts associated with the wireless radio devices. The final 1024 IBI intervals were taken from each time series and a segmented Fast Fourier Transform (FFT) with a Triangle window was performed on the data using Matlab software. The FFT segmented the data such that there were four unique 256 point analyses, and three overlapping 256 point analyses, with the average power at each frequency used for the final slope calculation. The resulting power spectral density for each participant was then plotted and β values were generated by calculating the regression slope for each participant.

Manipulation Check

As a manipulation check, the cognitive and physical tasks were examined for differences due to fatigue between the first and second experimental trials using separate paired sample *t*-tests (one for experimental task). Results of the analyses indicated there were no statistically significant differences between trials ($p > .05$ for all tasks), suggesting that fatigue did not influence spectral slope values in this experiment.

Task Difficulty

To examine spectral slope values for differences associated with task difficulty in the two cognitive tasks, a 2 (task) × 2 (task difficulty) analysis of variance (ANOVA) was computed. The results of the analysis revealed no statistically significant sources of variance in the analysis (all $p > .05$). These results indicate that the spectral slope values observed were similar across the cognitive tasks and task difficulties employed in this experiment.

Task Comparison

Given that the previously discussed analyses did not reveal statistically significant differences between trials due to experimentally manipulated factors, mean spectral slopes were calculated for each participant across experimental trials. As an index of the sensitivity of spectral slope values to task type, mean slope values for all experimental tasks were compared in a one-way ANOVA. The purpose of the analysis was to determine if spectral slope could be used to distinguish between primarily cognitive and physical sources of workload. Results supporting the sensitivity of spectral slope values would include a statistically significant main effect of task type, followed by statistically significant post hoc tests demonstrating differences between the cognitive and physical tasks.

Analysis of the spectral slope data revealed a main effect of task type, $F\ (3, 20) = 3.88, p < .05$. A follow-up post hoc Tukey HSD test indicated a statistically significant difference in spectral slopes between the treadmill and anagram tasks, and between the treadmill and mental arithmetic tasks. In addition, a trend in the data suggested a difference between the treadmill and card sort tasks ($p < .10$). No other comparisons were statistically significant (all $p > .05$). Mean spectral slope values for each condition are presented in Table 1.

Table 1. Mean spectral slope values for each experimental task.

Task	*M*	*SE*
Cognitive Tasks		
Anagrams	1.95	.17
Mental Arithmetic	1.88	.18
Physical Tasks		
Card Sort	1.86	.22
Treadmill	.97	.35

Note. M = mean, *SE* = standard error

Discussion

The goal of the present study was to examine the utility of a fractal analytic method, spectral slope analysis, as an index of workload. The results provide some interesting insight into the use of fractal measures of heart rate variability for assessing cognitive workload. We initially hypothesized that the fractal exponents associated with primarily cognitive tasks would be differentiable from those associated with physical tasks. This hypothesis was partially supported, in that both cognitive tasks had greater spectral slopes than those observed in the treadmill task. However, spectral slopes did not differ between the cognitive tasks and the card sort task. In addition, we hypothesized that task difficulty would influence fractal exponents such that more difficult tasks would be associated with increased fractal exponent values. This hypothesis was not supported; no differences were observed between the easy and hard task difficulty conditions for either of the cognitive tasks. It is again worth noting that these results are preliminary; our final results may differ from those discussed here.

The results concerning task type were more nuanced than initially hypothesized. Spectral slopes in the treadmill task differed from all other tasks, but no such differences were observed between the other three tasks. This seems to suggest that the fractal metric employed may not be sensitive to sources of workload (cognitive versus physical), but that it is sensitive to large differences in workload regardless of source. Although the current data seem to indicate that fractal techniques do not discriminate between physical and cognitive workload, the current data do suggest that light workload (i.e. using a computer, light office work) should not be a major concern in future experiments examining cognitive workload using fractal methods, as there were no significant differences between the card sorting task and the cognitive tasks examined in this experiment. This information is important for future research, as cognitive workload experiments are likely to include low levels of physical activity.

As mentioned previously, the results of the task difficulty manipulation employed in the current experiment did not match initial hypotheses in that the cognitive task conditions did not appear to generate changes in β values. This is at odds with the results reported by Nickel & Nachreiner (2003), who found mean IBI differences between verbal and other cognitive tasks. The results observed in the current experiment may be due to the manipulations themselves (i.e., the manipulations of task difficulty employed were insufficient to drive changes in β), or to relatively low power in these analyses (as overall N was relatively small). Alternatively, it may be that, unlike other measures of heart rate variability, β values do not lend themselves to observable differences based on workload manipulations (i.e., while IBIs follow power law scaling, β values may be relatively stable across individuals and most situations). Further research will be necessary to determine which of these alternative explanations is most accurate.

The results of this study indicate a strong need for future research. As previously mentioned, increasing the sample size reported here should provide greater statistical power for our analyses. To better understand the stability of β values under levels of workload, future research should explore the relationship of fractal heart rate variability with a greater diversity of workload levels and also in conditions more akin to "real world" tasks. More specifically, workload levels should include higher levels of cognitive workload than were employed in the present study, as our current results suggest that fractal measures appear to be sensitive only to large differences in workload. Furthermore, future research should also examine if fractal measures of workload are sensitive to changes between resting and working heart rates. Although baseline data was collected in this experiment, the duration of the baseline (5 minutes) provided an insufficient number of IBIs for a spectral analysis of the baseline period. Finally, future research should better address the issue of fatigue and cognitive workload. Although intertrial differences were not observed in the present experiment, it is likely that fatigue will exert some influence on fractal measures, provided that the task is of sufficient duration.

Overall, the (preliminary) results of the current experiment suggest that fractal analysis may yield useful metrics for understanding workload. At present, however, it appears that these metrics are unlikely to be more sensitive than those reported by previous researchers in the area (e.g., Kalsbeek, 1971; Kalsbeek & Ettema, 1963).

References

Gleick, J. (1999). *Chaos: Making a new science.* New York, NY: Penguin.

Goldberger, A., Amaral, L., Hausdorff, J., Ivanov, P., Peng., C., & Stanley H.E. (2002), Fractal dynamics in physiology: Alterations with disease and aging. *Proceedings of the National Academy of Sciences, 99*, 2466-2472.

Iannaccone, P. M. & Khokha, M. (Eds.) (1996). *Fractal Geometry in Biological Systems.* Boca Raton FL: CRC Press.

Ivanov, P., Amaral, L., Goldberger, A., Havlin S., Rosenblum, M., Struzik, Z., & Stanley, H. E. (1999), Multifractality in human heartbeat dynamics. *Nature, 399,* 461-465.

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.

Malamud, B.D. & Turcotte D.L. (1999). Self-affine time series: I. Generation and analyses. *Advances in Geophysics, 40,* 1-90.

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.

Nickel, P. & Nachreiner, F. (2003). Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload. *Human Factors, 45,* 575-590.

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.

Russell, S.M., Funke, G.J., Knott, C.C., & Knott B.A. (2009). Fractal time series analysis of human heartbeat intervals in a change blindness task. Proceedings of the International Society of Aviation Psychology. Dayton, OH, April 2009.

Ryan, S., Goldberger, A., Pincus, S., Mietus, J., & Lipsitz, L. (1994). Gender and age related differences in heart rate dynamics: Are women more complex than men?, *Journal of the American College of Cardiology, 24,* 1700-1707.

THE EFFECT OF MEASURING SITUATION AWARENESS ON PILOT BEHAVIOR

J. van Eijck, C. Borst, M. Mulder, M. M. (René) van Paassen
Delft University of Technology, Faculty of Aerospace Engineering
Kluijverweg 1, 2629HS Delft, The Netherlands

When developing new human-machine interfaces for aircraft cockpits, the contribution of the interface to the pilot's situation awareness is of interest because this might influence human decision-making. Besides subjective measures, a frequently used objective method of measuring SA is the freezing of the task, removing information from the displays and querying the pilot about the situation. Research has indicated that by doing so, the pilot's notion of what is important during this task may change, thereby potentially influencing decision-making. This paper describes an experiment in which pilots are given a terrain avoidance task, either with or without an interruption and SA query. The results showed a reorientation of the pilot's gaze without having a significant effect on the decision-making.

During the past decades, the number of automated systems in modern aircraft cockpits has increased significantly. Although the number of physical displays in the pilot's work environment decreased, the amount of information that needed to be presented on them increased. This has inevitably led to interfaces that are more complex and the question is often raised if the pilot is able to correctly interpret all the data that is presented. In other words: will the pilot still have sufficient Situation Awareness (SA)?

The concept of SA has been the subject of many discussions and therefore many different definitions have been developed (Uhlarik & Comerford, 2002). These definitions all seem to lack on some aspect of what is believed to be SA. Many definitions seem to neglect the dynamic nature of SA, while others provide no means of how to measure it. To further illustrate the disagreement about these different definitions: some researchers believe that more emphasis is required on the 'situation' aspect of SA (Flach, 1994). Others believe that the 'situation' part is well-defined, but 'awareness' requires more research (Sarter & Woods, 1991). A unified definition of SA appears to be far away, if not impossible. However, the most widely accepted definition of SA is the one by Endsley, who defines SA as *the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future* (Endsley, 1995). Although Endsley neglects to define what these meaningful elements are, she does provide the means to measure the SA using perception, comprehension and projection of the future status of the situation. Not only the definition of SA causes much discussion, but also its relation to human error is still not clear. Considering low SA as the cause of human error would only lead to circular reasoning (Flach, 1995). This also has implications on how to measure SA.

This paper describes an experiment that evaluates a variant of the freeze technique. The largest advantage of the freeze technique is that it is believed to be an objective method to assess the subject's SA. However, there is a risk of changing the subject's notion of what the 'meaningful elements' in a particular situation are. Therefore an experiment is performed to investigate the effect of the method on the subject's behavior. Similar research was performed in the past by McGowan and Banbury, but they tried to assess the effect of the method on the subject's SA, instead of behavior (McGowan & Banbury, 2004). Because there is no accepted measure for SA available, it is not possible to investigate the effect of the freeze technique on a subject's SA.

Experiment

The focus of the experiment lies on investigating the effect of a query on the attention distribution of the pilot and the decisions that the pilot makes. The experiment is designed around a terrain avoidance task. To perform this task, pilots will use a Synthetic Vision Display (SVD) enhanced with overlays that depicted the aircraft maneuvering opportunities relative to the terrain (Borst, Mulder, & Paassen, 2010). The items of the SA query addressed the elements that could be perceived from the flight display.

Subjects and instructions

A total of thirteen professional airline pilots participated in the experiment. With one of the pilots, the eye-tracker failed to produce reliable data, therefore data of only twelve pilots is used during the analysis. Half of the subjects was presented the SA query during the measurement runs, while the other subjects were not given this query.

The subjects were not briefed about the true goal of the experiment, as this might influence their behavior during the experiment. The pilots that were presented the SA query were given the task description to first observe the flight situation and avoid terrain obstacles. The pilots were instructed to select either a straight climb, climbing turn (left or right) or level turn (left or right) as escape maneuvers. Furthermore, the pilots were asked to deviate as little from their original course as possible. A straight climb was preferred above all other options and a level turn should only be made if no other solution is possible. Finally, pilots were not given any minimal separation with the terrain.

Independent variables

The presence of the SA query (QUERY) was introduced as a between-subjects independent variable with two levels. In addition, since multiple scenarios were used for every pilot, the scenario (SCENE) was introduced as a within-subjects independent variable with ten levels.

Dependent measures

Decision-making The pilots were instructed to select one of the five possible actions to avoid the terrain ahead of them and deviate as little as possible from their initial heading. They could either perform a straight climb, climbing turn (left or right) or a level turn (left or right). In addition, the pilot was instructed to deviate as little from the initial course as possible. This means that if a straight climb was possible to safely avoid the terrain ahead, this was considered the best option. With a level turn, the course deviation would be largest. This option should therefore only be chosen if no other possibility exists to avoid a collision with the terrain ahead. The quality of the decision was rated optimal or non-optimal. The decision taken by the pilot is of interest, because pilots that were presented a query may make different decisions than the ones that were not given this query. It is hypothesized that when the pilot is forced by the query to think through multiple options, this might have an effect on the resulting decision.

Pilot gaze A faceLAB eye-tracker is installed in the simulator for tracking the pilot gaze. The eye-tracker consists of a set of two cameras with infrared filters and three infrared sources. The infrared sources provide a constant illumination on the pilot's face, which can only be made visible with the infrared filters on the cameras. A schematic overview of the eye-tracker and screen set-up is given in Figure 1. Because it is known that attention is related to SA, the distribution of attention of the pilot is of interest as well. If the distribution of attention changes as a result of the SA query, this means that this method is influencing the SA that it intends to measure. The pilot gaze direction can be determined and combined with a model of the test environment. Although it is well known that it is dangerous to draw conclusions about direction of attention based on gaze direction, the SA query is hypothesized to force the pilot to consciously focus on certain aspects of the display to perform well on the query.

Eye-tracking needs to be done very accurately, because the part of the test environment that is of interest here is the screen with the interface, which contains a lot of information on a small surface area. The different aspects of the interface need to be distinguishable from each other. High accuracy with this eye-tracker is achievable when it is well calibrated. The different elements of interest are displayed in Figure 2. Most of the identified elements are static elements (e.g. altitude and speed indicators are always located on the same place), but some have a dynamic nature and these depend on the aircraft state at a specific point in time: the flight path vector and the expanding collision box. The calculated positions of these dynamic elements are continuously updated such that these data can be correlated with the output of the eye-tracker. The eye-tracker data considered in the analysis is the data obtained during the pre-freeze phase of every simulation run. If the query would affect the pilot's behavior, this would be best noticeable during the simulation phase before the query is presented. In addition, since the simulation runs do not have a fixed time span, this results in equally long measuring times for all runs.

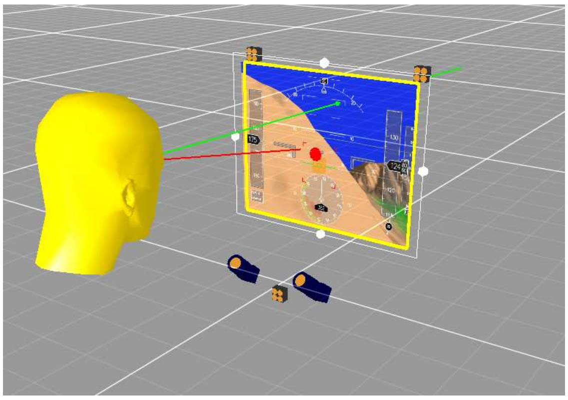


Figure 1: Test set-up with screen and eye-tracker.

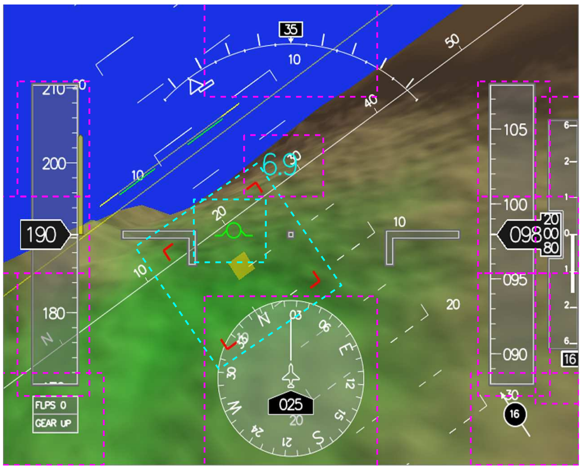


Figure 2: Relevant elements for eye tracking.

In order to eliminate learning effects and individual differences from the data, the average time percentage per interface element over the last three training scenarios was used as a reference for gaze direction without the influence of a query. The same was done for the last three measurement runs and the difference between the two values is computed as a measure for change of gaze direction during the experiment. For the pilots without the query, this would be influenced by learning effects, while for the pilots that are given the query this change in gaze direction might have been caused by both learning effects and the effect of the SA query. By measuring the change of gaze direction instead of actual gaze direction in terms of time percentage per element, both individual differences and learning effects will be removed from the analysis.

The accuracy of the eye-tracker was evaluated between every four runs. This was done by means of a moving square on the screen that the pilot was requested to track. Both gaze-screen intersection coordinates and square position coordinates were logged and from the difference, the accuracy of the eye-tracker could be computed. When the eye-tracker accuracy during this test appeared to be of low quality, the subsequent simulation runs were discarded from the data.

SA query performance The effect of conditioning a pilot with the SA query on the performance on this query was also researched. During the last run of the experiment, also the pilots of the group without the query were given the SA query to see if they performed worse on answering these questions than the group of pilots that had been given this query during all the runs. It is expected that the latter will perform better on the query, although both groups have the same amount of experience with the display.

Apparatus

The experiment was performed in the Human Machine Interaction (HMI) laboratory at the Faculty of Aerospace Engineering of the Delft University of Technology. The HMI-laboratory is a fixed-base flight simulator in a dark room. Two 18” LCD monitors were in front of the pilot, one showing the ESVD interface, the other showing the engine parameters. The aircraft model that was used during the simulation was a non-linear Cessna Citation 500 model. The aircraft was controlled using a control-loaded side stick on the right-hand side of the pilot and rudder pedals. The pilot was also able to adjust throttle and flap setting. The initial flap settings were randomly selected for every scenario as well as the wind conditions. Wind speed and direction remained constant during each run.

Results

Decision-making

The effect of QUERY on the pilot's decision-making was analyzed using Cochran's Q-test. There appeared to be no significant effect of QUERY on the decisions made by the pilot (Cochran, $Q(1, 60) = 2.0, p = 0.16$). Figure 3 shows that pilots performing the experiment with a query took slightly less optimal decisions than pilots that were not given the query.

Pilot gaze

The pilot gaze fixations on the elements defined in Figure 2 are expressed as a percentage of the thirteen seconds before performing a maneuver or filling in the query. The average gaze distribution during the last three training scenarios is used as the undisturbed gaze pattern for each pilot. The average gaze distribution during the last three measurement scenarios is influenced by learning effects and, for the group given the query, possibly the SA query. The questions in this query that test the SA level 1 of the pilots require them to remember values that are presented on the display. It is therefore hypothesized that the query will focus the pilot's gaze to the airspeed and altitude tapes, the compass rose and the wind speed and flaps/gear indicators.

The average gaze intersections for the sum of all these elements are calculated for both the training phase and measurement phase. Then, the difference between these two averages is calculated to see if the pilots who were given a query during the experiment have changed their gaze direction more than pilots that were not given this query. Based on these elements, the pilots with the query tended to adapt their gaze distribution significantly more during the experiment than the pilots without the query ($U(1, 6) = 3.0, p = 0.016$). When decomposing this again in these different elements, it can be observed in Figure 4 that pilots from the query-group were on average focusing more on all of these elements, while Figure 5 shows that pilots who were not given this query tended to focus less on all of these elements.

The differences between training and measurement runs that are visible in Figures 4 and 5 were also analyzed using the Mann-Whitney test. There appeared to be no significant effect between training runs and measurement runs for any of the elements when no query is given to the pilot ($U(1, 6) > 130.0, p > 0.30$), while the addition of a query results in significant effects for all of the elements ($U(1, 6) < 90.0, p < 0.015$). These results indicate that the query significantly influences the pilot gaze direction. Screen shots with the gaze intersection points for two pilots of different groups in Figures 6 and 7 visualize the gaze reorientation due to the query.

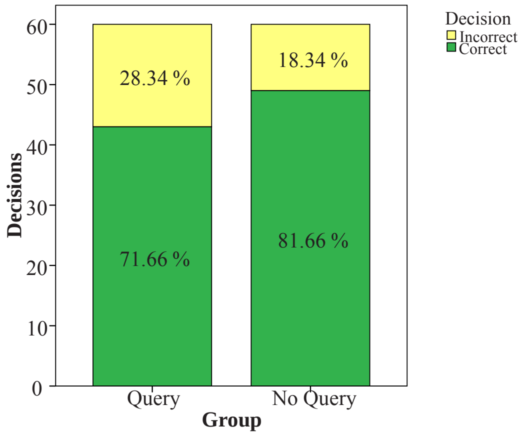


Figure 3: Decision-making per group of pilots.

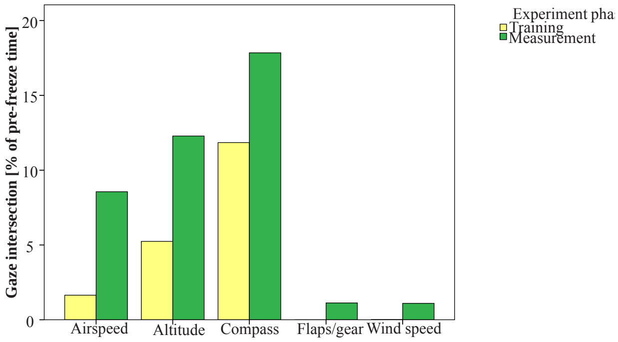


Figure 4: Gaze-screen intersection change during the experiment for pilots with the query.

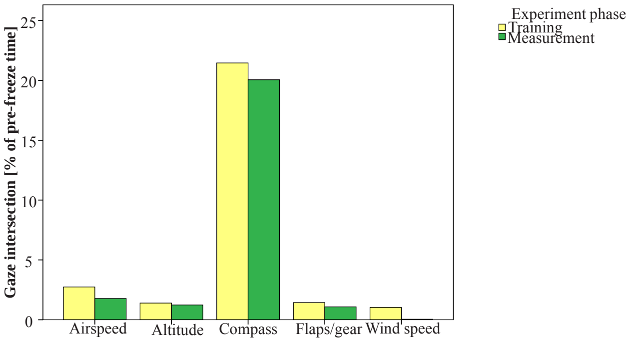
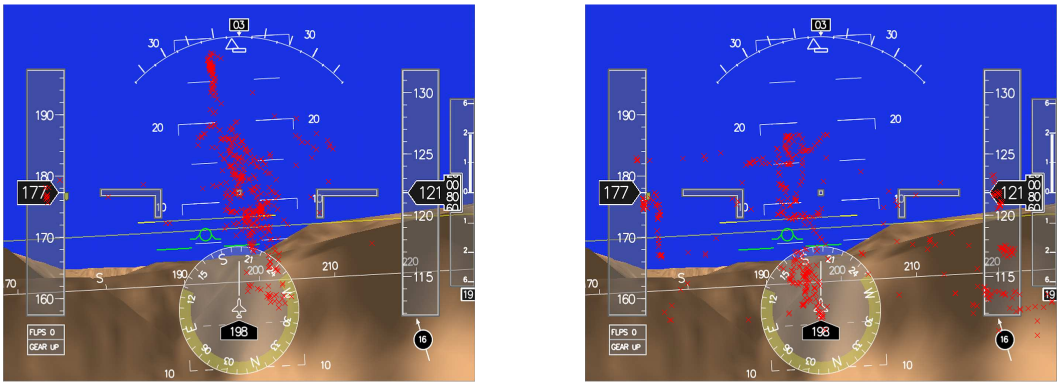


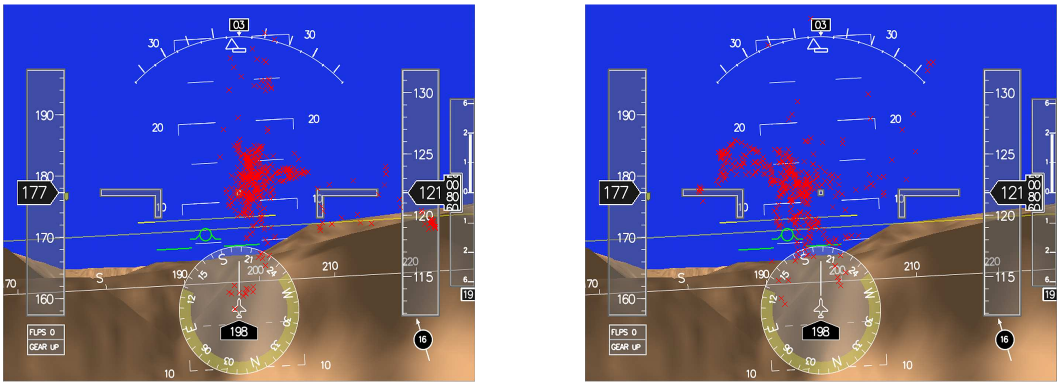
Figure 5: Gaze-screen intersection change during the experiment for pilots without the query.



(a) During last training scenario

(b) During last measurement scenario

Figure 6: Gaze distributions for a pilot who was subjected to a query.



(a) During last training scenario

(b) During last measurement scenario

Figure 7: Gaze distributions for a pilot who was not subjected to a query.

SA query performance

The scores on the SA query during the last measurement run showed no significant difference between the two groups of pilots ($U(1,6) = 18.0, p = 1.0$), but when decomposing the scores according to the levels of SA by Endsley, some differences can be observed. The average partial and total SA scores are presented in Figure 8. Although not significantly ($U(1,6) = 6.5, p = 0.056$), the scores for SA level 1 are higher when the pilot knows what information is requested by the query. The query has a negative, but much smaller effect on scores for SA level 2 ($U(1,6) = 13.0, p = 0.42$) and SA level 3 ($U(1,6) = 9.5, p = 0.16$). The small positive effect of the query on SA level 1 is apparently compensated by small negative effects on SA level 2 and SA level 3, such that no effect can be observed on the total SA score.

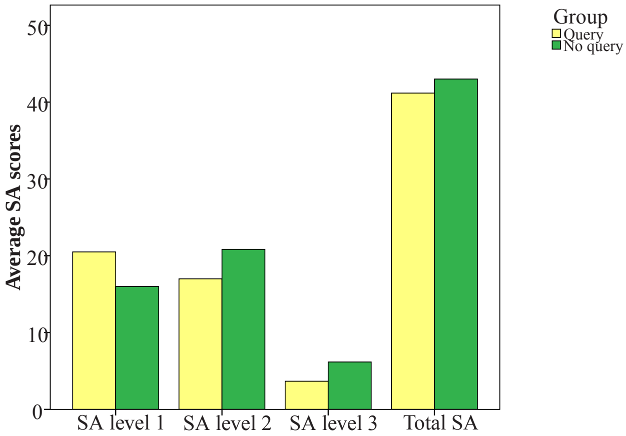


Figure 8: SA query scores for both groups of pilots.

Conclusions

An experiment was conducted to investigate the possible intrusive effect of freeze techniques to measure situation awareness on the behavior of the pilot. The results indicated that pilots were indeed reorienting their gaze on the flight display to improve their performance on the SA query. Despite this reorientation, decision-making was not significantly affected as well as the total measured SA score. The conclusion we can draw here is that queries can be intrusive in terms of pilot scanning behavior, but also that they do not necessarily affect decision-making.

References

Borst, C., Mulder, M., & Paassen, M. M. van. (2010). Design and Simulator Evaluation of an Ecological Synthetic Vision Display. *Journal of Guidance, Control, and Dynamics*, *33*(5), 1577–1591.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, *37*(1), 32-64.

Flach, J. M. (1994). Situation Awareness: The Emperor's New Clothes. In M. Mouloua & R. Parasuraman (Eds.), *Human Performance in Automated Systems: Current Research and Trends.* Hillsdale.

Flach, J. M. (1995). Situation awareness: Proceed with Caution. *Human Factors*, *37*(1), 149-157.

McGowan, A. M., & Banbury, S. P. (2004). Evaluating Interruption-Based Techniques using Embedded Measures of Driver Anticipation. In S. P. Banbury & T. S (Eds.), *A Cognitive Approach to Situation Awareness* (p. 176-192). Ashgate.

Sarter, N. B., & Woods, D. D. (1991). Situation Awareness: A Critical but Ill-Defined Phenomenon. *The International Journal of Aviation Psychology*, *1*(1), 45-57.

Uhlarik, J., & Comerford, D. A. (2002). *A Review of Situation Awareness Literature Relevant to Pilot Surveillance Functions* (Tech. Rep.). Manhattan KS Kansas State University.

USE OF DIGITAL PERFORMANCE DATA IN THE FLIGHT TRAINING ENVIRONMENT

Lauren M. Vala
Michael S. Nolan
Brian G. Dillman
John P. Young
Department of Aviation Technology, Purdue University
West Lafayette, Indiana, USA

The ability to record and monitor flight data in primary training aircraft has dramatically changed in the last decade. By taking advantage of digital data recording equipment on single engine aircraft, the implementation of Flight Operations Quality Assurance (FOQA) programs is now feasible for non-airline operations. The integration of FOQA data into training and evaluation will provide multiple opportunities for utilization of this data in a variety of applications. Differences between the operational environments of airlines and collegiate training programs must be addressed if FOQA is to be properly integrated into a collegiate training environment. Through interviews of key university, technology, and regulatory personnel, protocols will be developed that will aid in the establishment of a collegiate FOQA program.

Flight Operations Quality Assurance (FOQA) programs are voluntary safety programs which use aggregate flight data to identify unsafe flight conditions or deviations from policy. In recent years, airlines have developed and successfully implemented these programs (FAA, 2004). FOQA programs are crucial for safety systems and can enhance training operations, safety and efficiency (Ramsey, 2005). The planned research will suggest implementation methodology for a university to establish its own functional FOQA program.

Background

As digital aircraft enter the general aviation market, numerous safety advancements are now possible. Developing a collegiate FOQA program has the potential to optimize the use of the data collected from aircraft. Evidence-based training such as Advanced Qualifications Programs (AQPs) or similar initiatives are now possible. Efficiency may be improved by student-tailored training made possible from the collection of user-specific data. Instead of utilizing uniform training procedures prescribed under standard Federal Aviation Regulations (FARs), FOQA data allows for individual training and skill improvement. This process can enhance each student's flight training experience as well as increase efficiency for the university and individual student.

The development of FOQA programs for collegiate flight schools is of high importance when considering the improved safety initiatives currently undertaken by the FAA. Training redesign has already occurred for commercial pilots through successful FOQA implementation, but efforts have not yet been made to convert such programs to general aviation pilot training. After careful comparison of commercial pilot and general aviation pilot training requirements, this project provides the guidelines for the implementation of a collegiate FOQA program. This information can aid university flight programs in development of their own FOQA programs. Improved safety and cost savings may be realized by flight departments if they so choose to develop a FOQA program based on the guidelines provided in this project.

FOQA System Operation

While determining the implementation requirements of a collegiate FOQA program, many areas highlighted by airline FOQA programs must be given appropriate attention. Specifically, the importance of a functioning safety culture and the available integration of FOQA into Advanced Qualification Programs (AQPs) are important. Also, data uses, security, and analysis must be approved by the FAA (FAA, 2004).

FOQA is a voluntary safety program that intends to make commercial aviation safer through the recording of objective, quantitative data gathering and analysis (Wiley, 2007; Mitchell, K., Sholy, B., & Stolzer, A., 2006; FAA, 2004; FSF, 1998). FOQA programs function primarily through the immense amount of data that is collected onboard an aircraft during flight. More advanced than the traditional flight data recorders (FDRs), quick access recorders (QARs) gather flight information that is available for analysis by software on a personal computer post-flight (FAA, 2004). Specialized processing and analysis software called Ground Data Replay and Analysis System (GDRAS) is used to convert information from the QAR during flight to usable data that is relevant to managers, pilots, and maintenance personnel (FAA, 2004). FOQA data differs from that gathered from an FDR in the amount of data recorded and purpose for data use. A standard FDR typically collects the last 25 hours of flight information leading up to an accident, and the data is then only accessed in the event of an accident (Wiley, 2007). A QAR for FOQA use records parameters at one second intervals, with data available for collection and analysis upon upload at the user's request. This electronic upload usually occurs between 3 and 20 operating days after the flight during which it was recorded, or during scheduled maintenance (FAA, 2004; Wiley, 2007).

The aforementioned data gathering processes should not occur as a standalone process, but rather be built into a total data gathering and analysis program. For airline purposes, the FAA (2004) lists multiple set-up phases for FOQA programs. These include the integration of a FOQA program into other systems within the aviation operation. Before a FOQA program or further safety management system can be launched at a university flight school, it must be determined if the cultural environment is in place to support it (Wiley, 2007). The FAA (2006b) states that, "the principles that make up the [Safety Management System] functions will not achieve their goals unless the people that make up that organization function together in a manner that promotes safe operation" (p. 4).

An airline FOQA program development guideline is available in Advisory Circular 120-82, which discusses the benefits, organization, and maintenance of such a program (FAA, 2004). This document also provides a template for the Implementation and Operations (I & O) plan development as well as key definitions that must be addressed during program establishment (FAA, 2004). In order to be fully operational in a collegiate flight school setting, a FOQA program must fit into the safety program goals and be supported by the college or university.

In order for FOQA data to be of use for a collegiate flight program, baselines must be established and caution must be taken in trending (Wiley, 2007). Wiley (2007) cautions that pilots must all operate under the same rules using the same tools, or else data collection could cause an apples to oranges type comparison. Routine Operational Measure (ROM) identification is a capability of the GDRAS system, which is the ability to find trends from which to later measure deviations. ROMs provide a snapshot look of a chosen parameter from which statistics such as mean, minimum and maximum can be determined (FAA, 2004). This information can lead to the establishment of baselines for normal operation (FAA, 2004). Establishing user-specific ROMs is a necessary part of the FOQA program adoption and set up.

Airline FOQA programs may attribute some of their success to confidentiality. On the other hand, collegiate FOQA programs must address different protocols for data protection due to their educational operation. The Family Education Rights and Privacy Act (FERPA) of 1974 protects the privacy of student educational records (U.S. DoE, 2010). Data collected from students in the collegiate training environment may be subject to protection under FERPA, necessitating a review of the laws and their applicability. Van Dusen (2010) states that education records are, "records that directly relate to a student and that are maintained by an educational agency or institution or by a party acting for the agency or institution." These documents may include written documents and computer media, but data compilation and administrative records kept exclusively by the creator of the records that are not accessible to anyone else are not considered educational records. The latter falls outside of the FERPA disclosure guidelines (Van Dusen, 2010). As data collection for FOQA purposes is a subject not directly addressed in the FERPA laws, special care must be taken by a flight school establishing a FOQA program in order to ensure that data use complies with legal rights of students.

Airlines have discovered that once FOQA programs are in place, additional programs can be developed to improve training (FAA, 2006a). The most developed program which uses FOQA data is the Advanced Qualifications Program (AQP), which again has only been developed for use by airlines (Wright, 2003). According to the FAA (2006a), "AQP is a systematic methodology for developing the content of training programs for air carrier crewmembers and dispatchers. It replaces programmed hours with proficiency-based training and evaluation derived from a detailed job task analysis that includes crew resource management." The goal of an AQP is to create the "highest possible standard of individual and crew performance" (FAA, 2006a, p. i). Traditional Federal Aviation Regulations (FARs) are prescriptive, stipulating minimal levels of required performance, knowledge, or skills to be demonstrated before pilots may be certified. AQPs take a different approach by utilizing feedback and evaluation to conduct proficiency-based training (FAA, 2006a). This feedback, however, can only be made possible through the use of reliable quantitative data. Airlines that have established AQP programs have first gained FAA approval for the use of FOQA data (FAA, 2006a). As both programs are non-regulatory, airlines that have taken initiative to develop them and receive FAA program approval have successfully met or exceeded FAR requirements. As traditional FOQA programs have created additional safety and training opportunities for airlines, potential is shown for collegiate FOQA programs to do the same for university flight training operations.

The possibilities FOQA programs offer are too beneficial to be ignored by collegiate flight school operations. However, the process of adapting FOQA programs to college flight needs might prove daunting and cumbersome for traditional operators. Guidance from previous system implementations may assist with collegiate FOQA development, but attention must be paid to the legalities of data collection which relate to collection of student data. With support from management and a solid safety culture in place, a data collection system can be developed and standardized. Hopefully, collegiate flight schools realize many of the same benefits FOQA has provided to the airlines.

Design Process

For this project, two main information sources were used. The Federal Aviation Administration's (FAA) airline FOQA program establishment guide was analyzed and professionals in the flight training field were consulted. Airline FOQA program requirements were evaluated and tailored for suggested use in a university setting. Timelines and implementation schedules as recommended by the FAA were modified to reflect university flight training needs. Advisory Circular 120-82 was referenced as a main template. After airline FOQA establishment guidelines were analyzed, it was necessary to collect a listing of university-specific training requirements for pilots. This information was important for the next and final project step which was the creation of guidelines for the establishment of a collegiate FOQA program.

The most influential document for this research was Advisory Circular No. 120-82 published by the FAA. This document is the standard for airlines to use when developing a FOQA program, and best directed the formulation of guidelines for general aviation FOQA development. To address data security and student privacy issues, the Family Educational Rights and Privacy Act (FERPA) was examined.

Data collection hardware vendors were sought for their expertise regarding the proper selection of data capturing units. Similarly, the Purdue University Information Technology (IT) department was questioned as to data collection unit installation and integration with current university systems. To ensure compliance with legal requirements regarding student records and FERPA policies, the Purdue University Registrar's Office was consulted. The advice of maintenance department management was also used regarding data requirements in their operations. Advice for many program establishment topics was sought from the Purdue Aviation Technology Department Head and director of operations for flight training, as they serve supervisory roles and have most direct authority over faculty in the aviation department. Lastly, Flight Standards District Office (FSDO) FAA employees from Indianapolis were consulted for their view on program integration in accordance with established FAA regulations. All data collection for the purpose of this research was conducted through interviews.

Concluding Remarks

This project will be completed by May 15, 2011. Final results will be published in a Directed Project for degree completion as per Purdue University's requirements for a Master's of Science in Aerospace Management. After this date, professionals with a research interest in this field may obtain final results of this project by contacting the lead author, Lauren Vala, at lvala@purdue.edu.

The possibilities FOQA programs offer are too beneficial to be ignored by university flight school operations. However, the process of adapting FOQA programs to university flight needs proves daunting and cumbersome for traditional operators. Guidance from previous systems may assist with collegiate FOQA development, but attention must be paid to the legalities of data collection which relate to collection of student data. With support from management and a solid safety culture in place, a data collection system can be developed and standardized. It is anticipated that university flight schools would provide similar benefits that airlines have realized from FOQA programs.

References

Block, E.E., Sabin, E.J., & Patankar, M.S. (2007). The structure of safety climate for accident free flight crews. *International Journal of Applied Aviation Studies, 7*(1), 46-59.

Federal Aviation Administration. (2006a). *Advanced Qualification Program.* (DOT Advisory Circular No. 120-54A). Washington, DC: U.S. Government Printing Office.

Federal Aviation Administration. (2006b). *Introduction to safety management systems for air operators.* (DOT Advisory Circular No. 120-92). Washington, DC: U.S. Government Printing Office.

Federal Aviation Administration. (2004). *Flight operational quality assurance.* (DOT Advisory Circular No. 120-82). Washington, DC: U.S. Government Printing Office.

Flight Safety Foundation. (1998). Aviation safety: U.S. efforts to implement flight operational quality assurance programs. *Flight Safety Digest, 17*(7-9), 1-54.

Mitchell, K., Sholy, B., & Stolzer, A. (2006, December 21). General aviation aircraft flight operations quality assurance: Overcoming the obstacles. *IEEE Aerospace and Electronic Systems Magazine, 22*(6), 9-15

Ramsey, J. (2005, December 15). FOQA: Training tool, and more. *Avionics, 29*(12).

United States Department of Education. (2010). *Family Educational Rights and Privacy Act (FERPA).* (20 U.S.C. 1232g; CFR Part 99). Washington, DC: U.S. Government Printing Office.

Van Dusen, W. R. (2010). *FERPA: Basic guidelines for faculty and staff, a simple step-by-step approach for compliance*. Vancouver, WA: Clark College.

Wiley, J. (2007, June). C-FOQA: Has its time arrived? *Business & Commercial Aviation, 100*(6), 76-82.

Wood, M., Dannatt, R., & Marshall, V. (2006, June). *Assessing institutional resilience: A useful guide for airline safety managers?* (Australian Transport Safety Bureau No. B2004/0240). Canberra City: Australian Capital Territory.

Wright, R. (2003, July). Changes in general aviation flight operations and their impact on system safety and flight training. Proceedings from AIAA/ICAS International Air and Space Symposium and Exposition: *The Next 100 Y*. Dayton, OH.

Appendix 1: Research Interview Questions

1. Should a collegiate flight program considering the establishment of a FOQA program first establish a steering committee? Who should be on this committee? What should the steering committee's function be?

2. How should a university aviation program go about establishing goals and objectives? Who should be in charge of developing goals and objectives? What should the goals of a collegiate FOQA program be? How should a FOQA program fit into the operational environment of a collegiate aviation program? What safety improvements should be addressed in the formation of goals and objectives?

3. How should stakeholders in a collegiate FOQA program be identified? Who might be some of the stakeholders in a collegiate FOQA program? What are the roles of each identified stakeholder in a FOQA program's daily operation?

4. What is the most appropriate and cost effective technology (hardware and software) needed for a university aviation program to operate a collegiate FOQA program? What are the aircraft requirements for operating a collegiate FOQA program? What steps are necessary for integrating data collection and analyzing technology into the established university technology structure?

5. What personnel need to be assigned to daily FOQA operation tasks? Do personnel need to be solely assigned to FOQA tasks or may they also have other roles in the university flight program? Should additional personnel be hired to perform or manage daily FOQA tasks? Which personnel should have access to identified student pilot information? Which university personnel need to have a working knowledge of FOQA programs and operation? Should student pilots in the program be briefed or trained on FOQA data collection and use?

6. What safeguards must be developed for a collegiate FOQA program to meet university requirements? What are the FERPA law implications for the collection and use of student data in a collegiate flight program? Are there any other pertinent policies or regulations regarding the collection and use of student flight data? What data security requirements must be met in order to ensure compliance with university and other requirements? Is there any reason student information should be de-identified from pilot records? Is there any specific training that individuals who will come into contact with student flight information should complete before working with flight data?

7. What critical events must be defined for data collection in a collegiate FOQA program? How should normal operating parameters for student training aircraft be determined? What critical events are necessary to define for maintenance personnel to conduct appropriate aircraft health monitoring? What is the most appropriate method for a maintenance department to receive health monitoring data?

8. Do FERPA laws require a signed agreement be on file for each student for which data will be collected? Does the FAA require any student-university agreements for the collection and use of FOQA data? What are student's rights in dealing with collected flight data? What are the FAA's rights in using flight data for enforcement or administrative purposes?

9. Should the airline Implementation and Operations (I&O) plan format published by the FAA be followed when developing a collegiate FOQA program? Who should be responsible for developing the collegiate I&O plan? How often should the document be reviewed and/or updated? Who should be charged with reviewing/updating the I&O plan? What are the college's rights in reviewing/updating a formal I&O plan?

DEVELOPMENT OF A MODEL OF AIRLINE CONSUMER SATISFACTION

Dr. Erin E. Bowen
Department of Technology Leadership and Innovation, Purdue University
West Lafayette, IN

Dr. Brent D. Bowen
Department of Aviation Technology, Purdue University
West Lafayette, IN

Dr. Dean E. Headley
W. Frank Barton School of Business, Wichita State University
Wichita, KS

Previous research on perceptions, satisfaction, and attitudes regarding the major commercial air carriers in the United States has provided little more than an interesting descriptive “snapshot” of the average air traveler. Building upon 20 years of work with the National Airline Quality Rating, the present study attempts to move beyond basic descriptive information of air travelers to identify attitudinal patterns and relationships in the way consumers at varying levels of travel frequency view the commercial air industry. Development of such a model allows key players the ability to improve their understanding of the prime drivers and perceptions of passenger behavior. The modeling of attitudinal patterns and perceptions plays an important role in determining the need and priority, and potential consequences of such action. This study will exemplify the connectivity between subjective measures as reported by the survey respondents, and the formula driven weighted average that constitutes the Airline Quality Rating.

The Airline Quality Rating

Developed in 1991 by Drs. Brent Bowen and Dean Headley, the Airline Quality Rating (AQR) debuted as an innovative, objective method of comparing airline quality on combined multiple performance criteria. Two decades of data have since been reported and published. Prior to the AQR, there was no consistent method for monitoring the quality of airlines on a timely, objective and comparable basis. The introduction of the AQR resulted in a multi-factor, weighted average approach that had not been previously utilized in the airline industry. The outcome is a rating for individual airlines with interval scale properties that is comparable across airlines and across time.

The Airline Quality Rating is a summary of month-by-month quality ratings for U.S. airlines that have at least 1% of domestic passenger volume. AQR scores for the calendar year are based on 15 elements in four major areas that focus on airline performance aspects important to air travel consumers. Using the Airline Quality Rating system of weighted averages and monthly performance data in the areas of on-time arrivals, involuntary denied boardings, mishandled baggage, and a combination of 12 customer complaint categories, airlines' comparative performance for the calendar year is reported.

Elements considered for inclusion in the AQR rating scale are screened to meet two basic criteria; 1) an element must be obtainable from published data sources for each airline; and 2) an element must have relevance to consumer concerns regarding airline quality. Data for the elements used in calculating the ratings represent performance aspects of airlines that are important to consumers. This information is calculated monthly from US Department of Transportation statistical reports and reported annually in a resulting research monograph (Bowen & Headley, 2010).

All of the elements reported in the Air Travel Consumer Report are based on data maintained by the U.S. Department of Transportation. The Airline Quality Rating criteria and the weighted average methodology allow a focused comparison of airline domestic performance. Unlike other consumer opinion approaches that rely on consumer surveys and subjective opinion, the AQR continues to use a mathematical formula that takes multiple weighted objective criteria into account in arriving at a single, fully comparable rating for airline industry performance. The Airline Quality Rating provides both consumers and industry watchers a means for looking at comparative quality for each airline on a timely basis, using objective, performance-based data. Over the years, the Airline Quality Rating has often been cited as an industry standard for comparing airline performance. With the continued global trend in airline operations alliances, the argument becomes even stronger for the Airline Quality Rating to be used as a standard method for comparing the quality of airline performance for international operations as well (AQR, 2011).

Airline Passenger Survey

Background

In response to airline consumer disappointment and an increased interest in the relationship between consumer perceptions and objective industry performance measures in recent years, a new feature of the Airline Quality Rating was integrated beginning in 2008. This feature was a new survey launched to gauge additional, subjective data particularly from both business and leisure travelers, with a focus on frequent fliers. Resulting data is aimed at providing the flying public a new perspective on airline travel. In addition, these consumer opinions serve as a validation of the AQR annual report and its contribution to industry trend analysis.

The flying public has, in recent years, sought Congressional intervention to aid in the turmoil that the airline industry continues to experience. “The public is turning to Congress for action, and that’s why we have a member of Congress encouraging us to conduct further research” (Airline Quality Plummets, 2009) stated AQR co-creator Dr. Brent Bowen. Nebraska Congressman Lee Terry, a member of the Committee on Energy and Commerce, as well as a sponsor of Airline Passenger Bill of Rights legislation, responded by raising questions. Terry is seeking readily available data with a widespread base, stating “I’m sure many of my colleagues in Congress will be interested in this information.” AQR co-creator Dr. Dean Headley adds, “It’s no surprise that frequent fliers are disgruntled. All elements of the air travel experience are getting worse, and the price is going up" (Airline Quality Plummets, 2009).

Both the FAA and Congressional representatives are seeking legislative or regulatory changes to commercial air travel with the intent of improving customer satisfaction with the airlines. Additional key players who stand to benefit from this research include decision makers at air carriers, charter services, and those involved with the transportation of air passengers at a variety of levels.

The Airline Passenger Survey (APS) attempts to move beyond basic descriptive information of air travelers provided by the annual Airline Quality Rating to identify attitudinal patterns and relationships in the way consumers at varying levels of travel frequency view the commercial air industry (Bowen, Bowen & Headley, 2010a). It is the intent of AQR researchers that the collective voices of the flying public be heard. Survey data is to be shared with Congress and the Department of Transportation, among other organizations.

Since its official debut in 2008, AQR.Aero, Inc.'s web-based survey, the Airline Passenger Survey, has been issued annually in conjunction with the release of the Airline Quality Rating. The APS is anchored on the AQR's official web site, www.aqr.aero. Survey items gathered from the APS include information from the flying public on airline preferences, perceived passenger friendliness, proposed Congressional intervention, satisfaction with the flight experience, and other issues of critical relevance to passengers and industry leaders as it regards the U.S.-based airlines ranked in the AQR.

Airline Passenger Survey respondents are primarily U.S. residents who visited the website www.aqr.aero and voluntarily participated in the survey. Between February 2008 and February 2010, over 8,000 unique responses from airline consumers were collected. Thus far, this multi-year survey has revealed numerous significant findings regarding consumers' opinions and perceptions of the airlines and their performance. This conceptual model of airline consumer satisfaction and the critical variables associated with it continues to develop within the framework of the variables comprising the Airline Quality Rating (Bowen, Bowen & Headley, 2010b).

APS Methodology

Airline Passenger Survey elements are developed and revised each year via the utilization of rigorous scientific methodology, with the intention of capturing the most important data from passengers while reducing confusion or variability in comprehension of questions. APS items are a combination of demographic variables, categorical data, and Likert-type scale responses asking participants to respond to evaluative statements regarding their perceptions of the current state of the airline industry.

The selection of survey items was based on a review of extant literature on the subject of air passenger satisfaction, current events in the aviation industry that are likely to affect the traveling public, and impending wide-reaching regulatory changes to the aviation industry (Bowen, Bowen & Headley, 2010b).

Airline Passenger Survey Questions

Airline Passenger Survey questions are comprised of both open- and closed-ended questions aimed at gaining a better understanding of the current passenger environment. For example, respondents were asked to provide anecdotal information regarding the summer travel crisis in Europe and its global implications. In the most currently available survey, participants are then asked which of four factors (arriving on time, no denied boarding, bags arriving with me, customer service) are most important to consider when booking a flight. Respondents are asked to select which airline is preferred and which is most passenger-friendly. Opinions are also sought as to whether air travel over the past year has improved, remained the same, or gotten worse. A Likert-style question regarding potential Congressional regulation of the US airline industry in an attempt to protect the rights of the flying public is also included, as this has been a topic of growing discussion among industry and Congressional leaders (AQR, 2011).

The APS is designed in such a way that allows data to be delineated among very frequent fliers (those who have flown more than 20 times in the past year) and those who fly less often. Results are also able to be differentiated between those who fly for business purposes versus those who are primarily categorized as leisure fliers. Demographic data such as gender and age are also obtained and assessed for comparison.

Once the survey has been completed, survey respondents are directed to both the current Airline Quality Rating annual report as well as other timely airline industry information.

Airline Passenger Survey Results

Resulting Airline Passenger Survey statistics have been published and are available at the AQR.Aero Inc.'s website, www.aqr.aero. Data from 2009-2010 is posted in .pdf format at http://www.aqr.aero/consumerdata/AQRConsumerData.pdf.

Airline Passenger Survey 2010 Primary Findings

APS 2010 data indicate that Southwest Airlines continues to be the " 'darling' of the flying public" (Bowen, Bowen & Headley, 2010a, page 3). The airline received the highest consumer ratings in every category assessed, surpassing competitors by a wide margin in consumer perceptions. In fact, Southwest Airlines' consumer preference rating (18.6%) made it the only clear "airline of choice" for the traveling public (Bowen, et al, 2010a). Consumer preferences for the other airlines clustered together at a much lower level.

Result data lends itself to the fact that passengers indeed are in favor of Congressional intervention to improve travel experiences. Nearly three-quarters of all APS respondents (74.2%) support such governmental involvement (Bowen, et al, 2010b).

The APS has revealed that, if consumers cannot complain directly to the airlines, passengers want an online, third-party complaint system available via consumer advocates. In response to this call, AQR researchers are developing and will be implementing an airline passenger complaint system into the website www.aqr.aero. This action is even more critical and timely, given the fact that many airlines are eliminating complaint hotlines or providing responses to emails.

Air travelers have made clear their strong opposition to a la carte fees for added services, preferring higher fares to after-purchase added fees. In fact, such a la carte fees are the last choice among passengers (Bowen, Bowen & Headley, 2010a).

Airline Passenger Survey 2010 Secondary Findings

APS findings revealed that the nation's regional air carriers uniformly fell at the bottom of the survey results with regard to consumer preference, perceived passenger friendliness, and use by business or leisure travelers.

In all cases, when the percentage of consumers who reported both preferring a particular airline as well as perceiving it to be most passenger-friendly was below 50% for a particular airline, the majority of consumers who preferred the airline but did *not* think it was most passenger-friendly reported Southwest as the most passenger-friendly airline. Continental is perceived as the most passenger-friendly network carrier.

On-time performance ranks highest among consumers, with 47.9% respondents reporting that it outweighs customer service, bags not getting lost, and no denied boarding. Additionally, customer expectations of the airlines continue to decline: 93.6% of travelers reported that air travel has either stayed the same or gotten worse over the past year (Bowen, Bowen & Headley, 2010).

Very frequent fliers.

The APS revealed that those deemed very frequent fliers (those averaging more than 20 flights per year) made up 14% of total passenger respondents. Most are college-educated (79.7% have at least a bachelor's degree) and the majority (87.5%) are men who are flying primarily for business purposes. Additionally, most earn between $100,000 - $150,000 annually (Bowen, Bowen & Headley, 2010a).

Casual fliers.

Representing 43% of APS respondents, casual fliers fly one to five times per year. Most earn under $100,000 annually and fly primarily for leisure purposes. They overwhelmingly view Southwest as both their most preferred and most passenger-friendly airline (Bowen, Bowen & Headley, 2010a).

Airline Passenger Survey Overall Consumer Rankings

Again, overall consumer rankings are available for public viewing at www.aqr.aero. A detailed statistical breakdown of information regarding preferred airlines is provided. Information is further delineated into network, regional and low-cost carriers. Data analysis may also be made by comparing the responses of frequent fliers and casual fliers. Airline-specific findings from AirTran to US Airways are also provided.

Conclusion

Airline Quality Rating researchers have responded to the requests and concerns of the flying public by creating a means for subjective factors relating to air travel to be captured, analyzed, and related to objective industry-wide data. The Airline Passenger Survey is effectively identifying attitudinal patterns and relationships which are sure to impact the airline industry. This research is laying the groundwork for communicating the needs of the flying public to government as well as the airline industry, and in developing a model that evaluates the relationships between consumer perceptions and their impact on industry performance. Future incorporation of an online, third-party complaint system on www.aqr.aero will serve as a consumer advocate tool. As a result, it is intended that customer satisfaction be improved in the future.

Acknowledgements

The authors thank Mary Fink, MPA for many years of support in AQR research and the final research support of this manuscript.

References

Airline quality plummets in summer; Frequent travelers disappointed; U. S. Congressman discusses Airline Quality Rating; New study launched (2009, May 22). *Ascribe Newswire.* Retrieved from http://www.ascribe.org.

Airline Quality Rating (2011). Retrieved from http://www.aqr.aero/aboutaqr.htm.

AQR Consumer Survey (2011). Purdue University: West Lafayette, IN. Retrieved from https://purdue.qualtrics.com/SE/?SID=SV_2m0rF7CVUE3q1s8&SVID=Prod

Bowen, B. D., & Headley, D. E. (2010). *Airline Quality Rating 2010.* Research Monograph, W. Frank Barton School of Business. Wichita, KS: Wichita State University.

Bowen, B. D., & Headley, D. E. (2008, May 18). *Airline Quality Rating: Media advisory*. Retrieved from http://www.aqr.aero/pressreleases/aqrprerelease051908.htm.

Bowen, B. D., Bowen, E. E., & Headley, D. E. (2010a). *AQR consumer satisfaction and preferences. Survey results: 2009-2010*. Wichita, KS: Wichita State University. Retrieved from http://www.aqr.aero/consumerdata/AQRConsumerData.pdf.

Bowen, B. D., Bowen, E. E., & Headley, D. E. (2010b, April). *Airline passenger satisfaction report 2010.* Wichita, KS: Wichita State University.

THE DIGITAL MIGRATION OF RESEARCH DISSEMINATION IN AVIATION PSYCHOLOGY DISCIPLINES

Brent D. Bowen
Department of Aviation Technology, Purdue University
West Lafayette, IN

Erin E. Bowen
Department of Technology Leadership and Innovation, Purdue University
West Lafayette, IN

Henry R. Lehrer
Department of Aviation Technology, Purdue University
Indianapolis, IN

John H. Mott
Department of Aviation Technology, Purdue University
Indianapolis, IN

Charles T. Watkinson
Purdue University Press
West Lafayette, IN

Mark P. Newton
Purdue University Libraries
West Lafayette, IN

Jennifer Kirschner
Department of Aviation Technology, Purdue University
West Lafayette, IN

Innovations in research dissemination have emerged over the last decade in the movement toward on-line digital materials and distribution by increasingly environmentally-friendly processes. The access to scholarship has often been limited to major research organizations capable of funding subscriptions, the costs of which have escalated to prohibitive values. Demonstrated herein is a model for world-wide open access to the latest contributions to the foundations of our discipline. The development of a systemic process to cross boundaries so that overall progress can result through the integration of research and industry practice at the individual level is provided. The foundational relationships and targeted outcomes are represented in an open conceptual design construct with intent to disseminate and transfer new knowledge resulting from research worldwide. The Open Access model is applied and represented in this paper.

Development of aviation as an academic discipline hinges on the ability of scholars to collaborate, discuss, and publicize the results of developmental and scientific research. Historically, the aviation discipline has struggled to find appropriate outlets that serve these purposes and provide a high level of scholarly support for those who seek to expand aviation science and technology (see Truitt & Kaps, 1995 for an early review of these issues). A forthcoming article by Bowen, Dyrenfurth, and van Epps provides an updated review of scholarly outlets in the aviation disciplines, and preliminary findings suggest that progress in the ability of aviation researchers to engage in high-quality knowledge sharing has not progressed as far as Truitt and Kaps would have anticipated by this time.

One significant challenge facing scholars in aviation and aviation-related disciplines is the high level of both scientists and practitioners achieving progress in the field. Unlike more traditional scientific endeavors which may rely heavily on an established body of academic-based researchers, aviation relies on active collaboration between a vast, globally dispersed collection of collegiate flight administrators/educators, commercial aviation industry leaders and employees, general aviation industry leaders and enthusiasts, and government/regulatory entities. This dispersion inhibits collective knowledge-sharing on a large scale, and means that gains by a particular agency, program, or industry collaborative may not be shared with others for many years (in this discussion we are of course excluding the gains of knowledge for competitive advantage, and focusing on knowledge driven for safety, training, or developmental gains). For example, aviation industry leaders attending the fourth Safety Across High-Consequence Industries conference in 2008 jointly agreed to the ideal that "we do not compete on safety" (Bowen & Bigda-Peyton, 2011); however, safety information gains from this and similar meetings may be shared in only limited fashion and with limited accessibility to those not immediately connected to these types of groups.

Background on the Development of Aviation Refereed Scholarship

The University Aviation Association (UAA), a professional organization that promotes aviation education as a collegiate academic discipline, initially presented scholarly papers at the association's Fall Educational Meeting. These papers were peer-reviewed for the first time in the mid-1980s by members of the publication committee and then appeared in the *Proceedings*. By approximately 1996, the *Proceedings* had evolved into The *Collegiate Aviation Review* (CAR), a refereed journal sponsored by UAA; the CAR is published twice annually in CD format and is still composed of papers from the fall meeting.

In 1989, a referred journal with multiple annual issues was conceptualized and launched at Embry-Riddle Aeronautical University (ERAU) in Daytona Beach, FL. With Henry R. Lehrer, a faculty member in the Aeronautical Science Department as the founding editor, *The Journal of Aviation/Aerospace Education & Research (JAAER)* published its first issue in April 1990. JAAER, a scholarly publication for educators and researchers as well as for professionals in the aviation and aerospace industry, had a primary focus on how the educational process influences various segments of the aviation and aerospace community and how education affects the industry. The *Journal* has been in continuous publication since the founding date with three issues per year.

The *Journal of Air Transportation Worldwide (JATWW)* was founded by Brent D. Bowen at the Aviation Institute of the University of Nebraska at Omaha in 1996; initial funding for this publication was through a grant from the NASA Space Grant. The *JATWW*, published three times annually, was established to be the preeminent scholarly journal in the aeronautical aspects of transportation. With an international and interdisciplinary emphasis, the journal focused on articles in all areas of aviation and space transportation research, theory, case study, practice, and issues. In addition, a key concentration of journal articles was in aviation administration and policy. *JATWW*, which was later re-named the *Journal of Air Transportation*, has not been published for several years.

Rationale for the Journal of Aviation Technology and Engineering (JATE)

Aviation is a growing industry, particularly in underdeveloped regions of the world. The International Air Transportation Association (2011) predicts that by 2014, the number of air passengers will increase from 2.5 billion to 3.3 billion; about half of these new passengers are expected to come from Asia Pacific countries, primarily China. Asia Pacific overtook North America this year as the largest aviation market and is expected to increase in size to 30% of global traffic by 2014. The Middle East, Africa, and Asia Pacific are the fastest growing markets for international travel currently. As these markets continue to expand at a rapid rate, the infrastructure and support systems necessary to handle them will need a corresponding increase. According to the International Civil Aviation Organization (2010), the world-wide annual demand for pilots is just under 50,000 while there is an annual training capacity for 47,000. This difference becomes striking in Africa, where 1,600 pilots are needed annually, but there exists the training capacity for only 175 in that continent. Latin America needs 3,600 pilots, but can only train slightly over 1,000 pilots annually.

The sudden increase in the demand for air travel stems from increasing economic development, an increasingly mobile population, and a more open political climate. The ability to travel large distances by air is still a relatively new phenomenon in certain parts of the world. As the viability of such modes of transportation increases and they become more commonplace, the cost of doing business in aviation should decrease, due in part to more competition and an increased focus on sustainability initiatives, opening up air travel to an even larger demographic (Cambridge & Whitelegg, 2008).

The purpose of the present paper is to outline a strategy currently in process for the creation of an Open Access, all-digital format for disseminating research in aviation and aviation-related disciplines. The global dispersion of aviation researchers, practitioners, and stakeholders, the high degree of variation in accessibility to print-format scholarly journals (with their concomitant cost concerns for many non-academics), and the growing need to share safety and performance gains in aviation from more developed to underdeveloped parts of the world who are interested in aviation all drive the critical need for this new journal.

Strategic Advantage of Open-Access Format

While a consensus on what actually constitutes an Open Access publication has been elusive, it is safe to say in the context of this paper that the term implies that information published in the Journal of Aviation Technology and Engineering (JATE) is freely available online throughout the world, for readers to read, download, copy, distribute, and use (with attribution) in any manner they wish. The costs of production are incurred by the producer rather than the user. A 2011 assessment of the landscape of Open Access publishing estimated that there are probably just under 3,000 active, fully Open Access English-language journals, publishing almost 120,000 journal articles per year. This means that about 8 to10% of all journal articles published in 2010 were first published in Open Access format. Many more become Open Access after an embargo period. Chemistry, physics, and technology journals (the category that would include JATE) account for 28% of the total number of Open Access articles published (Dallmeier-Tiessen, et al., 2011).

Open Access publication strategies are particularly effective when the information published is international in relevance, interdisciplinary in character, and of relevance to practitioners as well as academics. Because aviation technology is a field which exhibits these three characteristics, an Open Access approach to publishing JATE seems particularly attractive. These characteristics are worthy of further examination.

When compared to subscription-based journals, Open Access titles extend the global impact of published research, especially to under-developed countries such as Brazil and China where most universities cannot afford subscriptions (Evans & Reimer, 2009). As a recent study of citation patterns by country by information scientist Yanjun Zhang shows, “Open Access could effectively improve the articles’ impact in developing countries and contribute to decreasing the academic gap between developing countries and developed countries,” (Zhang, 2006, p. 155). Since aviation is an international industry, and authoritative information on issues in areas such as aviation safety is as much (perhaps even more) crucial in the developing as the developed world, Open Access publication seems an especially appropriate strategy. While the tracking period is too short so far to produce statistically significant numbers, the top ten sources for visitors to JATE’s online publishing platform (www.jateonline.org) have come from the USA, Australia, Latvia, China, India, the Philippines, Barbados, Brazil, France, and Ethiopia. It is encouraging that seven of these ten countries are classified by the World Bank as “low income” and this pattern of interest suggests that JATE will achieve its goal of extending access to “must have” literature beyond the privileged few.

A journal such as JATE publishes articles of interest to many different disciplinary communities. Generally, each discipline has its own journals, to which its associated communities and institutions subscribe. However, it is often difficult to discover and read information published outside one’s immediate field of study, a disadvantage that tends to lead to silos within the disciplines. By making content openly accessible, immediately available full-text from a Google Scholar search, serendipitous discovery is facilitated and broader-based knowledge is advanced.

Open Access titles also extend the reach of scholarly information to practitioners, educators, and entrepreneurs outside the major universities or large corporations that can afford to pay for subscription access to a large number of journals for their employees. The positive impact of Open Access literature on professional communities has been shown through an analysis of the increased citation of Open Access vs. subscription-based literature in professional and trade publications (Zhang, 2006). Such an increase in citations is, again, a relevant finding for the field of aviation technology, which is characterized by the participation of a diverse community of professionals from a range of different sizes and types of organization.

Despite these benefits, two particular challenges face journals that adopt an Open Access publishing strategy: respectability and sustainability. The perception that Open Access journals are of a lower quality is gradually fading. Many established journal publishers now offer both full Open Access journals and an Open Access option to authors who want to disseminate their work more widely than a subscription-based journal might otherwise allow. Springer and Oxford University Press are two established publishers producing academic resources in a wide range of subjects who have particularly well-articulated policies (http://www.springer.com/open+access and http://www.oxfordjournals.org/oxfordopen). Newer publishers, particularly in the biomedical sciences, have established a number of journals with high and growing impact factors over the last five years. Two notable examples are BioMed Central and Hindawi. While it does not endorse impact factors, the Public Library of Sciences (PLOS) is another highly regarded "all Open Access" publisher, and is responsible for possibly the largest journal in the world, PLoS One, which published 6,749 articles in 2010.

After the three year "proving period" that Thomson Reuters imposes on all journals that aim to be included in the best known citation indices, JATE will apply to be included and will hopefully receive an impact factor. Until then, the journal must clearly communicate a commitment to quality through a transparent, double-blind, peer review process, a distinguished editorial board, a professional appearance, and association with a publisher of proven quality. With this aim, the Journal has formed a publishing partnership with Purdue University Press, a small but highly-regarded scholarly publisher with extensive experience in publishing Open Access titles. In return for a modest fee, the Press provides access to manuscript management and publishing software, copyediting, typesetting, marketing, and publishing advice. JATE staff, meanwhile, project-manage the publication. The emphasis throughout is on a streamlined workflow and low overhead costs.

Because Open Access journals usually lack subscription income, most tend to generate revenue from their contributors, often through page charges. Since JATE aims to not only appeal to users from developing countries but also to welcome their insights and contributions, this approach was deemed unacceptable by its editors. JATE has therefore adopted a sponsorship or "affinity" model, drawing on support from the Raisbeck Foundation and the College of Technology at Purdue University. Because the Journal publishes much applied research of interest to practitioners, such support is more feasible than it would be for similar journals in other disciplines. This means of support, however, requires a clear understanding of the importance of editorial independence by the sponsor, and close adherence to best practices (Crow, 2009).

Research Dissemination through Creation of a Multi-National Journal

JATE, published by the Purdue University Press, is a biannual, Open Access, refereed publication serving the needs of collegiate and industrial scholars and researchers in the multidisciplinary fields of aviation technology and engineering. The Journal is primarily available to its readership electronically through a professional online publishing system, http://www.jateonline.org, although hard copies of individual issues can be printed on demand. As previously noted, JATE is partially supported by a gift from the James D. and Sherry L. Raisbeck Endowment at Purdue University (Lehrer & Mott, 2010).

The Raisbeck Engineering Distinguished Professorship for Engineering and Technology Integration was established in 1999 by James Raisbeck, founder of Raisbeck Engineering, Inc. and Purdue alumnus, within the Purdue School of Aeronautics and Astronautics. The goal of that professorship is "to bridge the School of Aeronautics and Astronautics and the Department of Aviation Technology to teach graduate and undergraduate students the art and science of mixing theory and application in the design, build, and test process" (Purdue, n.d., para. 2).

In keeping with this goal, a key focus of JATE is the promotion of the bridging of these fields by publishing scholarly articles related to the integration of theory and application within the design, build, and test process. This process generally consists of various constituencies working toward a common goal of an end product which is properly designed through theory and made practical through application (Lehrer & Mott, 2010).

JATE publishes both quantitative and qualitative research articles. Topics on which the Journal focuses include the technological and operational aspects of air carriers, general aviation businesses and airports, issues related to aviation maintenance and engineering, and aviation human factors and applied training research. Significant developmental and historical topics related to these focus areas are of relevance, as well (Lehrer & Mott, 2010).

Journal workflow is entirely electronic, substantially improving convenience for authors, reviewers, and editors and reducing the time required for peer reviews and editorial decisions. All submissions sent for peer review are reviewed by at least one Associate Editor and two additional reviewers. In a recent example, the electronic workflow process allowed a submission to be completely reviewed and an editorial decision made within five business days after requests for review were distributed and within eight business days after the submission was first received. In a global industry such as aviation, with many practitioners as well as academic researchers, the founders of JATE regard rapid, free access to evidence-based research as essential for the development of the discipline (Lehrer & Mott, 2010).

Establishing a Formal Network Structure for a Journal-Based Collaborative

Originating as an informal grouping of concerned parties, a collaborative community of scholars is evolving to be built around the more structured, technologically-based networking solution that an all-digital, Open Access scholarly journal provides. By placing the network structure within the context of a scholarly journal, relationships and advantages that have been built through previous generations of ad hoc collaboration may be formalized and shared with a wider audience.

As described previously by Metz (2007) and Bowen and Lu (2004), collaborative networks are formed to bring together synergistic relationships for the purpose of provisioning an optimal foundation from which to pursue common goals. Collaborative networks often begin through ad hoc information sharing; for example, the Safety Across High-Consequence Industries (SAHI) conference program sponsored by Saint Louis University grew from such ad hoc and informal sharing (Bowen, Block, & Patankar, 2009). Bowen and Block (2008) then proposed a virtual organizational structure to support sustainable collaboration for the SAHI contributors.

What is missing from these preliminary attempts at network collaboration, though, is the application of existing scholarly constructs to the new, virtually-mediated environment. The present program in development takes a concept familiar to both aviation academics and industry leaders/stakeholders, that of the scholarly, peer-reviewed research journal, and re-imagines it to provide grounding for the evolution of a new method of scholarly communication. The economic feasibility of Open Access scholarly publication enables the JATE to initiate an optimal forum for a sustainable and long-term collaborative network intended to bridge the gap between the disciplines of aviation technology and engineering. 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.

References

Bowen, E. E., & Bigda-Peyton, T. (2011). 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, 1,* 6-21.

Bowen, B. D., Block[*], E. E., & Patankar, M. (2009). A network collaborative design construct for the dissemination of aviation safety research. In *Proceedings of the 15th International Symposium on Aviation Psychology.* Dayton, OH: Association for Aviation Psychology.

Bowen, B. D., & 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 pp.*

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 pp.*

---

[*] E. E. Block is now E. E. Bowen and has published scholarly articles in the aviation and psychology literature under both names.

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.

Cambridge, H., & Whitelegg, J. (2008). The growth of aviation in Asia and challenges to sustainability. In J. Cochrane (Ed.), *Asian tourism: Growth and change* (pp. 117-130). Oxford: Elsevier.

Crow, R. (2009). *Income models for Open Access: An overview of current practice.* Washington, DC: Scholarly Publishing & Academic Resources Coalition. Retrieved from http://www.arl.org/sparc.

Dallmeier-Tiessen, S., Goerner, B., Darby, R., Hyppoelae, J., Igo-Kemenes, P., Kahn, D., ... van der Stelt, W. (2011). *The landscape of open access publishing today*. Paper presented at the SOAP Symposium, Berlin, January 13, 2011. Retrieved from http://edoc.mpg.de/524966.

Evans, J. A., & Reimer. J. (2009). Open access and global participation in science. *Science, 323*(5917), 1025. doi: 10.1126/science.1154562.

International Air Transport Association (2011, February 14). *Successful Vision 2050 meeting concludes – Building a sustainable future* [Press release]. Retrieved from http://www.iata.org/pressroom/pr/Pages/2011-02-14-01.aspx.

International Civil Aviation Organization. (2010). Answering the industry's call: The 2010 NGAP Symposium. *The ICAO Journal, 65*(5), 5-16.

Lehrer, H. R., & Mott, J. H. (2010). *About this Journal*. Retrieved from http://www.jateonline.org/about.html

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).

Purdue University School of Aeronautics & Astronautics. (n.d.). Raisbeck Engineering Design/Build/Test. Retrieved from https://engineering.purdue.edu/AAE/Academics/Courses/Raisbeck/

Truitt, L. J. & Kaps, R. W. (1995). Publishing aviation research: An interdisciplinary review of scholarly journals. *Journal of Studies in Technical Careers, 15*(4), 229-243.

Zhang, Y. (2006) The effect of Open Access on citation impact: A comparison study based on web citation analysis." *Libri, 56*(3), September 2006, 155. Retrieved from: http://librijournal.org/pdf/2006-3pp145-156.pdf

Acknowledgements

The presentation of this paper was sponsored in part by the Raisbeck Foundation Endowment, Purdue University Press and the resources of the College of Technology at Purdue University. For more information on joining or submitting to the Journal of Aviation Technology and Engineering, go to www.jateonline.org.

Development of an Advanced Net-Centric Communication Management Suite: Multi-Modal Communication

Victor Finomore[1], Kelly Satterfield[2], Courtney Castle[2], Dianne Popik[1], Ron Dallman[3], & John Stewart[4],
Air Force Research Laboratory, Wright-Patterson AFB[1];
Oak Ridge Institute for Science and Education, Wright-Patterson AFB[2];
Ball Aerospace, Wright-Patterson AFB[3]; Independent Consultant, Dayton OH[4]

Even with advanced collaborative technology, communication remains a critical component to the success of a mission. Command and Control (C2) operators rely heavily on radio and chat communication to efficiently plan, direct, coordinate, and control assets. With the shift towards network-centric warfare, standard radio communication needs to meet the needs of today's warfighter. A net-centric communication management suite called Multi-Modal Communication (MMC) has been developed to increase the performance and situational awareness of the operator while also alleviating the workload and errors associated with this communication intensive environment. This integrated system captures, displays, records, and archives radio and chat-based communication to better equip the warfighters by providing instant access to past transmission as well as increasing the intelligibility of current messages. This poster and demonstration explores the development and testing of these advanced tools as compared to standard radio and chat interfaces. This study examined the performance associated with monitoring multiple communication channels with access to different tools. Performance was analyzed in regards to message detection, response accuracy and time. Data showed that MMC provides a balance between the speed of radio listening and the accuracy and data-capturing capabilities of chat displays. MMC can be a beneficial tool to C2 operators in its ability to increase intelligibly while providing a persistent, searchable visual display of voice and chat communication.

COLLABORATION TECHNOLOGIES DECREASE RELIANCE ON RADIO COMMUNICATION IN SIMULATED AIR BATTLE MANAGEMENT

Adam J. Strang
Consortium Research Fellows Program, Alexandria, VA

Greg J. Funke, Sheldon Russell, Brent Miller, & Benjamin Knott
Air Force Research Laboratory, WPAFB, OH

Radio has been the dominant collaboration technology in military aviation, but may not be optimal owing to channel overload and static interference. The current study sought to examine how team communication was changed when radio was augmented by text-based chat, a virtual whiteboard, and two ecological resource displays in a simulated air defense task. To accomplish this, twenty-one five-person teams performed the simulation with goals to protect friendly assets, eliminate enemy aircraft, and conduct refueling operations. The joint availability of chat and the whiteboard increased the total number of team communications, but reduced team reliance on radio. Access to a graphical resource display decreased team communications and radio reliance. A content analysis indicated that access to chat and the whiteboard, and to the graphical resource display, allowed teams to communicate more effectively. Overall, these results support the utility of advanced collaboration technologies for enhancing team communication in military settings.

Working in teams is ubiquitous in military aviation. Because of this, the ability to receive, process, and transmit information effectively within and between teams is critical for performance and safety in aviation operations (Alberts & Hayes, 2003; Burke, Stagl, Salas, Pierce, & Kendall, 2006). Historically, radio has been the dominant technology used for team communication in these settings (Vidulich, Bolia, & Nelson, 2004). While effective, radio is limited by serial transmission (e.g., only one person can talk at a time), no archival log (useful for refreshing understanding of past or missed dialogue), static interference, and channel overload. Also, a more general concern is that overreliance on any single communication modality could negatively impact teams responses to dynamic changes in the operational environment (e.g., if radio communications are disrupted or compromised; Alberts & Hayes, 2003).

As alternatives to radio, emerging collaboration technologies, such as email, instant messaging, and virtual whiteboards, have begun to be integrated into military operations (e.g., Heacox, Moore, Morrison, & Yturralde, 2004). The availability of these tools not only alters the ways in which personnel communicate and collaborate, but also composition and proximal location of team members. For example, teams can now be geographically and temporally disbursed, as opposed to constrained to a central location. Based on this, it has been suggested that collaboration technologies will enable a degree of command decentralization, resulting in greater flexibility of military teams (e.g., Alberts & Hayes, 2003).

Utilization of these new collaboration technologies typically entails different costs and benefits for operators. For example, email and chat feature a log of previous communications, allowing users to offload responsibility and workload associated with retention of communication (Berry, 2006). Such technologies also provide for simultaneous and asynchronous communication (Bolstad & Endsley, 2003). A virtual whiteboards allows users to convey information through a spatial medium, potentially enhancing comprehension (e.g., spatial information has been shown to be communicated more effectively through visual, rather than verbal, media; Wickens, Vidulich, & Sandry-Garza, 1984). Finally, emerging collaboration technologies may reduce the need for operator communication. For example, whiteboard displays may increase team situation awareness of ongoing operations, but may also reduce communication regarding those events. Consequently, collaboration technologies that afford operators the ability to represent and transmit information in supplemental modalities may positively impact team performance.

It is also possible that collaboration technologies may increase operator workload if information is not easily accessible, if it creates an additional monitoring burden (Parasuraman & Riley, 1997), or if insufficient practice with the tool is provided (Funke & Galster, 2009). Moreover, new technologies may distract users as they focus on composing messages and monitoring responses instead of engaging in and attending to task goals. This has negative implications for situation awareness, particularly in tasks that are heavily dependent upon visual attention (Funke & Galster, 2009). Finally, these technologies may be perceived unfavorably if composing and transmitting messages is significantly slower than oral communication.

Due to the increasing drive to implement collaboration technologies into command and control environments (Kaufman, 2005), it is vital to evaluate the impact these technologies exert on team performance, communication, and perceived workload. In a recent experiment, Strang and colleagues (under review) asked five-person teams to engage in a simulated air battle management (ABM) task. The task required participants to assume the roles of weapons directors (WDs), sweep operators, and tanker refueling operators while working together in a simulated air defense task to protect friendly assets, eliminate enemy aircraft, and conduct refueling operations under high and low task demands. While engaged in this task, teams were occasionally provided with supplemental decision aids (two types of ecological team resource displays) and communication technologies (chat and a virtual whiteboard). Their results indicated that the graphical resource display increased team performance and reduced workload. Access to chat and the virtual whiteboard did not influence team performance, and had mixed effects on perceived workload (increasing some team members' workload, while reducing others). Strang and colleagues concluded that new collaborative technologies, and domain-specific, ecologically designed resource displays, may be of benefit to team ABM operations as supplements to radio. However, due to space limitations, the work reported by Strang et al. (under review) did not include a thorough analysis of the communication data that were collected. The purpose of the current manuscript is to examine these data in order to explore how the collaboration technologies that were employed may have altered the quantity and quality of team communications observed in that experiment.

Methods

Participants

Seventy men and 35 women between the ages of 18 and 30 ($M = 21.94$, $SD = 3.16$) participated in the experiment. Participants, drawn from local universities and a temporary work agency, were partitioned into 21 five-person teams. Participants provided written informed consent prior to taking part in the experiment; all experimental procedures were approved by an Institutional Review Board prior to data collection. Participants were fiscally compensated for their participation.

Experimental Design

A 2 × 2 × 2 within-subjects design was employed. The factors were *task-demands* (standard, high), *collaboration technology* (radio-only, augmented), and *resource display* (tabular, graphical). Dependent measures included the total number of team communications and the cumulative duration of radio communications. In addition, an analysis was conducted to examine the semantic content of radio communications made by participants.

Apparatus

Workstations. Five computer workstations (one for each team member) located within the same laboratory space were equipped with Dell 1703FPs 17-inch LCD monitors, a standard optical mouse, and a standard keyboard. Oral communication was transmitted and recorded using radio headsets (Sennheiser HD250 Linear II headphones and a Sennheiser HMD 224 microphone) tuned to a universal frequency. To promote use of headsets and microphones, as well as to simulate the auditory constraints of real aviation environments, a 50 kHz pink noise projected at 55 dB was generated in the laboratory. To initiate a radio transmission, participants pushed down a pressure-sensitive foot pedal underneath each workstation. The instantiation and relief of each foot press was recorded, enabling experimenters to record the number and length of each radio transmission.

Tactical Simulation Environment. A simulated air defense task was created using Aptima, Inc.'s Distributed Dynamic Decision-making (DDD) software (version 3.0; MacMillan, Entin, Hess, & Paley, 2004) in which two WDs were responsible for matching friendly fighters with enemy targets, scheduling fighters for refueling and resupply, and communicating action plans to the team. The simulation also featured two sweep operators and a single tanker operator who maneuvered team assets as instructed by WDs, engaged enemy targets, and provided feedback to WDs concerning asset resources (i.e., weapon and fuel status).

The number of enemy targets present in each trial was dictated by the task-demand condition. The standard condition featured four enemy aircraft; the high condition featured six. Each time an enemy aircraft was intercepted and destroyed, a new one would appear. This generation rate of enemy aircraft ensured consistent task-demand in each trial. In addition, trials featured six fast-moving "high-threat" aircraft, which appeared at random intervals. All enemy targets entered the scenario from the right side of the tactical display (the gray zone in Figure 2), and proceeded on a random path to the left (the red zone in Figure 2). As they moved, enemy aircraft could attack fighter and tanker assets, as well as an Air Force base and four infantry units positioned in the red zone.

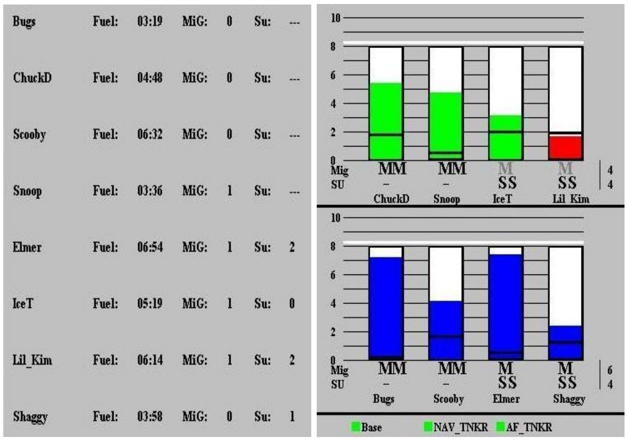
Team communication. Team communication was manipulated across two levels. In the radio-only condition, participants communicated orally via radio headset. In the augmented communication condition, participants could converse using radio or with two additional collaboration tools: chat and a virtual whiteboard. The virtual whiteboard allowed WDs to generate spatial images that could be distributed in real-time using pre-programmed "drag-and-drop" symbols or "free-hand" (examples appear in Figure 2).

Resource displays. Team resource display was also manipulated across two levels. The tabular display condition provided sweep and tanker operators with access to team weapon and fuel status information in a digital format (Figure 1). The graphical resource display (Figure 1) included the same basic information, but was arranged in a "user-friendly" analog format that was made available to all team members. The graphical display also conveyed additional information by alerting teams to low asset fuel reserves (indicated by a change to an amber colored fuel status), and a black bar indicating the minimum fuel required for each asset to rendezvous with a refueling tanker.

Procedure

Participants completed an eight-hour training session on the day prior to beginning the experiment. During that time, participants were told that the purpose of the study was to evaluate how teams use collaboration technologies in ABM operations. Next, participants were randomly assigned to team positions. Teams then completed fourteen practice trials (ten minutes each) to familiarize themselves with the task and technologies.

Teams returned the next day for the experimental session. Upon arrival, they were assigned an order of presentation for sixteen trials, blocked and counterbalanced across eight unique conditions (representing all combinations of factors). Participants were given a rest period (up to 20 minutes) each time they completed four trials. Sessions were completed in eight hours. All dependent measures were averaged across block prior to inferential analysis.



Bugs	Fuel:	03:19	MiG:	0	Su:	---
ChuckD	Fuel:	04:48	MiG:	0	Su:	---
Scooby	Fuel:	06:32	MiG:	0	Su:	---
Snoop	Fuel:	03:36	MiG:	1	Su:	---
Elmer	Fuel:	06:54	MiG:	1	Su:	2
IceT	Fuel:	05:19	MiG:	1	Su:	0
Lil_Kim	Fuel:	06:14	MiG:	1	Su:	2
Shaggy	Fuel:	03:58	MiG:	0	Su:	1

Figure 1. Tabular (left) and graphical (right) resource displays. Both displays included information concerning remaining fuel and weapons of team assets.

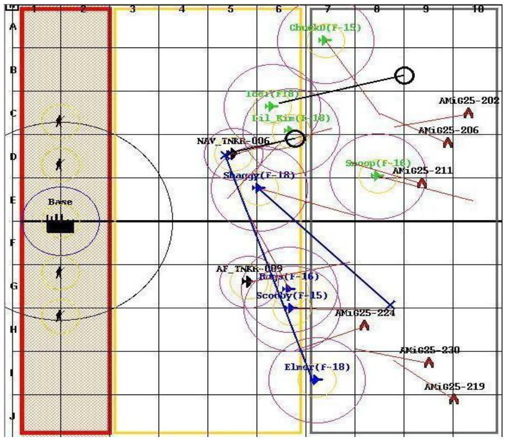


Figure 2. An image from the tactical display with participant-created whiteboard marks (blue and black lines) indicating asset and target route information.

Results

Following completion of the experiment, audio recordings, chat logs, and whiteboard logs were compiled and examined. When supplemental collaboration tools were available, teams sent an average of 3.61 chat messages and 71.44 whiteboard annotations per trial. These means were tested against zero using one sample *t*-tests to establish that teams were, in fact, using these tools. Results confirmed significant uses of both chat, t (20) = 5.94, $p < .05$, and the whiteboard, t (20) = 5.94, $p < .05$.

To test the effects of the experimental manipulations on team communication, separate 2 (task-demands) × 2 (collaboration technology) × 2 (resource display) within-groups analyses of variance (ANOVAs) were computed for the frequency of team communications and the cumulative duration of radio communications.

Frequency of communication. Main effects of *task-demand*, F (1, 20) = 5.53, $p < .05$, $\eta_p^2 = .22$, *collaboration technology*, F (1, 20) = 15.62, $p < .05$, $\eta_p^2 = .49$, and *resource display*, F (1, 20) = 13.05, $p < .05$, $\eta_p^2 = .40$, were detected. As illustrated in Figure 3, teams increased their frequency of communication in the high *task-*

demands condition compared to the standard condition. Teams also communicated more in the augmented *collaboration technology* condition compared to the radio-only condition, which was driven by utilization of the virtual whiteboard since radio communications declined substantially in this condition. Finally, teams communicated more when they had access to the tabular, as compared to the graphical, resource display.

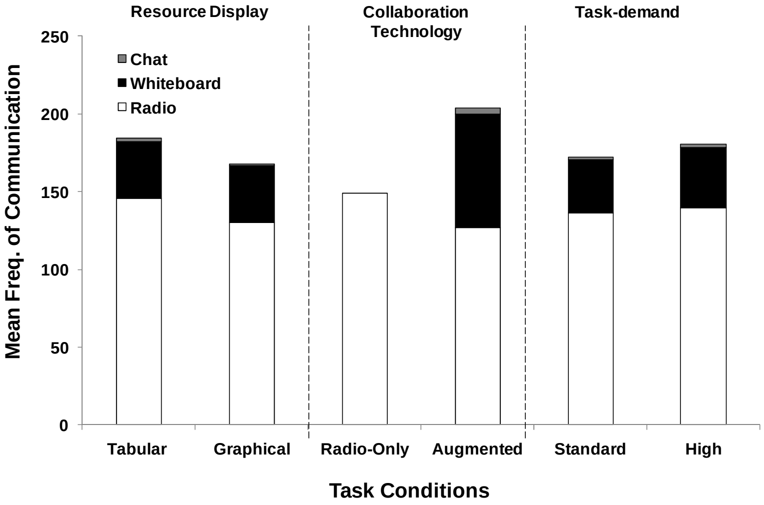


Figure 3. Mean frequency of team communication for each task condition.

Duration of radio communication. Main effects of *collaboration technology*, F (1, 20) = 35.86, $p < .05$, $\eta_p^2 = .64$, and *resource display*, F (1, 20) = 10.92, $p < .05$, $\eta_p^2 = .35$, were detected. These results indicate that the availability of the graphical *resource display* (M = 387.28 s, SE = 22.75 s) caused a decrease in the duration of communications compared to the tabular display (M = 426.40 s, SE = 23.03 s). In addition, duration was less in the augmented condition (M = 364.10 s, SE = 24.06 s) than in the radio-only condition (M = 449.58 s, SE = 22.38 s).

Radio communication content analysis. A subset of all radio communications between participants were hand-transcribed (four trials from each team, 84 trials total). From this set, 10 trials from each experimental condition (80 trials total) were selected at random for a content analysis that resulted in a database of 11,661 transcribed communications. The categorization scheme for coding these communications was developed specifically for this experiment (Table 1).

Table 1: Short descriptions of the nine categories used for content analysis in this experiment.

Category	Description	Example Statements
Clarification / Confirmation	Statements either complying with or clarifying an order or request.	*“Copy that.”*
Coordinate	Statements which reflect planning, back-up behavior, or assisting teammates, but which were not directives for action.	*“I need some help down here.”*
Directive – Maneuver / Attack	Statements concerning maneuvering fighters or tankers, or intercepting enemy aircraft.	*“Intercept MiG 227 at G6.”*
Directive – Resupply	Statements tasking assets for refueling or resupply.	*“Restock and refuel your fighters.”*
Resource Status Request	Questions concerning asset fuel or weapons loadings.	*“Who has low fuel?”*
Resource Status Update	Statements providing information about fuel or weapons loadings.	*“I’ve got 1 minute of fuel left.”*
Simulation Dynamics	Statements or questions concerning the nature or functions of the DDD simulation environment	*“You have to right click to bring up the menu.”*
Situation Update	Statements or questions concerning scenario events and developments.	*“There are still two MiGs at I5.”*
Social / Emotive	Statements which reflected emotion, social interaction, or performance feedback.	*“Good job!”*

Using this protocol, two judges independently classified the transcribed radio communications into the prescribed categories. Inter-rater reliability of the judges was deemed sufficient (proportion of overall agreement = .82; Cohen's kappa = .77, $p < .05$; Ubbersax, 2000; Cohen, 1960). The percentage of radio communications in each category for each condition is presented in Table 2.

Table 2. Percentage of radio communications by category as a function of task condition.

	Radio-Only Comm.				Augmented Comm.				
	Tabular Display		Graphical Display		Tabular Display		Graphical Display		Percentage of Total[b]
Category[a]	Stand.	High	Stand.	High	Stand.	High	Stand.	High	
Directive – Maneuver / Attack	26.17	32.09	42.23	38.25	25.43	18.77	24.86	26.27	29.07
Situation Update	26.33	20.70	18.91	22.19	30.93	24.92	27.34	30.03	24.86
Clarification / Confirmation	20.51	16.96	17.19	17.42	14.61	21.03	20.93	19.31	18.77
Directive – Resupply	6.76	8.33	10.87	11.04	8.71	7.03	11.27	11.21	9.22
Resource Status – Update	9.68	11.00	4.95	6.32	8.11	13.18	6.18	4.58	8.30
Social	2.58	2.90	3.03	3.16	3.00	5.49	4.09	5.73	3.77
Resource Status – Request	5.44	6.42	.96	.31	5.01	6.72	.69	.08	3.37
Simulation Dynamics	.88	.68	.76	.57	1.60	1.63	3.33	1.56	1.33
Coordinate	1.65	.92	1.10	.74	2.60	1.23	1.31	1.23	1.31

[a] Categories are presented in their order of predominance, from largest to smallest, in the complete 11,661 item data set.

[b] Indicates the prevalence of communications in each category from the complete data set, collapsed across experimental conditions to facilitate cross-condition comparisons.

As can be observed in the table, access to the virtual whiteboard and chat in the augmented communication conditions resulted in decrements in the percentage of radio communications classified as *directive – attack*, consistent with the types of communication the whiteboard was designed to convey. Utilization of these tools also resulted in an increase in the percentage of *situation update* communications. This suggests that access to chat and the whiteboard may have supported team situation awareness by providing increased opportunities to discuss task strategies. Access to the graphical resource display resulted in a relatively substantial decrease in radio communications classified as *resource status – update* and *resource status – request*, which is also consistent with the information conveyed by the display, and an increase in *directive – attack* communications, indicating that the display allowed participants to better manage critical resources and intercept enemy aircraft. Substantial differences in communication were not observed with manipulation of trial task demands, suggesting that demands did not impact the types of communication participants engaged in.

Discussion

The purpose of the current manuscript was to explore the influence of a pair of collaboration technologies and team decision-aids on the quality and quantity of team communications in a simulated air defense task. Results indicated that changes in task demands, the availability of the graphical resource display, and access to supplemental collaboration technologies altered the dynamics of team communication in this task.

With regard to the effects of task demands, results indicated that high demand caused an increase in the number of team communications. This is not surprising given that participants were tracking and responding to an increased number of enemy aircraft in that condition. Though task demands and collaboration technologies did not interact in the current experiment, it is possible that under "real-world" circumstances, the ability to offload communication to additional media, such as chat or whiteboard, may prevent radio overload and increase team effectiveness when task demands are high.

Findings related to the effects of the resource displays clearly favored the utility of the graphical display since it decreased the total number and duration of team communications, and those reductions allowed teams to more fully focus on intercepting enemy aircraft, rather than on resource management. When these findings are paired with those of Strang et al. (under review), which showed that the graphical display also reduced perceived workload and improved team performance, it suggests that such a monitor may be a very useful tool for ABM operations. It is also worth noting that the graphical display was developed from feedback the authors received provided by subject matter experts (Russell et al., 2009).

Finally, the joint availability of chat and the virtual whiteboard caused an increase in the total number of team communications, but decreased the duration of radio communication. This seems to imply that the tools were effective for reducing reliance on radio while simultaneously increasing overall team communication. Interestingly, these findings might explain why Strang et al. (under review) found that the additions of chat and whiteboard resulted in an increase in perceived workload for some team members (WDs), but not others (sweep operators). Specifically, since only WDs needed to monitor all communications, given their important role in issuing commands, an increase in overall communication across multiple modalities may have led to perceptions of increased workload. However, sweep operators, who had fewer responsibilities, may have found the advantages of chat and whiteboard useful and manageable. These results illustrate that, under varying circumstances, some team members may be benefitted by the availability of collaboration technologies while others are unchanged (or hindered) by the same tools. This suggests that teams may be better served by *adaptive* collaboration technologies, which may be tailored to the needs of individual team members and roles (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. Determining the nature and behavior of such tools is likely to be a fruitful area of future research.

References

Alberts, D.S., Hayes, R.E. (2003). *Power to the edge*. Washington, DC: CCRP Publication Series.

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.

Berry, G.R. (2006). Can computer-mediated asynchronous communication improve team processes and decision making: Learning from the management literature. *Journal of Business Communication, 43*, 344–366.

Bolstad, C.A., & Endsley, M.R. (2003). Tools for supporting team collaboration. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 47*, 374-378.

Burke, C.S., Stagl, K.C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: A conceptual analysis and model. *Journal of Applied Psychology, 91*, 1189-1207.

Cohen, J. (1960). A coefficient for agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Funk, G.J., & Galster, S.M. (2009). The effects of cognitive processing load and collaboration technology on team performance in a simulated command and control environment. *International Journal of Industrial Ergonomics, 39*, 541–547.

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.

Kaufman, A., 2005. Caught in the network: How the doctrine of network-centric warfare allows doctrine to dictate military strategy. *Armed Forces Journal, February 2005*, 20-22.

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., & Riley, V. (1997). Humans and automation: Use, misuse, abuse. *Human Factors, 39*, 230-253.

Russell, S.M., Knott, B.A., Galster, S.M., Funke, G.J., & Dukes, A.W. (2009, April). *Validation of synchronous iconographic digital whiteboard interface for team collaboration in a simulated command and control task*. Poster presented at the 15th International Symposium on Aviation Psychology, Dayton, OH.

Strang, A.J., Knott, B.A., Funke, G.J., Russell, S.M., Miller, B.T., Dukes, A.W., Hyson, J., Courtice, A.M., Brown, R., Lyons, J., & Bolia, R.S. (under review). Collaboration technologies improve performance and communication in air battle management. *Military Psychology*.

Uebersax, J. (2000). Statistical methods for rater agreement. Retrieved from http://ourworld.compuserve.com/homepages/jsuebersax/agree.htm.

Vidulich, M.A., Bolia, R.S., & Nelson, W.T. (2004). Technology, organization, and collaborative situation awareness in air battle management: Historical and theoretical perspectives. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 233-253). Aldershot, UK: Ashgate Publishing.

Wickens, C.D., Vidulich, M., Sandry-Garza, D. (1984). Principles of S-C-R compatibility with spatial and verbal tasks: The role of display-control location and voice-interactive display-control interfacing. *Human Factors, 26*, 533-543.

AN INTEGRATED ALERTING AND NOTIFICATION SYSTEM UTILIZING STAGES OF AUTOMATION AND UNCERTAINTY VISUALIZATION

Tiffany N. Saffell, Amy L. Alexander, Alan S. Carlin, Andy C. Chang, & Nathan Schurr
Aptima, Inc.
Woburn, MA

While NextGen operations are still under development, several key issues have already emerged, including increased information demands on the flight deck. The ALerting And Reasoning Management System (ALARMS) was designed as a strategic, automated system for combining and evaluating alert-related outputs from current and proposed NextGen systems. The model-driven interface integrates the status of the environment, pilot, and system to automatically present the most critical information at the right time, augmenting existing flight deck technologies. The current level of uncertainty in the environment and system as a whole is also evaluated and represented within the display. The four stage model of information processing presented by Parasuraman, Sheridan, & Wickens (2000) was used to guide the development of the underlying ALARMS automation. This document provides a brief overview of the ALARMS development process. Examples of the interface are included, and we discuss its implications.

Introduction

Although the technologies to support Next Generation Air Transportation System (NextGen) operations are still under development, it is clear that the airspace and flight deck of the near future will raise new challenges in how pilots interact both with components of the airspace and with the existing and advanced technologies of the cockpit.. While alerting systems have existed in aircraft cockpits since their instantiation, the changing role of the pilot within NextGen is expected to fundamentally alter the way these systems will need to operate. Current systems and NextGen systems will both provide critical flight information; however the information will need to be presented to the pilot in an accessible manner, so the pilot does not become overwhelmed by data.

We developed a new model-driven interface to address the future needs of pilots in NextGen airspace while using current and NextGen technologies. The **AL**erting and **R**easoning **M**anagement **S**ystem (ALARMS) is an Integrated Alerting Notification (IAN) system which combines environment and aircraft hazard warnings in an intelligent manner to provide the pilot with the right information at the right time, taking into account the state of the environment, the aircraft, and the pilot. ALARMS consists of a user-centered model to drive alerting functions and a dedicated display to portray these recommendations to the pilot. ALARMS was not designed to replace current alerting technologies but instead to augment them by providing information to support strategic planning. This paper gives a brief overview of the interface development process, focusing on the utilization of Parasuraman, Sheridan, and Wickens' (2000) model of information processing and stages of automation. The presentation of uncertainty information is also discussed.

Information Processing and Stages of Automation

Parasuraman and colleagues (2000) outlined four stages of automation to support four stages of information processing. In particular, they focused on the following stages of information processing: sensory processing, perception/working memory, decision making, and response selection. Sensory processing is the act of information gathering and initial processing of information. This stage includes orienting sensory receptors, sensory processing, and selective attention. The next stage, perception/working memory, refers to conscious perception of the sensory information. During this stage, the information is manipulated along with information from long term memory. This stage can also include rehearsal, integration, and inference. These processes lead to the third stage, decision making, or the selection of a possible action. The fourth stage, response selection, is the embodied action of the decision.

Each of the four stages of automation supports a corresponding stage of information processing. Stage 1 automation, Information Acquisition, supports sensory processing. Stage 2 (Information Analysis), Stage 3 (Decision and Action Selection), and Stage 4 (Action Implementation) correspond respectively with perception/working memory, decision making, and response selection. We used these stages of automation to guide the development of the ALARMS automation. All levels can be associated with greater or lesser levels of automation.

Stage 1 automation (supporting Information Acquisition), can have multiple levels of automation. At a low level of automation, the machine can direct the position and action of the sensors. At a greater level, the automation can prioritize the presentation of information. At the highest level of automation, only selected information is presented to the operator. During the lowest level of stage 2 automation (Information Analysis) algorithms can make predictions of future actions, such as future trajectories. Higher levels of stage 2 automation can include data presentation or data summaries that are context relevant. Stage 3 automation (Decision and Action Selection) can be partial or complete replacement of human selection of a decision from decision alternatives. Decision automation can range from presenting intelligent decision alternatives to making an action selection without human intervention. Finally, during stage 4 (action implementation), the machine or system executes the action choice, at different levels of automation. Stage 4 automation can completely replace the person in the system. On the other hand, the person in the system can make a selection that sets off a series of automatic events.

Method

Aptima used a multi-faceted approach to develop the ALARMS interface. First, we consulted documentation covering current and potential NextGen technologies to develop a catalogue of flight deck systems. This information was used as a catalyst during both subject matter expert (SME) interviews as well as model creation for interface information presentation. SME input was specifically used to define interface requirements. Finally, we adapted Parasuraman and colleagues (2000) stages of automation to drive the display of information on the ALARMS interface.

Cataloging Existing and NextGen Technologies

In order to develop an accurate understanding of the types and nature of the alerts pilots may face, we created a catalogue of current and potential NextGen technologies. The Aptima team consulted manuals and FAA documentation to create the catalogue. We focused on the cockpit environment and avionics systems typically associated with FAR Part 121 operations, and also consulted with aviation SMEs. Each current and potential NextGen system was mapped including its input sources and output parameters. Most notably, we mapped the types of alerts (i.e., advisory, caution, warning), their criticality, and the alerting outputs (i.e., aural, verbal, tactile). Therefore, we were able to categorize the importance of potential alerts and what types of information would be presented. For more information on the cataloging efforts, including examples of consulted references, please see Alexander, Alaverdi, Geiselman, Galster, and Schurr (2010).

SME Knowledge Elicitation Sessions

Knowledge elicitation sessions with pilot SMEs were conducted in order to create prototype interface design requirements. Our first subject matter expert was a First Officer on a commercial carrier. Aptima conducted a Work Domain Analysis (Vicente, 1999) in order to capture the objectives, constraints, and information required during nominal and off-nominal situations. We used this information along with our catalogue of current and NextGen technologies to compile an Abstraction/Aggregation hierarchy. The draft Abstraction Hierarchy was reviewed and corrected during interviews with three additional pilots (a retired commercial Boeing pilot, a military cargo pilot, and a business jet pilot). In addition, we conducted a Control Task Analysis (ConTA) with all four SMEs. Two use cases were developed to guide the ConTA, one describing an Optimized Profile Descent and another describing a Trajectory Based Operation. Questions asked during the ConTA assessed what pilots would think, look for, and do during an incident. These use cases were also used to orient the design of ALARMS. For more information on the Abstraction Hierarchy development, please see Alexander, Chang, Saffell, and Schurr (2010).

Model Creation

Aptima developed a model to drive the information presented by the ALARMS interface. The model is primarily composed of a Hazard State Estimator, a Pilot State Estimator, and Planning Module. Finally, information is passed to a Stages of Automation module. The Hazard State Estimation module is actually a Bayesian Network that model's the airplane's environment via the information being passed by the alerting systems. It weighs these inputs and associates them with the certainty of hazards. The Hazard State Estimation module outputs a probability distribution of the "Environment State," that is, an estimate on the probability and severity of the various hazards.

The Pilot State Estimation module receives information about the performance of the pilot as well as the ALARMS interface and outputs the expected performance quality and the expected duration of pilot actions. It predicts the pilot's capability to handle each hazard at each state of automation. The Pilot State Estimator can also accept inputs for measures of mental effort, task demands, and ongoing task performance.

The Planning of Interface Actions module leverages a TMDP (Time-Dependent Markov Decision Processes) (Boyan & Littman, 2000) model that is used to capture both state uncertainty in the environment as well as duration uncertainty in human (pilot) actions (Schurr, Marecki, & Tambe, 2008). Its input is the Hazard and Pilot states, as well as a Markov model of the effectiveness of the pilot and automation in handling the hazards, given various levels of alerts. Its output is a time-dependent plan for addressing the alert.

The plan is passed to the Stages of Automation module, which interprets the level of automation and decides what level of alerts and options to send to the pilot. This decision is then sent to the ALARMS interface, which displays the information to the pilot. More information about the modeling methods can be found in Carlin, Alexander, and Schurr (2010) and Carlin, Schurr, and Marecki (2010).

ALARMS Stages of Automation

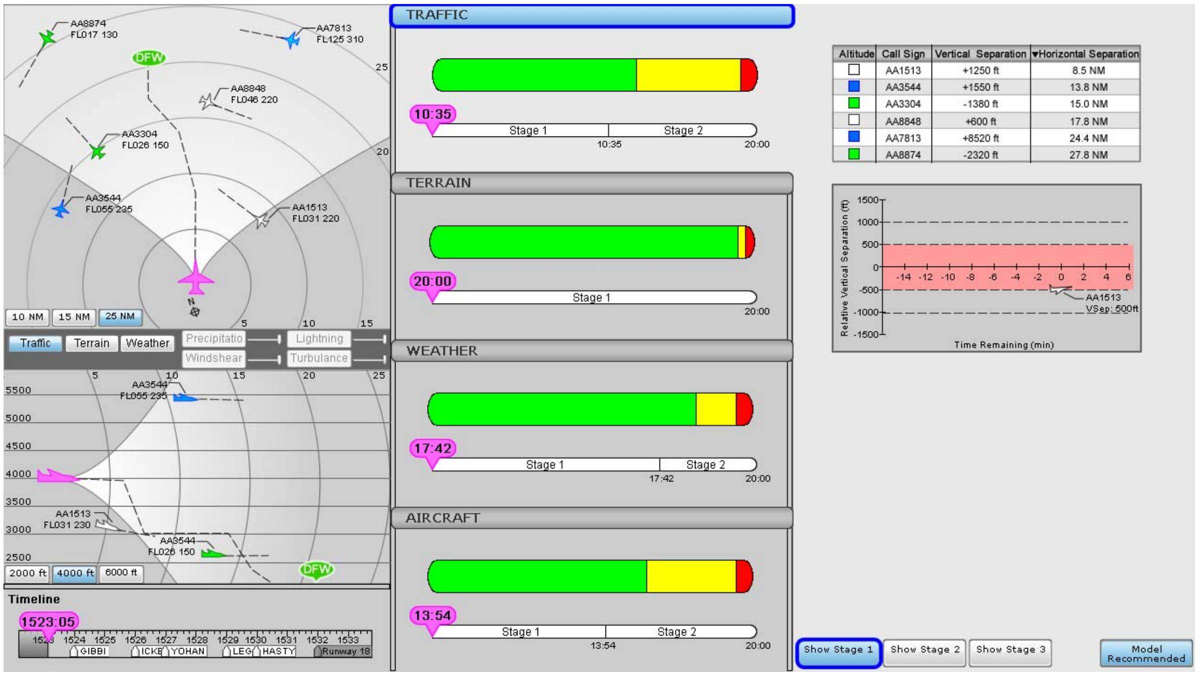
We used Parasuraman, et. al's (2000) stages of automation to guide the development of the ALARMS automation. Automation stages 1 through 3 were slightly redefined to better support the interface and the task environment. Stage 1 automation was designed to support Information Acquisition. This stage supports the perception of, and attention to, relevant information in the environment. Within an IAN context, information acquisition automation can support the allocation of operator attention to relevant hazards within the environment. Stage 2 automation supports Information Analysis. The interface displays concepts to assist in the integration and understanding of information. Within an IAN context, information analysis automation can be used to convey the present and predicted future locations of hazards relative to ownship. Finally, we used stage 3 automation to support Decision Selection. The interface displays decision support for identifying feasible actions and/or making optimal decisions. Within an IAN context, decision selection automation can provide recommendations for trajectory changes to avoid impending hazards.

Note that ALARMS does not currently address stage 4 – Action Implementation within our notification methodology. This is intentional because at this stage in aircraft and automation technology adoption, ultimate decisions often remain in the hands, and the heads, of the pilot. However, our current approach can be easily extended to incorporate stage 4. Often advisories appear during stage 1 automation, warnings appear in stage 2, and cautions appear in stage 3, however there is not a direct 1-to-1 mapping between hazard or alert type and stage of automation.

ALARMS Interface

The ALARMS interface is shown in Figure 1. The full functionality of ALARMS will not be discussed in this paper due to space constraints. It is important to note that ALARMS is not just an information display, but a tool to display categorized information at any stage of automation based on the pilot's choosing. An information display would only show the pilot what the automation calculated as important; however we designed an interface so that the pilot can choose which information to display.

ALARMS is composed of three main panels. The left hand panel supplies broad situational awareness information, the middle panel shows the status of different alert categories (traffic, terrain, weather, and aircraft), and the third column displays alert specific information (i.e. traffic information for possible incursion). The model described in the Methods section drives what information is displayed. However, at any time the pilot could touch a radial button and select other information to display. For example, the pilot can choose to view weather information by pressing the Weather button, even if the display defaults to traffic information. There is also a button in the lower right hand corner labeled "model recommended." At any time the pilot can return to the model recommended state display by pushing this button.



Altitude	Call Sign	Vertical Separation	▼Horizontal Separation
☐	AA1513	+1250 ft	8.5 NM
🟦	AA3544	+1550 ft	13.8 NM
🟩	AA3304	-1380 ft	15.0 NM
☐	AA8848	+600 ft	17.8 NM
🟦	AA7813	+8520 ft	24.4 NM
🟩	AA8874	-2320 ft	27.8 NM

Figure 1. The ALARMS interface in stage 1 automation, supporting Information Acquisition

Left Column: Map of External Elements

The leftmost panel of ALARMS presents the external elements. It contains a vertical and horizontal situation display (VSD and HSD, respectively). The aircraft's possible operation envelope is shown as a white flare emanating from the nose of ownship. The HSD and VSD also graphically represent external entities that may pose potential hazards or provide critical information, such as traffic, terrain, weather, flight path, and airports. This information can be filtered by selecting the associated information button (i.e., traffic, terrain, weather) presented between the HSD and VSD, automatically changing it to "on" or "off." The timeline at the bottom of the panel presents upcoming waypoints in the preprogrammed flight plan.

Middle Column: Hazard State and Stages of Automation

The center column displays information about the hazard state and stage of automation associated with four alert status categories: traffic, terrain, weather, and aircraft. Based on the level of alert, the information displayed in these categories will dynamically change to aid the pilot in maintaining situation awareness about the respective hazard states (discussed in depth in the right column discussion below). The middle column provides a broad overview of the hazard status to the pilot. It shows the pilot how certain a hazard is, the time the hazard might advance to a later stage, and how the hazard relates to other potential hazards. The column also shows what category of information is being displayed in the right hand column. If the "Traffic" button is highlighted or selected (as it is in Figure 1), then traffic information is being displayed in the right column.

Under each of the hazard state categories is a stage planner timeline presenting the current stage of automation as well as a countdown to any anticipated future stages. To ensure that the proposed interface serves as an effective tool for the pilot, the ALARMS interface was designed to reflect the output of the integrated model in order to display each of the first three stages of automation as well as their level of uncertainty.

Hazard State Uncertainty Display. The Hazard State Estimator information is presented as large horizontal bars in each of the four categories. Each bar consists of up to three colored subsections (green for hazard stage 1, yellow for hazard stage 2, and red for hazard stage 3), and was designed to have a total fixed width so that the subsections combine to fill the entire bar. The more certain a stage, the larger its color block. As hazards appear and priority changes, the three colored bars will vary in width to indicate the uncertainty associated with each hazard state. Therefore, even though an alert can consist primarily of a certain hazard state, there is a possibility for the other two hazard states to simultaneously exist. Using this convention, the current hazard state is represented by the bar with the largest width, while bars with equal widths may indicate equal uncertainty of two hazard states, such as during a transition period. This design allows the pilot to quickly and visually make a qualitative assessment of the current hazard state and the uncertainties associated with the other two.

Right Column: Detailed Contextual Information

The right column of the ALARMS interface provides additional information to the pilot by displaying only relevant information about the most critical hazard based on the stage of automation determined by the models. The pilot can chose to display the information about any hazard by selecting the radial button, however ALARMS defaults to display the most critical hazard at any given time.

Altitude	Call Sign	Vertical Separation	▼Horizontal Separation
🟥	AA1513	+1250 ft	8.5 NM
🟦	AA3544	+1550 ft	13.8 NM
🟩	AA3304	-1380 ft	15.0 NM
☐	AA8848	+600 ft	17.8 NM
🟦	AA7813	+8520 ft	24.4 NM
🟩	AA8874	-2320 ft	27.8 NM

SUGGESTED TRAJECTORY CHANGE

1. UAL438, Fly Heading 330. Decrease Speed to 210 knots.
2. UAL438, Fly Heading 310. Decrease Speed to 220 knots.
3. UAL438, Fly Heading 174. Maintain Speed at 195 knots.

Figure 2. Two right hand columns of the ALARMS interface in stage 2 (left image) and stage 3 (right image) automation.

Stage 1 Automation. Stage 1 automation is used to support information acquisition, in this case to provide information about the presence of potential hazards (see Figure 1). At this stage, the interface simply indicates possible types of hazards and allows the pilot to assess the situation without further assistance. Hazard type indication takes the form of a status display that provides computational offloading and contextual information to enhance the pilot's ability to acquire information.

Stage 2 Automation. During stage 2 automation (Figure 2, left image), the interface supports the pilot's ability to integrate information by highlighting that aircraft AA1514 is about to breach vertical separation. This is displayed in a number of ways: (1) in the right hand column, top table - tabulated information about ownship's protected airspace is highlighted in red; (2) in the right hand column, the middle graph - a graphical representation of ownship's vertical separation zone is shown with the encroaching aircraft's past and projected trajectory and an estimate of time of impact shown in red; (3) the middle column - a large yellow bar in the Hazard State Estimator for Traffic and stage 2 automation in the stage planner; (4) in the left hand column - the symbol for the encroaching aircraft turns to red in both the HSD and the VSD in the very left hand column (the left panel can be seen in Figure 1, however no planes are currently displayed in red); and (5) in the left hand column - a red marker is added to the main timeline to indicate the projected time of collision (the main timeline is shown in Figure 1 without the red marker). These projections are based on the assumption that both aircraft continue on their respective trajectories. In effect, the pilot is sufficiently alerted to the problem while simultaneously being shown the same information in different contexts. Therefore, instead of having to mentally integrate one piece of information into different contexts, the pilot is shown the integrated information to support their decision making process, reducing their cognitive load.

Stage 3 Automation. If a hazard is not addressed in sufficient time, the model invokes stage 3 automation (Figure 3). This stage of automation provides the same information as stage 2 automation but also offers decision support. This may manifest as one (or multiple) suggested courses of action which the pilot can follow to avoid the hazard. In the example shown in Figure 2 (right image), the suggested mitigation strategy is for ownship to alter its horizontal trajectory and decrease its speed. On the HSD and VSD, the result of such a change is shown graphically so that the pilot can immediately visualize the outcome with no added cognitive workload. Time-stamped location markers are also supplied to give the pilot additional contextual information about the 4D trajectory. Note that in this case, the right hand column has only had the "Suggested Trajectory Change" added.

Conclusions

The ALARMS interface is a significant step forward in intentionally applying a variation of Parasuraman et. al's (2000) stages of automation when developing an interface. By designing the automation to support various stages of information processing, the pilot is provided with the right information at the right time given the dynamic nature of the flight environment.

The current instantiation of the ALARMS prototype is limited in its functionality such that it is driven by simulated data as opposed to real-time sensor/system inputs. In addition, the underlying models are limited to a small set of particular situations, such as weather hazards, traffic incursions, engine fire alerts, and landing gear indicators. However, the system was designed for extensibility and further development could focus on modifying the models as needed. Finally, while the interface has been reviewed by subject matter experts, the interface itself and the models driving the information display have not been rigorously tested. The models should be tested for situation appropriateness and the interface should be tested in simulated flight scenarios prior to transitioning the alerting logic and/or representations to actual systems.

Acknowledgements

This research was supported by NASA Langley Research Center under Contract NNL08AA20B. Kara Latorella 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. The authors would also like to thank Omeed Alaverdi, Kevin Durkee, Scott Galster, Eric Geiselman, and Christopher Wickens for their significant contributions to the ALARMS project.

References

Alexander, A. L., Alaverdi, O., Geiselman, E., Galster, S., & Schurr, N. (2010). A system-hazard matrix for the NextGen flight deck. Proceedings of the International Conference on Human-Computer Interaction in Aerospace. Cape Canaveral, FL: HCI-Aero

Alexander, A. L., Chang, A., Saffell, T. N. & Schurr, N. (2010, July). Work domain analysis to support flight deck integrated alerting and notification. Poster session presented at the meeting of the Proceedings of the Applied Human Factors and Ergonomics Society International Meeting, Miami, Florida.

Boyan, J. & Littman, M. (2000). Exact solutions to time-dependent MDPs. *Proceedings of Neural Information Processing Systems* (pp. 1026–1032). Denver, CO: NIPS.

Carlin, A., Alexander, A. L., & Schurr, N. (2010). Modeling pilot state in next generation aircraft alert systems. Presented at the MODSIM World 2010 Conference & Expo, Virginia Beach, VA.

Carlin, A., Schurr, N., and Marecki, J. (2010). ALARMS: Alerting and Reasoning Management System for Next Generation Aircraft Hazards. Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI 2010). P. Grünwald and P. Spirtes (Editors). AUAI Press.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model of types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics*, *30*, 286-297.

Schurr, N., Marecki, J., & Tambe, M. (2008). Resolving Inconsistencies in Adjustable Autonomy in Continuous Time (RIAACT): A robust approach to adjustable autonomy for human-multiagent teams. *Proceedings of 7th International Conference on Autonomous Agents and Multiagent Systems.* AAMAS.

Vicente, K.J. (1999). *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*, Mahwah, NJ: Erlbaum Associates.

ENHANCED SCENARIO VISUALIZATION FOR SIMULATION-BASED TRAINING

Jonathan Borgvall, Martin Castor & Stephen D. O'Connell
Swedish Air Force Combat Simulation Centre, FLSC
Swedish Defence Research Agency, FOI
Stockholm, Sweden

This paper describes the development and initial experiences of an enhanced scenario visualization system recently introduced at the Swedish Air Force Combat Simulation Centre (FLSC). FLSC provides team training of fast-jet pilots, performs research on training effectiveness and human performance, and simulation-based development and acquisition. The system has been developed in-house, based on an extensive set of experiences and needs among various user groups, to support and enhance the effectiveness of simulator-based team training, as well as research and development programs. It provides enhanced presentations of scenarios, enabling pilots, instructors, training designers, researchers, and operational analysts to observe, reflect, and analyze sorties during and after execution. Initial experiences and reactions are promising, and the paper will address and relate those to other relevant efforts and elaborate on future interesting development paths.

The Swedish Air Force (SwAF) runs a program for simulator-based training, research, and development at the SwAF Combat Simulation Centre (FLSC). FLSC mainly conducts training of fast-jet pilots and the facility is designed to provide experiences that develop the trainees' knowledge and skills in decision making, planning, communication, tactical behavior, and situational awareness. The emphasis is on developing skills and knowledge on a team and inter-team level using eight fast-jet cockpits and four fighter controller stations (ground control intercept, GCI). The research activities are mainly focused on training effectiveness and human performance, while the development program conducts simulator-based acquisition studies and tactics development.

Training and Feedback

The importance of feedback and guidance has been identified as an essential part of any training since many years back (e.g., Holding, 1987), and it seems to have grown in relevance over the past decade following an increased use of simulators (e.g., Cannon-Bowers & Bell, 1997; Wickens & Hollands, 2000; Freeman, Salter, & Hoch, 2004). Considering modern simulator-based training, one effective way of delivering immediate feedback is to replay critical situations of a scenario, visualizing the events while discussing different situations. Especially when dealing with teams, the importance and effect of scenario replays during debriefs become critical.

The training concept at FLSC is based on Kolb's Experiential Learning Theory (ELT) and his four-stage cycle of learning: concrete experience, reflective observation, abstract conceptualization, and active experimentation (Kolb, 1983). At FLSC, this basically means that the trainees (i.e., pilots in this case) have a high degree of responsibility and influence on their own training. For example, they plan their training together with the instructors and run the debriefings themselves (even though the FLSC instructors may support with additional comments). By doing so, the pilots, with the support of the instructors, are provided with concrete experiences in the simulator, leading to observations and reflections during execution (pilots flying and pilots not flying) and debriefs. The pilots' influence over the training aims to stimulate them to assimilate their reflective observations into abstract concepts which they can actively test and elaborate with. The experience is that this model supports motivation and deliberate practice among the pilots, which in several domains has been found to be a crucial ingredient for developing expertise (Ericsson & Charness, 1994).

Another important part of simulator training is debriefs and after-action review (AAR), since providing feedback and knowledge of results has been proven important for both motivation and performance improvement (Holding, 1987). Freeman, Salter, & Hoch (2004) declare that feedback in debriefs is particularly important in team training since teamwork itself does not necessarily produce immediate feedback from which the team members can learn during task execution. Further on, feedback offered during performance is normally processed less well than feedback received immediately after completion since attention must be divided in the former case (Wickens & Hollands, 2000).

The experience at FLSC is that effective debriefs can substantially enhance the effect of simulator-based team training. The development of the enhanced scenario visualization is one example of that. Since the pilots hold a high level of responsibility for their own training it provides the opportunity for objective critique and reflections based on an audio-visual replay of the scenario. Examples of proficient or poor performance can be exposed and discussed with the support of the replay and the visualization aids. Several studies have shown that allowing trainees to practice without specific feedback and guidance of good and bad performance can be detrimental (Wickens & Hollands, 2000) and may produce sub-optimal decision-making skills (Cannon-Bowers & Bell, 1997). In that sense, the tool supports the pilots learning process since it facilitates critical thinking about their own performance (Freeman, Salter, & Hoch, 2004) and supports the development of common ground (e.g., Clark, 1996) among the team members while discussing the audio-visual replay of the events.

Even though feedback is critical for nearly any form of skill acquisition, Wickens and Holland (2000) point out that there are characteristics of decision making that can prevent feedback from providing maximum training value. Two of them are that (1) feedback is often ambiguous, and (2) feedback is often delayed. One way of overcoming this could be to immediately provide debriefs using embedded scenario visualization aids making as much information as possible available for observation and critique in an objective manner.

User Needs for Scenario Visualization

Following a growing experience of providing training and conducting simulator-based research and development at FLSC over the past ten years, an emerging need for augmented scenario visualization was identified a few years ago. A previous tool (Figure 1) served some of the basic visualization needs, but was deemed insufficient and inflexible to suit different emerging needs for various user groups. The needs emerged not only from trainees and instructors but also from engineers, operational analysts, and researchers. Experiences from these user groups were collected during workshops and during actual training, research and development efforts, over several years. This section lists the specific needs and intended usage patterns of the scenario visualization system for the three user groups mainly involved with training: pilots, instructors, and training designers.

Figure 1. The FLSC instructor and operator station, and in the background the debriefing facilities using the old scenario visualization tool.

The pilots require a visualization system for several purposes. Pilots not flying may follow the scenario in real-time from the debrief area, including the verbal communication via head-sets and desired visualization presentation aids. The pilots flying would use it for immediate debriefs, as well as for conducting delayed analyses of performance on the recorded sorties from a week of training at FLSC. Hence, the pilots would have a chance of conducting delayed analyses of individual or team performance, using the various visualization aids. This could play an important role for enhancing the effectiveness of delayed debriefs or analyses of individual performance since the memory of execution normally declines if feedback is delayed, hence making it less effective (Wickens & Hollands, 2000).

The instructors at FLSC would primarily use the visualization system to monitor and supervise the scenarios during execution. Further, it could be used for pre-training purposes, before squadrons arrive at FLSC. The idea is that an instructor from FLSC visits the squadron scheduled for training the week before the actual training itself. Such pre-practice tools have been supported by Burke, Salas, Wilson-Donnely, and Priest (2004) who included this approach as one of their ten guidelines for effective team training. The instructor would provide replays and examples of scenarios and events of sorties the squadron or another squadron has previously performed, and a preview of what they can expect during the coming training. The purpose is to prepare the pilots for the training in order to increase the initial effectiveness of the training. Further on, and similar to the delayed analyses of performance for the pilots, the replay and visualization aids also could support the instructors delayed performance assessment. Since many of the exercises at FLSC include eight manned pilot stations, time constraints make it difficult for the instructor to conduct reliable performance assessments during execution, especially at an individual level. Combining performance assessment tools or mission blueprints might enable the instructors to conduct reliable judgments after execution. This also provides enhanced in-depth assessment since the visualization aids embedded in the tool can be used to a higher extent due to lower time constraints.

The instructional designers would use the visualization tool to monitor, analyze, and evaluate the training, for example the exercise management, and the vignettes (i.e., trigger events) that the white force (e.g., role players) and constructive forces (computer-generated forces, CGFs) make. By using a audio-visual replay and the embedded visualization aids they could conduct in-depth analyses of the effectiveness of different vignettes (i.e., if they actually triggered desired reactions).

Needs and intended usage patterns vary between the different user groups. Common for all user groups is the need for visualization of non-physical parameters and composed measures. Non-physical parameters here refer to information or composed measures that are not normally available for observation. One example of a non-physical parameter would be the predicted point of impact (PRIMP). A visualization of the PRIMP quickly shows a pilot how his maneuvers forces the missile to recalculate its PRIMP and change course, thus draining the missile of its kinetic energy. An example of a team related composed measure would be a visualization of a fourship's current radar coverage (given a certain stipulated radar cross section of a target), combined with remaining own armament ranges (given certain target characteristics and behavior), which would represent the fourship's offensive potential at any given moment. The pilots can thus easily see when their offensive potential drops due to their behavior. Focusing on training applications, one could argue that these non-physical parameters represent knowledge that typically differentiates a more experienced individual from a less experienced. Enhancing the scenario visualization with non-physical parameters and composed measures would in this sense make some knowledge normally limited to experienced pilots available and observable, and hence provide a potential for more effective training through enhanced debriefs.

Based on this analysis of user needs and intended usage patterns, the enhanced visualization was developed as a flexible and expandable package using a plug-in architecture.

Enhanced Scenario Visualization System

The enhanced system presents a 3D "God's Eye" view of the simulated world, together with the head-down displays from all eight cockpits. The graphical user interface gives the user full control of the view of the world, with full zooming and control of the view angle. The temporal control of the replay provides up 16 times fast forward or backward playback speed. The functionality to go directly to a specific timestamp also exists. However, the 3D-view not only shows all the aircrafts in a selected area, but also in real-time visualizes data such as status parameters and measures of each aircraft, for example its weapon and sensor systems. Emphasis have been put to the development of visualization of the non-physical parameters (e.g., verbal and datalink communication), and information that wouldn't be visible to pilots in reality (e.g. the missile envelope; or PRIMP, mentioned above). These are a few examples:

- Visualization of radio communication, including statistics such as total air time, percent of air time, and number of step-ons (i.e., when two or more pilots are verbally communicating over the radio at the same time).
- Visualization of tactical data link messages (Link 16) between aircraft and with ground controllers.
- Radar coverage, radar lock-on and target sorting symbols.

- Missile envelope, missile lock-on status and PRIMP.
- Shot logs for missiles, including hits/misses and reasons for missile termination/abort.

During replay, the visual presentation of every parameter or function can be turned on and off as different users have different needs and preferences. The tool also provides the possibility to enter time-stamped observer comments that are shown during replay. It is also possible to web-cast the replay which is a useful feature during distributed simulation exercises.

After a training week at FLSC the visiting squadron receives all the logged files from the week, enabling them to review the recorded material of the sorties on any secure PC. The current implementation represents a modular concept where new plug-ins (e.g., the logging and presentation of Link 16 tactical data link messages) are developed as required. Figure 2, 3, and 4 provide some examples of enhanced visualizations in the new system.

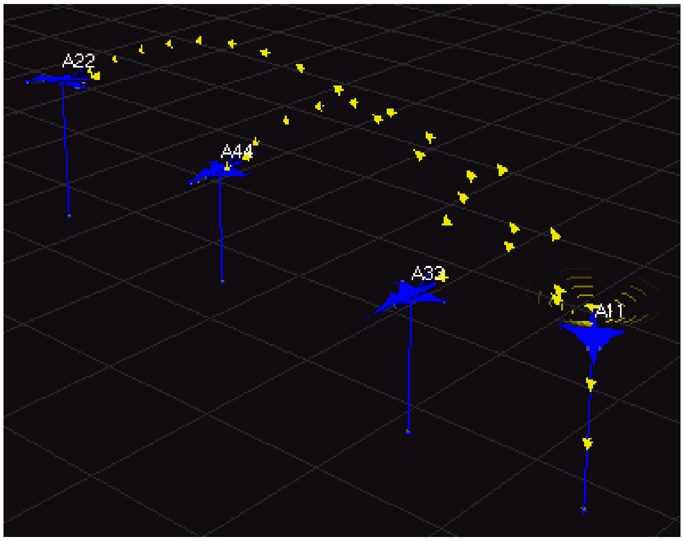


Figure 2. Visualization of verbal communication transmission (yellow arrows) from A11 to A22, A33, A44 and to the ground control interceptor.

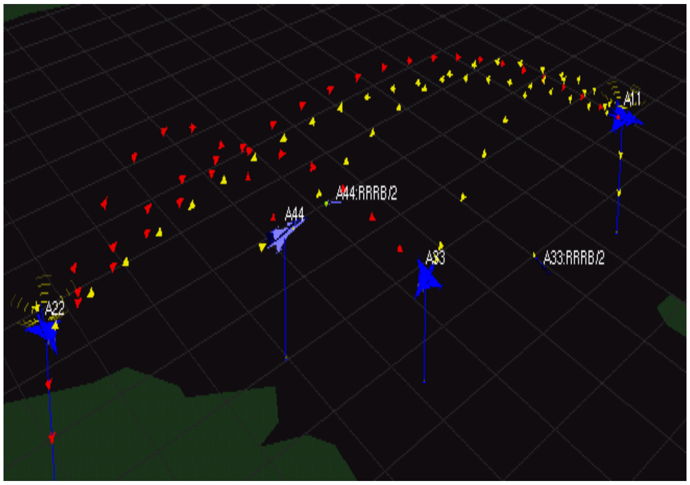


Figure 3. Visualization of verbal communication (yellow arrows) and step-on (red arrows). A step-on occurs when two or more pilots are verbally communicating over the radio at the same time. In this situation A11 is already communicating with the rest of the 4-ship and the ground control interceptor via radio when A22 steps-on.

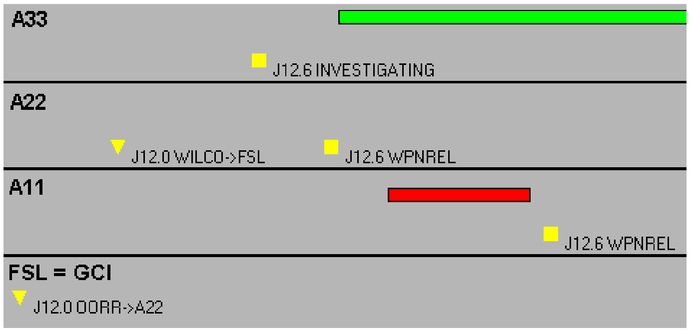


Figure 4. Visualization of verbal communication and data link communication (A11, A22, and A33 represent three different aircraft as can be seen in Figure 2 and 3 above, GCI is the Ground Control Intercept). The green line represents speech, the red line step-on, and the yellow boxes and triangles Link 16 messages.

Development Paths

A number of paths are envisioned for the continued development of this tool. One such path is to visualize historical data from 10 years of previously conducted exercises at FLSC. By processing, filtering and aggregating this data, typical cases of pilot and team behavior in specific situations can be generated. Visualizing such typical cases, and in particular illustrating the deviation from the planned data, could help instructors and training designers in understanding, analyzing and refining the scenarios. In addition, human factors researchers could benefit from such visualizations for analysis of pilot and team performance. Providing such insight into previous experiences and outcomes, could prove beneficial in optimizing the planning and pre-training stages.

Similarly, pilots may benefit from comparative visualizations of their individual performance and such aforementioned typical cases in specific situations. Through direct visual presentation of spatial and temporal deviation from the expected outcome, pilots may be further aided in reflecting on their own performance and decisions throughout the executed sortie.

There are plans to incorporate psychophysical sensors, such as eye trackers, in the pilot stations. This would enable many additional non-physical parameters, e.g. a pilot's physical or cognitive state, in the visualization. Various measures, including fatigue, workload, and attention, can be extracted from such sensors, and be analyzed online or offline in a human factors research context. Additional complex composed measures, such as a pilot's current *intent, aggressiveness* or *awareness*, could conceivably be extracted from a combination of the aircraft state (e.g. "determined" or "uncertain" motion pattern), pilot gaze (what the pilot is paying attention to, e.g. surrounding objects or specific data in head-down display devices), and control input (e.g. distinct and clear adjustments vs. constant corrections). With such measures reliably and intuitively visualized, scenarios could be altered in real time to increase difficulty for under-stimulated pilots or ease workload for stressed ones, as appropriate.

Including such additional parameters to the visualization can add significantly to the already visually complex scenarios. Therefore, another direction of investigation is how stereoscopic 3D rendering can aid users in perceiving the visualization. Air combat scenarios are inherently three-dimensional: combat aircraft, missiles, civilian traffic, and CGFs, fly at various flight levels in various directions and attitudes. These items should be unambiguously visualized in 3D space, along with the aforementioned non-physical and composed measures. When this information is rendered on a conventional 2D display, the objects are projected onto a flat image plane, potentially creating visual overlap of multiple objects from various depths in the original 3D scene. This information loss may cause confusion in the interpretation of spatial relationships. The problem is generally reduced by rotating the scene back and forth, generating motion parallax – one of two ways to provide absolute depth information (Cutting 1997). It is not as effective for moving scenes, however. The other way is to provide binocular disparity of the objects in the scene, which is accomplished through stereoscopic rendering techniques. Not only can spatial relationships potentially be better understood; the stereoscopic depth differences themselves may aid in visually isolating overlapping or clustered objects, further reducing clutter and confusion (Parrish, 1994).

Concluding Remarks

This enhanced scenario visualization system is currently being integrated into the training program. Continued development based on operational needs will emerge as the experience of the tool grows, and the development paths outlined in the previous section will be further explored. The effect of the system, at its current state, for the various applications and user groups, is still to be empirically investigated through controlled experiments and studies. Since the development has been carefully driven from user needs and contemporary research findings on the topic, the system is predicted to have a high potential of supporting the training program in particular, but also the research and development programs, and the initial experiences and reactions have indeed been very promising.

The use of the system for research and development is under way. For research purposes it will primarily be used by human factors researchers for studies of human performance, human behavior, and training effectiveness. Researchers working on CGFs have also expressed the potential of using replays and the visualization aids for knowledge elicitation processes. For simulator-based development, operational analysts, and to some extent engineers, plan to use the tool for in-depth analyses of behavior and performance of weapon and sensor models, as well as their effects and implications on tactical behavior, both on an individual and team level.

Acknowledgments

This work was funded by the Swedish Armed Forces Research and Development Program. The views presented in this paper reflect those of the authors and not necessarily the Swedish Armed Forces, the Swedish Defence Research Agency or the Swedish Air Force Combat Simulation Centre.

References

Burke, C.S., Salas, E., Wilson-Donnelly, K., & Priest, H. (2004). How to turn a team of experts into an expert medical team: guidance from the aviation and military communities. *Quality and Safety in Health Care*. 13 (Suppl. 1), i96-i104.

Cannon-Bowers, J. A., & Bell, H. H. (1997). Training decision makers for complex environments: Implications of the naturalistic decision making perspective. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 99-110). New York, NY: Lawrence Erlbaum Associates.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Cutting, J. E. (1997). How the eye measures reality and virtual reality. In *Behavior Research Methods, Instrumentation, and Computers*, volume 29, 29-36.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*(8), 725-747.

Freeman, J., Salter, W. J., & Hoch, S. (2004). The users and functions of debriefing in distributed, simulation-based team training. *Proceedings of the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC) 2004*. Orlando, FL: National Security Industrial Association.

Holding, D. H. (1987). Concepts of training. In G. Salvendy (Ed.), *Handbook of Human Factors* (pp. 939-962). New York: John Wiley & Sons, Inc.

Kolb, D. A. (1983). *Experiental learning: Experience as the source of learning and development*. Upper Saddle River, NJ: Prentice Hall Inc.

Parrish, R. V., Williams, S. P., & Nold, D. E. (1994). Effective declutter of complex flight displays using stereoptic 3-d cueing. *Technical Report 3426*, NASA.

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

A HUMAN PERFORMANCE MODEL OF COMMERCIAL JETLINER TAXIING

Michael D. Byrne, Jeffrey C. Zemla
Rice University
Houston, TX
Alex Kirlik, Kenyon Riddle
University of Illinois Urbana-Champaign
Champaign, IL
Amy L. Alexander
Aptima, Inc.
Boston, MA

There is a critical gap in attempting to predict aviation system performance between large-scale engineering-oriented simulations and human-in-the-loop experiments. In order to bridge this gap, we have constructed a model of a human pilot taxiing a commercial jetliner using ACT-R, a computational theory of human cognition and performance. The model was constructed on the basis of a task analysis that was synthesized from a mixture of prior literature, official procedures, and consultations with SMEs. The model taxis a simulated 737 in the X-Plane flight simulation environment. Our approach to validation, which we believe to be unique, will be to validate the model against actual taxi trajectories recorded by real pilots at DFW airport in actual operations. The model can ultimately be used to provide higher-fidelity pilots in large simulations or used to populate the environment in human-in-the-loop experiments.

Surface traffic management is a critical concern for NextGen. The task of optimizing the timing and route of each plane from the gate to the runway is computationally difficult, and ground controllers do not have the proper resources to do such optimization. This task becomes even more complex as the amount of surface traffic increases, which leads to delays that cost airlines time, fuel, and money (FAA, 2010).

In recent years, this task has been made easier with the introduction of a variety of technologies that, taken together, provide controllers with precise, real-time positions of all nearby planes. This helps ground controllers perform their job more efficiently. Futhermore, researchers have begun experimenting with computer algorithms that calculate the optimal sequencing and routing of planes as they move about the taxi surface (Malik, Gupta, & Jung, 2010). However, the current methods for testing these algorithms are limited in several ways.

One common method is to employ human-in-the-loop (HITL) experiments. In order to perform experiments with ground controllers, the simulation environment must be populated with aircraft that respond flexibly and in real time. This means human "pilots" are necessary, because the capabilities of real pilots play an important role in determining the validity of the algorithm. For instance, an algorithm may produce high throughput by closely spacing planes together, but human pilots may not be able to safely implement the required procedures. In addition, the reaction times of pilots can add latency to the system that is not apparent otherwise. While HITL testing can provide realistic results, it suffers from certain drawbacks. First, it is expensive, as thousands of man-hours can be required to test new changes. This limits the scale of issues that can be considered. For instance, predicting the rate of runway incursions that arise from several nearby airports over the span of a few months is simply not tractable.

Another common method for testing these algorithms is to use computer simulations, such as the Surface Operations Simulator and Scheduler (SOS^2; Wood, Kistler, Rathinam, & Jung, 2009). Computer simulations overcome the major concerns of HITL testing: they are both fast and comparatively inexpensive. However, current computer simulations have their own limitations. SOS^2 does not dynamically simulate human pilot behavior. Responses to ground controllers are predetermined, meaning that the planes in these simulations always react to air traffic controllers without error and in zero time. Furthermore, off-nominal situations are neither detected nor corrected by the simulated pilots, since they lack the cognitive capabilities of true human pilots. While such omissions are not uncommon in the early stages of research on a problem, they expose a serious gap in our ability to accurately predict the outcome of changes to the surface management systems.

A computational cognitive model, once developed, has the benefits of being both fast and inexpensive, while also integrating key components of human cognition and behavior that may affect the simulations, such as pilot errors, response times, and detection of off-nominal conditions.

Model Platform

We constructed our cognitive model using ACT-R 6.0 (Anderson, 2007), a computational cognitive architecture that simulates human performance through the interaction of lower-level psychological processes, such as memory retrieval and visual attention. ACT-R has proven capable of modeling complex tasks in both aviation (Byrne & Kirlik, 2005) and driving (Salvucci, 2006) domains. ACT-R models are created by specifying the domain-specific procedural and declarative knowledge of the human being modeled. For our model, this was derived from a task analysis derived from multiple sources, such as subject matter experts as well as airline procedural documentation. This knowledge is then provided to the ACT-R architecture, which interacts with a simulated world to produce a timestamped stream of behavior, which can be slowed down to real time if necessary.

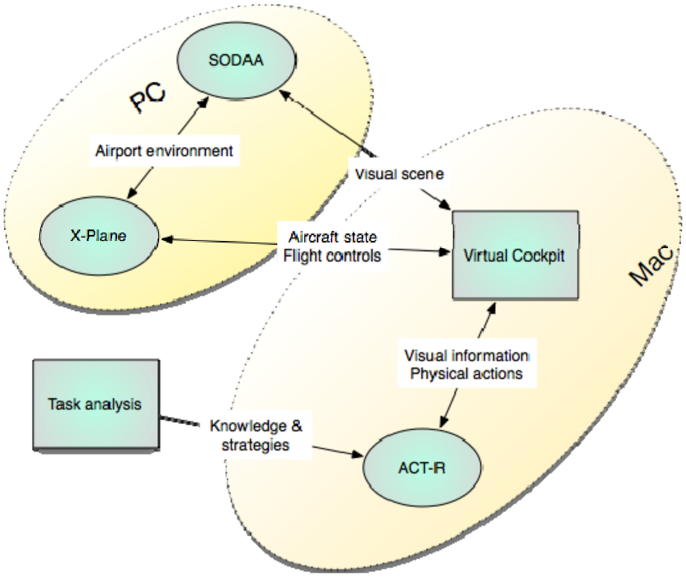


Figure 1. ACT-R communicates directly with the virtual cockpit, both of which run on one machine. In turn, the virtual cockpit communicates with X-Plane, which runs on a separate machine.

X-Plane 9, a commercial flight simulator, acts as the external environment for our model. The model communicates with X-Plane using a plug-in infrastructure, which allows our model to read state variables, such as position and velocity. However, since ACT-R does not contain a machine vision component, visual aspects that are crucial to the model's performance must be redrawn on a proxy interface in a manner that our model can "see." This proxy window takes the form of a Lisp window, with visual objects marked up such that they can be encoded by ACT-R's visual system.

X-Plane handles the physics necessary to make the simulation realistic. For instance, when the model decides to increase the thrust of the plane, X-Plane determines the acceleration and velocity depending on the type of plane the model is currently piloting. In addition, X-Plane provides detailed maps of airports worldwide, including signage on the taxiways. This enables us to simulate real clearances at real airports, which produces concrete predictions about how well these systems work at any particular airport. The resulting system runs on two machines, a PC running X-Plane and a Macintosh running Lisp and maintaining both the virtual cockpit and ACT-R. The system is depicted in Figure 1.

Model Overview

Prior to constructing the model, we surveyed airline procedural documentation and questioned pilots in order to determine what domain-specific information was necessary to create the model. With this information, we conducted a task analysis that defined the sequence of operations a pilot must perform to taxi a plane. The task analysis identified several key components that are required for a pilot to successfully taxi a commercial jetliner. These components include navigating the taxiways, steering the plane, maintaining the speed of the aircraft, and scanning the taxiway for incursions. Each of these components represents a high-level goal that the pilot is responsible for. The details of each component are described in the sections below. There are, of course, additional responsibilities of the pilot that are not accounted for by these four components. Notably absent are goals for processing incoming and outgoing audio transmissions to air traffic control, as well as a variety of pre-flight items (including checklists). These tasks are absent primarily for tractability, however we plan to integrate aspects of these tasks in later versions of the model.

Navigation

The model keeps a representation in memory that maintains the current location (taxiway) of the model, the next taxiway in the clearance instructions, and the action to perform at that taxiway (e.g., hold, turn right, turn left). In order to navigate, the model begins scanning the visual scene for signs located on or near the taxiways. When the model reads a sign, the content of the sign is compared to the navigational chunk stored in memory, and the model decides what action is appropriate (if any). For example, when seeing a sign indicating the current taxiway, the model checks its memory to determine if the plane is on the correct taxiway. If this is the case, no action is taken. If the plane is on the wrong taxiway, however, the model must take corrective action, such as radioing ground control, coming to an immediate stop, or attempting to find its way back on track. On the other hand, when seeing a sign designating a crossing taxiway, the model checks to see if it corresponds to the upcoming taxiway expected in memory. If it does, the model must then look at the action required at that taxiway to decide what to do next. If the plane is to come to a hold, the model sets the target speed to zero. The actual process of decreasing the throttle and hitting the brake is taken care of by the maintain-speed goal. If the plane is to perform a turn, the model begins looking at the intersection to determine the distance to the turn. When the plane reaches a critical distance to the intersection, the turning subgoal (described in the next section) is initiated.

Steering

The model has two distinct steering procedures, one for intermittent corrective steering, and one specialized for turning.

Corrective steering. This goal is responsible for small steering adjustments, which are necessary to drive straight down a taxiway. Essentially, the purpose of this goal is to minimize the distance of the plane to the centerline of the taxiway. This involves small-angle corrections and can be modeled similarly to how Salvucci's (2006) model handles highway steering of an automobile (though obviously the physics are substantially different).

Turning. This goal is invoked only when the navigation goal signals that a turn is imminent. Steering a commercial jetliner through a turn is a complex perceptual-motor operation, one for which ACT-R did not contain adequate motor capabilities. Based on data from the Surface Operations Data Analysis and Adaptation (SODAA) tool (Brinton, Lindsey, & Graham, 2010), we had access to the turn trajectories of multiple commercial jetliners, and were able to fit those data using a series of motor adjustments based on the speed of the plane and the approximate distance to the hypothetical point where the turn is expected to be completed. The expected heading of the plane can then be calculated as a function of the tangent line at different points on this curve and the model then adjusts the yoke accordingly to match the new heading value. When the yoke adjustments become sufficiently small, the plane is stable and the turn is complete.

Maintaining Speed

The maintain-speed goal controls the speed of the aircraft. When this goal is initiated, the model reads the current speed off of the speedometer, and compares this value to the value of the target speed in memory. If the current speed is too high, the model may apply the brakes. This behavior is stochastic, such that the probability of applying the brakes increases as the speed of the aircraft increases. Typically, the throttle remains in the idle position for the majority of the taxiing, though this also may be adjusted if the speed of the aircraft is too low.

Scanning the Taxiway

As the model taxis the aircraft, it scans the visual environment for possible incursions. Currently, this is limited to other planes present on the taxiway, but this will be expanded to include other possible incursion targets. If another plane is encountered, the model must decide how to act. If the other plane is in front of the model's plane on the taxiway, the model checks its current speed and the distance to the other plane, and determines whether it is necessary to reduce speed or even come to a halt.

Model Validation

For the ACT-R model to be valuable in HITL experiments or computer simulations, it has to be a valid model. Conceptually, the ACT-R model should be on relatively solid ground in terms of validity due to the validation done on the basic components of the architecture and to the extent that the task analysis correctly captures the taxiing task. However, further validation is crucial and we have a unique opportunity in the case of this particular modeling effort. Rather than bringing pilots into a lab to perform the same task as the model, we can use real world taxiing data to compare to our model's results.

This is possible because we have access to data collected using SODAA at Dallas Fort-Worth (DFW) airport. The SODAA tool dynamically records the position of each plane on the taxiways and nearby airspace, thus fully capturing the real world data for the taxiing jetliners. X-Plane can "play back" those data, which provides an opportunity for operational validation of the model using historical data (Sargent, 2010).

Thus far, we have only performed face validation as a qualitative assessment of the model's performance by comparing a video of the model performing a specific taxi sequence to a video of the same taxi sequence recorded in the SODAA data in X-Plane. See Figure 2 for a frame of what the running system looks like. We can simultaneously observe the ACT-R model as well as the X-Plane environment that shows the behavior of the controlled aircraft. The model now performs well enough that it is difficult to determine simply from watching the X-Plane view whether it is a replay or whether it is ACT-R in control. This is, in some sense, a form of "Turing test" for the ACT-R model.

However, more quantitative validation is necessary. We are currently in the process of developing the underlying framework that will allow historical data validation. This framework involves letting one jetliner to be controlled by the ACT-R model while all the other jetliners are replays from the SODAA data stream. We can then record the trajectory in both time and space of the jetliner controlled by the ACT-R model and compare it to the data it replaced from the SODAA stream. This will enable a quantitative assessment of our model's performance, though it is not entirely clear exactly what measures or metrics are most appropriate for measuring the degree of deviation between model and data. If the model takes a wrong turn, for instance, that is clearly inappropriate. However, what if the model drives almost identical spatial trajectory, but a few seconds slower or faster than the human pilot? Is that valid enough? Obviously, there are some open issues with respect to validation. However, unlike other human performance modeling efforts, we are fortunate in that we have a large volume of data against which to validate model performance.

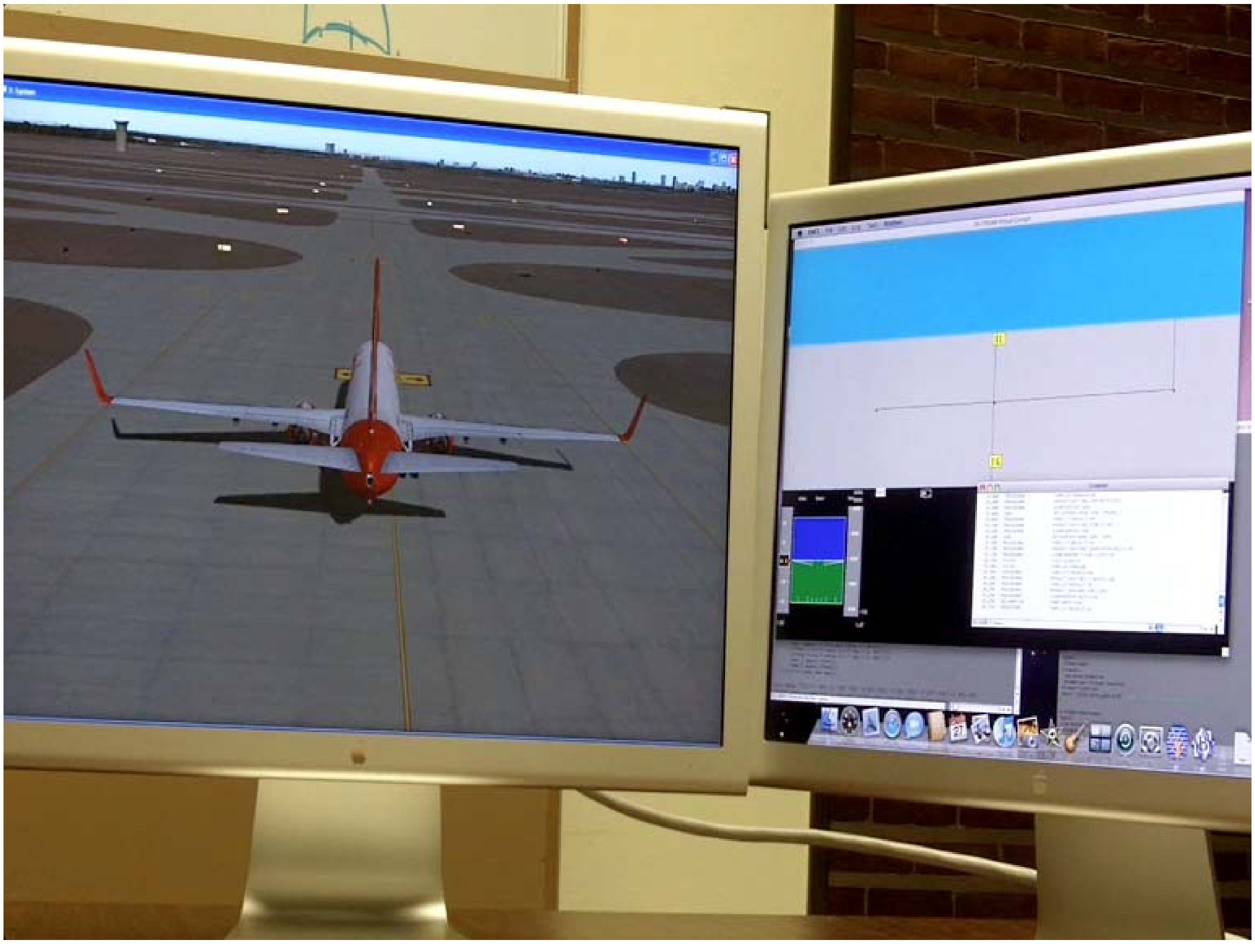


Figure 2. X-Plane is shown on the left monitor, and the virtual cockpit and ACT-R trace are shown on the right monitor.

Discussion

The current model has several possible applications. One potential use is to integrate the model with other computational models such as SOS^2 to allow for rapid prototyping of surface taxiing algorithms. This may be possible by having ACT-R models participate directly in the simulations or indirectly through provision of human performance data. That is, the model may be used to provide estimates for human responses time distributions that are not documented in the literature. Thus, if a researcher needs to know how long it takes for a pilot to react to another plane in a particular scenario, and the empirical literature does not provide adequate guidance, the ACT-R model may be used to estimate human response times in the required situation.

Alternatively, the current model may be used to replace humans in HITL experiments. Essentially, the HITL experiments may remain the same as they are now from the perspective of the ground controller, but instead of having humans controlling the aircraft participating in the simulations, we can use the ACT-R model to do so. This will substantially reduce both the financial and logistical burden on experimenters.

There are other avenues for extending the model in the future. For instance, audio communication with ground control is likely to be displaced by data link communication in near future. Data link provides a textual transcript of instructions and communications with ground control to the pilot, so that he is able to rely less on his working memory. While this technology is likely to make taxiing safer, the addition of a new cockpit display may influence other aspects of the pilot's task (Byrne et al., 2004). With an ACT-R model, we can predict how this new technology will affect a pilot's ability to perform the task prior to deploying it on a wide scale.

Additionally, the model's decision-making capabilities can be augmented. Byrne and Kirlik (2005) investigated how pilots decide when to make a turn based on time constraints. Following an incorrect clearance can increase the probability of a runway incursion. Though the current version of the model is capable of navigating the taxiways, it overemphasizes the role of working memory in this task and is likely to under predict wrong or missed turns, and provides no guidance once a wrong turn has been made. By augmenting the decision-making capabilities of the model, we can better predictions of runway incursion rates.

Overall, the model has potential implications for the way new surface management systems are designed, tested, and implemented. By providing a fast, inexpensive, and accurate method for simulating traffic management, we can help NextGen achieve its goal.

Acknowledgements

We would like to thank Ron Carbonari and Jonathan Sivier at the University of Illinois for their assistance with X-Plane integration and the SODAA data.

This research was supported by the National Aeronautics and Space Administration. The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies or endorsements, either expressed or implied, of NASA, the U.S. Government, or any other organization.

References

Anderson, J.R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

Brinton, C., Lindsey, J., & Graham, M. (2010). The Surface Operations Data Analysis and Adaptation tool: Innovations and applications. *Proceedings of the 29th Digital Avionics Systems Conference.*

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

Byrne, M.D., Kirlik, A., Fleetwood, M.D., Huss, D.G., Kosorukoff, A., Lin, R., & Fick, C.S. (2004). A closed-loop, ACT-R approach to modeling approach and landing with and without synthetic vision system (SVS) technology. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting,* 2111-2115. Santa Monica, CA: Human Factors and Ergonomics Society.

Malik, W., Gupta, G., & Jung, Y.C. (2010). Managing departure aircraft release for efficient airport surface operations. *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, Toronto, Canada.

NextGen Integration and Implementation Office, FAA (2010). NextGen implementation plan. Retrieved December 3rd, 2010, from http://www.faa.gov/about/initiatives/nextgen/media/ngip_3-2010.pdf

Salvucci, D.D. (2006). Modeling driving behavior in a cognitive architecture. *Human Factors*, *48*, 362-380.

Sargent R.G. (2010). Verification and Validation of Simulation Models. *Proceedings of the Winter Simulation Conference*, Baltimore, MD.

Wood, Z., Kistler, M., Rathinam, S., & Jung, Y. (2009). A simulator for modeling aircraft surface operations at airports. *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, Chicago, Ill.

Applying the Reliance-Compliance Model to System-Wide Trust Theory in an Aviation Task

Kasha Geels
Stephen Rice
Jeremy Schwark
Gayle Hunt
Joshua Sandry
New Mexico State University
Las Cruces, New Mexico

System-wide trust (SWT) strategy can occur when operators encounter multiple aids of differing reliabilities. Keller and Rice (2010) have shown effects of one unreliable aid influencing a perfectly reliable aid; participants had a tendency to treat both aids as one entire unit (SWT) rather than as two separate aids (component-specific trust). One limitation was that the use of only two diagnostic aids may not have been enough to generalize their results. This study seeks to further explore SWT with additional aids. Participants performed a 4-gauge monitoring task augmented by a diagnostic aid that provided recommendations of failures. The aids were either 70% or 100% reliable. Data revealed that although providing information and feedback about the aids benefited overall performance, agreement rate data showed that participants still employed a largely SWT strategy. The results from these data are applicable to the design and use of systems which contain multiple aids.

It is often useful to employ the help of an automated aid in environments with complex systems in order to increase safety and efficiency (Sheridan, 1987). This is the reason for the implementation of automated aids in places such as aircraft cockpits, surface transportation systems, and unmanned aerial systems. Parasuraman, Sheridan, and Wickens (2000) have proposed four stages of automation: synthesis, diagnosis, response selection and response execution. For the purposes of this study, we will focus on the diagnosis stage because it provides a recommendation while still leaving the decision-making to the human operator.

Diagnostic automation is used in complex systems so concurrent tasks may be performed while maintaining a specific workload. It is difficult for the human operator to perform concurrent tasks efficiently (Dixon & Wickens, 2006), including monitoring system gauges. Operators also lack the ability to perform cognitively demanding tasks while maintaining the same efficiency as an automated aid due to insufficient cognitive resources (Maltz & Shinar, 2003). Thus, the performance of the system is restricted unless automation is implemented. Even if one task is augmented by automation, multiple tasks can be performed at much higher levels of difficulty while increasing performance levels. The goal of diagnostic automation is to alert the operator of important information only when it is necessary (e.g., Wogalter & Laughery, 2006), such as warnings in an aircraft cockpit, which only alert the pilot of potential problems, so their attention can be focused on other tasks.

Diagnostic aids are not completely reliable. When a diagnostic aid errs, it produces either a false alarm or a miss (Green & Swets, 1966). A false alarm is produced when the aid detects an event that has not actually occurred, while a miss is produced when the aid does not detect an event that actually occurred. Both false alarms and misses negatively affect operator trust in the automated aid (Parasuraman & Riley, 1997; Rice, 2009).

Although previous research has shown how a single automated aid impacts trust (e.g. Dixon & Wickens, 2006; Dixon, Wickens, & Chang, 2005; Dixon, Wickens, & McCarley, 2007; Lee & Moray, 1994; Parasuraman, Molloy, & Singh, 1993; Parasuraman & Riley, 1997; Parasuraman, Sheridan, & Wickens, 2000; Rice, 2009; Rice, Clayton & McCarley, in press; Rice & McCarley, 2008; Rice, Hughes, McCarley & Keller, 2008; Rice, Trafimow, Clayton & Hunt, 2008; Wiegmann, Rich, & Zhang, 2001), not much is known of how trust is impacted by multiple aids. Keller and Rice (2010) have shown that operators tended to treat two separate automated aids in a task as one unit (system-wide trust) rather than as two individual units (component-specific trust), despite the fact that each aid differed in reliability. Therefore, errors produced by one aid negatively affected trust in the entire system.

Keller and Rice (2010) had participants fly simulated unmanned aerial vehicle missions while monitoring two separate gauges, each with its own diagnostic aid. Reliability of the aids varied from 70% to 100%. Trust in the 100% reliable aid was affected by the less reliable aid; thus, participants used a system-wide trust (SWT) strategy.

Six limitations to the previous study will be taken into account in the current study. One limitation to Keller and Rice (2010) is that participants only dealt with two diagnostic aids. Since one gauge makes up 50% of the system, it could be that the unreliability of one aid has a larger effect on trust in the entire system because it makes up half the entire system. The current study further employed the hypotheses to the use of four gauges with one unreliable aid; set to 70%, and three 100% reliable diagnostic aids. It is possible that one unreliable aid that only makes up 25% of the system would not have as much of an effect as one that accounts for 50% of the system. The second limitation to Keller and Rice's (2010) study is that they did not inform participants of the reliability of each aid. Perhaps a SWT strategy is only employed when participants are unaware of each aid's reliability. It is possible that making participants aware of the reliability of each aid may decrease their use of a SWT strategy. A third limitation is that participants were not given feedback on their performance after each trial. It is not known whether feedback is beneficial or not, so providing them with feedback could either help them or potentially confuse them, possibly damaging their trust in the automated aid. This is usually seen when first failures occur (Molloy & Parasuraman, 1996; Wickens & Xu, 2002). The fourth limitation is that Keller and Rice (2010) only used systems that erred with false alarms. Both false alarms and misses can affect operator trust, so the current study used misses in place of false alarms to further generalize to both types of diagnostic errors. Fifth, because the task employed by the participants in the previous study included two concurrent tasks which were unrelated, it may have confused participants as to the reliability of the diagnostic aids. The current study uses only a single-task paradigm, without the worry of a second task. The last limitation is that dependence measures were not taken into account. Dependence on the aids often positively correlates with overall accuracy; although this is not always true (Rice, 2009). So participants' accuracy may be high even if they ignore the aid. Dependence is measured by agreement rates with the aid and dependence is especially high if participants agree with the aid when it is in error.

The current study had two conditions which differed in the level of information: a) no information about reliability levels of the aids and no feedback on performance; and b) information on reliability in addition to feedback on their performance. Participants were given four gauges, with one set at 70% reliability and the remaining gauges set at 100% reliability.

Hypotheses for this study were that a) operators would employ a SWT strategy by combining all four gauges into one system, resulting in similar dependence across all four; b) the distance of each aid from the 70% reliable gauge would not make a difference in the reduction in trust; creating a systematic pull-down effect; and c) participants would have higher overall accuracy rates when given the reliability of each aid.

Method

Fifty-four (32 female, 22 male) undergraduates from a large Southwestern university participated in the experiment for partial course credit. The mean age was 19.82 (*SD* = 3.93). All participants had for normal or corrected-to-normal vision. The experiment was run on a Dell computer with a 3.3 GHz processor and a 22" monitor, with the resolution set to 1024 x 768. Viewing distance was controlled for using a chin rest centered at approximately 21" from the screen. The experimental display consisted of four gauges lined up from left to right; each displayed a randomly assigned 4-digit value. A range indicator was located on the top of each gauge, giving an ideal range for each gauge; which was also randomly assigned. For example, the 4-digit value and range might be 1836 (332). This means that the gauge's safe range is 1836 +/- 332 (between 1504 and 2168), if it went over or under the range, the gauge failed. Underneath each gauge, the aid provided a recommendation of either "Safe" or "Failure" and the left-most aid was 70% reliable. Errors were only misses; determining the gauge was "Safe" when a failure had actually occurred. The other aids were always 100% reliable. The position of the 70% reliable aid stayed constant so participants had to respond to the unreliable gauge first and to determine whether the pull-down effect would still occur even with the differences in distance of each 100% reliable aid from the unreliable aid (whether the right-most gauge would be as affected by the unreliable aid as the second gauge would).

There were two conditions in which the amount of information given to participants regarding their, and the aids', performance was manipulated. In the first condition (NI-NF: No Info, No Feedback), participants were told that the reliability levels of the aids were unknown. There was no feedback after each trial. In the second condition (I-F: Information, Feedback), participants were told the reliability of each aid at the beginning of the experiment, and were also given feedback after each trial.

Participants first signed consent forms and were then seated comfortably in a chair facing the experimental display. They read instructions on screen and were informed that they may ask questions before beginning the experiment. Each trial began with a fixation display that lasted 500 msecs. Following this, the task display presented the Gauge information for 10 seconds, after which a Choice display required participants to determine if the value in each gauge fell within the safe range. They responded by pressing the appropriate key for “Safe” or “Failure”. They could agree with the automation if they desired, but were told the final choice was decided by them. Participants were required to respond to the gauges from left to right, individually. In the Feedback condition, a feedback display was provided for 1000 msecs after each trial. In the No-Feedback condition, a blank screen was presented for 1000 msecs. Participants completed 100 trials. The experiment lasted approximately 20 minutes. Upon completion of the experiment, participants were debriefed and dismissed. A mixed design was employed, whereby the Information factor was between participants, and the Reliability factor was within participants. Participants were randomly placed in each of the between-participant conditions.

Results

The between-participants Information factor referred to whether participants were given: a) no information and no feedback, or b) information and feedback. The within-participants Reliability factor referred to the reliability level of each aid (70%, 100%, 100%, and 100%). These data are presented in Figure 1.

Accuracy

Accuracy was measured by the proportion of successes divided by the total number of trials. An overall 2-way ANOVA using Information and Reliability as factors found a significant main effect of Information, $F(1, 52) = 14.29$, $p < .001$, and a main effect of Reliability, $F(3, 156) = 53.54$, $p < .001$, with a significant interaction between the two factors, $F(3, 156) = 5.40$, $p < .05$. These data indicate that providing more information to participants improved their performance across all reliability levels. Furthermore, performance in the 100% reliable aids was generally superior to performance in the 70% reliable aid.

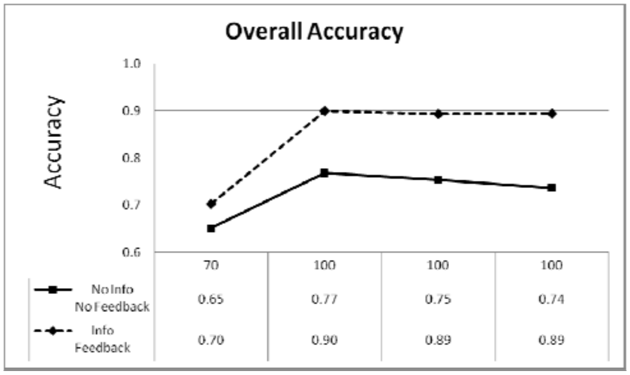


Figure 1. Overall Accuracy.

Dependence

Dependence was measured by the agreement rates between participants and the diagnostic aids (higher agreement rates equal stronger dependence) and is shown in Figure 2. When the aid correctly determined that the gauge had a “Failure”, there was a significant main effect of Information, $F(1, 52) = 14.21$, $p < .001$, with no main effect of Reliability, $F(3, 156) = 1.68$, $p > .10$; however, this was qualified by a significant interaction between the two factors, $F(3, 156) = 3.75$, $p < .05$. Even though providing information and feedback benefited overall agreement rates, it did not appear to cause participants to treat the 100% reliable aids as more reliable than the 70% reliable aid.

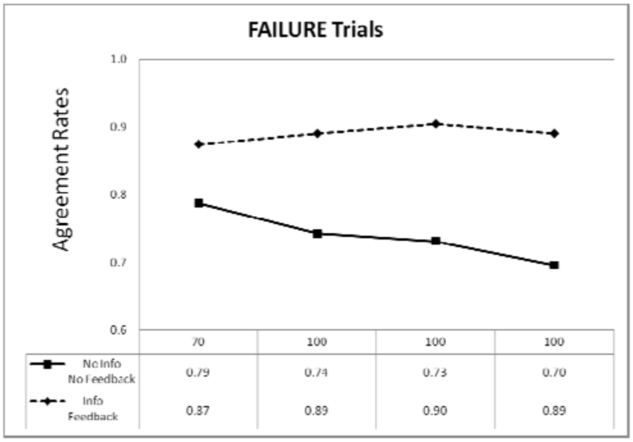


Figure 2. Failure trials.

When the aid correctly determined that the gauge was "Safe", there was a significant main effect of Information, $F(1, 52) = 6.33$, $p < .05$; however, there was no main effect of Reliability, $F(3, 156) = 2.04$, $p > .10$. This was qualified by a significant interaction between the two factors, $F(3, 156) = 6.04$, $p < .01$. As Figure 3 shows, only when information and feedback was presented were participants better able to differentiate between the reliabilities of the aids, and employ a more component-specific trust strategy (although not perfectly so).

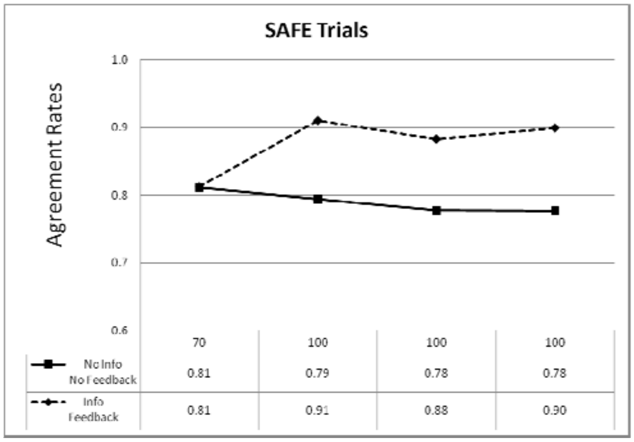


Figure 3. Safe trials.

During trials when the aid incorrectly determined that the gauge was safe (this only occurred for the unreliable aid), adding information and feedback reduced dependence on the aid for these trials, improving overall performance (Agreement Rates: NI-NF = 71%; I-F = 59%; marginally significant: $t(52) = 1.58$, $p = .06$); however, the agreement rates were still quite high given that the aid had erred, causing the drop in overall accuracy for that corresponding gauge.

Discussion

The purpose of this study was to further the findings of Keller and Rice (2010) as it applies to multiple aids while using only a single-task paradigm. Keller and Rice found that dependence on a perfectly reliable aid suffered when it was paired with an unreliable aid. Interestingly, they found that performance for both aids was almost equal. This showed that participants treated these two aids as having the same reliability and demonstrated the use of a SWT strategy.

Data in the current study further support the findings of Keller and Rice (2010). When one aid failed, the other three were treated similarly. Having additional information and/or feedback improved overall performance, but did not increase dependence on the reliable aids as compared to the unreliable aid. Low agreement rates fall in line with previous research which shows that miss-prone automation has a propensity to affect operator trust for both alerts and non-alerts (Dixon & Wickens, 2006; Dixon, Wickens & McCarley, 2007; Rice, 2009). When participants were given information and feedback, they then agreed with the 100% reliable aids more than they did with the unreliable aid. It is possible that they trusted these aids more because they realized the aids would never err (produce misses).

Although a purely system-wide trust strategy did not occur in the present experiment, component-specific trust strategy was rarely used. The trust strategy employed depended upon whether the aid determined a gauge was safe or not and if they were given information and feedback. A pull-down effect occurred in such a way that all three aids were almost equally affected by the unreliable aid. The increase in distance between the gauges did not prevent the pull-down effect from occurring all the way to the right-most aid.

As predicted, overall performance increased when participants were given information as to the reliability of each aid and when they were given feedback of their performance during each trial. This increase was seen equally on all the aids so it did not affect their trust strategy, except for the one situation in which the aid determined the gauge to be “Safe” and participants were given information and feedback.

Findings from the data provide important considerations when designing or using systems which contain multiple aids. Both designers and operators must be aware of the implications of interacting with flawed diagnostic aids and their effect on trust. If one diagnostic aid fails, an operator may lose trust in the other diagnostic aids when, in fact, the aids are performing optimally. This results in a loss of trust that is not necessary and could lead to falsely ignoring important automated aids (e.g. alarms, engine light, computer warnings, etc.). It could also lead to cognitive overload for the user because they take on the tasks normally performed by the diagnostic aid. Overall performance would also decrease because the operator’s attention is taken away from the tasks they normally perform while they try to override the automated task.

Conclusion

Results from this study show the possibility of an operator treating multiple aids as one entire system, rather than as individual aids. When designing such systems, designers must be wary of the implementation of system-wide trust strategies which operators may employ. As shown in both the previous study (Keller & Rice, 2010) and the current one, all automated aids in a display can be affected by a loss of trust in only one aid. The designers of a system should try to avoid this decrease in overall performance by reducing the dependence on only the failed diagnostic aid while maintaining trust in the reliable diagnostic aids.

Acknowledgements

The authors wish to thank Eric Johnson, Audrey Rosenblatt, David Flemming, Sandra Deming, Eduardo Rubio, Crystal Sandy, Rachael Currier, Crystal Garcia, and Natasha Devries for their help in collecting data.

References

Dixon, S. R., & Wickens, C. D. (2006). Automation Reliability in Unmanned Aerial Vehicle Control: A Reliance-Compliance Model of Automation Dependence in High Workload. *Human Factors*, *48*(3), 474-486.

Dixon, S. R., Wickens, C. D., & 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.

Green, D.M., Swets J.A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Keller, D. & Rice, S. (2010). System-wide versus component-specific trust using multiple aids. *Journal of General Psychology*, *137(1),* 114–128.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40,* 153-184.

Maltz, M., & Shinar, D. (2003). New alternative methods in analyzing human behavior in cued target acquisition. *Human Factors, 45*(2), 281-295.

Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors, 38,* 311-322.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency". *International Journal of Aviation Psychology, 3*(1), 1-23.

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

Parasuraman, R., Sheridan, T., & Wickens, D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 30*(3), 286-297.

Rice, S. (2009). Examining single and multiple-process theories of trust in automation. *Journal of General Psychology, 136*(3), 303-319.

Rice, S., Clayton, K., & McCarley, J. (in press). The effects of automation bias on operator compliance and reliance. *Human Factors Issues 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. *Proceedings of the XXIX International Congress of Psychology*.

Rice, S., Trafimow, D., Clayton, K., & Hunt, G. (2008). Impact of the contrast effect on trust ratings and behavior with automated systems. *Cognitive Technology Journal, 13*(2), 30-41.

Sheridan, T.B. (1987). *Handbook of Human Factors.* New York: Wiley.

Wickens, C. D., & Xu, X. (2002). *Automation trust, reliability and attention*. (AHFD-02-14/MAAD-02-2). Savoy, IL: University of Illinois, Aviation Research Lab.

Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science, 2*(4), 352-367.

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.

STORYBOARD DEVELOPMENT AS A METHODOLOGY TO DEVELOP INTEGRATED OPERATIONAL CONCEPTS

Demetrius O. Madrigal, Metric Lab
Nicole J. Stasio, San Jose State University
Daniel N. Peknik, San Jose State University
Bryan U. McClain, Metric Lab
Bettina L. Beard Ph.D, NASA Ames Research Center

We utilized a visualization method known as storyboarding, a design and simulation methodology that is effective at evaluating system implementation concepts in operational context. Using the NextGen Midterm Operational Improvements as a platform, we used human factors and design principles in coordination with input from Air Traffic Controllers and FAA project managers to provide a script and visualization of the Midterm operational environment for Tower and TRACON. The storyboards include a range of both nominal and off-nominal arrival/departure scenarios. The nominal storyboards provide a conceptual baseline of NextGen capabilities and procedures in an ideal scenario. The off-nominal storyboards will serve as a means to generate and evaluate integration concepts that address more challenging and complex situations. This poster session will present storyboard development methods and anticipated application.

The Next Generation Air Transportation System

The Next Generation Air Transportation System (NextGen) is a modernization effort of the National Airspace System (NAS) consisting of a set of Operational Improvements (OIs) that describe added capabilities and procedural changes intended to increase capacity, efficiency, and safety. Like many other large, complex systems, individual NextGen components are being developed in relative isolation and with a phased implementation (both legacy and new systems are planned to operate concurrently). This could result in inconsistencies in user interfaces that would impact the performance of the ANSP, increase workload, decrease overall situation awareness and create the potential for critical operational errors.

Current controller workstations are composed of disparate and variable systems that were developed independently without consistency in user interfaces. The non-integration of systems forces the air traffic controller (ATC) to divert attention away from his or her main focus to acquire needed information such as weather or location codes from separate systems as well as adjust to the use of a different user interface, which can slow response times. Non-integrated systems can increase workload, increase fatigue, reduce performance, and contribute to operational errors and poorly support efficient control of high-volume air traffic (Perry, 1997; Wickens, Mavor, Parasuraman, and McGee, 1998).

The goal of the current project is to develop a concept of systems integration that will guide development of workstations that integrate functionality provided by various current and NextGen systems. These workstations will implement a consistent user interface across air traffic control facilities and a consistent and highly usable interface will support NextGen OIs. The unified user interface should eliminate the need to adjust to different interfaces while performing operator duties. The integration of various functions into a single workstation should reduce the need to divert attention from primary focus to acquire needed information. These benefits should allow the ATCs to more easily manage higher-volume air traffic.

Storyboards

Storyboards are a way of visually conceptualizing a process that incorporates operational context in a way that is easy to understand. Storyboards are intended to illustrate the effect of integrating the different OIs into ATC operations and to explore the impact of different technological integration and implementation concepts. Storyboards for design purposes share commonalities with storyboards in the motion picture industry. In both cases, storyboards are a method to explore concepts through images. Also, in both cases, they rely upon a narrative (telling a story) in order to convey complex concepts and processes in a way that is easily consumable by the reader. In both cases, storyboards are developed from shooting scripts, in which events are described in moderate detail using a formal structure. The formal structure supports conversion to visualization, ease of editing, and allows a modular format in which events can be easily inserted or removed.

The intended consumers of storyboards are system designers, who use concepts and processes explored in the storyboards to make design decisions. Storyboards are conceptual by nature and so should be viewed as propositions of task flow and interaction concepts. They can help to generate discussion by presenting ideas in a concrete format that generates a common understanding. At early points in development, storyboards are very general and somewhat simplified in order to determine system functions and general processes. As the design process progresses, storyboards become filled with detail in order to guide more specific interface design decisions such as information display location and iconography.

For the purposed of our current project, we focused on Air Traffic Control Tower and TRACON operations. Specifically, we examined departure and arrival operations with various nominal and off-nominal events in order to explore how these operations could occur with NextGen technology.

In order to develop the storyboards, the following steps were completed:

1. A thorough review of data source background materials
2. Development of scenarios and events through collaboration with ATC SMEs
3. Identification of differences between current and NextGen operations
4. Preparation of storyboard scripts

5. Completion of scripted event and interface concept visualizations

Data Sources

A thorough understanding of current and planned NextGen air traffic control operations is essential to developing meaningful visual depictions of integration concepts. In order to acquire this understanding, we conducted a thorough review of the following data sources:

- Operational Improvements from the NAS EA
- Task Analyses (cognitive and job task analyses)
- JO7110.65T ATC procedures
- Concept of Operations – these often include operational scenarios that can be helpful in determining task flow for storyboard scripts. These scenarios are typically extremely high level.
    - Midterm, STBO, TFDM, Go Button (XFS), DataComm, IADS, Surface Conformance Monitor, 3D Path Arrival Management (PAM), Collaborative Airspace Constraint Resolution (CACR), Required Time of Arrival (RTA), Flight Object, Time Based Flow Management (TBFM)
- OV5 and OV6 Scenarios
- Input and collaborative feedback from experienced Air Traffic Controllers and Commercial Pilots
- Communication with FAA stakeholders

Scenario Development

A scenario typically includes a general description of events that will occur within a hypothetical situation. A storyboard script is constructed using a scenario as a foundation to establish events and should provide a step-by-step description of the tasks that are undertaken in response to the events. To establish a useful scenario we collaborated with air traffic control Subject Matter Experts (SMEs. This ensured an accurate understanding of current-day operations, processes, and task flow. In order to establish the scenarios, we took the following actions:

- Identified the scope of the script (e.g., integrated/arrival departure, ground taxi operations)
- Identified relevant perspectives (e.g., ground control, local control, supervisor)
- Identified the setting of the script (i.e., location, conditions, timeframe)
- Created a list of events planned for the scenario
- Reviewed the list of events with SMEs to identify additional events

- Collaborated with SMES to identify operations undertaken in response to planned events
- Collaborated with SMEs to identify processes and current-day task flows that compose operations. (For example “what actions would you take to reroute aircraft on the surface following an airport reconfiguration?”)

Current Day and NextGen Operations

To accurately depict operations in the established NextGen timeframe, we identified differences between current-day and NextGen operations. The primary source of information regarding NextGen operations were the OIs provided on the NAS EA website. Additional sources of information included the ConOps and OV6 scenarios. SMEs also provided ideas of how NextGen technology may alter current day operations. Once differences were identified, we described them in the script through the use of an OI *Impact Statement*.

An OI *Impact Statement* specifies, for a given step, how the current-day operation is expected to change once the OI is implemented. OI *Impact Statements* were written to reflect the specific difference between current and NextGen operations.

In addition to OI *Impact Statements*, we also constructed statements called *Integration Points*. *Integration Points* occur when the combination of OI-based capabilities implies capabilities not explicitly stated in the operational improvements. *Integration Points* describe how the OIs or systems interact and what benefit the interaction provides to operations.

Preparing Scripts

Storyboard scripts were organized like a motion picture shooting script . The scripts are structured into short paragraphs (or steps), with each step denoted by a number. Each step and its associated number correspond to a single frame in the eventual storyboard. Each step includes the following content: what action is taking place, where the action is taking place (may be omitted if the same as the previous step), and who are the actors. Each step describes a single action, so if a step is the transmission of a data communications route assignment, the next step might be acknowledging a notification from the system that the message was successfully transmitted. As the different steps link together, the action in each step leads to the next task and establishes a task flow. All scripts were reviewed by several SMEs before being approved for visualization.

Visualization

Visualizations were completed by an Interaction/Visual Designer using a process of hand-sketches, translated into higher resolution images in a free 3D software package named Blender. Visual cells were vetted in the low-fidelity form prior to conversion into three dimensions. Entire Tower, TRACON and En Route facility representations were populated with current field systems such as ASDE-X and CARTS-III. These were then modified to suit ideas represented in the OI *Impact Statements* and *Integration Points*. The 3D cells that resulted were placed into an Adobe Illustrator template designed to scale for multiple uses and best tell a complex story while maintaining a consistent look-and-feel.

Operational Improvements and *Integration Points* were called out in each cell as needed to directly connect visuals to the task at hand. See Figure 2 for an example of a visualized storyboard.

SCRIPTED DIRECTION

1. Interior of an air traffic control tower. Air traffic controller sits at a workstation consisting of two displays, with a slewball and a keyboard with an unspecified layout. The workstation is integrated into a low counter beneath windows.

2. Controller views schedule of upcoming departures which indicates anticipated time until pushback. As a departing flight meets a pre-determined time before scheduled pushback, system indicates anticipated pre- departure clearance request.

3. Interior of commercial airline flight deck. First officer transmits pre-departure clearance request through data communications.

4. Pre-departure clearance request on screen accompanied by flight plan information. The controller can view the RNAV/RNP route on screen. Indicators represent zero errors found in the flight plan, no Traffic Management Initiative restrictions, no anticipated severe weather, and zero predicted constraint conflicts.

5. Controller performs cursory review of flight plan for common errors before performing a single-action approval and pre-departure clearance transmission through data communications.

Figure 2: Example Storyboard Visualization

In order to create visualizations of systems and interfaces still under development, many of the screen designs, equipment layouts and tasks performed by controllers were created for the storyboards. These concepts were generated in accordance with Interaction Design best practices in such a way that they represented NextGen ideas with a level of specificity that avoided narrowly defining a particular concept or idea. Approximately eighty 3D cells and one hundred and twenty screen designs were created, culled down and fitted into template form. The visualizations were designed to be presented in comic-cell layout. Each image was also tagged with keywords in order to be retrieved from a requirement's database when appropriate relational search terms were entered.

Application

The primary application of the storyboards is the exploration of various design and user interface concepts, identification of key operational differences between current and NextGen Midterm operations, and identification of needed integration of systems. This process can also be used for exploration of other operations (En Route, Traffic Management operations, etc.).

Storyboard Architecture and Modularity

The storyboard development process is designed to be modular and flexible. The scripts are designed so that nominal and off-nominal events (e.g., TMI dissemination with ground stop) can be inserted by replacing numbered steps in a baseline script. Likewise, the visualization process includes a library of pre-constructed images and screens that can be adjusted in such a way to insert nominal and off-nominal events relatively quickly. This visual architecture allows for agile production of storyboards that illustrate operational concepts of various nominal and off-nominal events in response to requests from FAA stakeholders. Limitations to this modular process include examination of different types of operations (e.g., meteorological operations) or in different facilities (e.g., ARTCC). For these types of scenarios, no baseline scripts have yet to be developed nor have visual libraries been created, thus they would require significantly more time to develop. In addition, changes to underlying assumptions could take more time to implement.

Cognitive Walkthroughs

Cognitive walkthroughs are currently planned, using the storyboards as stimulus materials. These cognitive walkthroughs will be conducted with a team of air traffic controllers and will describe controller thought processes, information needs, and decision-making processes for each action depicted in the storyboards using a structured interview format. This information can lead to generation of Human Systems Integration requirements, validation of user interface concepts, and additional controller needs.

Additional Research

Storyboards can also be used to structure additional research efforts including part-task and full task simulations of air traffic control operations as well as full demonstrations of operational systems. By providing a systematic set of detailed and validated operational scenarios used across different research efforts, storyboards can support a standardization of research methodology that will allow for quantitative comparison and aggregation in the form of advanced analytics such as meta-analysis. This research standardization fits a major need in evaluating operational and system design concepts. Without this standardization, independent studies will remain difficult to interpret or determine how they contribute to an overall body of knowledge.

Conclusion

Storyboards are a powerful method to generate and evaluate operational concepts. They are commonly used in the design industry to evaluate a product in the context of actual use. The storyboards that we have developed are being used to evaluate user interaction concepts associated with planned NextGen capabilities. Thanks to a modular visual architecture, the storyboards can be modified relatively quickly to accommodate requests for the depiction of additional events in air traffic control operations. The storyboards also provide very detailed scenarios that can be used to support standardization of research efforts. Research standardization would allow for more rapid testing of technology, operational concepts, or user interfaces. This would also assist the FAA with interpretation and consumption of aviation research.

References

Perry, T. S., (1997). In search of the future of Air Traffic Control. *IEEE Spectrum, 34* (8), 18-35.

Wickens, C. D., Mavor, A. S., Parasuraman, R., McGee, J. P. (1998). *The Future of Air Traffic Control: Human Operators and Automation.* Washington D.C.: National Academy Press.

AN EXAMINATION OF A CONCEPTUAL FRAMEWORK FOR PROCESSING AND INTEGRATING INFORMATION

Asha H. Smith, Ph.D.
San Jose State University Research Foundation
Moffett Field, California

Immanuel Barshi, Ph.D.
NASA Ames Research Center
Moffett Field, California

Flight crews rely on displays to assess the state of individual aircraft systems and to remain cognizant of how those systems interact. Degani, et al. (2009) suggest that understanding how humans routinely interpret complex environments should aid in creating displays that help flight crews gain holistic understandings of their vehicles. They propose a six-level hierarchy representing how humans integrate large amounts information. The current experiment sought to understand the costs and benefits of solving classic logic problems when details are presented at key levels of this hierarchy. The results showed that displays representing the highest level of the hierarchy yielded the highest accuracy across a diversity of task types. This effect was strongest when participants only allotted a small amount of time for understanding the problem and display before reading any questions.

Through real-world interactions, humans are faced with a wealth of information that they are able to process and use to solve problems and make decisions (e.g., Barlow, 2001). For example, when we enter a crowded room, we encounter details ranging from the color of the floor tiles to the disposition of the people with whom we interact. Despite the bustling confusion, we are still quite successful at integrating relevant stimuli to reach our goals. Such adaptations are less reliable when it comes to interacting with information on flight displays. One example is the case of Air Transat Flight 236, an Airbus A330-200 that had to make an emergency landing as the result of an undetected fuel leak (Government of Portugal, 2004; reviewed in Degani, Jorgensen, Iverson, Shafto, & Olson, 2009). Indicators notified crewmembers that the airplane was experiencing low oil temperature, high oil pressure, and low oil quantity. Through the relationship between the oil and the fuel systems, these indicators ultimately pointed toward a crack in the fuel line. This relationship was not apparent to the flight crew until quite some time after the landing. While in our day-to-day experiences, we are able to make sense of an abundance of cues, this task is particularly challenging when interacting with information presented in flight vehicles.

Controlling flight vehicles differs from everyday encounters in that flight crews are not fully immersed in the mechanics of their aircraft. Instead, they interact with displays intended to show information that is relevant for completing flight tasks. Degani, Shafto, & Olson (2006) and Degani, et al. (2009) propose that in order to maximize a flight crew's ability to interpret and make inferences about their vehicles, display designs should lend themselves to the cognitive processing we use in day-to-day encounters. Such an adjustment should make it easier for people to intuitively detect patterns and relationships as they do in natural settings. To do so, it is important to understand how it is that people are able to make sense of complex real-world environments. Degani and his colleagues propose a framework that outlines this process. The experiment in this paper examines this framework and its design implications.

The Hierarchy

Degani, Shafto, & Olson (2006) and Degani, et al. (2009) propose a hierarchical processing framework that represents how humans make sense of their surroundings. There are six levels of the hierarchy: *physical quantities, signals, data, information, structures of information*, and *order and wholeness*. Physical quantities are fundamental forms of stimuli in the environment, such as light waves, that may or may not reach our sensory organs. We can extract interpretable *signals* by activating designated neurons in response to certain light wavelengths or line orientations. We then transform these interpretable signals into meaningful *data*. Data includes any stimuli that are available for interpretation. Data is typically abundant and not filtered for relevance.

The remaining three levels are most important in the context of display design because they represent what is likely to be depicted on a display. After signals are transformed into data, we then abstract *information* from the data. Information differs from data in that it is "(1) relevant for the task and (2) meaningful and well suited to the users who need to perform the task" (Degani, et al., 2009; p. 5). Once we have isolated the relevant pieces of information, we integrate that information into *structures of information.* Beyond understanding what is important in a scene, structures of information allows us to also understand how the individual pieces of information are related to one another. Finally, the structures of information are organized into *order and wholeness.* The order and wholeness level represents having a holistic understanding of an environment and high-level patterns become apparent. It is at the order and wholeness level that people are best at making broad inferences and solving complex problems.

Testing the Hierarchy

According to the hierarchy, the optimal display for complex situations should present content at the level of order and wholeness (OW). This arrangement would give users the content and organization necessary to make inferences similar to the ones they make in natural settings. It is important to determine whether or not this is in fact the case. As shown in research examining how to best match displays with task demands (Bennet & Flach, 1992), one potential cost to presenting material at the level of OW could be decreased access to the structures of information (SI) level. Likewise, presenting material as SI could result in less access to information (I). The experiment described in this paper examines this potential cost by testing the hierarchy in its most basic form. As a precursor to directly applying the hierarchy to display design, we are interested in assessing the costs and benefits of applying the framework to solving standard logic problems.

Participants were given logic problem premises that briefly described the problem context. For example, one premise stated "Three of our Olympic swimmers (Mary, Tracy, and Nancy) posed for pictures in three different sports' magazines (Swimming, Splash, and Fast Lane). For their photographs each wore a different colored swimsuit (red, purple, and blue)." Each premise was supported by an external representation (ER) that corresponded with one of the top 3 levels of the hierarchy. Each ER would show, for example, that Tracy appeared on the cover of Splash. To represent the I level of the hierarchy, participants were given unintegrated lists of features that represented the variables needed for each question. ERs representing the SI level displayed unintegrated lists that were aligned in a way that lent themselves easily to integration. OW ERs presented groups of features in an integrated grid that outline the holistic relationship between features (see Figure 1).

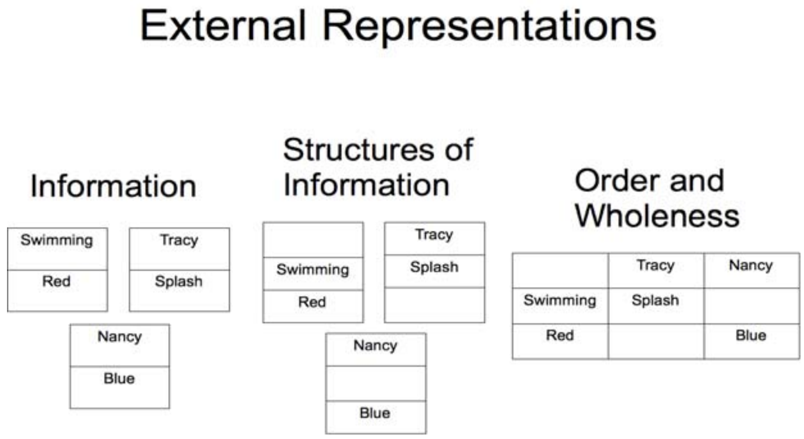
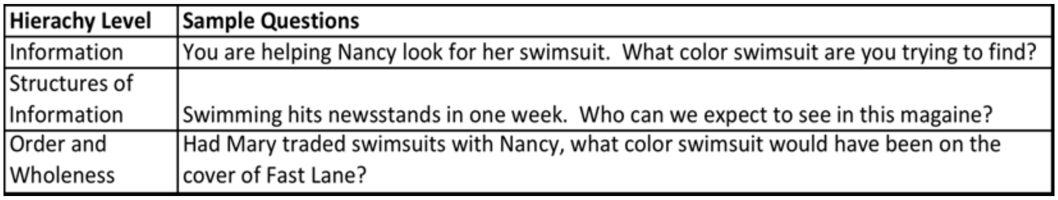


Figure 1: The types of external representations (ER) used to represent the top three levels of the hierarchy.

Following the ER presentations, participants were given questions that required them to interpret the situation at a level that also corresponded with one of the top three levels of the hierarchy. Questions about I required participants to draw from explicit information provided in the problem. Questions about SI could only be solved if participants integrated features in order to deduced one piece of implicit information. The final question type, representing OW, required participants to integrate group features in order to deduce two pieces of implicit information. An example of each question type is presented in Figure 2.



Hierachy Level	Sample Questions
Information	You are helping Nancy look for her swimsuit. What color swimsuit are you trying to find?
Structures of Information	Swimming hits newsstands in one week. Who can we expect to see in this magaine?
Order and Wholeness	Had Mary traded swimsuits with Nancy, what color swimsuit would have been on the cover of Fast Lane?

Figure 2: Sample questions used to represent each level of the hierarchy.

Drawing on the proximity compatibility principle (e.g., Wickens & Carswell, 1995), which states that performance should be optimized when displays parallel task demands, participants should have the most difficulty solving problems when ERs and questions do not align in the hierarchy. If, OW is the optimal presentation level, that ER should work best for all question types.

Methods

Participants

There were 32 participants in this study. Twenty-eight participants were summer psychology students at San Jose State University, and four participants were interns in the Human Systems Integration Division at NASA Ames Research Center, in California. One participant was omitted because he reported misunderstanding the instructions for the task. One participant was excluded because her accuracy level was at floor for all of the questions for one problem scenario. Of the 30 remaining participants, 12 were male and 18 were female. Their ages ranged from 19-years-old to 30-years-old, with a mean age of 22-years-old.

Materials

All stimuli were presented on a computer monitor using Super Lab software. Six experimental logic problem premises were created. Each premise description was made up of two sentences and provided context for the group of features outlined in each ER. Of the six experimental logic problem premises, half contained three groups of three features and half contained four groups of four features. Under the assumption that the 4X4 problems would be more difficult than the 3X3 problems, this variation was created as an exploratory measure of any effects of problem difficulty on performance. For each difficulty level, each premise was accompanied by an ER representing a different level of the hierarchy (see Figure 1).

Six questions were used for each problem premise. Two questions (see Figure 2 for examples) were designed to represent each of the top three levels of the hierarchy. To provide a response for each question, participants had to select one of four multiple-choice answers. The questions in each group ranged between 13 and 20 words (mean length = 16 words). Each multiple-choice response was one word long.

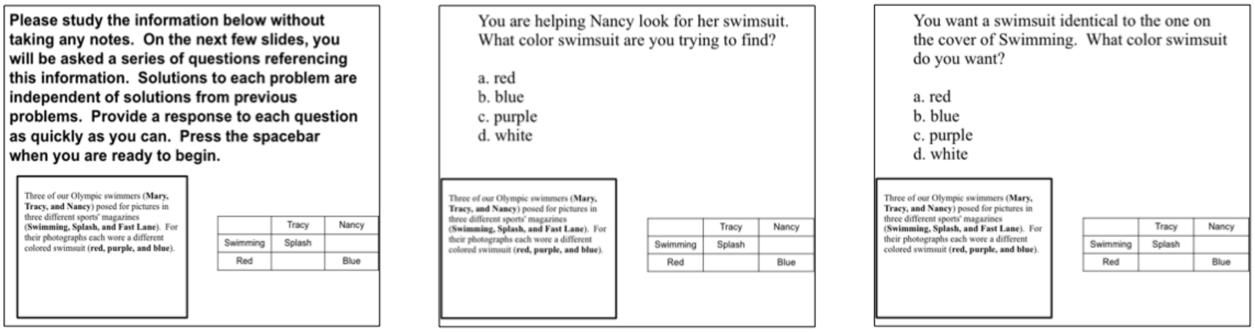
In addition, seven experimental filler questions were used. The filler questions were three and four-term series deductive reasoning problems adapted from Knauff and Johnson-Laird (2002), requiring a true-false response choice. The filler questions were created in order to break the monotony of the experimental logic problems and to divert attention from the experiment goals. Finally, a brief survey was prepared asking participants to report their background information, such as age and gender.

Design and Procedure

Questions were randomized within each problem. Half of the participants were given the 3X3 problems first and half were give the 4X4 problems first. Within each of the problem difficulty blocks, each of the problems was paired with an ER that represented one of the top levels of the hierarchy. Participants saw each ER type twice: once with a

3X3 problem and once with a 4X4 problem. The order of the ER types was randomized between participants. For consistency, the order for the 3X3 ER types always matched the order for the 4X4 ER types.

Once familiarized with the nature of the task, participants pressed the spacebar to advance to the first problem. They read the task instructions and reviewed the problem and corresponding ER. When they were ready, they pressed the spacebar and were presented with the first question. The problem and ER remained on the screen throughout the presentation of all of the questions. Figure 3 shows an example of an instructions and problem scenario screen, followed by two questions. Every set of six questions (one logic problem scenario) was followed by a filler question. The same structure repeated for each of the six problem scenarios. The amount of time spent on each screen (including the time to read each problem and study each ER) and proportion of accurate responses were recorded.



Please study the information below without taking any notes. On the next few slides, you will be asked a series of questions referencing this information. Solutions to each problem are independent of solutions from previous problems. Provide a response to each question as quickly as you can. Press the spacebar when you are ready to begin.

Three of our Olympic swimmers (**Mary, Tracy, and Nancy**) posed for pictures in three different sports' magazines (**Swimming, Splash, and Fast Lane**). For their photographs each wore a different colored swimsuit (**red, purple, and blue**).

	Tracy	Nancy
Swimming	Splash	
Red		Blue

Instructions and Problem Scenario

You are helping Nancy look for her swimsuit. What color swimsuit are you trying to find?

a. red
b. blue
c. purple
d. white

Three of our Olympic swimmers (**Mary, Tracy, and Nancy**) posed for pictures in three different sports' magazines (**Swimming, Splash, and Fast Lane**). For their photographs each wore a different colored swimsuit (**red, purple, and blue**).

	Tracy	Nancy
Swimming	Splash	
Red		Blue

Question 1

You want a swimsuit identical to the one on the cover of Swimming. What color swimsuit do you want?

a. red
b. blue
c. purple
d. white

Three of our Olympic swimmers (**Mary, Tracy, and Nancy**) posed for pictures in three different sports' magazines (**Swimming, Splash, and Fast Lane**). For their photographs each wore a different colored swimsuit (**red, purple, and blue**).

	Tracy	Nancy
Swimming	Splash	
Red		Blue

Question 2

Figure 3: The first screen presents participants with task instructions and initial introduction to problem scenario and ER. They are then given each question accompanied by the problem scenario and ER.

Results

A 2(difficulty level: 3X3 vs. 4X4) X 3(ER type: I vs. SI vs. OW) X 3(question type: I vs. SI vs. OW) repeated measures ANCOVA was run to test for interactions in accuracy performance between type of question and type of external representation, as well as to observe any performance differences in problem difficulty level (3X3 versus 4X4 problems). There was a great deal of variability in the amount of time participants spent reading each problem and studying the corresponding ER before viewing any questions (range of mean reading time across problems for each participant = 3740.34 - 78480.67 milliseconds; mean = 26844.12 milliseconds; standard deviation = 15920 milliseconds). To account for any advantages in allotting more time to understanding the problem scenario, this ANCOVA controlled for the amount of time it took participants to initially read each problem before beginning to answer the questions (referred to as "reading time").

The ANCOVA revealed no significant effect of difficulty level (F (1, 28) = .565; p =.46), so responses for 3X3 and 4X4 problems were collapsed. There was no interaction of ER and question type for accuracy (F (3, 79) = .451; p = .71, Greenhouse-Geisser assumed). There was a main effect of ER type (F (2, 56) = 3.416; p < .05). Participants were more accurate overall using the OW ERs than using the I ERs and the SI ERs (see Figure 4). The ANCOVA also revealed an interaction between ER type and reading time (F (2, 56) = 3.424; p < .05). To examine this interaction and localize the main effect of ER on accuracy between those with fast and slow reading times, the same 2X3X3 repeated measures ANCOVA was run isolating participants who took more time than average (more than 26844.12 milliseconds) and less time than average (less than 26844.12 milliseconds) to read each problem and study the display before proceeding to the first question.

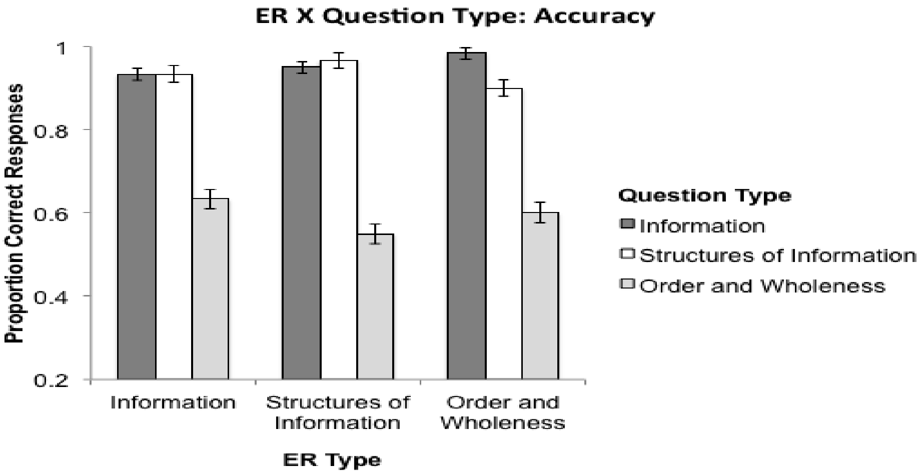


Figure 4: Accuracy responses showing an overall performance for ER and question types.

For participants who took more time than average to study the problem and the display (n = 12, minimum reading time = 26899.8 milliseconds, maximum reading time = 78480.64 milliseconds, mean reading time = 40112.1 milliseconds, standards deviation = 15730 milliseconds), there was no significant main effect of ER type F (2, 20) = .185; p =. 82). Eighteen participants took less time than average to read each problem scenario and examine each display (minimum reading time = 3740.34 milliseconds, maximum reading time = 26822.17 milliseconds, mean reading time = 17988.8 milliseconds, standard deviation = 8070.39 milliseconds). For this group of participants, there was a significant main effect of ER type (F (2, 32) = 7.884; p < .003). Participants were more accurate when answering questions using the OW ER. The interaction between reading time and ER type and accuracy is displayed in Figure 6.

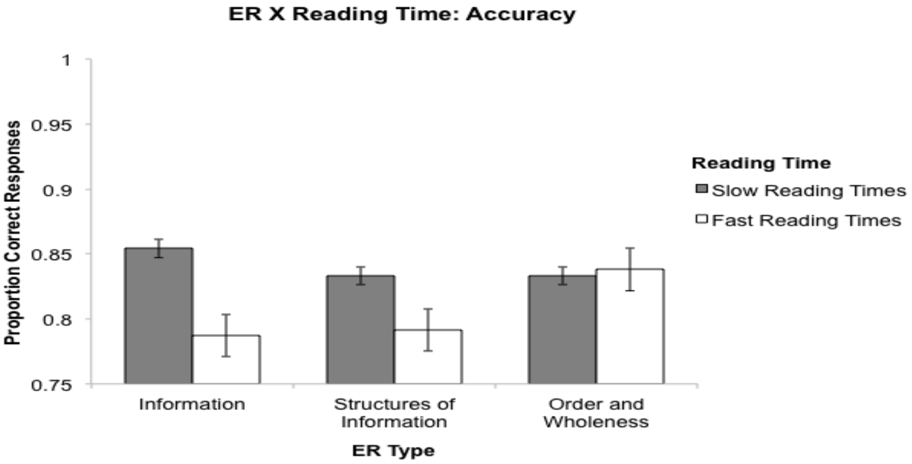


Figure 6: Accuracy responses between ER types for participants with slower than average and faster than average times reading times.

There was a main effect of question type (F (2, 56) = 31.778; p < .001). Participants were most accurate in answering I and SI questions and least accurate with answering OW questions.

Discussion

The current experiment was designed to test the hierarchical framework presented by Degani, Shafto, & Olson (2006) and Degani, et al. (2009) by examining the costs and benefits of presenting material at each level of the hierarchy. There was an overall advantage to using order and wholeness ERs across questions types. The OW advantage was driven by participants who spent less time than average examining each display and problem scenario before viewing any questions. This finding suggests that OW ERs are optimal in novel, fast-paced environments where committing time to fully understanding a scenario may not be possible. There was no interaction between ER and question type. Participants were significantly less accurate when answering questions representing OW (most likely because the OW questions were more complex), but this was true across all three ER types.

One limitation of this experiment was that only one approach to integration was used to represent the different levels of the hierarchy. Participants may have had an easier time using the OW ERs because matrices are commonly used to solve logic problems (e.g. Novick & Hurley, 2001). To account for this limitation, future studies should examine alternative ways of creating integrated ERs beyond traditional matrices. Independent of this limitation this experiment identifies an advantage of using integrated displays.

The Order and Wholeness level of the hierarchy respresents having a high-level understanding of the existing patterns in an environment. For the types of problems used in this experiement, the advantages of having a display represent high-level patterns was strongest when less time was spent examining ERs or understanding the problem premise. The findings from this experiment have positive implications for using OW displays to improve performance in time-constrained situations. When information about high-level relationships is available, problems can be solved more accurately without large time allottments for prior problem scenario and display understanding.

Acknowledgements

We would like to thank Lynne Martin, Dorrit Billman, and Angela Kessell for their feedback in developing this paper. This study was funded by the National Aeronautics and Space Administration (NASA) through a grant to the San Jose State University (grant number NNX08AX13A).

References

Barlow, H. (2001). The exploitation of regularities in the environment by the brain. *Brain and Behavioral Sciences,* 24(3), 602–607.

Bennett, K. B. and Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors,* 34(5), 513–533.

Degani, A., Jorgensen, C., Iverson, D., Shafto, M., and Olson, L. (2009). *On organization of information: Approach and early work.* NASA Technical Memorandum #215368. Moffett Field, CA: NASA Ames Research Center.

Degani, A., Shafto, M., and Olson, L. (2006). Data abstraction and integration of information of display: Analysis of pilot-automation interaction. *Proceedings of the 11th International Conference of Human-Computer Interaction in Aeronautics*, 154-157. Seattle, WA.

Knauff, M. and Johnson-Laird, P. N. (2002). Visual imagery can impede reasoning. *Memory and Cognition*, 30(3), 363-371.

Novick, L. R. and Hurley, S. M. (2001). To matrix, network, or hierarchy: That is the question. *Cognitive Psychology*, 42, 158-216.

Wickens, C. D. and Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, 37(3), 473-494.

HUMAN FACTORS CHALLENGES FACING UAS INTEGRATION INTO THE NAS

Lisa Fern
San Jose State University Research Foundation
Moffett Field, CA

The need to fly Unmanned Aircraft Systems (UAS) in the National Airspace System (NAS) is increasing at a rapid pace. In order to address some of the issues impeding regular UAS access to the NAS, the National Aeronautics and Space Administration has begun a new program to assist the FAA and stakeholder community in establishing requirements for routine operations. One of the technical pillars of this program is the Human Systems Integration (HIS) effort which will work toward two major objectives: development of a research test-bed and database to provide data and a proof of concept of a Ground Control Station (GCS) for UAS integration into the NAS; and work with standards organizations to develop human factors guidelines for GCS operation in the NAS. In addition to a brief overview of the HSI program and objectives, members from the HSI group will present what they consider the biggest human factors challenges to the integration of UAS in the NAS, and how these challenges will be addressed in the program. The authors have been chosen for their individual expertise in differing areas relevant to the issue such as: military and civil UAS operations, small UAS operations, air traffic control and airspace management, pilot/operator challenges, and guidelines development. This paper therefore attempts to provide a comprehensive overview of the wide range of human factors challenges facing the integration of UAS into the NAS for regular operation.

Continuing demand for the use of Unmanned Aircraft Systems (UAS) has put increasing pressure on airspace operations in civil airspace. The need to fly UAS in the National Airspace System (NAS) in order to perform missions vital to national security and defense, emergency management, and science, is increasing at a rapid pace. In addition to limiting UAS usage for civilian applications, current Federal Aviation Administration (FAA) restrictions on UAS access to the NAS constrain the U.S military's ability to fulfill regular training requirements to prepare UAS operators for combat. The National Aeronautics and Space Administration (NASA) has begun a new program to support the UAS community in developing a national strategy for the integration of UAS into the NAS. Program leaders aim to transition design guidelines, algorithms, technologies, operational concepts, and knowledge to the FAA and stakeholder community to assist them in establishing requirements for routine operations.

As one of the technical pillars of this new NASA program, the HSI effort will apply human factors research principles to achieve two major objectives: development of a research test-bed and database to provide data and a proof of concept of a Ground Control Station (GCS) for UAS integration into the NAS; and collaboration with standards organizations to develop human factors guidelines for GCS operation in the NAS.

The purpose of this paper is to provide a brief overview of the HSI program and how it fits within the larger program, proposed methodology, and research objectives, presented by the technical element's Project Engineer. In addition, four other members from the HSI group will present what they consider the most important human factors challenges to the integration of UAS in the NAS, and how these challenges will be addressed within the scope of the program. They will detail previous research efforts and findings, and present how this knowledge will be applied to future research in this new area of UAS integration into civil airspace. The authors have been chosen for their individual expertise in differing areas relevant to the issue, thus providing a comprehensive overview of the wide range of human factors challenges being faced. The authors' areas of expertise and backgrounds include: military and civil UAS operations, small UAS operations, air traffic control (both current and NextGen) and airspace management, pilot/operator challenges, and guidelines development.

UAS in the NAS: Human Systems Integration

R. Jay Shively
NASA Ames Research Center
Moffett Field, CA

NASA is initiating a project to address the formidable technical barriers related to safety and operational challenges associated with routine access of UAS to the NAS. The program will address an increasing need to fly

UAS in the NAS to perform missions of vital importance to Homeland Security, Defense, Emergency Management as well as commercial applications, which directly supports goal six of the National Aeronautics R & D Plan (2010) to "develop capabilities for UAS NAS integration". The overall effort will cover four key technical areas; 1) Separation assurance, 2) Human Systems Integration, 3) Communication and 4) Certification. While these technical areas will necessarily interact, the current discussion will focus on the Human-Systems Integration (HSI) element.

The focus of the HSI effort is specifically on safe access to the NAS. It seeks to provide the capabilities, meet the information requirements and allow interactions with current and NextGen airspace operations. This effort will NOT address overall human factors problems concerning Ground Control Station (GCS) design, nor offer specific solutions for a particular UAS. Instead, the primary goals are to define the information requirements for a UAS to operate in the NAS (now and future), and how to present that information in an integrated/intuitive fashion. This last step is critical to ensure that operator workload remains manageable while increasing situation awareness of the airspace and the entities operating in that airspace. A prototype display suite will be integrated with an existing ground station to serve three purposes; 1) serve as a test-bed for procedures/ displays, 2) provide data input for guideline development, and 3) serve as an instantiation/ proof of concept of those guidelines. Another key goal of this effort is for NASA to work with standards organizations to capture the lessons learned from this effort and others as guidelines for the design of GCS access to the NAS.

A workshop that includes stakeholders from industry, academia, and government will be conducted to ensure that all critical issues are addressed, that no unnecessary duplication of effort takes place, and that strategic partnering is pursued to maximize limited resources. In addition to the workshop, two key tasks will be completed in the first year of the project. An information requirements task will be the crux for all the future work. Through analytical evaluation of the Federal Air Regulations (FARs), interviews with manned aircraft pilots, ATC and UAS operators, we will determine what information is required for UAS operators in a ground station to operate in the NAS, including update rates, latencies and accuracies of the information. In addition, a literature survey and review of all existing ground stations will be completed, including any fielded systems as well as R & D efforts, to ensure that the project will build on, and not duplicate, on-going efforts for operation in the NAS.

Taking the findings of these efforts into the later years of the project will serve as the basis of simulations and flight tests to develop, test and document a prototype candidate display suite that allow integration into the NAS. This work will, in turn, serve as the database with which to work with standards organizations to develop guidelines for UAS GCS operation in the NAS. The project will endeavor to use this and other forums to keep the UAS and Aviation Psychology communities abreast of our efforts.

Challenges to the Seamless Integration of UAS into NextGen: Supporting Pilot Roles and Responsibilities

Walter Johnson
NASA Ames Research Center
Moffett Field, CA

A major challenge to routine operations by UAS in the NAS is the need for special handling by air traffic control (ATC). This need is due to the discrepancies between how UAS vehicles and manned aircraft are presently managed. Today these systems are handled on a case-by-case basis and require special permission for almost all operations in positively controlled airspace (where separation is a direct responsibility of the air traffic controller), such as flight levels above 18000 ft and terminal control areas (TRACONS). In these airspaces, for a UAS vehicle to look and feel to the controller like other vehicles operating in positively controlled airspace the vehicle would have to behave like other manned vehicles. This will require UASs to:

1. File a normal flight plan: departure and arrival locations, en route routing, altitude, etc.
2. Conduct normal departure, climb-out and landing procedures; no special handling
3. Maintain see and avoid and collision avoidance from other traffic (both IFR and VFR traffic)
4. Establish and maintain routine communication with the controller, where lost comm. is a rare event.
5. In the event of lost comm. follow standard radio out procedures so that the controller can predict what control actions are needed to maintain safety of flight for the UAS and others operating near the UAS.
6. Respond to control instructions in a timely manor
7. When requested expedite responses to control instructions
8. Maintain a minimum safe distance from convective weather, icing, and turbulence

Unfortunately, UAS vehicles are not uniformly equipped with standard communication and navigation equipment, do not have see and avoid capability, and their pilots (we use the generic term 'pilot' without taking a stand on required training) are often not in regular direct radio communications with ATC. As a result of these and other differences between UASs and manned aircraft operations (both VFR and IFR) in the NAS, the FAA typically requires sterile corridors for UAS arrivals and departures, and blocked airspace during en route flight.

However, fitting into today's NAS is both an easier and a harder proposition than fitting into the NAS as it evolves over the next 5-10 years. Over the next decade the system for management of positively controlled airspace will undergo a transition. This new system, called NextGen, is being designed around operational concepts that require trajectory-based operations (TBO). In TBO an aircraft is expected to maintain a 4D trajectory, or flight plan, that is sufficiently detailed to support ground-based conflict detection and resolution. The primary reasons for this is to provide automated separation assurance systems, managed by groundside automation, with intent information that will in turn enable automated conflict detection and resolution support to air traffic controllers; and to ensure schedule conformance. The job of the pilot will be to monitor and ensure conformance to the 4D trajectory. This method of controlling the aircraft is an outer-loop supervisory control function. Inner loop control will continue to be handled primarily, by automation embedded in the flight management systems and autopilots, and for lesser equipped, by the pilot, with the goal of keeping the aircraft on its flight plan.

To the extent that the operation of future UAS systems will also tend to be an outer-loop function, the HSI challenge for these systems will very likely parallel those being encountered in the non-UAS elements of NextGen. There are positives and negatives to this new situation. On the plus side, the GCS is not volume and weight limited in the same way as is typical for a flight deck. Subject to basic ergonomic limits, there is a lot of space in which to put displays, automation (computers), and controls (a mouse!). Also, you needn't worry about hardening systems to deal with turbulent encounters (the aircraft may bounce but the ground station does not). So the UAS pilot is not encumbered by in-flight environmental factors that make interacting with advanced controls and displays difficult for pilots. Then there are the challenges, including those to immediate situation awareness (e.g., no window for direct viewing of traffic, weather, terrain, and other environmental hazards); no direct vestibular, visual, or auditory information that pilots typically use to monitor such things as aircraft accelerations, rotations, turbulence, loss of lift, and engine performance (sputtering, engine spikes, etc.); and those that can be traced to loss of continuity and lags in communications between the aircraft and the ground station (radio communication delays and lags, lost link).

Not all of the above challenges affect the ability to fit seamlessly into the present day and future NAS, but several elements can be identified and need to be worked. First, traffic and hazard awareness, and the ability of the UAS pilot to respond to these in a manner similar to a pilot of a manned aircraft. For present day operations this means a flight management system (FMS) and associated interface that keeps aircraft on planned trajectories, displays for weather and terrain, and conformance to the current flight plan, and an on-board system that can handle the see-and-avoid maneuvers. For NextGen operations this means displays of weather and terrain to allow pilots to effectively plan trajectory modifications; and probably displays of traffic, with conflict detection and resolution tools, that will allow them to plan conflict free modifications. Also, if there are times when the UAS must act autonomously to deal with the "see-and-avoid" traffic hazards, the problem of smooth transition of roles and responsibilities back and forth between the UAS and the operator will need to be supported.

Research Steps Towards Operating Small Unmanned Aircraft Systems in the National Airspace System

Anna Trujillo
NASA Langley Research Center
Hampton, VA

Work within the Human Systems Integration element of the NASA Unmanned Aircraft Systems program at the NASA Langley Research Center (LaRC) will evaluate candidate "on-the-loop" Ground Control Stations that will allow UAS operators to maintain safe separation between the unmanned aircraft and other vehicles in the National Airspace System. The LaRC work has a focus on small UAS (sUAS) which present unique requirements due to portable GCS that are often equipped with point-and-click interfaces, with inner loop control handled by vehicle software. "On-the-loop" control, as opposed to "in-the-loop" control, incorporates outer-loop control, or control of higher level and typically slower changing vehicle dynamics, which requires a high level of vehicle autonomy. This high level of autonomy leads to a central problem of integrating UAS into the NAS – a potential automation mismatch between UAS and ATC. In the current system, immediate execution of ATC commands is possible because the commands typically include functions quickly accessible to the pilot on the flight deck, especially through manual control. This ability to easily execute ATC commands is likely due to evolution of piloted flight decks in concert with the NAS. sUAS, however, have evolved in user communities outside of the NAS such as the military. This evolution of system design coupled with portability requirements from both the military and hobbyists, led to operator systems with limited display space, limited vehicle system status, and limited trajectory management. Hence, quickly executing typical ATC commands can be difficult since the operator must often traverse several command menus, and possibly perform mental calculations to translate the ATC command into actions using current day interfaces, which are typically more "point and click" with minimal trajectory definition rather than "stick and rudder."

The above characteristics – an automation mismatch between the UAS and ATC, a high level of autonomy, and portable GCS – will drive the Concepts of Operations (CONOPS) and information requirements for sUAS. The primary driving force behind the CONOPS and information requirements for the GCS will be the need for the UAS to respond directly to ATC in a timely manner. This operating condition will help define the information UAS operators need in order to aviate, navigate, communicate, and manage the systems of the vehicle in the NAS. Initially, research will establish response time requirements for ATC to the vehicle operator, and operator to the UAS for on-the-loop UAS operations. ATC to UAS command response was chosen because this response time is seen as generating the most restrictive requirements in controlled airspace. In addition, operator hand-off protocols (a change in GCS for a given vehicle) will be defined to accommodate differences in operational endurance and in some cases line-of-sight operational requirements. This effort will define procedures necessary for UAS response to vehicle specific operations during operator handoff while maintaining high levels of safety through situational awareness of the vehicle, mission, and the characteristics of the aerospace in which the vehicle is operating. Lastly, the HSI effort will incorporate off-nominal operations into any recommended guidelines. This will aid in showing the robustness and limitations of proposed concepts.

The UAS in the NAS research steps in HSI at LaRC include batch simulations with a pilot model, simulations with human-in-the-loop (HITL) or on-the-loop, and flight-testing. The batch simulations will allow for refinement of GCS concepts before HITL testing. The testing of the concepts generated will follow an incremental process such that initial testing will be with a single UAS pilot operating a vehicle alone in the NAS, then a single UAS pilot operating a vehicle with other simulated vehicles in the NAS, and finally multiple vehicles operating in the NAS that will include vehicles with varied missions and performance characteristics. This testing will provide a proof of concept and design requirements for sUAS operations in the NAS.

Flying NASA Unmanned Aircraft: A Pilot's Perspective

Mark Pestana

NASA Dryden Flight Research Center

A century of aviation evolution has resulted in accepted standards and best practices in the design of human-machine interfaces - the displays and controls which serve to optimize safe and efficient flight operations and situational awareness. The current proliferation of non-standard, aircraft-specific flight crew interfaces in UAS, coupled with the inherent limitations of operating UAS without in-situ sensory input and feedback (aural, visual, and vestibular cues), has increased the risk of mishaps associated with the design of the "cockpit". The examples of current non/sub standard design features range from "annoying" and "inefficient", to those that are difficult to manipulate/interpret in a timely manner, as well as "burdensome" and "unsafe". A concerted effort is required to establish best practices and standards for the human-machine interfaces, for the pilot as well as the air traffic controller.

For example, the presentation format of information is critical. Any teacher knows that the digital clock is no way to teach a classroom to tell time; children barely know the relationships between the numerical symbols, much less their values. In contrast, the traditional analogue display offers all the numbers at once, in order of their relationship, and the hands point to current time. Moreover, the ''trend'' of time is indicated by the movement of the second hand. Similarly, traditional cockpit displays —mostly analogue gauges—have needles that point at numerical values. Many gauges are labeled with ''green arcs'' and ''red lines,'' indicating the normal range of values, or limits, respectively. During typical flight the pilot routinely devotes time and attention to the assessment of information. A quick glance across analogue gauges affords the pilot a ''normal'' assessment if gauges are pointing in normal directions, without needing to read the actual numbers. In the same glance, the pilot can detect a needle pointed in ‘‘abnormal’’ or ‘‘unsafe’’ directions, and assess the condition by noting the value. In digital presentations typical of UAS , precise numbers are displayed, but the pilot must take precious moments to determine whether a number is in the normal range or is trending toward abnormal.

Terminology used in the displays is also vitally important. In these software-intensive systems, development engineers refer to interface control documents to develop the menu-driven commands and displays. When pilots are not part of the development process, the resulting terminology can be baffling. Words such as ‘‘enable’’ or ‘‘inhibit,’’ probably derived from the software coding, supplant standard ‘‘on’’ or ‘‘off’’ commands. Sometimes, clicking the cursor on a command results in the familiar, ‘‘Are you sure?’’ dialogue box common to certain PC-based operating systems. Sometimes, the results are less benign. An infamous case involved a fuel heater switch labeled, ‘‘FUEL HEAT INHIBIT.’’ The pilot was given two choices for the fuel heater command: ‘‘ENABLE’’ or ‘‘DISABLE.’’ To turn on the fuel heater required the pilot to ‘‘disable’’ the ‘‘inhibit’’ … a double negative! This protocol has been changed to, ‘‘FUEL HEATER,’’ ‘‘ON’’ or ‘‘OFF.’’

In addition, roles, responsibilities, knowledge and skill sets are subject to redefining the terms, “pilot” and “air traffic controller”, with respect to operating UAS, especially in the Next Gen NAS. The knowledge, skill sets, training, and qualification standards for UAS operations must be established, and reflect the aircraft-specific human-machine interfaces and control methods.

NASA’s recent experiences with flying its MQ-9 Ikhana in the NAS for extended duration, has enabled both NASA and the FAA to realize the full potential for UAS. Ikhana is a Predator-B/Reaper UAS modified for research. After several years of planning and negotiation, in 2007 the FAA granted a Certificate of Authorization (COA) to NASA which enabled the use of Ikhana for wildfire geo-location missions while flying in Class A airspace in the NAS. The technology which was demonstrated, coupling a NASA infrared sensor with the aircraft satellite data-link system, provided unprecedented information to fire-fighting agencies in near real-time. The concerted planning effort involved detailed analyses to optimize expected flight routes without significant impact to air traffic lanes, or flight over dense population centers, and demonstrate a safe return to home base in the case of loss of the command-and-control link. Additionally, to prepare for systems emergencies (e.g. generator failure, engine failure) which would prevent Ikhana from returning to its primary landing site at Edwards AFB, detailed analyses were performed to locate remote landing sites and selected military bases within a 50-mile glide distance of the flight route. Four initial demonstration missions, of up to 20 hours, were flown in August and September, reaching as far as Montana, Wyoming, Idaho, and Washington, covering several fires on each mission. In October, when Southern California “exploded” in multiple wildfires, Ikhana flew four missions in a five-day period. The overall success of this campaign is measured best by the feedback from the fire incident commanders- e.g. “…lives and property saved… because of NASA’s Ikhana”. The efforts by NASA and the FAA brought about a greater understanding of the implications of current limitations (e.g. lack of sense-and-avoid capabilities) in mitigating the risks and tackling the challenges associated with integrating UAS in the NAS.

Alan Hobbs
San Jose State University
Moffett Field, CA

Ground control stations of a UAS range from small commercial off-the-shelf laptops, to sophisticated purpose-built shelter trailers or control facilities. The GCS, which is essentially a ground-based cockpit, must provide the pilot with the information needed for safe flight, and enable the pilot to make the necessary control inputs in a timely and accurate manner. A challenge for the designers of GCS is to enable the UAS pilot to maintain situational awareness in the absence of the rich perceptual cues available to the pilot of a conventional aircraft (Pestana, 2008).

Many of the display requirements for a GCS are similar to those in conventional aircraft, such as airspeed, attitude and the performance of on-board systems. In other cases however, the GCS must provide the UAS operator with unique classes of information. Examples are: the strength of the communication link, the frequency band in use for aircraft commands, radio spectrum activity, and the status of the GCS itself. The pilot may also require visual information on the surrounding environment, either for landing or takeoff, collision avoidance, or weather avoidance (Cooke, et al. 2006, Hobbs, 2010).

Since the early days of aviation, design principles were identified for cockpit displays and controls (e.g. Fitts and Jones, 1947). With the introduction of "glass cockpit" aircraft, much was learned about the optimal interface between human and automation (Wiener, 1988). Traditional cockpit displays and controls will have a reduced place in GCS, as unmanned aircraft are increasingly designed to be controlled via automation, with the operator making command inputs using "point and click" devices, text-based menus, and dialog boxes. Already, significant design deficiencies are being identified in GCS built using computer interfaces. Problems have included error-provoking control placement, non-intuitive automation interfaces, an over-reliance on text displays, and the need for complicated sequences of menu selection to perform minor or routine tasks (Pedersen, Cooke, Pringle, & Connor, 2006).

In summary, the design of the GCS presents several sets of challenges for NAS integration. These relate to the changed experience of the pilot (notably the reduction or complete elimination of direct perceptual cues), the unique information requirements associated with teleoperation, the heavy reliance on automation, and the introduction of displays and controls adapted from consumer electronics devices. Human factors guidelines for GCS design will help to reduce design-induced errors, maintain pilot situational awareness and ensure that pilot workload is manageable.

There are several on-going efforts to develop standards for GCS by groups external to NASA. These include RTCA special committee SC203, EUROCAE Work Group 73, and workgroups reporting to ICAO, and NATO. A difficulty faced by these groups is the wide variety of UAS and the "moving target" nature of the field as rapid technological developments continue to occur. The current project will involve coordination with these existing external groups to avoid duplication, while also coordinating with NASA research groups, notably the "GCS Database and Proof of Concept" team. While other workgroups focus on airworthiness or operational problems, the HSI workgroup will bring NASA's unique human systems integration perspective to focus on the human factors issues of GCS design.

Guided by CONOPS information, our objective is to develop best practice design guidelines for GCS, initially focusing on the human/system interface issues relevant to Medium Altitude Long Endurance (MALE) systems. This will commence with the development of broad guiding design principles for human system interface, under which will nest more specific guidelines. For example, a broad principle may relate to requirements for avoiding adverse weather. Under this would be guidelines concerning the weather challenges facing the pilot, and the information and control capabilities that must be available to the pilot to enable these challenges to be overcome.

The work will draw on existing guidelines and standards documents, including design standards for conventional aircraft cockpits and also industrial control systems, where relevant. The project will also draw on lessons learned from UAS operational experience, incident and mishap reports and direct observations of pilot interactions with existing GCS.

References

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

Fitts, P. M. & Jones, R. E. (1947). *Analysis of factors contributing to 460 'pilot error' experiences in operating aircraft controls*. Report TSEAA-694-12, Aero Medical Laboratory, Air Materiel Command, Wright Patterson Air Force Base, Dayton Ohio.

Hobbs, A. (2010). *Unmanned aircraft systems*. In E. Salas & D.Maurino (Eds.). Human Factors in Aviation, 2nd edition. Elsevier: San Diego.

Pedersen, H. K., Cooke, N. J., Pringle, H. L, & Connor, O. (2006). UAV human factors: operator perspectives. In, N. J. Cooke, H.L. Pringle, H.K. Pedersen, & O. Connor (Eds.), *Human factors of remotely operated vehicles* (pp. 21–33). Oxford: Elsevier.

Pestana, M. (2008). *NASA MQ-9 Ikhana Human Factors: Pilot Perspective*. Paper presented at NTSB Forum on the Safety of Unmanned Aircraft Systems. Washington, DC.

HUMAN FACTORS CONSIDERATIONS FOR INTEGRATING TRAFFIC INFORMATION ON AIRPORT MOVING MAPS

Michelle Yeh and Scott Gabree
United States Department of Transportation
John A Volpe National Transportation Systems Center
Cambridge, MA

The purpose of this research effort was to identify human factors considerations in the integration of traffic information and surface indications and alerts for runway status on airport moving maps for flight deck displays. The information is primarily intended to support the development of Federal Aviation Administration (FAA) policy and guidance for surface conflict detection and alerting (e.g., runway incursion alerting). The Volpe Center gathered information in two ways: (1) from observations made during FAA-sponsored demonstrations of Automated Dependent Surveillance – Broadcast (ADS-B) surface conflict detection algorithms, and (2) collecting pilot feedback on the use of ADS-B in the general aviation environment. Four human factors issues pertaining to the integration of traffic information and alerting are addressed in this paper: use of color, the presentation of indications and alerts, the design of traffic symbols, and position accuracy.

The Federal Aviation Administration (FAA) has requested human factors input to support the development of policy and guidance for surface conflict detection and alerting on airport moving maps. A surface conflict detection function has two components: an airport moving map that depicts a dynamic image of the airport along with the aircraft's current position (see Yeh and Eon, 2009 for examples); and a traffic function to overlay aircraft traffic based on Automatic Dependent Surveillance – Broadcast (ADS-B) or other surveillance technologies. Figure 1 presents an example of an airport surface moving map that shows traffic information. In the figure, ownship is represented via a magenta triangle (in the center of the figure), and traffic aircraft are represented via the chevron, diamond, and bullet-shaped symbols.

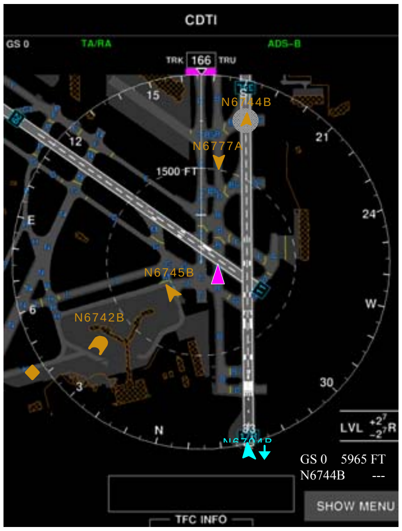


Figure 1. Photo courtesy of ACSS (Excerpted from Yeh and Eon, 2009).

The FAA provides guidance for the design and approval of airport moving maps and traffic surveillance applications in Technical Standard Orders (TSOs) and Advisory Circulars (ACs). TSO-C165, *Electronic Map Display Equipment for Graphical Depiction of Aircraft Position*, (FAA, 2003), and Advisory Circular (AC) 20-159 (FAA, 2007), *Obtaining Design and Production Approval of Airport Moving Map Display Applications Intended for Electronic Flight Bag Systems*, address airport map displays. Both TSO-C165 and AC 20-159 reference RTCA DO-257A, *Minimum Operational Performance Standards for the Depiction of Navigational Information on Electronic Maps*, which defines the minimum standards for equipment that is intended to provide ownship position on an electronic map display, whether it is on the airport surface, in-flight, or vertical situation display. TSO-C195, *Avionics Supporting Automatic Dependent Surveillance – Broadcast (ADS-B) Aircraft Surveillance Applications (ASA)* provides guidance for a traffic function (FAA, 2010). TSO-C195 references RTCA DO-317, *Minimum Operational Performance Standards (MOPS) for Aircraft Surveillance Applications System (ASAS)*, which addresses requirements for the presentation of traffic information.

Although existing FAA guidance addresses many human factors considerations associated with individual traffic and airport surface map depictions, guidance is needed for the integration of this information, as well as advanced capabilities such as the presentation of surface indications and alerts. The performance standards specified in RTCA DO-257A and RTCA DO-317 define the *minimum* capabilities required for an airport moving map and traffic display, respectively, but higher performance standards may be needed to support more advanced capabilities. Industry input for the airport surface conflict detection concept is documented in RTCA DO-323, *Safety, Performance and Interoperability Requirements Document for Enhanced Traffic Situational Awareness on the Airport Surface with Indications and Alerts (SURF IA)*(RTCA, 2010). RTCA DO-323 defines a concept for airport surface conflict detection consisting of the presentation of *indications* and *alerts*. RTCA DO-323 defines indications for "a normal operational condition that could become a runway safety hazard" (p. 6) and *alerts* for "non-normal operational situations where collision hazard exists or a collision appears imminent" (p. 2). The concept for displaying indications and alerts consists of highlighting the potentially incurring (i.e., conflict) traffic aircraft or conflict runway, showing symbols for traffic aircraft that are offscale or outside of the current display range, presenting text information, and presenting an auditory message for alerts. The design of each of these elements is specified by the manufacturer.

The FAA has requested additional human factors guidance to support the development of minimum performance standards for surface conflict detection. To support the FAA, the Volpe Center gathered information in two ways:

1. Observing demonstrations of ADS-B surface conflict detection algorithms sponsored by the FAA Surveillance & Broadcast Services Office.
2. Observing demonstrations of the airport moving map and traffic display at Embry-Riddle Aeronautical University (ERAU) in Daytona Beach, Florida, and interviewing their flight instructors to understand the impact of the airport moving map and traffic display in a general aviation operating environment.

The methodology used to gather information is described in the next section. The preliminary findings from these efforts are then described. This paper presents a brief overview of this research effort; more detail is provided in Yeh and Gabree (2010).

Method

Observations during FAA-sponsored ADS-B Demonstrations

In 2008, the FAA Surveillance and Broadcast Services (SBS) Office sponsored a program to demonstrate surface conflict detection algorithms and alerting using ADS-B, and to test and validate requirements for an ADS-B surface application. The Volpe Center observed four demonstrations. The first three were conducted in November and December, 2009, at Philadelphia International Airport (PHL) by ACSS in partnership with US Airways. An US Airways Airbus A330 and an ACSS–owned Beechcraft King Air were equipped with ACSS-prototype software. The software was displayed in the A330 on two Electronic Flight Bags (EFBs), one installed under the left and right side windows. In the King Air aircraft, the software was displayed on an EFB, temporarily mounted for the demonstration in between the pilots' seats. Both aircraft were used to simulate potential conflict scenarios in night-time and day-time conditions.

The fourth observation occurred in January 2010 and was conducted by Honeywell at Seattle-Tacoma and Paine Field airports. Honeywell equipped two aircraft (a Cessna Citation Sovereign and a Beechcraft King Air) with their prototype software. In the Citation Sovereign, the software was presented as part of an integrated navigation display located in front of the captain. In the King Air, the software was shown on a temporary display, installed for the demonstration, and located in front of the first officer.

The demonstrations were scheduled at times when there would not be many operations of other aircraft to minimize the impact of the demonstrations on the airports' operations. During the demonstrations, the Volpe Center took the opportunity to observe the airport moving map and traffic display. The primary purpose of these demonstrations was to evaluate the technical feasibility of surface conflict detection and to examine the human factors and safety impacts. During the ACSS-US Airways demonstration, there was an opportunity to interview the A330 pilots to gather their opinions on the EFB, airport moving map, and the display of traffic, indications, and alerts. Three of the ACSS-US Airways demonstration participants completed a short questionnaire on the usability of the displays. Honeywell also conducted human-in-the-loop simulator evaluations and flight tests of their display concepts in conjunction with their demonstrations (see Khatwa and Lancaster, 2010a, 2010b).

Use of ADS-B on Airport Moving Maps in a General Aviation Environment

The Volpe Center visited Embry-Riddle Aeronautical University (ERAU) in Daytona Beach, Florida in March, 2010 to gather information on the use of an airport moving map and traffic display in a general aviation operating environment. ERAU has equipped 63 aircraft with airport moving map displays as well as technologies to transmit and receive ADS-B. The airport moving map is provided via a Garmin 1000 display (installed in 51 ERAU aircraft) or MX20 (installed in 12 aircraft), which have different capabilities with respect to depicting the airport surface. The Garmin G1000 provides the SafeTaxi application, which shows ownship position on a database-driven airport surface depiction (as shown in Figure 1), if an airport database is available and installed. Some of ERAU's aircraft did not have the full airport database. If no airport database is installed, SafeTaxi will show an airport depiction of runways only (with centerline markings and runway labels). The Garmin MX20 provides a runways-only (with centerline markings and runway labels) depiction of the airport surface.

Information was gathered in three ways. First, ERAU provided a demonstration of the Garmin G1000 and Garmin MX20, both with traffic capabilities. Second, interviews were conducted with eight of ERAU's instructor pilots. The interview sessions had two parts: a paper questionnaire, in which the pilots rated their opinions on the usability of the airport moving map and traffic information on the ground, and a discussion of human factors concerns related to the use of the airport moving map and/or traffic display. The questionnaire consisted of Likert-scale items, in which pilots indicated their level of agreement on a scale of 1 (strongly disagree) to 5 (strongly agree). Pilots who completed both the interview and the questionnaire received a $30 gift card for their participation. Because some pilots were interested in providing input but could not participate in the interviews, ERAU coordinated the distribution of the paper questionnaire to all their instructor pilots. The questionnaires were distributed in April, and pilots were given two weeks to complete it. The Volpe Center provided 150 questionnaires, and 44 were returned. (The eight pilots who participated in the interviews were excluded.) Pilots who completed only the paper questionnaire received a $10 gift card for their participation.

In total, 52 Embry Riddle Aeronautical University (ERAU) flight instructor pilots provided input to this effort, but one of the pilots' feedback was excluded because his primary aircraft was not equipped with an airport moving map. Most of the pilots (48) primarily used the Garmin G-1000. The other three indicated they primarily used the Garmin MX-20. To gauge pilots' flight experience, participants selected one of four categories corresponding to their total hours flown: 24 pilots had 1,500 flight hours or less, 18 pilots had between 1,501-3,000 flight hours, 6 pilots had between 3,001-7,000 flight hours, and 3 pilots had more than 7,000 flight hours.

Results

In general, pilots perceived that the use of an airport moving map showing traffic aircraft, indications, and alerts was a positive enhancement to the safety of surface operations. The general aviation pilots interviewed reported that an airport moving map showing ownship supported their position awareness on the airport surface better than a paper chart alone. The display of traffic information was an additional benefit.

However, several potential human factors concerns were identified through observations of the implementations of the airport moving map and traffic displays and through interviews with pilots using the displays. The concerns noted here focus only on those issues that affect the integration of traffic information and alerts on airport moving maps. General human factors concerns for airport moving maps are addressed in the full report and in other research reports (Gabree and Yeh, 2010; Yeh and Gabree, 2010). Please note that the intent of this effort was not to compare manufacturer displays. Additionally, specific details regarding a manufacturer's implementation of the airport moving map, traffic function, or alerts are not included.

Use of color

Color can be used to call attention to information on the display, but in some cases, color may attract attention when none is warranted. For example, color is often used to code aircraft traffic symbols to indicate whether the aircraft is in the air or on the ground. Aircraft on the ground are often colored as tan/brown, whereas aircraft in the air are cyan. The change in the color of the traffic symbol when an aircraft transitions from the air to the ground (or vice versa) is salient. During one of the demonstrations, this color change indicating the change in aircraft state was observed to be more attention-getting than the onset of a traffic indication. The use of color to indicate the transition from air to ground was proposed to be an effective means to distinguish the two states without increasing the complexity of the symbol set, but it is important to understand the implications of such designs. Color may be an effective way to convey this information if the design can be implemented so that the color changes does not attract attention inappropriately. It will be important to understand the implications of this color change, particularly during routine operations when aircraft are repeatedly taking off and landing.

There are also issues in the colors used for indications and alerts. In the demonstrations, the color blue was often used to indicate a potential conflict, by outlining the active runway or traffic symbol to highlight it. The use of the blue on avionics display may be problematic. First, many airport moving maps have a black background, and the presentation of pure blue on a black background may not be salient at all map ranges. Second, from a physiological standpoint, blue is the shortest wavelength, so it is difficult to bring blue display elements into focus when it is used in combination with other colors. Third, the use of a blue border to highlight the runway led to a yellow afterimage – the illusion of a yellow border surrounding the runway, when the blue border disappears. Afterimages were observed during the demonstrations even upon glancing at the display.

Indications/Alerts

A surface conflict detection algorithm may have several states, including normal operations, indications, cautions, and warnings. Each of these states is represented differently. Additionally, each of the manufacturers participating in the surface conflict detection demonstrations developed their own algorithm defining the operating conditions for presenting a surface indication or alert. As a result, it is conceivable that slight differences in the way each state is represented as well as in the output behavior may be observed. Such differences may make it difficult for pilots to understand the operating conditions under which indications and alerts may be presented, thereby reducing the usability of that information. RTCA DO-323 provides recommendations for the general behavior of the surface conflict detection algorithms, which may ensure consistency. However, the complexity of the algorithms for presenting runway indications and alerts will also require manufacturers and operators to consider how to optimize the training so that pilots understand the symbology, the meaning of the attributes used, and the rules in which the indications and alerts are presented. More training will be needed as the complexity of the algorithms increase.

During the demonstrations, the airport moving map was located on an EFB mounted to the left or right side of the pilot. Visual indications and alerts were presented directly on the EFB and were out of the pilot's primary field of view. Information is most quickly detected if it is presented within the pilot's primary field of view, an area is generally considered to be approximately 15° horizontal and ±15° vertical in front of the pilot (Cardosi and Huntley, 1993). Research is needed to understand the usability and effectiveness of surface conflict indications and alerts depending on the location in which they are presented on the flight deck to ensure that the presentation of alerts is sufficient to attract attention during non-normal operating conditions (e.g., with an auditory alert or a separate visual alert in the primary field of view). Discussion regarding the location of surface conflict alerting has included integrating their presentation to the aircraft's master caution and warning systems, which are presented in the pilot's primary field of view. There is general consensus, however, that the master caution and warning systems are reserved for aircraft-specific failures; consequently, the presentation of a traffic or runway incursion alert in the master caution and warning panel would be inconsistent with current flight deck philosophy.

Traffic Symbols

Symbols for depicting traffic are intended to convey several attributes, including whether the aircraft is in the air or on the ground, its directionality, and its reliability. These symbol attributes must also be interpreted with respect to the different potential aircraft states (i.e., the display of indications and alerts). There are a limited number of symbol attributes for conveying this information (e.g., shape, color, fill). The properties of aircraft that are depicted and how they are depicted may differ from one display to another, and sometimes the depiction may be inconsistent on displays developed by the same manufacturer. For example, one attribute which has been used inconsistently across avionics displays is the fill of a symbol. Symbol fill was used on one display to indicate the aircraft that is the selected target whereas another display used it to indicate those aircraft that are in close proximity to ownship. Color was generally used to indicate whether aircraft is in the air (cyan) or on the ground (tan/brown), but the application of this color varied; on one display, this coding scheme was applied only for aircraft within a certain distance from ownship, but on another display, this coding scheme was applied to all aircraft traffic. Inconsistency in the properties of traffic symbols can make it more difficult for pilots to learn the symbol set, and increase the potential for confusing the meaning of the attribute. Pilots who fly different types of airplanes may use different avionics systems and may not know what information is readily available. When designing new methods for presenting symbology, it is important to consider consistency with applicable standards, as well as related standards such as for TCAS (Traffic Alert and Collision Avoidance System). Recommendations for conveying different traffic symbol attributes are provided in RTCA DO-317.

One open issue is how traffic aircraft should be depicted with respect to the airport moving map when the traffic aircraft falls outside the current display range; that is, when the aircraft is off-scale. The position of the traffic aircraft symbol may be depicted using its relative bearing with respect to ownship, which would be consistent with TCAS conventions for depicting off-scale traffic (when ownship is in the air). However, when an airport moving map is used for reference, the depiction of traffic position using relative bearing could provide misleading information regarding that aircraft's actual position. For example, the relative bearing of a traffic aircraft that is on approach to an airport could lead to the aircraft being depicted as approaching on one runway when it will in fact land on a different runway. Depicting the projected track of the aircraft would offer a more accurate representation of where the selected aircraft will be, although current position may not be precise. Research is needed into the level of precision required for operations on or near the airport surface and how to ensure consistency in the presentation of traffic information with current technologies, such as TCAS. Other open issues for symbols include whether and how to depict traffic that does not meet the performance required for indications and alerts.

Position Accuracy

The position accuracy with which ownship, traffic, and airport information are depicted on an airport moving map must support the intended function. Several errors can contribute to the accuracy of ownship or traffic depiction, including but not limited to position error, latency, survey error, and display resolution. RTCA DO-257A provides accuracy requirements for ownship position on an airport moving map, RTCA DO-317 defines the accuracy requirements for the depiction of traffic on an airport moving map, and RTCA DO-323 recommends accuracy standards for the presentation of ownship and traffic to support indications and alerts. Traffic aircraft, indications, and alerts are *not* intended to be presented if aircraft do not meet their respective accuracy and other data quality requirements.

The demonstration observations and general aviation pilot interviews provided examples of position *in*accuracy to consider in the implementation and integration of traffic information on the airport moving map. During the demonstrations, there were a few instances when the ADS-B signal was lost, likely due to reflection from nearby buildings. There was also one instance where a "false" target was observed on the active runway when ownship was on final approach, because that aircraft target was transmitting erroneous values (zero) for the Navigation Integrity Code (NIC), the Navigation Accuracy Code for Position (NACp), and the Surveillance Integrity Level (SIL). In the general aviation domain, pilots indicated that the information shown by the airport moving map usually matched the out-the-window view, but 25% noted position errors of ownship, traffic, or the airport map. The types of errors reported included the depiction of ownship on or near the edge of the taxiway (2 reports); ownship drawn in the grass (3 reports); and "other," a category which included shadow "ownship" targets, errors in ownship headings, and incorrect depictions of traffic aircraft (6 reports). Of significance is that one pilot noted that a traffic aircraft holding short of one runway was depicted on the airport moving map as being *on* the runway. The participants noted that these errors were rare. Some tended to occur when the system was first turned on or because the airport moving map database was not up-to-date; for example, several of the participants noted a closed taxiway that was still shown on their airport moving map.

Finally, it is important to consider the accuracy/consistency of traffic symbols shown on the airport moving map with the view out the window. All traffic aircraft may not be shown on the airport moving map, presenting an incomplete picture. There are several reasons why a traffic aircraft may not be displayed. First, the introduction of a traffic function will lead to a mixed equipage operating environment, so all aircraft may not yet be equipped. Second, even if all aircraft are equipped with surveillance technologies, some aircraft may not have their transponders on whereas others may not be visible via the surveillance technology. Third, technical limitations can affect the completeness of the traffic picture. During the demonstrations, several instances were noted in which an ADS-B signal was lost and aircraft on the airport surface did not appear or disappeared from the airport map.

Conclusions

This paper provides a preliminary glimpse into human factors concerns with the integration of traffic information, indications, and alerts on an airport moving map. It is important to understand issues faced in the current state of implementation and identify where additional guidance may be needed to support this functionality. This information was gathered in support of the FAA, but manufacturers may also find the information useful for their design and evaluation process.

Acknowledgements

This report was prepared by the Behavioral Safety Research and Demonstration Division of the Human Factors Research and Systems Applications Center of Innovation at the John A. Volpe National Transportation Systems Center (Volpe Center). It was completed with funding from the FAA Human Factors Research and Engineering Group (AJP-61) in support of the Aircraft Certification Service Avionic Systems Branch (AIR-130) and the Technical Programs and Continued Airworthiness Branch (AIR-120). Our FAA program manager is Dr. Tom McCloy, and our technical sponsors are Colleen Donovan and Dr. Bill Kaliardos. We would like to thank Jim Duke and David Gray from the FAA SBS Program Office, Dr. Peter Moertl at MITRE, and ACSS, Honeywell, and US Airways for allowing us to attend and observe the SURF IA demonstrations. We would also like to thank Blake Kelly at Embry-Riddle Aeronautical University for demonstrating ERAU's airport moving map and ADS-B technologies. The views expressed herein are those of the authors and do not necessarily reflect the views of the John A. Volpe National Transportation Systems Center, the Research and Innovative Technology Administration, or the United States Department of Transportation.

References

Cardosi, K. and Huntley, M.S. (1993). *Human Factors for Flight Deck Certification Personnel* (DOTVNTSC-FAA-93-4; DOT-FAA-RD-93-5). Cambridge, MA: USDOT Volpe National Transportation Systems Center.

Federal Aviation Administration. (2003). *Electronic Map Display Equipment for Graphical Depiction of Aircraft Position*. (TSO-C165). Washington, DC: Federal Aviation Administration.

Federal Aviation Administration. (2010). *Avionics Supporting Automatic Dependent Surveillance – Broadcast (ADS-B) Aircraft Surveillance Applications (ASA)* (TSO-C195). Washington, DC: Federal Aviation Administration.

Federal Aviation Administration. (2007). *Obtaining Design and Production Approval of Airport Moving Map Display Applications Intended for Electronic Flight Bag Systems* (AC 20-159). Washington, DC: Federal Aviation Administration.

Gabree, S. and Yeh, M. (2010). Common Human Factors Issues in the Design of Airport Moving Maps. *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics.* 3-5 November, Cape Canaveral, FL.

RTCA. (2003). *Minimum Operational Performance Standards for the Depiction of Navigational Information on Electronic Maps* (RTCA DO-257A). Washington, DC: RTCA.

RTCA. (2009). *Minimum Operational Performance Standards (MOPS) for Aircraft Surveillance Applications System (ASAS).* (RTCA DO-317). Washington, DC: RTCA.

RTCA. (2010). *Safety, Performance and Interoperability Requirements Document for Enhanced Traffic Situational Awareness on the Airport Surface with Indications and Alerts.* (RTCA DO-323). Washington, DC: RTCA.

Yeh, M., & Eon, D. (2009). *Airport moving map industry survey*. (DOT-VNTSC-FAA-09-15). Cambridge, MA: US DOT Volpe National Transportation Systems Center.

Yeh, M. and Gabree, S. (2010). *Human Factors Considerations for the Integration of Traffic Information and Alerts on an Airport Surface Map* (Technical Note; DOT-VNTSC-FAA-10-16 & DOT/FAA/AR-TN10/28). US DOT Volpe Center: Cambridge, MA.

DISPLAY REQUIREMENTS AND ALERTING MODALITIES OF A FLIGHT DECK BASED RUNWAY SAFETY ALERTING SYSTEM

Peter M. Moertl, Ph.D.
Kathleen McGarry, Ph.D.
MITRE/CAASD
McLean, VA

This paper presents the results of a human in the loop simulation that evaluated the display of runway safety relevant traffic and runway indications and alerts on a cockpit display of traffic information (CDTI). The simulation investigated differences between directive versus non-directive alert types and between an airport map with and without taxiway information. 24 pilots evaluated the CDTI in 18 airport surface scenarios that contained conflict opportunities. Findings indicate that with directive alerts, pilots avoided all conflict opportunities, while with non-directive alerts 90 % of conflicts were avoided. Response to directive alerts was generally faster than to non-directive alerts. Limitations for non-directive alerts became apparent in scenarios where pilots had to respond under time pressure. While pilots preferred taxiway information to be displayed on the CDTI, no performance differences were found compared to CDTI's with taxiway information. The paper concludes with implications for the development of avionics standards.

Runway incursions have been a major area of concern for the worldwide aviation community. 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 (FAA 2008). RIs at airports in the United States have been a major area of concern for the U.S. National Airspace System (NAS) for the past several years.

Extensive human factors research has been performed to understand the causes leading to runway incursions and identified the primary causes for runway incursions by pilots, controllers (e.g. FAA 1998, Adam and Kelly 1996, Bales, Gillian, & King, 1989; Steinbacher, 1991), or surface vehicle operators.

Numerous methods have been applied to reduce runway incursions and collisions in the United States (see e.g. FAA, 2003a, 2003b, 2005, 2006a, 2006b, 2007, Honeywell, 2010). A significant amount of research and development activities has been performed on flight deck-based airport surface safety systems (e.g. Jones 2002, 2005; Jones & Prinzel, 2006; Young & Jones 2000) and avionics standards have been defined (RTCA, 2003, 2009). International efforts have included the development of an Advanced Surface Movement Control Guidance System (A-SMCGS) to provide surface traffic management, guidance, and alerting functionality to Air Traffic Control (ATC) and pilots (see IFATCA 2003; Roeder et al., 2008 and Vernaleken, Urvoy, Klingauf, 2007).

In 2000 the National Transportation Safety Board (NTSB) has recommended the development of a ground movement safety system with direct pilot warning capabilities to prevent runway incursions (NTSB, 2000). This recommendation motivated the formation of an RTCA sub-working group under the Special Committee 186 for Automatic Dependent Surveillance – Broadcast (ADS-B) with stakeholders from aviation industry, user communities, and governmental organizations to develop a flight deck application to provide direct runway safety alerts to the flight crew on a CDTI. The application is named Enhanced Traffic Situational Awareness on the Surface with Indications and Alerts (SURF IA). The application development was completed and approved in Dec 2010 (RTCA 2010, DO-323). This paper presents the results of a human in the loop simulation that investigated critical human factors questions associated with the development of the SURF IA application.

SURF IA Application Description

The SURF IA application enhances CDTIs to increase their effectiveness in preventing runway incursions. While basic CDTIs increase the situation awareness of pilots under many situations, Moertl, McGarry & Nickum (2009) identified situations where pilots were not able to use a CDTI to prevent runway incursions. Specifically, during take-off operations, and while on final approach and landing on a runway, pilots were sometimes unable to detect runway safety problems even with a CDTI available to them. In addition, while CDTIs display large amounts of information under many situations, this abundance of information on a relatively small display, requires pilot cognitive resources to extract the relevant information. During times when pilots have less “spare” cognitive resources available, such as during critical flight phases, the usefulness of a CDTI for incursion preventions appears limited.

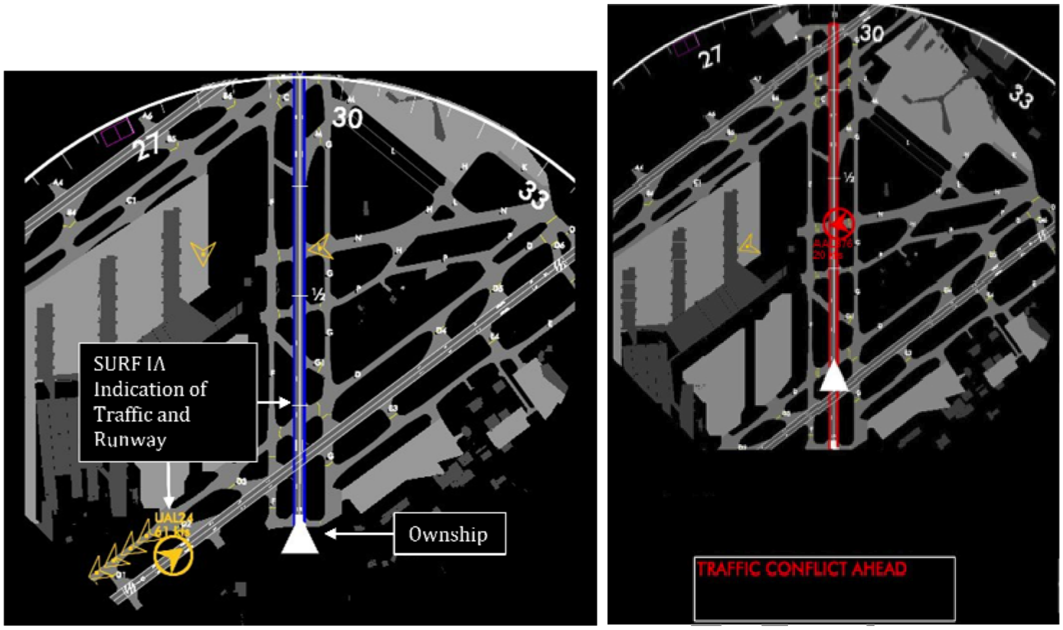


Figure 1 Example for a SURF IA Indication (left) and a SURF IA Warning Alert (right)

The SURF IA application addresses these limitations in two ways. First, the SURF IA application “indicates” to pilots safety relevant traffic. SURF IA indications increase saliency of runway safety relevant traffic under normal operational situations. Second, SURF IA alerts facilitate pilots’ immediate awareness and immediate response once an actual collision hazard has been detected. SURF IA provides caution and warning type alerts that are compatible with FAA guidance on the design of alerts on the flight deck (FAA 2011). Both alert types provide immediate flight crew awareness but require different pilot responses. While warning alerts require an immediate flight crew response, caution alerts require a subsequent pilot response. See *Figure* ***1*** for example SURF IA indications and alerts.

The simulation the MITRE Corporation’s Center for Advanced Aviation System Development (CAASD) performed were preceded by a series of three human in the loop simulations that assessed various aspects of SURF IA and are described elsewhere (see Moertl, McGarry & Nickum, 2009, and McGarry & Helleberg, 2011). These initial simulations evaluated various design characteristics of SURF IA indications and alerts that were implemented for the simulation described below.

Simulation Research Questions

This simulation focused on two primary research questions:

1. Should SURF IA alerts be directive or non-directive? Directive alerts include information for the flight crew on how to resolve a conflict, whereas non-directive alerts inform the flight crew only about the existence of a detected conflict. It was hypothesized that directive alerts would be more effective and preferred by pilots over non-directive alerts because they would decrease pilots’ demand on cognitive resources during response selection. Also, it was hypothesized that pilots would seek more confirming information when responding to non-directive alerts than to directive alert where they may just follow the resolution. Accordingly, pilots should be better able to recover from a false non-directive alert than from a false directive alert because they are expected to utilize additional visual or auditory information.
2. Is it permissible for the SURF IA system to be displayed on a CDTI with just runway information or does it require a complete airport surface map (including taxiways)? It was hypothesized that CDTIs with taxiway information would be more effective and preferred by pilots over CDTIs without taxiway information.

In the remainder of this paper, the two display conditions are referred to by their names as avionics standards as defined in RTCA DO-317, RTCA (2009). The term FAROA (Final Approach Runway Occupancy Awareness) is used for a CDTI that displays runway but not taxiway information. The term ASSA (Airport Surface Situational Awareness) is used for a CDTI that includes runway and taxiway information.

Method

The simulation was performed in MITRE/CAASD's fix-based simulator with a 120 degree out the window view and configured with a primary flight and navigation displays. The simulator did not replicate a specific aircraft type but resembled a large, transport category aircraft.

The CDTI was shown on an Electronic Flight Bag (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. Both displays could be seen without need for extensive head turning. Pilots listened to radio communication with other traffic from which they could form a mental image of the traffic environment. A former air traffic controller gave them clearances. The crew used a standard printed checklist that prompted questions and responses between the two pilots. Preflight preparations were simplified and did not include weight and balance calculations or programming the Flight Management System (FMS).

Experimental Design

Two independent variables were used. The variable "alert type" consisted of two levels, "directive alerts" versus "non-directive alerts" and was a within-subjects variable. The second variable "display type" had also two levels, "ASSA" (runway and taxiways) versus "FAROA" (runway only) and was a between-subjects variable. Alert type and display type conditions were counterbalanced and randomized across participants to account for order effects.

Scenarios

After having completed two initial familiarization scenarios, participants completed eight scenarios in each alert type condition. One group of pilots saw these 16 scenarios with the ASSA display type, the other group of pilots saw them in the FAROA display type. Visibility conditions were such that pilots could not determine conflict traffic by looking out the window. After completion of all 16 scenarios, participants completed two additional scenarios. The first additional scenario showed pilots the alternate display type to what they had seen previously. The final scenario was a "false alert" scenario where a warning alert was presented to pilots after take-off initiation, without the existence of an actual conflict. In that scenario, visibility conditions were such that pilots could visually determine that no conflict existed by looking out the window.

Participants

Twenty-four pilots participated between February and March, 2009. All pilots had experience as airline pilots and averaged 11,000 logged flight hours (ranging from 1,300 hrs to 21,800 hrs). Pilots flew a variety of airplanes including Boeing 717, 747, 777, 767, 757, MacDonald Douglas 80, Airbus 320/319, CRJ 700/900, EMB145, and CL-65. Two of the pilots were female; the rest were male. Participants were assigned by the study director to their role as pilot flying or pilot not-flying, depending on their qualifications and prior experience. The pilot flying took the left seat and the pilot not-flying took the right seat.

Results

Directive versus Non-Directive Alerting

The effectiveness of pilot response to the alerts was measured by counting flight crew responses that were appropriate to remove the safety hazard that the scenarios introduced. Safety hazards included conflict aircraft moving onto the runway ahead of ownship, accelerating toward ownship, or failing to clear a runway. Pilots could avoid such safety hazards by aborting their take-off, initiating a go-around, or clearing the runway. Such responses were counted as "effective." Any other response was counted as "ineffective."

Overall, pilots responded effectively to non-directive alerts in 53 of 59 conflict trials (90 %). Originally 60 conflict trials were presented but in one trial the auditory alert did not sound and this trial was dismissed. In comparison, the same pilots responded effectively to directive alerts in all 60 conflict trials (100%). The difference of 10 % is statistically significant between the two types of alert responses (Fisher's exact test, $p < 0.05$). This supported the hypothesis that directive alerts would lead to more effective alert responses than non-directive alerts.

Four of the ineffective responses to the non-directive warnings occurred during departure scenarios, when pilots had little time to decide on an action and therefore had to response under time pressure. These pilots continued their take-off even with the non-directive alerts coming on. Those flight crews reported that they were aware of the conflict aircraft ahead, but were avoiding an abort at high speed. The alerts sounded when ownship was above 80 knots but before reaching V1 speed (takeoff decision speed). Two crews mentioned they thought the remaining runway length was insufficient to stop the aircraft.

The two remaining ineffective alert responses occurred during approach and landing scenarios. In these cases, pilots either did not initiate go-around maneuvers, or did not attempt to exit the runway in time to clear the runway for an approaching aircraft from behind. During the final phase in an approach scenario, one of the pilots misattributed the alert to a proximate aircraft that in fact had not caused the conflict. In another scenario, the flight crew had landed and was taxiing on the runway when the alert was triggered by an aircraft behind them. The alert surprised the flying pilot who did not immediately comprehend that the conflict was caused by an approaching aircraft from behind. Therefore, the pilot decided to stop on the runway.

Ineffective responses did not appear to result from lack of familiarity with the alerting system. Half of the ineffective responses occurred during the first half of the simulation and the other half occurred during the second half. Therefore, run order apparently did not have an impact on response effectiveness.

Though overall, there were no differences in response times to directive versus non-directive alerts, response times were different only between types of operations. Response times were defined as the time between the onset of the alert and changes in throttle position resulting from a go-around or aborted take-off. Over all types of scenarios combined, there were no significant differences in response times to directive and non-directive alerts (3.7 sec vs. 4.5 sec, respectively). For departure operations it was found that pilots responded to directive alerts significantly faster than to non-directive alerts ($F\ (1, 34) = 9.63$, $p<.01$, 2.9 sec vs. 4.0 sec respectively). No response time differences were observed during arrival scenarios. Average alert response times in arrival scenarios were slower (4.8 sec) than in departure scenarios (3.5 sec) but this was not a statistically significant difference. In the arrival scenarios, pilots tended to have more time available to respond when the alert came on, as compared to in the departure scenarios.

Pilots generally thought directive alerts were more useful than non-directive alerts, though this trend only reached statistical significance for landing and exit scenarios where flight crews reported directive alerts to be more useful than non-directive alerts (Fisher's exact test, $p < 0.05$). Also, pilots indicated that directive alerts were easier to respond to than non-directive alerts (Fisher's exact test $p < 0.001$). Based on the described performance, pilot perceptions, and response time differences, it appeared that directive alerts led to faster, easier, and more effective responses than non-directive alerts.

Responses to False Alerts. To measure the extent to which flight crews relied on the displayed alert information versus seeking other confirming information (e.g., via out the window, CDTI, or radio) when selecting their response, flight crews saw a false alert in their last simulation scenario. Only two flight crews did not abort their take-off in response to the alert. One aborted in the directive alerting condition and the other one in the non-directive alert condition. This contradicts the hypothesis that pilots would be better able to recover from false non-directive alerts than from false directive alerts.

ASSA versus FAROA Display

Most pilots (21 of 24) preferred the ASSA display over the FAROA display. Pilots thought that the ASSA displays supported the task of navigating on the airport surface better with taxiway and runway information. Accordingly, the majority of pilots (20 of the 24 pilots, i.e. 83 %) found taxiway information on the ASSA display was useful when specifically asked about the value of taxiway information. A few pilots indicated during interviews that while taxiway information was useful when operating on the ground, it was not needed when flying an approach to the runway.

While pilots subjectively preferred the ASSA display, no evidence was found that pilots exhibited superior performance over the FAROA display. There was no significant difference between the number of ineffective responses between the ASSA and FAROA display conditions (3 ineffective responses using ASSA versus 2 using FAROA). Also, self-reported workload as measured by questions from the NASA task load index (TLX) showed no differences between ASSA and FAROA and ranged between 3 and 4 on a scale from 1 (lowest) to 10 (highest). The average situation awareness rating was 5.5 on a scale from 1 (lowest) to 10 (highest), and there were no differences between the display conditions.

Conclusions

The pilots in this simulation were able to successfully avoid 90 % of the presented runway safety conflicts using non-directive alerts, less than when presented with directive alerts when they avoided all conflicts. Limitations of non-directive alerts only became apparent in situations when pilots had to respond under time pressure or the appropriate response was not immediately apparent. Response times to non-directive alerts tended to be slower than to directive alerts and responses, but were only significantly slower during departure scenarios. During takeoff, pilot responses to a false alert were similar for both directive and non-directive alerts, with most pilots aborting their take-off. Pilots did not retrieve other confirming visual or auditory information (e.g. via out-the-window, CDTI, or radio) that would have let them know there was no actual conflict. This could be caused by the fact that, under time pressure, information search cannot be exhaustive. These findings suggest that the tested directive alerts were more effective than non-directive alerts for the prevention of runway incursions in the given scenarios and conditions, and that pilots did not recover any better from false non-directive alerts than they did from false directive alerts. System implementations that opt for directive alerting will need to consider the technical feasibility of directive alerting which was not part of this study.

The pilots in this simulation subjectively preferred a CDTI with taxiway information over a CDTI without taxiway information. However, there were no significant differences in observed pilot workload, situation awareness, and performance found between the two different CDTI display conditions. Therefore, while an ASSA display with taxiway appears desirable, a FAROA display without taxiway information appeared to be acceptable for the tested scenarios.

The results of this simulation were used as input into the development of safety and performance standards for the SURF IA application (RTCA DO-323, Dec 8, 2010). To support generalization of these simulation findings and to further validate the SURF IA application requirements, these simulation results should be combined with a larger body of research that evaluates the certification requirements for a surface cockpit alerting capability.

Acknowledgements

This simulation was funded by the FAA runway safety office. The authors thank numerous reviewers and contributors for their input that shaped this research. In particular, we would like to recognize the contributions by Sam Clark (Boeing), Rick Ridenour (ACSS), Christine Haissig (Honeywell), Mark Reed (ALPA), John Helleberg, Dave Domino, and Peter Stassen (MITRE CAASD), Doug Arbuckle (FAA), Denise Jones (NASA), and Michelle Yeh (VOLPE).

References

Adam, G., & Kelley, D. (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. *MITRE MTR 94W0000060.v2*.

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

Federal Aviation Administration (1998). Airport Surface Operations Safety Action Plan to Prevent Runway Incursions and Improve Operations, U.S. Department of Transportation.

Federal Aviation Administration (20011) *Title 14 Code of Federal Regulations (CFR) §§ 25.1322 and accompanying advisory circular*, U.S. Department of Transportation.

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

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

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

Federal Aviation Administration (2006a). *Design and Installation Details for Airport Visual Aids*, Advisory Circular 150/5340-1J.

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

Federal Aviation Administration (2007). *Runway Status Light Engineering Briefing 64a*. Department of Transportation: Washington, DC.

Federal Aviation Administration (2008). *Runway Safety Report*. Department of Transportation: Washington, DC.

Honeywell (2010). Product *Description SmartRunway TM / Smart Landing TM Functions of the Enhanced Ground Proximity Warning System*. [Online]. Available at: http://www51.honeywell.com/aero/common/documents/egpws-documents/raas-documents/SmartRunway_SmartLanding_description.pdf.

International Federation of Air Traffic Controller's Association (2003). *A SMCGS Implementation in Europe*. Publication of the International Federation of Air Traffic Controller's Associations.

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

Jones, D. (2005). Runway incursion prevention system testing at the Wallops Flight Facility. Presented at the SPIE Defense & Security Symposium, Orlando, Florida, March 28, 2005.

Jones, D., and Prinzel, L. J. (2006). Runway Incursion Prevention for General Aviation Operations. Presented at the 25th Digital Avionics Systems Conference.

McGarry, K., Helleberg, J. (2011). *Using a CDTI with Indications to Prevent Runway Incursions*. Accepted for the proceedings of the International Symposium on Aviation Psychology. May 2-5, 2011, Dayton, Ohio.

Moertl, P. M., McGarry, K., & Nickum, J. (2009). *Simulation Results for Highlighting Runway Safety Critical Information on Cockpit Displays of Traffic Information*. Proceedings of the International Symposium on Aviation Psychology. April 27 • 30, 2009, Dayton, Ohio.

National Transportation Safety Board (2000). *Safety Recommendation to the FAA*, A-00-66, Washington, DC.

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

RTCA (2010). *Safety, Performance, and Interoperability Requirements for Enhanced Situation Awareness on the Airport Surface with Indications and Alerts*, RTCA DO-323.

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

Steinbacher J. G. (1991). *An Analysis of ATC-Related Runway Incursions in the National Airspace*. MITRE WP 91W00234.

Young, S. D. & Jones, D. R. (2000). *Runway Incursion Prevention Using an Advanced Surface Movement Guidance and Control System (A-SMGCS).* Presented at the 19th Digital Avionics Systems Conference.

RTCA (2009). *Minimum Operational Performance Standards for Aircraft Surveillance Applications*. DO-317: Washington DC.

Vernaleken, C., Urvoy, C., Klingauf, U. (2007). *Prevention of Runway Incursions due to closed runways or unsuitable runway choices by enhanced crew situational awareness and alerting*. Enhanced and Synthetic Vision 2007. Edited by Jacques G. Gerly, Jeff J. Guell. Proceedings of SPIE Vol. 6559, 65590I-1.

DISCLAIMER

Approved for Public Release; Distribution Unlimited. Case Number: 11-0965

This work was produced for the United States Government under Contract DTFA01-01-C-00001 and is subject to Federal Aviation Administration (FAA) Acquisition Management System Clause 3.5-13, Rights In Data-General, Alt. III and Alt. IV (Oct. 1996).

The contents of this document reflect the views of the author and The MITRE Corporation and do not necessarily reflect the views of the FAA or the Department of Transportation (DOT). Neither the FAA nor the DOT makes any warranty or guarantee, expressed or implied, concerning the content or accuracy of these views.

PRELIMINARY HUMAN FACTORS FINDINGS FROM THE FAA CAPSTONE 3 ELECTRONIC FLIGHT BAG – AIRPORT SURFACE MOVING MAP OPERATIONAL EVALUATION

Michelle Yeh
United States Department of Transportation
John A Volpe National Transportation Systems Center
Cambridge, MA

Juliana Goh
The MITRE Corporation
Center for Advanced Aviation System Development
McLean, VA

The Federal Aviation Administration (FAA) Office of Runway Safety is interested in understanding the impact of an airport moving map with ownship position on operational usability and safety. To gather data on the use of this technology, the FAA is sponsoring airlines to equip revenue aircraft with an airport moving map on an Electronic Flight Bag (EFB) and to provide evaluation data to the FAA to help understand the safety impact. Understanding the human factors implications of an airport moving map is a key component of this effort. The FAA, the US DOT Volpe Center, and MITRE CAASD developed a process for gathering human factors feedback on the technologies throughout the operational evaluation using surveys, interviews, and observations. This paper will present the current status of the operational evaluation and the preliminary findings.

In 2008, the Federal Aviation Administration (FAA) Office of Runway Safety announced an initiative to evaluate the use of an airport moving map on an Electronic Flight Bag (EFB) and/or an approved aural runway safety alerting system on operational safety. An airport moving map provides a rendering of the airport's runways, taxiways, and buildings, based on information in a database that contains positional information for the location of airport surface elements. Own-aircraft position (i.e., ownship) may be superimposed on the airport moving map and updated in real-time. An airport moving map may be presented on avionics equipment that is installed or mounted in the flight deck (e.g., as one application on a multi-function display or on an EFB display). The purpose of the operational evaluation is twofold: (1) to assess the safety impact of an airport moving map with ownship position during airport surface operations, and (2) to gather information on the usability of the airport moving map software, EFB hardware, and other EFB software. Several FAA offices are also providing support for this effort, including the Office of Aircraft Certification; Flight Standards Services; and Human Factors Research and Engineering Group. Researchers from the John A. Volpe National Transportation Systems Center (Volpe Center) and the MITRE Corporation's Center for Advanced Aviation System Development (CAASD) are assisting in the human factors implementation and data collection for this effort.

The FAA has awarded contracts to four airlines: Atlas Air, CommutAir, Shuttle America, and US Airways. Each airline proposed to equip approximately 20 revenue aircraft with two EFBs hosting the airport moving map software. The aircraft type in which the EFB will be installed differs by airline, but each airline will equip the same type of aircraft in their fleet for the purposes of the operational evaluation. The airlines were allowed to select their own EFB hardware and/or software manufacturer(s) based on their specific needs. Therefore, there are slight differences in the EFB installation, EFB hardware, and EFB software from one airline to another.

Four different EFB hardware platforms (from three different EFB manufacturers) will be used during the operational evaluation. All the airlines will mount the EFB by the pilot's side window (one on the left side of the flight deck, and the other on the right side), although the specific mounting device differs for each airline. For example, two airlines have selected mounting systems that allow the EFB to be rotated (e.g., from portrait to landscape mode and vice versa), whereas one airline has selected a mounting system that locks the orientation of the EFB in landscape mode. In regards to EFB software, all the airlines selected the same manufacturer for the airport moving map but different software versions. Several airlines will also use electronic chart, electronic document, and logbook software, but the manufacturers for those products differ. These differences in the EFB hardware and software used from one airline to another will offer an opportunity to identify and understand the usability of a range of EFBs and EFB products.

To gather human factors information on the technologies, surveys, interviews, and observations will be administered and/or conducted throughout the operational evaluation with each airlines' pilots. The next section describes each data collection technique. The operational evaluation is in its inception stages, and each of the airlines is in different stages of installation and equipage. The third section of this paper describes the current state of data collection, and the fourth presents the preliminary findings.

Data Collection Techniques

Three techniques have been developed to gather human factors data throughout the operational evaluation: surveys, interviews, and observations. Each of the techniques is described below.

Surveys

Two surveys, each suited to meet different goals, will be administered to gather information throughout this evaluation. One is the *Capstone 3 EFB Survey*, which is a survey administered on the EFB and captures information specific to a particular taxi segment. This survey is intended to obtain pilots' opinions of whether the airport moving map with ownship position and/or EFB offered a perceived operational or safety benefit based on pilots' familiarity with the technology and the airport. The survey also collects feedback on areas where the technologies can be improved, for example in terms of database accuracy or in terms of pilot interface/usability. Pilots who used the airport moving map during taxi answer questions on airport moving map safety. For pilots that did not use the airport moving map (e.g., because the pilot was taxiing and was primarily looking out-the-window or if the airport moving map was not available or not needed), questions addressing the impact of the EFB are to be completed instead. The Capstone 3 EFB Survey is intended to be completed twice for each flight operation: the first time enroute after the pilot had used the airport moving map or EFB at the departure airport, and a second time after landing at the arrival airport, when the aircraft is parked at the gate. The number of questions is limited so that it can be completed during revenue operations in approximately one to two minutes. Airlines committed to have their pilots complete the survey at least once per flight segment.

The other survey will be administered on the web. The *Online Survey* gathers feedback regarding pilots' interaction with the airport moving map and EFB. This survey is intended to address a more comprehensive set of safety/usability issues than can be captured by the short Capstone 3 EFB Survey. For the Online Survey, a database of 64 human factors questions was developed addressing topics pertaining to the airport moving map software, EFB, training, and demographics. The 64 questions are not intended to be presented all at once due to concerns about completion time, but rather used as the source for creating several online surveys, each of which contain only a subset of the items. The different surveys can be rotated throughout the operational evaluation, so only one survey is "live" at a time. Questions related to the use of the airport moving map include pilots' perceptions regarding the role of the airport moving map in supporting position awareness, the legibility of the airport surface depiction (e.g., runways, taxiways), the ease of making adjustments to map range and orientation, and the perceived overall impact on workload and heads-down-time. The questions examining the use of the EFB include the ease of accessing information, the consistency of information presentation, the readability of that information, and the usability of buttons and controls. Responses to items on this survey are not expected to change from one flight to another, so the survey can be completed when pilots have more time than they would during line operations.

The survey questions on both the Capstone 3 EFB survey and Online Survey were developed from previous research evaluating the use of an EFB during line operations or examining how an airport moving map display with ownship position may support position awareness on the airport surface (e.g., FAA, 2001). It is important to note that the purpose of these surveys is to develop a general understanding of pilots' perceptions regarding these technologies and not to compare performance across airlines. The EFB and airport moving map, the operational procedures regarding their use, and the training provided will differ from one airline to another. In administering the surveys, it is expected that pilots will be appropriately trained in the use of the airport moving map application and EFB so that they can use it properly and provide meaningful feedback.

The surveys offer pilots from all the participating airlines the opportunity to provide feedback on the usability of the EFB and/or airport moving map. The surveys can be used in conjunction with interviews of limited number of pilots from the participating airlines (as described next) to fully understand the impact of the EFB and/or airport moving map. For example, survey responses can be probed during interview to gather more information on a usability concern. Additionally, information gathered from interviews can be used to develop new questions for the Capstone 3 EFB and online surveys.

Interviews

Several airlines indicated they would provide researchers an opportunity to speak with pilots directly about their experiences with the airport moving map and EFB display. Such a setting would allow for more detailed responses and could provide a framework for interpreting the responses provided by the Capstone 3 EFB and Online surveys. Sample questions were developed to structure the interviews; for example:

- Did the airport moving map show the information you needed to establish, maintain or regain position awareness on the airport surface? In general, how does your position awareness with airport moving map on EFB compare to your position awareness when using a paper airport chart only?
- Did you encounter any problems or confusing issues with the airport moving map? For example, was ownship position ever shown incorrectly, was the information shown on the map located in incorrect locations, or did you have any problems interpreting the information shown on the airport moving map display?
- Were there any surprises when you were using the EFB (did the EFB ever do something different than you expected)? If so, please describe the situation.

Additional questions and discussion points may arise during the interviews based on pilots' responses.

Each interview is expected to last approximately 30 minutes. The interviews are intended to be conducted with pilots from each airline throughout the operational evaluation.

Observations

Several airlines have provided the opportunity to observe the EFB installation and pilots' use of the airport moving map and EFB in their simulators. These observations provide an opportunity for conducting human factors assessments of the installation of the EFB (e.g., mounting issues) as well as an informal usability assessment of the hardware and software. The information gathered through the observations will be used to revise and refine the questions presented on the surveys and asked during the interviews. The observations will also provide insight into the feedback gathered through the surveys and interviews.

Assurances of Confidentiality

To encourage airline support for the human factors data collection, both MITRE CAASD and the US DOT Volpe Center provided assurances to the airlines that the raw information they provided in support of the operational evaluation would be kept confidential and that the airlines would not be mentioned specifically by name in the reporting of any results. Because the purpose of the interviews were to understand the potential human factors concerns associated with this technology and not to compare the different technologies or the differences in the implementation across the airlines, the findings from the interviews and simulator visits are presented in this report without specifically identifying the airline or the EFB manufacturer.

Data Collection Status

Surveys

1032 responses were received for the Capstone 3 EFB Survey, and 16 pilots completed the Online Survey as of February 2011.

Interviews

Two airlines have so far provided the opportunity to conduct interviews.

Airline 1. Two interview sessions have been conducted with this airline. The first was conducted in October, 2009, during the airline's internal test trial of the EFB hardware and software. In this session, the Volpe Center and MITRE CAASD led an informal focus group with 20 pilots. The pilots were using the EFB for viewing electronic charts and documents but had not started using the airport moving map software yet. During the focus group, each pilot was asked to indicate what they liked most about the EFB and what they liked least about the EFB. Pilots were then provided the opportunity to raise and discuss other issues. Separate from the focus group, an opportunity to interview pilots two at a time in a training simulator was provided; these discussions focused on gathering pilot opinions on the impact of the EFB on their operations.

After the test trial ended, Airline 1 chose to make extensive modifications to their EFB hardware and software configurations, and the airline invited the Volpe Center and MITRE CAASD to conduct a second interview session in February 2011. Airline 1 was in the early stages of re-integrating the EFB technology into their fleet, but 13 of the pilots in the focus group had used the EFB in its new configuration. Additionally, 9 of the pilots had previous experience with an EFB (from the first test trial). The airline had not yet introduced the airport moving map. Similar to the first interview session, each pilot was asked to describe what they liked most about the EFB and what they liked least about the EFB. The opportunity to interview pilots two at a time in the training simulator was again provided.

Airline 2. In June 2010, the Volpe Center visited Airline 2 and conducted informal usability evaluations of the EFB and EFB software with the airlines' pilots. The purpose of these sessions was to identify and understand any potential human factors concerns with the software suite the airline was using for the Capstone 3 operational evaluation. During the session, the Volpe Center met with 5 pilots. The pilots, working in groups of 2 or 3, were asked to view and use the different software on the EFB. For example, pilots were asked to develop a logbook describing a typical flight. Pilots were also asked to use the electronic charting software to look up the charts needed for a flight, pull specific charts for a flight, and view them. It is important to note that unlike the pilots from the previous interview session, none of these pilots had been trained on the EFB or the software. Rather, the purpose of the review was to gather the pilots' first impressions of the EFB and to understand the intuitiveness of the software.

Observations

Team members from the FAA, MITRE CAASD, and US DOT Volpe Center have conducted three simulator visits so far. During each simulator visit, team members were provided the opportunity to taxi and/or fly the simulator using the airport moving map. The Volpe Center defined four scenarios to facilitate the simulator visits: two involved an aircraft taxiing out and then taking off; the other two involved the aircraft landing and then taxing in. However, in most cases, team members elected to taxi using their own scenarios.

Preliminary Findings

Pilots liked the idea of an "electronic flight bag", particularly as a replacement to their traditional flight bag. One feature they especially liked was a "push/pull" functionality, which can be used to "send" a current image of the information on one EFB to the other EFB. Several pilots indicated that the push/pull function facilitated communication between the Captain and First Officer. In fact, one Captain commented that he sometimes has his First Officer pull up the charts needed and "send" them to him. For example, one pilot can highlight the departure frequency on a chart and send it to the other pilot, which facilitates the briefing. During a demonstration trial of this software, this pilot noted that when flying into KBOS, he received four runway changes, and the ability to highlight the necessary information on the 10-9 chart and send the chart back and forth on the flight deck was beneficial.

Despite the potential benefits of an EFB and airport moving map, pilots' ratings in response to the question "How useful was the EFB" tended towards the negative. On a scale of 1 (negative) to 5 (positive), pilots, on average provided a rating of between 1 and 1.5 across the first three months of preliminary data collection. The preliminary results from the surveys, interviews, and observations were examined to identify potential usability issues with respect to the EFB hardware and software that may have led to these low ratings. These are discussed below.

Usability of EFB Hardware

Several human factors issues were noted with respect to the EFB's installation, touch screen sensitivity, and brightness.

Installation. Pilots at one airline noted that the installation (i.e., location and mounting system) of their EFB could impede their movement on the flight deck. This issue was noted primarily by captains, who bumped into the EFB while accessing the tiller, although a few first officers mentioned they also sometimes bumped into the EFB as well. These events occurred primarily during taxi operations when the pilot's seat is adjusted close to the flight deck for approach and taxi. Pilots' body size is a contributing factor, such that larger pilots tended to note this as an issue. Pilots mentioned that they keep the EFB in landscape mode (i.e., horizontal orientation) during taxi, take-off, and landing as a workaround. However, in some cases, the locking mechanism for the mounting system was failing; consequently, the EFB could not be locked into landscape mode.

Touch screen sensitivity. All the EFBs used during the operational evaluation will have a touch-screen display. One recurring issue noted throughout the interviews concerned the responsiveness of the touch screen. Some pilots found that some EFB touch screens did not always respond immediately to a touch, whereas others noted that the same touch screens could be too sensitive. Additionally, several pilots at one airline noted that touching the display with their finger did not always produce the desired input; one example given was that the touch screen did not work with cold fingers. To better interact with the EFB, some pilots indicated they use a pen as a stylus; upon hearing this, other pilots noted that they found pen markings on the EFB displays. Unfortunately, using a pen as a stylus could lead to damage of the EFB display (e.g., scratches) and make finger input more difficult. Consequently, one pilot indicated that he was concerned about the durability of the touch-screen display in the long-term.

Brightness. All the pilots at one airline indicated that glare on the EFB was an issue that affected display readability. In particular, the EFB display was not bright enough to be readable during daylight without having the shade down. Some pilots mentioned that on bright days, the EFB can reflect an uncomfortable amount light off of its screen and into the pilot's eyes. Additionally, the night mode was considered to be too bright. In fact, one pilot noted that rather than use the night mode, he used the brightness function at the display to adjust the contrast of the EFB. These issues with readability were reflected in responses to the online survey. On a 5 point scale (1: bad, 3: neutral, 5: good), the average pilot rating for readability was as follows: readability under various lighting conditions (mean = 1.7), readability from sitting position (mean = 2.2), and readability under night mode (mean = 2.5), color readability under all lighting conditions (mean = 2.7).

Software Usability

Human factors observations regarding the use of the airport moving map and other EFB software was collected primarily from the simulator visits. Although the focus of the operational evaluation is the airport moving map software, usability feedback was gathered on other software used on the EFB (e.g., electronic charts and electronic documents) and discussed in this section. Note that the purpose of the evaluations was not to compare the software applications but rather to develop a general understanding of common human factors considerations for EFB software. Thus, specific details for a software application are distinguished only when necessary for understanding the issue.

Airport moving map. Several pilots noted that the aircraft position represented by the ownship symbol on the airport moving map differed from that depicted on their navigation display. In one instance, the nose of the aircraft on the airport moving map was represented by the center of the ownship symbol. For the navigation display, however, the nose of the aircraft is represented by the nose of the ownship symbol. The location within the ownship symbol that corresponds to the position of the aircraft should be consistent across the flight deck.

Electronic Charts. Feedback on two different electronic chart software applications was gathered during the interview sessions. Two primary issues were noted. First was inconsistency between the paper and electronic charting information. At one time during the operational evaluation, pilots at one airline used paper charts from one manufacturer and electronic charts from a second manufacturer. During this time, pilots indicated that the information on the paper charts was laid out differently from the electronic charts, so that they sometimes could not find the information they needed on the electronic chart quickly. One pilot noted that on one flight, he received a runway change and could not easily find a new frequency on the electronic chart. Consequently, two pilots indicated that they relied primarily on their paper charts, although they continue to load the electronic charts on the EFB so that they stay current and familiar with the procedure for doing so. (Note: Since that interview session was conducted, that airline has switched its electronic chart manufacturer so that both the paper and electronic charts are provided by the same manufacturer. However, this concern was included since it calls attention to the potential impact of the inconsistency in the presentation of charting information.)

A second issue was a failure to preserve the modifications made to electronic charts. Pilots noted that if they left an electronic chart (e.g., by displaying another chart or viewing another page) and then return to it, the zoom settings (map range) is reset to a default level. At one airline, pilots noted that they commonly zoom in on to a chart and then rotate it, but when the chart is rotated, the zoom level is reset to a default level. In some cases, pilots encountered EFB speed and reliability issues, such that these two actions (rotate and zoom) would lock up the EFB.

Electronic documents. Feedback on two different electronic document software applications was also gathered. In general, pilots considered electronic document library an improvement over their paper documents, because it was much more comprehensive than what they could carry in their traditional flight bag. However, pilots noted two examples of when the electronic document did not match the paper one: one was the availability of titles for each document chapter, and the other was in the page numbering. The pages of the electronic document are numbered according to the page in the pdf file, whereas the pages in the paper document are numbered by section number (e.g., page 7-35). This inconsistency may be confusing and increase the time it takes to find the appropriate section or text.

Conclusions

The Capstone 3 operational evaluation intends to examine the potential for safety gains on the airport surface with an airport moving map and to understand the human factors implications involved with the integration of airport moving map displays into the flight deck. The Capstone 3 operational evaluation is still in its inception, so the findings reported here are preliminary. In Fiscal Years (FY) 2011 and 2012, MITRE CAASD and the Volpe Center will continue to work with the FAA to coordinate with the participating airlines as they continue and complete their equipage and submit data in support of the Capstone 3 operational evaluation.

The preliminary feedback has identified potential human factors areas of concern that may be of interest to FAA evaluators or manufacturers developing EFBs and/or airport moving maps, and these findings may inform the design and evaluation process. As with any new technology, the functions and capabilities for EFBs and airport moving maps will continue to evolve, and it will be important to stay abreast of this evolution to understand the human factors implications.

Acknowledgements

This report was prepared by the Behavioral Safety Research and Demonstration Division of the Human Factors Research and Systems Applications Center of Innovation at the John A. Volpe National Transportation Systems Center (Volpe Center) and the MITRE Corporation Center for Advanced Aviation System Development. Funding support for the Volpe Center was provided by the FAA Office of Runway Safety and Operational Services and FAA Human Factors Research and Engineering Group (AJP-61) in support of the Aircraft Certification Service Avionic Systems Branch (AIR-130) and the Technical Programs and Continued Airworthiness Branch (AIR-120). The FAA program managers are Dr. Pradip Som, Dr. Tom McCloy and Colleen Donovan. The authors would like to thank Dr. Scott Gabree, Greg Orrell, and Dr. Julian Sanchez for their assistance during the data collection. Many thanks also to the participating airlines and their pilots for supporting and participating in the human factors data collection effort.

The views expressed herein are those of the authors and do not necessarily reflect the views of the John A. Volpe National Transportation Systems Center, the Research and Innovative Technology Administration, or the United States Department of Transportation.

References

FAA (2001). *Safe Flight 21. Operational Evaluation-2. Final Report*. Washington, DC: Federal Aviation Administration

USING A CDTI WITH INDICATIONS TO PREVENT RUNWAY INCURSIONS

Kathleen McGarry
John Helleberg
MITRE/CAASD
McLean, VA, USA

A human-in-the-loop simulation was performed to evaluate an advanced Aircraft Dependent Surveillance – Broadcast (ADS-B) application that provides runway safety indications on a Cockpit Display of Traffic Information (CDTI) implemented on a Class 2 Electronic Flight Bag (EFB). There are display limitations associated with the use of a Class 2 EFB including when the ownship symbol is shown, and how the surface moving map is displayed at various stages of flight. Nineteen pilots viewed the CDTI in a baseline condition without indications, and then in two conditions with traffic and/or runway indications while operating a medium-fidelity flight simulator. Subjective results indicate that pilots preferred a CDTI with runway indications over a baseline CDTI without such indications. Pilots reported that it was difficult to determine the location of traffic relative to ownship when the display was in North-up mode. The objective performance results showed few performance differences across display types.

The Federal Aviation Administration (FAA) has made reducing the number and severity of Runway Incursions (RIs) one of their top priorities. These efforts have produced a significant reduction in the number of serious RIs (Category A and B), lowering the number of these incursions from 67 total events in Fiscal Year (FY) 2000 to only 6 events in FY 2010 (FAA, 2010). While the number of serious incursions has declined, the rate for all incursions went from 12.3 per million operations in FY 2005 to 17.2 per million operations in FY 2008 (FAA, 2009). To address this problem, extensive human factors research has been performed. This research has generally indicated that human behavior is a root cause for RIs (FAA, 1998; Bales, Gillian & King, 1989; Steinbacher, 1991; Adam and Kelly 1996). This research has identified several factors contributing to RIs, including airport characteristics such as signage, markings, lighting, runway geometry, as well as lack of pilot familiarity with the airport surface and procedures. Other factors include the communication of control clearances over the radio, which can represent an information bottleneck, as well as factors concerning crew and air traffic control (ATC) operational procedures.

In an effort to increase pilot situation awareness of surface hazards, flight decks have begun to be equipped with moving maps that can display the airport and airport surface movement. Standards for the CDTI are currently being developed for different platforms on the flight deck such as electronic flight bags (EFBs) to support retrofit applications as well as integration with the Navigation display for forward fit aircraft. Research has been conducted in support of RTCA (formerly Requirements & Technical Concepts for Aviation; and formerly Radio Technical Commission for Aeronautics) Special Committee (SC)-186 Automatic Dependent Surveillance-Broadcast (ADS-B), “SURF IA” subgroup. This working group is developing the concept of enhanced traffic situation awareness (ATSA) on the surface (SURF) with indications and alerts (IA). Several simulations have been conducted to examine the effectiveness of indications and alerts presented on a CDTI as a way to reduce the risks of RIs and runway conflicts.

SURF IA Application Description

The goal of the SURF IA application is to enhance the information on the CDTI to prevent runway incursions and avoid collisions. To do this, relevant traffic and runways are indicated or alerted on the CDTI. During normal operational conditions, safety relevant traffic is highlighted on the display. Once the situation becomes non-normal, alerts are provided to help avoid collisions. Previous research has explored the use of both indications and alerts provided on a Class 3 EFB, which is considered to be installed equipment and are certified to a level that allows both indications and alerts to be presented. Another type of flight deck display that could host the SURF IA application is a Class 2 EFB. Class 2 EFBs are considered Portable Electronic Devices (PED) and are not certified to present caution/warning level alerts. On Class 2 EFBs, indications provide safety-relevant information during normal operations and persist when the situation transitions from normal to non-normal (i.e., no alerts are provided). There are also some display limitations when using a Class 2 EFB. When ownship is taxiing on the airport surface at a speed below 40 kts, the surface moving map (SMM) is ownship-centric. The map will display as track-up, and the ownship symbol will be shown on the display, corresponding to its actual position. Ground and airborne traffic that is within five miles and up to 1000 feet Above Ground Level (AGL) of the airport will be shown on the display. When ownship exceeds 40 kts (usually, this would be during a takeoff roll), the ownship symbol is removed from the display. The SMM continues to move and rotate according to ownship position and heading, but the symbol depicting the ownship location on the map is no longer shown. Once ownship reaches the airport map boundary (after takeoff), the SMM freezes and awaits a configuration change. When ownship is on approach, the SMM is a north-up map centered on the airport. Ownship position is not shown, and ground and airborne traffic within five miles and 1000 feet AGL of the airport are shown on the display once ownship is within three miles of the airport center. Prior to ownship being within three miles of the airport center, no traffic is shown; only the north-up map is shown. Once ownship touches down, but is still above 40 kts, the north-up map transitions to a track-up, ownship-centric SMM. However, the ownship symbol is not displayed on the map at touchdown. When ownship slows to below 40 kts, the ownship symbol appears on the map, and the display exhibits the same behavior as taxi operations below 40 kts.

CDTI and Indication Styles

The current research compared a previously used set of indication components (termed ATSA 2 style) with a new modified set of indications (termed ATSA 3 style) that was developed based on pilot feedback and performance in the two previous studies. Also, as limited work had been done in the previous research examining the effectiveness of a CDTI without indications at preventing RIs, this study included a third condition with no indications. In all three display conditions, all traffic within five miles and 1000 ft AGL of the airport, and within the pilots' currently selected display range was shown regardless of whether it was considered relevant to the current operation (i.e., there was no filtering of traffic). Airborne traffic not causing an indication was depicted with a cyan chevron and included a data tag depicting the traffic's relative altitude. Surface traffic not causing an indication was depicted with an unfilled brown chevron with a dot in the center of the symbol.

This simulation used a Class 2 EFB to present the traffic and indications to the pilots. Due to the certification limitations with a Class 2 EFB, only indications (no alerts) were presented to participants. The indications highlighted safety relevant traffic to the flight crew. Because alerts cannot be shown on a Class 2 EFB, if a situation became a conflict, the indications remained on and there is no display change between normal and non-normal (i.e., conflict) events. There were two levels of indications provided to the flight crew: secondary and primary. A secondary indication was presented if there was a low potential threat for runway safety, such as when there is a low speed convergence between ownship and the indicated traffic, or a high speed divergence between them. A primary indication was presented if there was a high potential threat for runway safety, for example, when there is a high speed convergence in front of the departing aircraft.

In the ATSA 2 indication style, when a secondary indication was active, the relevant traffic was highlighted and a message was provided in the text box on the display. For secondary indications, the text box provided the relevant runway number followed by the word "PREDICTED." The traffic was highlighted by enlarging the chevron, filling it in, and providing flight ID and ground speed for ground traffic, and flight ID and relative altitude for airborne traffic, see Figure 1.

17R OCCUPIED

Figure 1. **ATSA 2 Style Indications**

When a primary indication was active in the ATSA 2 style, the traffic was highlighted in the same way as the secondary indications, but in addition, the affected runway was also highlighted with a blue and white outline (see Figure 1). The text box provided the relevant runway number followed by the word "OCCUPIED" at the bottom of the display indicating that the runway was occupied. The text box messages filled the function of providing indications for off-scale traffic, which are aircraft that precipitate the indication (either secondary or primary), but were not in range to be visible on the display at its current zoom setting. The message provided the flight crew with awareness of relevant traffic that they otherwise would not be able to see due to it being outside the current display range. Note that the text box runway status information was always provided regardless of whether the indicated traffic was on or off-scale.

The ATSA 3 style of indications kept some of the styling from ATSA 2, and made several modifications. The runway highlighting was kept, as was the flight ID and groundspeed of indicated traffic. A circle (either dashed or solid, depending on whether a primary or secondary indication was being given) was added around the indicated traffic. The indicated traffic was a filled in chevron, but it was not enlarged. In addition, an off-scale traffic indication was provided to show what traffic was triggering the indication, even if the zoom setting was such that the traffic was not shown on the display. The ATSA 3 secondary indication was similar to the ATSA 2 style in that it consists of a filled traffic chevron and provides the flight ID and ground speed/absolute altitude. However, in addition to these features, the ATSA 3 style secondary indication also included highlighting of the runway. For a secondary indication, the runway is outlined with a dashed blue line, and a dashed blue circle outlines the traffic that is triggering the secondary indication (see Figure 2). When a primary indication was presented, the traffic was filled in, the flight ID and ground speed/absolute altitude were shown, and the runway outline and traffic symbol circle become solid.

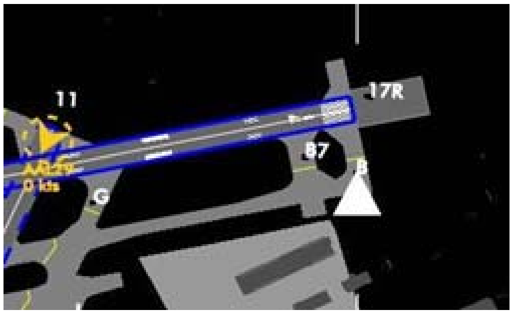


Figure 2. **ATSA-3 Style Indications**

With both secondary and primary indications, off scale traffic was represented by a filled arrow shape symbol, with the ground speed/absolute altitude information displayed, and the symbol surrounded by a circle (dashed for secondary and solid for primary indications). The color of the symbol and surrounding circle reflected the traffic's current air/ground state such that cyan symbols were airborne and brown symbols were on the surface. The off-scale traffic was shown on the display in a dedicated "margin" area outside of the other map information, at relative bearing to ownship, and the arrow pointed in the direction of the traffic's ground track. This off-scale symbol provided the flight crew with relevant traffic information when the traffic was beyond the currently selected display range. Once the zoom level was changed so the traffic was within the range of the display, or if the traffic moved within range of the current zoom setting, the off-scale indication changed into a regular traffic indication.

Method

This simulation focused on three main research questions. The first was: Will the limitations of a Class 2 EFB have an impact on pilot acceptance of the ATSA SURF IA system? The second question the research explored was would there be a difference in pilot preference between the three indication display styles? The final question the research focused on was if the three display types would have an impact on conflict avoidance.

Ten Airline Transport Pilot (ATP) rated crews participated in the study. Nineteen were external pilots who volunteered to participate, and were compensated for their participation. One pilot was a confederate who filled in for an external pilot that was unable to make the scheduled date. Fifteen of the participants were male, and four were female with an average age of 34.6 years (range of 22-48). The participants had an average of 15.9 years of piloting experience, ranging from 5-30 years.

The cockpit simulator used for this study was an enclosed, fixed based, medium fidelity transport aircraft simulator, configured as a generic twin-engine, large weight category, jet aircraft. An autopilot and auto-throttle system was used to control flight path and speed. The cockpit provided for two standard flight crew and observer positions. A retired Air Traffic Controller (ATC) acted as the confederate ATC and provided normal ATC instructions for all operation types to the pilots through a simulated radio communication system. Another confederate acted as the pseudo-pilot and provided realistic background radio communications for the other traffic aircraft depicted in the simulation.

All scenarios were flown in and around a representation of Louisville Standiford International Airport (SDF). The traffic presented in the scenarios did not replicate normal operations at SDF; instead normal traffic levels were inflated to increase the use of the CDTI. CDTI use was also encouraged by using ~1/2 mile visibility in all experimental scenarios. The CDTI was presented on a simulated Class 2 EFB with a touch screen interface. One was located to the left of the Primary Flight Display (PFD) for the left seat pilot flying (PF), and one was located to the right of the PFD for the right seat pilot monitoring (PM). Each traffic display could be configured independently (e.g., the PF and PM could have different display ranges set simultaneously).

Experimental Design

The study was a 3 (display type) x 3 (operation type) x 2 (conflict) factorial design. The three display types were: *CDTI-only* (CDTI with SMM and traffic); *ATSA 2* (CDTI with SMM, traffic and ATSA 2 style indications); and *ATSA 3* (CDTI with SMM, traffic and ATSA-3 style indications). The three operation types were *approaches*, *departures*, and *taxi* scenarios. Each scenario could be either a *conflict scenario*, in which a traffic aircraft would lead to a RI if ownship continued, or a *non-conflict scenario*, in which ownship could continue the operation and a RI would not occur. All scenarios were presented on a Class 2 EFB. Pilots were shown eight experimental scenarios on each of the three display types for a total of 24 experimental scenarios for each crew. Within each block of eight scenarios, four contained conflicts. The eight scenarios consisted of four approaches (two contained conflicts), two departures (one contained a conflict) and two taxi scenarios (one contained a conflict). Pilots saw each scenario once per display type, resulting in the viewing of each scenario a total of three times over the course of the experiment. The scenarios were randomized within display type. Pilots saw all eight scenarios using one display type at a time, and the order in which the display types were presented was counterbalanced.

Experimental Procedure

Pilots completed a consent form and demographics form when they arrived. They were then shown a short introductory briefing which described the technology they would be experiencing throughout the course of the simulation. Pilots were trained on the flight simulator characteristics and given several practice scenarios to become accustomed to flying the simulator and interacting with the CDTI. Once pilots were comfortable with the simulator, they were trained on the experimental display type they would be seeing in the first block of scenarios. The blocks were counterbalanced, so participants saw only the CDTI-only, ATSA 2, or ATSA 3 displays during each block. After the training, pilots were presented with the experimental scenarios, followed by a survey after each trial. Once they completed all the scenarios within the experimental block, they were given a block survey to provide their overall feedback on the experimental display type they just completed. Pilots were then trained for the next display type, were presented with the scenarios, and responded to the surveys. Once all three blocks of experimental scenarios were completed, there was debrief session to capture any feedback or comments from the pilots on the technologies they saw throughout the day. The entire experiment took about eight hours to complete.

Results

Subjective Data

Class 2 EFB Limitations. Pilots were asked to rate how difficult it was to determine where surface traffic was in relation to ownship position. This was asked because while using a Class 2 EFB on approach, the surface map was in a North-up position. Once the pilots touched down, the map transitioned to a Track-up display, though ownship was not shown on the display until aircraft speed dropped below 40 kts. Pilots reported that it was significantly more difficult to determine where the surface traffic was in relation to ownship when the display was in North-up mode than when it was in Track-up mode (Wilcoxon Signed Ranks Test = -3.83, $p < .01$). This was true across all three display types (see Figure 3). When comparing North-up and Track-up modes between display types, there were no reported differences. Pilots rated the difficulty similarly for both North-up mode and Track-up across display type.

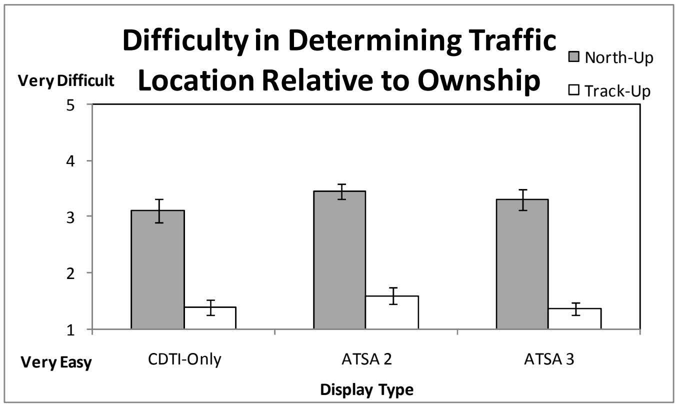


Figure 3. **Pilot Ratings of Difficulty in Determining where Surface Traffic was in Relation to Ownship**

Indication Ratings. Pilots were asked to rate the usefulness of the complete suite of indications they saw. For the analysis, this was broken down by operation type (approach, departure, and taxi). Pilots rated the suite of indications on a five-point scale from "completely useless" to "completely useful." There was no significant difference in the average ratings between ATSA 2 and ATSA 3 styles of indications.

Workload Ratings. Pilots completed the Bedford Workload Rating Scale after each scenario. Average workload ratings for all three display types were low, and there were no significant differences in pilot workload ratings between the three display types.

Effectiveness of CDTI. Pilots were asked to rate the effectiveness of the three display types (CDTI-only, ATSA 2 and ATSA 3) on incursion avoidance. On a four-point scale, analysis found that pilots rated the ATSA 3 display significantly higher than the CDTI display in effectiveness (Friedman Chi Square (2) = 8.9, $p < .05$). Pairwise comparisons found that ATSA 3 was rated significantly more effective than CDTI-only (Wilcoxon Signed Ranks Test, $p < .01$). There were no pairwise differences between ATSA 2 and ATSA 3, or between ATSA 2 and CDTI-only (see Figure 4).

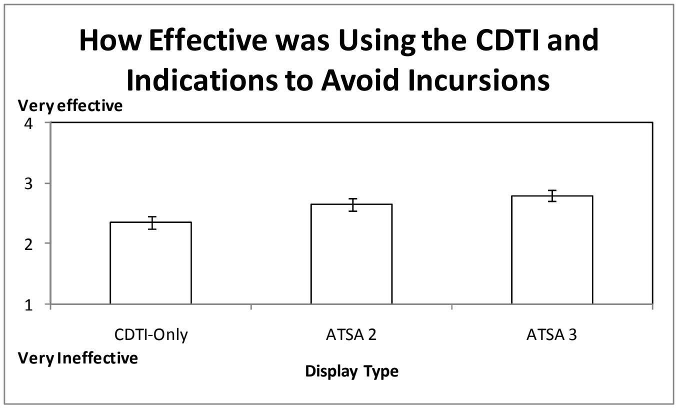


Figure 4. **Effectiveness of Display Type on Incursion Avoidance**

Objective Data

Conflict Avoidance. Each crew flew 24 experimental scenarios for a total of 240 experimental scenarios across all crews. In half of those scenarios the crews were presented with a conflict scenario in which they would have to respond in some way (dependent on operation type) in order to avoid a RI. This resulted in a total of 120 conflict scenarios across three operation types (approach, departure, and taxi scenarios) for analysis. Each crew was presented with six approach conflicts (two per display type), three departure conflicts (one per display type) and three taxi conflicts (one per display type). The experimenters monitored the crew's reaction to each conflict scenario and recorded whether or not they were able to avoid the conflict. For approach scenarios, avoiding the conflict meant not touching down on an occupied runway, for departure scenarios, avoiding the conflict meant not initiating a takeoff while the runway was still occupied, and for taxi scenarios, avoiding the conflict meant not crossing the intersecting runway hold line until after the conflict departure had passed. When all three operation types were combined, there were no significant differences in number of conflicts avoided between display types. All three display types avoided a high number of conflicts presented in the scenarios (34/40 in CDTI-only, 37/40 in ATSA 2, and 38/40 in ATSA 3).

However, when each operation type was analyzed separately, there was a significant difference between display types for approach (Friedman Chi Square (2) = 6.3, $p < .05$) and departure operations (Friedman Chi Square (2) = 6.0, $p = .05$). For approach operations (20 scenarios), the crews avoided all of the conflicts with the ATSA 2 display, but were unable to do so with the CDTI-only display (15 conflicts avoided) and ATSA 3 displays (18 conflicts avoided). Pairwise comparisons showed that there was a significant difference in conflicts avoided during approach scenarios between the ATSA 2 display and the CDTI-only display (Wilcoxon Signed Ranks Test = -2.24, $p < .05$), but the other pairwise comparisons were not significant. For departure operations (10 scenarios), the opposite trend was found, in that the crews avoided all of the conflicts with the CDTI-only and ATSA 3 displays, but were unable to do so with the ATSA 2 display (7 conflicts avoided). However, pairwise comparisons showed that there were no significant differences in conflicts avoided during departure scenarios between the ATSA 2 display and the CDTI and ATSA 3 displays.

Conclusions

In conclusion, the Class 2 EFB limitation that requires a North-Up display while on approach may not be acceptable to pilots when traffic is included on the SMM. However, lack of an ownship symbol was more acceptable to pilots as they rarely noticed it was missing. While differences in pilot performance due to display types were not observed, pilots did report that they felt that the ATSA 3 style of indication was more effective at preventing incursions. This is a preference, however, and was not supported by the performance data which suggests that indications may not provide significant benefits beyond what can be gained with a CDTI-only display. There was no evidence of any performance or preference costs associated with indications as compared to the CDTI-only display as the indications did not adversely impact surface efficiency. It is possible that performance benefits may be realized through further refinement of the indication style and pilots showed a preference for the ATSA 3 style indications over the CDTI-only display. Therefore, it is suggested that additional research be conducted.

References

Adam, G. & Kelley, D. (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* (MTR 94W0000060, v2). Mclean, VA: The MITRE Corporation.

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

FAA. (2010, October 8). *Serious Runway Incursions Cut in Half for Second Straight Year* [Press release]. Retrieved from http://www.faa.gov/news/press_releases/news_story.cfm?newsId=11959

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, (2009). *FAA Runway Safety Report*, Washington DC, Federal Aviation Administration.

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

Approved for Public Release: 11-0796. Distribution Unlimited

AIR TRAFFIC CONTROL INTERFACE FOR CREATING 4D INBOUND TRAJECTORIES

Rolf E. Klomp, M. M. (René) van Paassen, Max Mulder
Aerospace Engineering – Delft University of Technology
Delft, The Netherlands

Mariska I. Roerdink
Air Traffic Control The Netherlands (LVNL)
Schiphol-Oost, The Netherlands

It is to be expected that the task of an air traffic controller will change with the introduction of 4D (space and time) trajectories for aircraft. To support this new work, an interface for the manipulation of 4D trajectories has been created. The interface uses a time-space diagram – which shows progress of the aircraft along its planned track, in which conflict zones for conflict with other aircraft are presented, and a vertical path display, which presents the altitude along the planned track. In combination with a traditional plan-view display, and additional elements and tools in the displays to create a visual link between the paths displayed in the three displays, the interface enables the modification of a 4D planned trajectory for inbound aircraft.

An experiment has been performed with a PC-based simulation. Two variants were tested, one in which the aircraft could only be controlled in the current sector, and one in which the speed (but not the path) of aircraft could be modified some time before these aircraft entered the sector; a practice expected to produce improvement of efficiency in small sectors. The results validate that the enhanced interface can be used to manage the air traffic safely and efficiently. It was also shown that the ability to manipulate the speed of an aircraft in the adjacent sector can significantly *increase* situation awareness and *reduce* controller workload.

Introduction

Currently, air traffic controllers (ATCo's) perform a sector-based tactical form of control. They are responsible for planning and managing traffic within their assigned airspace, often with little help from automated tools (Oprins & Schuver, 2003; Oprins, 2008). In the coming decades, the task of an air traffic controller is predicted to undergo a large transformation. The pull for transformation comes from the increasing demands which are placed on the air traffic management (ATM)-system (Dlugi et al., 2007, 2008; Anon., 2007). A push is provided by technological advances on the air- and ground side of the ATM-system, which make a new form of air traffic control (ATC) possible (Slattery, 1996; Ballin, Hoekstra, Wing, & Lohr, 2002). This is expected to result in a situation where 4D (space + time) trajectories stored in automated support tools form the basis for the ATCo's work (Whiteley & Wilson, 1999; Jorna, Pavet, Van Blanken, & Pichancourt, 1999; Heesbeen, Hoekstra, & Valenti Clari, 2003; Prevot, Lee, Smith, & Palmer, 2005).

In a previous project (van Dijk, van Paassen, Mulder, & Roerdink, 2009), a human-machine interface (HMI) has been developed to visualize these 4DT's and support the air traffic controller in planning and managing the air traffic in this future situation. An initial validation experiment with this interface showed ATCo could safely and efficiently manage air traffic, however, a number of areas for improvement were identified. This paper describes the re-design and validation of the interface. The option to manipulate the vertical path of the aircraft has been added, and visual links between the different views presented to the ATCo were created.

Operational Concept

The interface is designed for a situation where the path of aircraft is define by means of 4D trajectories (Ballin et al., 2002; Wichman, Lindberg, Kilchert, & Bleeker, 2004), and it focuses on control in the CTA (Controlled Terminal Area) sector. The CTA is an air traffic zone where most traffic either comes from upper airspace or neighboring CTA's and enters the TMA (Terminal Maneuvering Area) around an airport, or vice versa, and control there is supplied by the Area Control Center (ACC). It is expected that instead of the tactical control performed today, where an ATCo shapes the aircraft's trajectory by issuing speed and heading instructions, the future ATC work will involve more strategic control, and a large part of the ATCo's task will be to create and define conflict-free trajectories *before* aircraft enter the CTA. It is also assumed that in the future, traffic in the TMA will largely follow fixed routes to the runway, to enable aircraft to perform low-noise continuous descent approaches (Wubben &

Businink, 2000; Meijer, de Gelder, Mulder, van Paassen, & in 't Veld, 2009). To make that possible, the timing of the aircraft entering the TMA has to be adjusted in the CTA.

As aircraft enter the radar screen and the ACC control system, the initial planned trajectory runs from the entry in the ACC straight to the intended entry point for the TMA. Such a path might have conflicts with aircraft currently in the ACC, and the time of arrival in the TMA might not be correct yet. The ATCo responsible for the ACC will have to create a 4DT that ensures the timing of arrival in the TMA is correct, and that is free from conflicts with traffic in the ACC. In our set-up the ACC is modeled after a Schiphol ACC and two entry gates to the TMA are used, see Figure 1. The schedules for the two entry points to the TMA are linked; the use of a single runway is assumed, and the aircraft have to enter the TMA such that separation at the runway threshold is achieved. The path length in the TMA is different for the two entry points, and when a change in entry point is selected a different entry time is required. In many cases, the size of the ACC airspace (in many places in Europe and on the US West Coast) limits the options for correcting the timing of an arriving flight. One optional extension considered is the possibility to control speed of aircraft still in neighboring sectors.

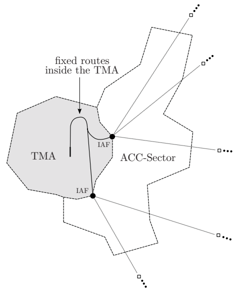


Figure 1: Three-dimensional fixed routes inside the TMA, starting at the Initial Approach Fixes. Outside the TMA, aircraft arrive from all directions and are merged by ACC ATCo.

Display Design

The display is a further development of a display used in a previous project(van Dijk et al., 2009). The design is focused on presenting the constraints of the work domain – i.e. planning and de-conflicting 4D trajectories and planning the arrival sequence for the TMA.

The planning interface is formed by a time-space diagram (TSD), which displays the planned times versus the along-track distance of an aircrafts 4DT, (1) in Figure 2. The time-space diagram is linked to a display of the inbound traffic sequence, bars linked to each aircraft's arrival time represent wake vortex separation minima, by "stacking up" the bars – indirectly, through speed and path changes – the correct entry times into the TMA are achieved. For the aircraft currently selected, the TSD provides a display of conflict areas, which corresponds to coloring of the ground track in the plan view display (PVD) when the track is crossed by other aircraft. In one of the display configurations, a time slider is added to the TSD. This slider can be used to display "ghost positions" for the aircraft in the system on all three views, visualizing the temporal development of the traffic pattern. Below the TSD, a vertical situation diagram (VSD) shows the vertical track, again against the planned ground track. The VSD shows where other aircraft cross the planned trajectory. For a selected aircraft, potential conflict areas will also be shown. In this way constraints on vertical maneuvers (in the VSD) or on velocity maneuvers (in the TSD) can be seen in the display. The linked combination of these two displays forms the Time-Space and Vertical Display (TSVD). The visualization of conflict with other aircraft support part of the ATCo's task, namely the de-confliction of the 4DT. To support the planning of arrival times to the TMA, additional visualization has been added to the TSVD. Given constraints on speed instructions, an arrival range prediction can be calculated for a selected aircraft. The prediction shows the fastest and the slowest possible flight, given that the lateral path is not changed.

Evaluation

With the PC-based implementation of the TSVD, an experimental validation has been performed to answer the following questions:

- Does the addition of the new tools (conflict visualization on the PVD, the ghost view tool) improve situation awareness and reduce controller workload?
- What is the effect of the possibility to plan speed changes for aircraft still in a neighboring sector; while the lateral path there is fixed (under the responsibility of the ATCo of the other sector), a speed change may be instructed.

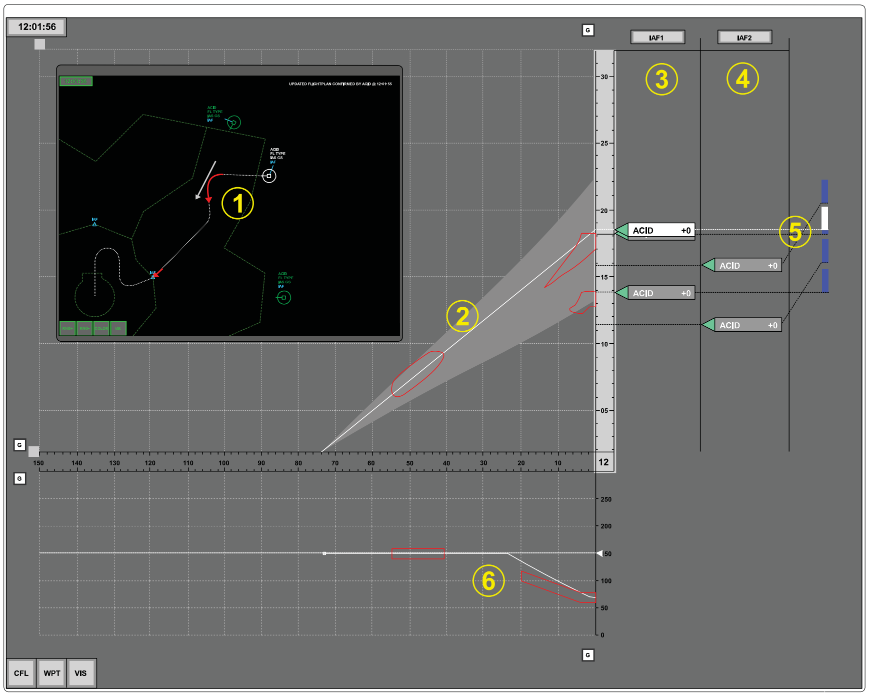


Figure 2: TSVD when an aircraft is selected

1. **1** Conflict zone in PVD with part of track of conflicting aircraft
2. **2** Conflict zones in TSD
3. **3** Arrival list for entry point IAF1
4. **4** Arrival list for entry point IAF2
5. **5** Arrival sequence planning, horizontal bars are for IAF1, slanted bars represent extra travel time for aircraft arriving at IAF2
6. **6** Conflict zones in VSD
7. **7** Range of possible arrival/travel times for selected aircraft (light grey zone)

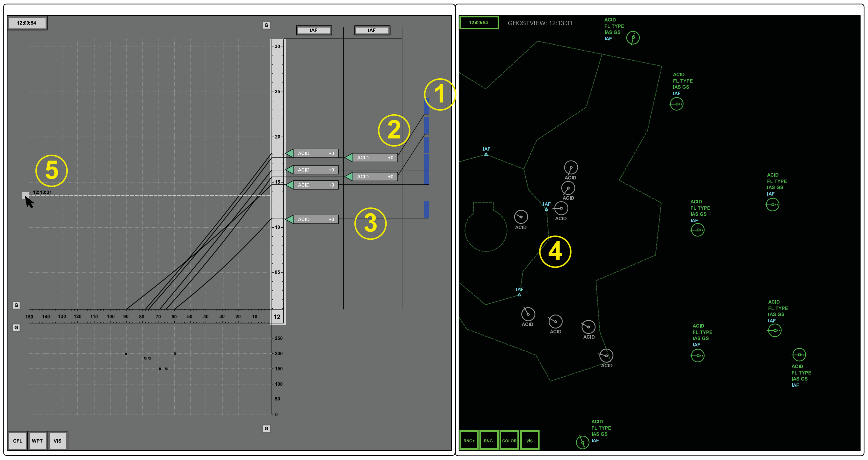


Figure 3: TSVD and PVD when no aircraft is selected, with the ghost view tool active

1. **1** Wake vortex separation bars (blue)
2. **2** Additional travel time for aircraft arriving at IAF2
3. **3** No additional travel time for aircraft arriving at IAF1
4. **4** Ghost images for the aircraft in the PVD
5. **5** Ghost slider, positioned at a future time

Four different scenarios, with different traffic of approximately equal difficulty were used. Three types of aircraft were simulated in clean configuration, each with different performance characteristics which influence the feasible range of 4DTs: The Airbus A319-114, the Airbus A330-301 and the Airbus A340-313. The Eurocontrol GAME (Anon., 2009, 2002) aircraft performance model has been used to calculate the relevant elements of aircraft performance such as minimum/maximum speed and idle rate of descent. Aircraft entered the controlled sector between FL120 and FL230, with an Indicated Air Speed (IAS) varying between 220kts and 280kts, and were given an initial 4DT leading to one of two Initial Approach Fixes (IAFs). For reasons of simplification, a fixed turn radius of 5NM has been assumed, and the fixed minimum landing interval between two aircraft (length of the 'blue bars') has been set to 1.7 minutes for all aircraft. The initial conditions were set such that at a certain point the TMA would become congested and the subject would have to issue changes to the 4DTs to maintain a safe and efficient sequence of aircraft.

Twelve subjects participated in the experiment. The subjects were either active ATCo, retired ATCo, or researchers with initial training in air traffic control. Two independent variables were varied, *TOOLS:* The possibility to use the ghost view tool and the conflict visualization on the PVD was switched on (T_E) and off(T_D) and *SECTOR:*

The possibility to issue speed requests to aircraft in the adjacent sector was enabled (S_E) or disabled (S_D). The four resulting conditions were combined with the scenarios in a latin square design. The following dependent measures were used to analyze and compare the four conditions:

Loss of separation Loss of separation was measured by logging the number of separation violations between aircraft. For the experiment, the minimum separation criteria were set to 5NM horizontally and 1,000ft vertically.

Workload The controller workload was measured by using a digital version of the NASA-Task Load Index (NASA-TLX (Hart & Staveland, 1988)) subjective mental workload questionnaire. The TLX-questionnaire returns a score ranging between 0 and 100. Here, a higher score indicates a higher workload.

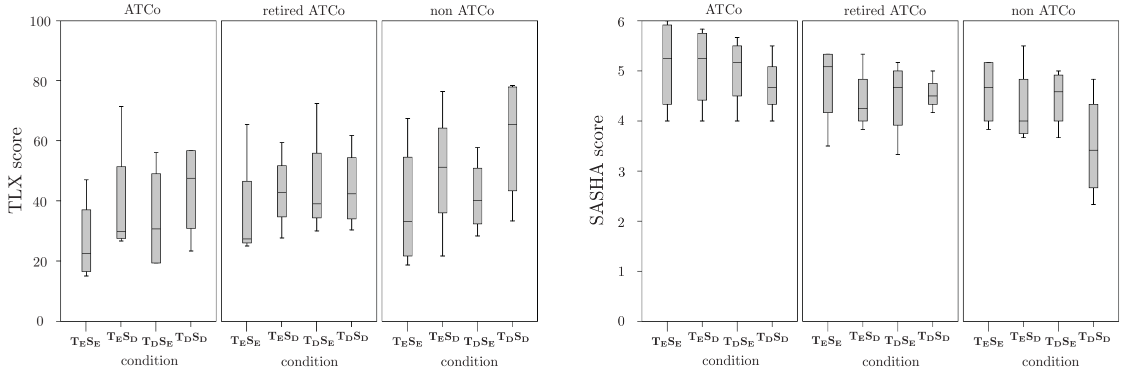
Situation Awareness The controller situation awareness was measured using the Eurocontrol SASHA (Straeter, Woldring, Barbarino, Skoniezki, & Philipp, 2003) questionnaire, which provides an overall subjective score for SA. The scores range between 0 and 6. A higher score indicates a higher (better) level of SA.

Experiment Results

Loss of separation. An analysis of all aircraft trajectories showed that no loss of separation occurred in the experiment.

Workload. Figure 4(a) shows a boxplot of the measured TLX scores per condition, and per subject group. Using a Kolmogorov-Smirnov test, the TLX data were found to be normally distributed ($D(48) = 0.120$, $p > .05$). A repeated-measures Analysis of Variance (ANOVA) of the TLX scores showed that the between-group effect was not significant ($F(2,9) = 0.627$, $p = 0.56$). Therefore, the within-subject effects due to the TOOLS and SECTOR option have been tested for all participants combined.

The TOOLS variable did not have a significant effect on measured TLX ($F(1,9) = 1.760$, $p = 0.17$). The SECTOR variable did have a significant effect ($F(1,9) = 11.087$, $p < 0.01$). Enabling speed control over aircraft in the neighboring sector significantly reduces workload.



(a) Boxplot of the measured TLX scores

(b) Boxplot of the measured SASHA scores

Figure 4: Boxplots of the measured NASA-TLX and SASHA scores per condition, clustered per subject group

Figure 4(b) shows a boxplot of the measured overall SASHA scores. The figure shows that the ATCo group rate their situation awareness the highest (closest to the maximum score of six), and the non ATCo group the lowest. Furthermore, the figure shows that most situation awareness scores are located in the higher range of the scale. A Kolmogorov-Smirnov test was performed, and showed that the measured SASHA data were normally distributed ($D(48) = 0.085$, $p > .05$). Again, no significant between-group effect was found by using a repeated-measures ANOVA of the SASHA scores ($F(2,9) = 1.439$, $p = 0.29$). By evaluating the within-subject effects of all participants combined, no significant increase of situation awareness was found for the conditions with the TOOLS enabled compared to the conditions with the TOOLS disabled ($F(1{,}9) = 3.416$, $p = 0.098$). However, situation awareness was significantly higher for the conditions with the SECTOR option enabled, compared to the conditions with the SECTOR option disabled ($F(1{,}9) = 6.153$, $p < 0.05$).

Conclusion

Experimental results show that the TSVD can support an ATCo in the new task of planning inbound 4D trajectories safely and efficiently. No loss of separation has been recorded. The addition of two new tools to the interface did not show significant effects on measured situation awareness and workload. Participants did like the option to dynamically project the time-evolution of the planned future traffic situation on the PVD (ghost view tool). The importance of multi-sector coordination and the size of the controlled sector has been underlined by the significant increase of situation awareness and decrease of controller workload, which has been measured for conditions in which speed changes could already be issued to aircraft in the adjacent sector. However, this option would require that the responsibility for these aircraft is to be shared by the two adjacent sectors.

For further development of the TSVD, it is recommended to investigate the form in which outbound and transit aircraft could be incorporated. The TMA planning (blue bars) should also be improved, making it more apparent where merge problems will occur. Furthermore, when no aircraft is selected, cues should be given for the existence and the priority of conflicts. When manipulating the horizontal path of an aircraft, forbidden areas could be indicated on the PVD in which it is unsafe to reroute, although this visualization is expected to require considerable calculation. Disturbances such as wind, bad weather, inaccuracies in trajectory prediction and other unforeseen events were not included in this experiment. These will have to be taken into account when calculating trajectory predictions. The size of the controlled sector has been shown to have significant influence on controller workload, situation awareness and to the efficiency of the traffic flow, but was not taken as an explicit experiment variable. Therefore, further research is required to determine this relationship. Finally, the TSVD is based upon a mature state of four-dimensional operations in which all aircraft have full 4D capabilities. In order to gain operational acceptance, solutions should be found for a transition period in which both the air- and ground segment are only partially equipped.

References

Anon. (2002, September). *G.A.M.E. aircraft performance model* (Report No. ver. 2.0). Brussels, Belgium: Eurocontrol Headquarters. Unpublished.

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

Anon. (2009, April). *Standard aircraft modeling interface* (Users guide (draft) No. rev. 4.2). Brussels, Belgium: Eurocontrol Headquarters. Unpublished.

Ballin, M. G., Hoekstra, J. M., Wing, D. J., & Lohr, G. W. (2002, October 1-3). NASA Langley and NLR research of distributed air/ground traffic management. *AIAA's Aircraft Technology, Integration and Operations 2002 Technical, Los Angeles, California, USA.*

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. www.eurocontrol.int/sesar/gallery/content/public/docs/DLM-0612-001-02-00.pdf.

Dlugi, O., Astheimer, T., Baldoni, C., Brown, R., Cato, R., Erb, M., et al. (2008). *The atm deployment sequence* (Deliverable 4 No. DLM-0706-001-02-00). Toulouse, France: SESAR Consortium. www.eurocontrol.int/sesar/gallery/content/public/docs/DLM-0706-001-02-00.pdf.

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. 139–183). Amsterdam: Elsevier Science Publishers (North-Holland).

Heesbeen, W. W. M., Hoekstra, J. M., & Valenti Clari, M. S. V. (2003, August 11-14). Traffic manager, traffic simulation for validation of future ATM concepts. *AIAA Modeling and Simulation Technologies Conference and Exhibit, Austin, Texas, USA*.

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. www.eurocontrol.int/phare/gallery/content/public/documents/99-70-02ghmi.pdf.

Meijer, L., de Gelder, N., Mulder, M., van Paassen, M. M., & in 't Veld, A. C. (2009, August). Time-based spaced continuous descent approaches in busy terminal manoeuvring areas. In *AIAA Guidance, Navigation and Control conference* (p. 23). Chicago IL. (AIAA-2009-6165)

Oprins, E. A. P. B. (2008). *Design of a competence-based assessment system for air traffic control training*. Ph.D. dissertation, Faculty of Economics and Business Administration, Maastricht University, Maastricht, The Netherlands.

Oprins, E. A. P. B., & Schuver, M. (2003). Competentiegericht opleiden en beoordelen bij LVNL in Dutch: Competence based training and assessment at LVNL. *HUFAG nieuwsbrief* (6), 2-4.

Prevot, T., Lee, P. U., Smith, N., & Palmer, E. A. (2005, August 15-18). ATC technologies for controller-managed and autonomous flight operations. In *Aiaa guidance, navigation and control conference, san francisco, california, usa.* San Francisco, (CA).

Slattery, R. A. (1996, July 29-31). Examination of Possible Future Terminal Area Procedures. *AIAA Guidance, Navigation, and Control Conference, San Diego, California*.

Straeter, O., Woldring, V. S. M., Barbarino, M., Skoniezki, A., & Philipp, W. (2003, June). *The development of situation awareness measures in atm systems* (Tech. Rep. No. HRS/HSP-005-REP-01; 1.0). Brussels, Belgium: Eurocontrol Headquarters. www.eurocontrol.int/humanfactors/gallery/content/public/docs/SHAPE_questionnaires/SHAPE User Guide v0.1.pdf.

van Dijk, E. S., van Paassen, M. M., Mulder, M., & Roerdink, M. I. (2009, April 27–30). An interface for inbound traffic route planning. In *Proceedings of the 15th international symposium on aviation psychology* (pp. 44–49). Dayton OH: Wright State University.

Whiteley, M. J., & Wilson, I. (1999, August). *Phare advanced tools problem solver* (Final Report Nos. PHARE/EEC/PAT-6.5/FR; 1.0, DOC 98-70-18). Brussels, Belgium: Eurocontrol. www.eurocontrol.int/phare/gallery/content/public/documents/98-70-18-v8_ps.pdf.

Wichman, K. D., Lindberg, L. G. V., Kilchert, L., & Bleeker, O. F. (2004, August 16-19). Four-Dimensional Trajectory Based Air Traffic Management. *AIAA Guidance, Navigation and Control Conference, Providence, Rhode Island, USA*.

Wubben, F. J. M., & Busink, J. J. (2000, May). *Environmental benefits of continuous descent approaches at schiphol airport compared with conventional approach procedures* (Report No. NLR-TP-2000-275). Amsterdam, The Netherlands: National Aerospace Laboratory NLR, division Air Transport. www.nlr.nl/id-4297/l en.pdf.

TRANSITIONING FROM DIGITAL TO ANALOG INSTRUMENTATION

Geoffrey Whitehurst
Western Michigan University
Kalamazoo, Michigan
William Rantz
Western Michigan University
Kalamazoo, Michigan

Technically Advanced Aircraft (TAA) have seen an increase in manufacturing within the last decade. The growing use of these aircraft will present unique challenges to the aviation infrastructure; as well as flight training. With the large number of analog aircraft remaining in the general aviation fleet, transitions between digital and analog will become more numerous and perhaps more precarious. A recent survey of flight instructors at one college highlighted situational awareness problems for 95% of TAA trained students when exposed to analog equipped instrument panels. Perhaps two options are available to study this problem on the ground: flight simulators or a Personal Computer - Aviation Training Device (PC-ATD). The initial challenge to any study of this issue was to select the option that would minimize, or would allow for control of, extraneous factors, so that the causal factors influencing any decrement in performance and/or situational awareness could be isolated. A comparison of the two options available showed that the PC-ATD was the better option for the study of this issue and a pilot study was carried out using the PC-ATD. The results of the pilot study suggested that the transitioning from digital to analog equipped aircraft produced degradation in performance and that further research was required.

There are many advantages to train new pilots using the latest technically advanced aircraft (TAA). Most believe that the advanced avionic displays, autopilots, and moving maps, which emulate larger commercial aircraft flight decks, are required to give new student pilots a training advantage. Workload, situational awareness, and systems management and integration of these elements will all be enhanced by using TAA. Aircraft were once only equipped with analog instrumentation. Today's general aviation flight schools may have a variety of new generation, digital instrumentation and pilots take their first lesson in digitally equipped aircraft. Once a pilot earns a flight certificate, regardless of whether or not the training aircraft used digital or analog instrumentation, there is no regulation requiring any type of transition training between the different types of instrumentation. Lack of instrumentation display formalization and layout may lead to impaired skills and decreased situational awareness. A related situation maybe expressed using digital and analog clocks for an example. What if an individual learns to read time only based on digital clocks and having never seen another style clock. This individual is then asked to read the time from an analog clock. It is highly likely that the individual's response rate will be reduced and may even be in error from lack of familiarization and practice with the analog time piece. In the early 21st century analog aircraft far outnumber their TAA counterparts. Given the large disproportionate number of analog aircraft, what transitional trap awaits those who lack transitional training?

Although a large number of aircraft accidents include situational awareness as a probable cause, information recorded by the National Transport Safety Board, in their accident data base (http://www.ntsb.gov/ntsb/query.asp), does not contain data of recent flight history. The inclusion of this data would allow analysis of the type of flight instrumentation used and reveal any transition between flight instrumentation types. A future requirement of accident investigation may be the inclusion of this data to help provide a clearer picture of this probable cause.

A recent survey of flight instructors at one college highlighted situational awareness problems for 95% of TAA trained students when exposed to analog equipped instrument panels. Of these 95% who experienced problems 34% had an initial struggle, 33% had a moderate struggle, 21% had a significant struggle, and 7% were still struggling at the end of this flight phase. The 5% who did not experience any situational awareness problems were students who had previous experience flying with analog instrumentation.

While it is assumed that pilots learn new rules or mental models in early practice to master the highly technical skills of instrument flight. It has yet to be determined the depth and complexity by which those mental models are formed and maintained. Given this unknown process, it becomes difficult to assume that transitional adaption of digital training to analog flying will be as reciprocally easy as has been analog training to digital flying. Especially given the fact that most digital displays have been developed to adapt and consolidate representations of dispersed analog instruments.

The purpose of the pilot study is to determine if there is performance degradation for pilots who have only experienced digital flight instrumentation when exposed to analog instrumentation for the first time.

Review of Existing Literature

The transition of pilots from a traditional cockpit to a modern-glass cockpit has been a training challenge for the last two decades (Dahlstrom, Decker &Nahlinder, 2006) and many studies have been conducted on how this transition training should be carried out (Reigner & Decker, 1999; Casner, 2003a, b; Fanjoy & Young, 2003). However, a review of the literature has uncovered no empirical research examining the transition of pilots from a modern-glass cockpit to a traditional analog cockpit and the possible risks involved. TAA can be defined as those aircraft equipped with new-generation avionics that take full advantage of computing power and modern navigational aids to improve pilot awareness, system redundancy, and depending upon equipment, improve in-cockpit information about traffic, weather, and terrain (AOPA Air Safety Foundation, 2005). TAA have seen an increase in manufacturing within the last decade. The growing use of these aircraft will present unique challenges to the aviation infrastructure; as well as flight training. With the large number of analog aircraft remaining in the general aviation fleet, transitions between digital and analog will become more numerous. According to the Federal Aviation Administration regulations in Title 14 part 61.31 which refers to additional training, there is no mention of the need or requirement to obtain transition training between digital and analog cockpits aircraft. (FAR AIM, 2010) Therefore as the fleet of TAA continues to expand, the potential for transitional incidents and accidents is likely to increase.

Initial research has shown that student pilots can be trained in technically advanced aircraft that will meet or exceed current training standards (Craig P. A., Bertrand J. E, Dornan W., Gosset S., Thorsby K. K., 2005). However, one study by Rantz W. G. & Van Houten R. (2011), found that using technically advanced aircraft as a primary trainer did nothing to improve student performance skills in checklist usage between the digital and paper checklists when flying technically advanced aircraft. Hamblin C. J., Gimore C. & Chaparro A., 2006 asserts that pilots armed with new technology, without proper training or understanding, can actually decrease safety. Given this same preface, pilots transitioning from digital to a different technology, such as analog, will likely experience a decrease in safety as well.

Methodology

When considering the options available to study this problem on the ground two possibilities were considered, a flight simulator, or a Personal Computer - Aviation Training Device (PC-ATD). The issue was to select the option that would minimize, or would allow for control of, extraneous factors, so that the causal factors influencing this decrement in performance could be isolated. For each of the two options (flight simulator or PC-ATD) two phases of the study needed to be considered; the simulation of a TAA with digital flight instrumentation, and the simulation of an aircraft with analog flight instrumentation.

For the first phase, the TAA with digital flight instrumentation, the flight simulator option would provide a true representation of the aircraft used in the participant's flight training (Cirrus SR20). The PC-ATD would emulate the Cessna 182 Skylane Glass, and the set-up would provide a limited representation of the cockpit environment.

For the second phase, the aircraft equipped with analog flight instrumentation, the flight simulator option would require a move to a flight simulator equipped with analog instrumentation. The only analog instrumented simulator available would be for a Piper PA-34 Seneca, which is a two-engine aircraft simulator. The PC-ATD would emulate a Cessna 182 Skylane, the analog instrumented version of the aircraft used in the first phase, which would only require a change of display not setting.

The PC-ATD allowed for better control of extraneous variables than the flight simulator and was therefore selected as the better option for this study.

Method

A pilot study was completed using a PC-ATD set up to emulate the Cessna 182 Skylane Glass for the digital equipped aircraft, and the Cessna 182 Skylane RG for the traditional analog aircraft. Participants were 6 college students recruited from junior and senior level aviation courses at Western Michigan University (WMU) who have completed the instrument rating course. The participants were randomly allocated, 3 to the treatment group and 3 to the control group. The experimental task consisted of flying different designated flight patterns using a PC-ATD emulating a Cessna 182 Skylane Glass and, for the treatment group, a Cessna 182 Skylane RG. During the simulated flights, participants were asked to fly a radar vectored flight pattern and to complete an instrument approach.

The performance of the flight student was measured in two ways, (a) their flight skills during the radar vectored flight pattern, and (b) their flight skills during the instrument approach. The dependent variables for comparing flight skills consisted of the number of times the aircraft deviated from the criteria listed in the Practical Test Standards for instrument flight check rides.

The experimental design for this study was a two group control group design. The participants were randomly allocated to either the control group or the treatment group. The pre-test for both groups consisted of a two-hour session flying 4 trials in the simulated Cessna 182 Skylane Glass. The post-test for the treatment group consisted of a two-hour session flying 4 trials in the simulated Cessna 182 Skylane RG and the post test for the control group was a two-hour session flying 4 trials in the simulated Cessna 182 Skylane Glass.

Setting

The experimental setting was a 12 by 16 foot room that is used as the PC-ATD flight and driving simulator laboratory. The laboratory is located in Wood Hall on WMU's Main Campus in Kalamazoo, MI USA.

Apparatus

The PC-ATD equipment consists of a Dell Optiplex SX260® computer with a Pentium (R) ® 2.40 gigahertz processor, and 1.0 gigabytes of SDRAM memory. Operating software is Microsoft Windows XP and simulation software is On-Top version 9.5. Flight support equipment for the PC-ATD will include a Cirrus yoke, a throttle quadrant, an avionics panel, and rudder pedals. On-Top software permits the simulation of several different aircraft types including the two that will be used in this study, the Cessna 182 Skylane RG and the Cessna 182 Skylane Glass. The technical flight parameters, which depict how well participants fly the designated flight patterns, vertically and horizontally, will be recorded for each flight on an external Seagate 1.0 terabyte hard drive. The On-Top simulation software automatically records these technical parameters and enables them to be printed.

Flight Patterns

In an effort to minimize any practice effects, a different flight pattern was used for each of the 4 trial flights. Participants were told that the PC-ATD aircraft was not programmed for any system failures and that the flight pattern would be a radar-vectored instrument flight, with an instrument landing system approach to a full stop landing. By using vectored instrument approaches and not having system faults, the flight environment should have allowed for consistent flight performance. The approach patterns used should not have provided the participant with any adverse stress or pressure to perform, as these patterns were typical of their existing training environment. The flight pattern that participants flew were divided into two segments for analysis: (a) cruise; consisting of take-off, climb and radar vectored flight (b) instrument approach; consisting of localizer interception, instrument approach and landing. The flight pattern took approximately 30 minutes to complete. To realistically simulate an actual flight pattern and ensure that it was flown in a consistent way across trials and participants, the experimenter provided typical air traffic control instructions throughout the flight pattern. These instructions were transmitted using a commercially available intercom system. The speaker was placed in the PC-ATD and the experimenter, who was in an adjacent area, used the push-to-talk feature on the monitor to transmit the air traffic control instructions.

Observation Equipment

The participants were observed remotely via EzWatch Pro Version 4.0 HiDef surveillance equipment as well as a dual computer monitor arrangement. The observing equipment consisted of 1 indoor/outdoor IR night vision bullet camera and 1 resolution indoor dome camera. The observer recording computer was a Dell Latitude D510® with a 5.7 gigabyte hard drive, a Pentium M® 1866 megahertz processor, and a plug and play monitor with 128 megabytes of memory. Other PC equipment included a Dell Microsoft Natural® PS/2 keyboard and a Sigma Tel C-Major® audio adapter. The observer occupied a room that was adjacent to the participant's room. One camera was mounted on the wall in front of the participant to capture hand and arm movements. The other camera was mounted on the wall behind the participant to observe the participant's interaction with the flight panel. All flights will be recorded and stored digitally for the purposes of conducting inter-observer agreement.

Analysis of Data

To reduce error variance an Analysis of Covariance (ANCOVA) with the pre-test scores as the covariate was used to analyze the data for both performance measures; flight skills during cruise and flight skills during instrument approach.

Results

Analysis of the pilot study data, see Table 1 and Table 2, suggests that there are differences between the control and treatment groups in the cruise and instrument approach phases of the flight. Further research, using a larger sample size, is required to provide the statistical power required for conclusive evidence of this difference.

Table 1

ANCOVA of Cruise Data

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	42.6667	1	42.6667	4.9621	0.0365	4.3009
Within Groups	189.1667	22	8.5985			
Total	231.833333	23				

Table 2

ANCOVA of Instrument Approach Data

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	28.1667	1	28.1667	12.1107	0.0021	4.3009
Within Groups	51.1667	22	2.3258			
Total	79.3333	23				

Benefits of Research

The full study may identify significant performance differences in digital and analog instrumented aircraft and provide empirical evidence of practice time needed to reach the required criteria using analog instruments.

The full study may identify instructional methods to increase flight safety by recommending transitional training objectives and practice time, thereby reducing the risk of errors associated with digital to analog transition.

Participants may improve their flight and instrument landing approach skills with repeated simulated flights and technical and vocal feedback.

References

AOPA Air Safety Foundation (2005). Technically advanced aircraft - Safety and training. AOPA Air Safety Foundation Special Report. AOPA Air Safety Foundation, Frederick, MD.

Dahstrom N., Dekker S. & Nahlinder S. (2006). Introduction of technically advanced aircraft in ab-initio flight training. Technical Report 2006-02. Lund University, Sweden

Casner S. M. (2003a). Teaching cockpit automation in the classroom. NASA report NASA/TM-2003-211865. Moffett field, CA: Ames Research Center.

Casner S. M. (2003b). Learning about cockpit automation: From Piston trainer to jet transport. NASA report NASA/TM-2003-212260. Moffett field, Ca: Ames Research Center.

Craig P. A., Bertrand J. E., Dornan W., Gosset S., & Thorsby K. K. (2005). Ab initio training the glass cockpit era: New technology meets new pilots. *Proceedings of the 13th International Symposium on Aviation Psychology*. Columbus, OH: The Ohio State University.

Fanjoy R. O. & Young J. P. (2003). Advanced collegiate flight automation training: What is the needed and at what cost? *International Journal of Applied Aviation Studies* 3(2), pp. 215-225. Oklahoma City, OK: FAA Academy.

Federal Aviation Administration (2010). Federal Aviation Regulations – Aeronautical Information Manual (FAR-AIM).

Hamblin C. J., Gilmore C., & Chaparro A. (2006). Learning to fly glass cockpits requires a new cognitive model. Proceedings of the Human Factors and Ergonomics Society: 50th Annual Meeting, 1977-1981.

Rantz W. G. & Van Houten R. (2011). A feedback intervention to increase digital and paper checklist performance in technically advanced aircraft simulation. *Journal of Applied Behavior Analysis, 44.*

Rignér J. & Dekker S. W. A. (1999). Modern flight training - Managing automation or learning to fly? In Dekker S. W. A. & Hollnagel E. (Eds.), *Coping with computers in the cockpit,* pp. 145-151. Aldershot, UK: Ashgate.

ASSISTING PILOTS IN THE TRANSITION FROM SELF-SEPARATED TO CONTROLLED AIRSPACE

P. Panman, J. Ellerbroek, A. C. in ‘t Veld, M. M. van Paassen, M. Mulder
Faculty of Aerospace Engineering, Delft University of Technology,
Kluyverweg 1, 2629 HS Delft, The Netherlands.

A concept for a four-dimensional planning interface is proposed, designed to assist pilots with the transition from unmanaged to managed airspace. The interface concept visualizes constraints on aircraft trajectory, that result from future separation and timing requirements, and facilitates direct manipulation of waypoints, speed profiles and altitude profiles. A preliminary evaluation of the concept showed that subjects were able to replan their 4D trajectory successfully in 71 out of 72 scenarios. Recommendations for future research are the addition of separation affordance overlays to the vertical and speed profiles, and adding timing affordance overlays in all profiles.

As airports and airways are becoming more crowded, pressure to reduce costs is increasing and standards for safety are set higher (FAA, 2009). In addition, the demand for air traffic is expected to be 2.4 times higher in 2025, compared to 2006 (SESAR, 2006). One approach to resolve some of these issues is to designate parts of the airspace as self-separated airspace. In these parts of the airspace, the responsibility of separation is delegated to individual aircraft. This will allow pilots to plan optimal routes and to reduce costs and delays, with the aid of automated separation assurance systems to resolve conflicts with other traffic. It is assumed that in dense traffic situations, such as the airspace around airports, separation will remain the responsibility of Air Traffic Control (ATC). The transition between these types of airspace will require a transfer of control. To ensure a safe and orderly transition, the air-traffic controller could assign merging aircraft a Required Time Over (RTO) at an entry point, see Figure 1. This additional four-dimensional constraint in the final phase of the flight will add to the complexity of the pilots’ tasks. As an aid for this merging task, the current study proposes an interface concept, designed to assist pilots in the task of merging from self-separated to controlled airspace. The remainder of this paper presents the interface design, followed by a summary of the results from a preliminary evaluation experiment. The paper concludes with a short dicussion on the results and current issues with the display, and reccommendations for future work.

Interface Design

A basis for a work-domain model for the merging application is provided by several earlier publications (Amelink, Mulder, van Paassen, & Flach, 2005; Borst, Suijkerbuijk, Mulder, & van Paassen, 2006; Borst, Mulder, & van Paassen, 2010; van Dam, Mulder, & van Paassen, 2008; Heylen, van Dam, Mulder, & van Paassen, 2008; Ellerbroek, Visser, van Dam, Mulder, & van Paassen, 2011). A number of additional constraints and observations can be made for the current application: 1. Merging requires aircraft to be lined up closely which implies reducing separation. 2. Separation between aircraft is determined by the full 4D trajectories of all involved aircraft. 3. The 4D trajectory must be achievable within the flight envelope and the energy constraints. 4. Separation is maintained when no other aircraft enter the own aircraft’s Protected Zone (PZ). 5. Creating a conflict-free 4D trajectory requires the full 4D trajectory of the other aircraft.

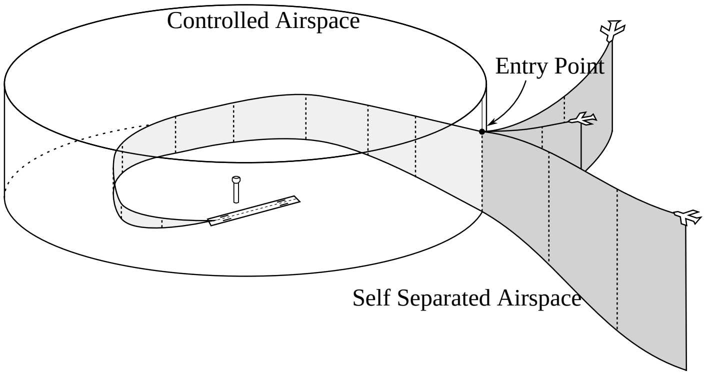


Figure 1: Airspace layout, entry point and 4D aircraft trajectories

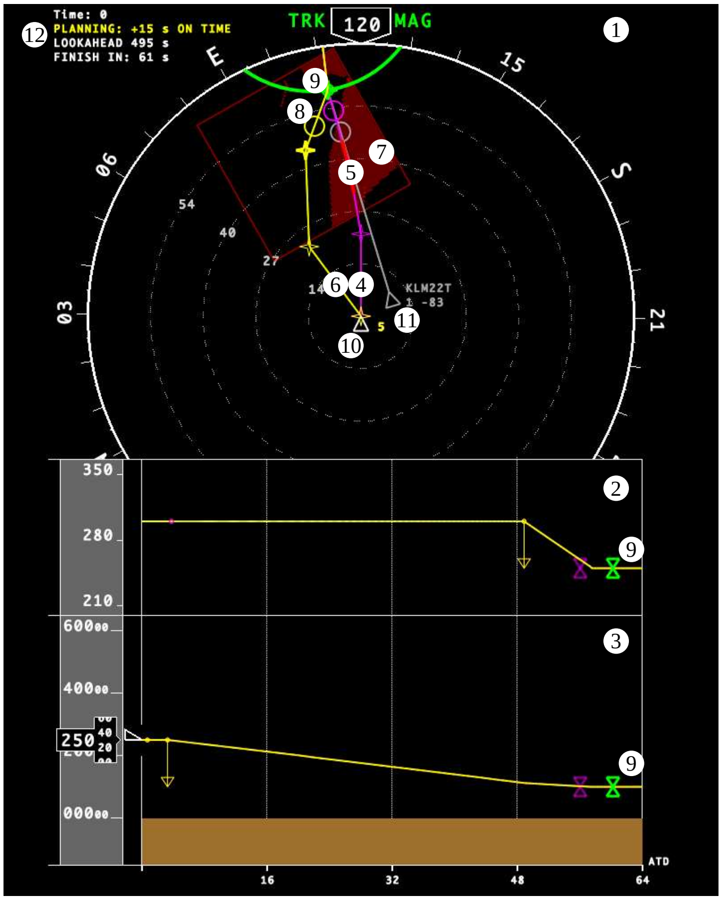


Figure 2: The interface concept. 1. Navigational Display (ND), 2. Speed Profile Display (SPD), 3. Vertical Situation Display (VSD), 4. Own current trajectory (magenta), 5. Conflict on trajectory (red line), 6. Own alternative trajectory (yellow), 7. Alternative waypoint locations (red causes conflict), 8. Separation circles of 2.5 NM radius, 9 Entry point, 10. Current own aircraft position, 11. Current other aircraft position, 12. Anunciation of current time, ∆RTO, and prediction lookahead

Several assumptions were made to narrow down the boundaries of this merging study. First, the airspace near the entry point is assumed to be for destination traffic only, to avoid conflicts with unrelated traffic. Separation minima are defined as a flat cylindrical area around each aircraft, with a horizontal diameter of 5 NM and a vertical height of 1,000 ft. (International Civil Aviation Organization, 1996). Also, a speed limit of 250 kts is imposed below 10,000 ft, based on restrictions in the Dutch airspace(AIS Netherlands, 2009). As a requirement for the constraint visualization, it is assumed that each aircraft shares its planned 4D trajectory. Also, each arriving aircraft has previously been assigned a required time of arrival, and a sequence number.

The main task is to fly the aircraft to the merge point while maintaining separation, meeting the entry point requirements and performing the flight in an efficient manner. In this situation, the planned spacing at the entry point will be the deciding factor for the traffic density upstream. Additionally, the proximity to the entry point will determine the amount of flexibility when replanning to assure separation, or meet constraints. Obviously, ownship performance limits and and efficiency requirements will further constrain maneuverability. Within these constraints, changes to the 4D trajectory can be made using waypoint modifications such as dog legs tromboning or holding patterns, and changes to the altitude and speed profiles.

Figure 2 shows the proposed interface concept. It contains three main parts. On top, the Navigational Display (ND) shows the horizontal trajectory, its waypoints, and the airspace layout. In the middle, the Speed Profile Display (SPD) shows the Calibrated Airspeed (CAS) commands and the resulting profile plotted against the Along Track Distance (ATD). On the bottom, the altitude commands and the resulting profile are plotted on the Vertical Situation Display (VSD) against the ATD. The altitude and speed commands will be executed at the specified ATD, as if entered in the Mode Control Panel (MCP) at that time. Together, the waypoints and commands form the intent. Combined with the ground track defined by the waypoints, atmospheric conditions and the aircraft performance, this defines the full 4D trajectory in terms of position, altitude and time.

The interface concept supports the pilot in his merging task in several ways. In order to make changes to the current intent, pilots must enter the planning mode. In this mode, the pilot can make an alternate trajectory, based on the planned trajectory, to resolve any conflicts. When these modifications are executed, the alternate trajectory becomes the active trajectory, which is also communicated to the other aircraft. Modifications can be done by direct manipulation of the waypoints. Speed and altitude commands are indicated by arrows, and can be altered by scaling or moving the arrows. The target speed or altitude is determined from the vertical location of the arrow head on the screen.

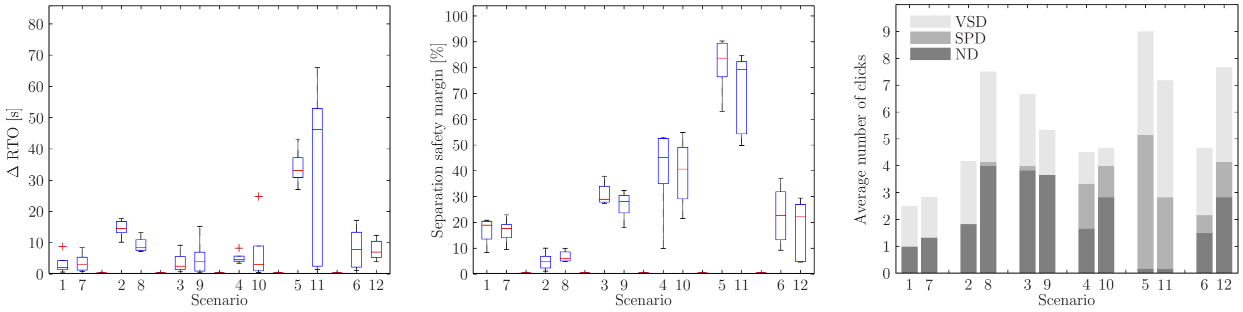
The interface also has the possibility to show a prediction of a future situation, based on the published trajectories. This enables pilots to visually inspect conflicts, and the consequences of trajectory modifications. In this prediction, aircraft are shown on the ND as circles with half the separation margin as its radius. This way, separation is guaranteed, as long as the circles do not overlap.

The segments of the own trajectory where ownship violates separation with one or more other aircraft are colored red, to allow pilots to identify problems with the current plan. In order to further assist pilots with replanning, a grid appears around a waypoint when it is being dragged on the ND. For each location in this grid, it is determined whether placing the waypoint on that location will cause one or more conflicts. The locations in the grid yielding a conflict will be colored red, others will not be colored. The grid forms a separation affordance overlay in which the selected waypoint can safely be placed. Multiple conflicts along the trajectory result in brighter red colors to indicate the conflict severity. Only locations in the square around the originally selected waypoint location are calculated. The limits of the conflict grid are indicated by the thin square around it. Finally, the interface also shows the estimated time of arrival at the entry point, and the difference with the required time of arrival.

Evaluation and Results

An evaluation was performed, to evaluate how well pilots could use the interface to perform the task of maintaining separation and meeting entry point requirements. Objective measures consisted of time to last accepted modification, The number of times a modification was made (and how), deviation from the RTO, and minimum separation during scenario. On a subjective level pilots were asked how they experienced working with the interface, which features they found useful, which were lacking and which could be improved. They were also asked to give an indication of required effort in terms of Rating Scale Mental Effort (RSME)(Zijlstra, 1993). Six professional airline pilots (min. 500hrs. glass cockpit exp.) evaluated the interface using twelve measurement scenarios.The evaluation results are gathered by observing the subjects during the runs, from the logged run data and from the questionnaire answers. They are presented and discussed in this section.

Figure 3(a) shows the absolute RTO error at the entry point. It can be seen that all scenarios except for 5&11 were solved within the 30 seconds RTO error margin. Furthermore, scenarios 2&8 were very densely spaced at the entry point, and the aircraft ahead in the sequence was also late. Therefore, flying over the entry point 15 seconds later was actually desired to stay out of conflict and to build in a little extra separation, which is what the subjects did. With scenario 11, the problem occurred that not enough commands were available on the speed and altitude profile.



(a) Absolute error in RTO (b) Separation safety margin (c) Clicks on ND, SPD and VSD per scenario

Figure 3: Processed run data

Due to this, a few runs for this scenario were solved sub-optimally, while the others were solved with much less time error. This explains the large spread in RTO error. The separation safety margin that pilots achieved with the other aircraft is shown in Figure 3(b). It shows that the separation requirement was never violated. The final objective measure was the number of times a waypoint or command is dragged before obtaining the final trajectory. These modifications are spread across the three trajectory profiles, the ND, SPD and VSD. Figure 3(c) shows these per scenario, where again scenarios 1&7 were replanned with very little modifications, and scenarios 5&11 with many. All scenarios except for 5&11 are a mixture of mostly VSD and ND modifications, while 5&11 barely used any ND modifications. This can be explained with the fact that scenarios 5&11 were very late. Since the ground tracks of 5&11 already headed directly towards the entry point, not much could be changed on the ND to gain more time. The other scenarios were early, so speed changes and path stretching were an option to arrive later. Most scenarios needed some fine-tuning to their altitude profile.

Subjects answered questions in the three parts of the run questionnaire as well as the questions in the post-evaluation questionnaire. In the pre-run questionnaire, subjects indicated their observations on the scenario and how they intended to resolve it. Some subjects tried to predict whether a change in only one of the trajectory profiles would be sufficient, or whether a combination would be necessary. In some cases, several subjects also realized that making a change in one profile would affect the other profiles.

After each run, subjects rated the effort required in terms of the RSME scale and they judged their own performance. The RSME scores are shown in Figure 4. It can be seen that scenario 5 required much effort, while its mirror scenario, scenario 11, resulted in a far lower effort rating. Scenarios 1&7 were perceived to require only little effort.

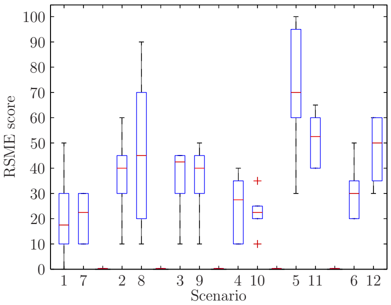
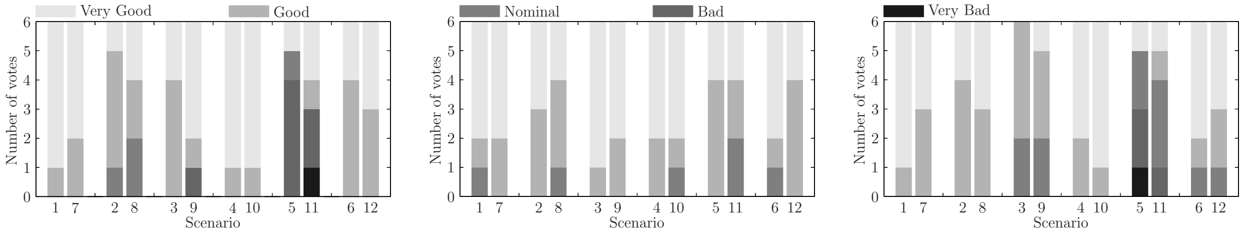


Figure 4: RSME scores

Figure 5(a) shows that scenarios 5&11 also score worse in terms of timing performance. Also scenarios 2&8 were rated low, possibly due to the fact that the leading aircraft was late at the entry point. As the separation was maintained in the end for all the runs, the subjects scored their separation nominal or higher in Figure 5(b). Scenarios 2&8 featured a high density at the entry point and scenarios 5&11 started close to another aircraft, most likely these were contributing factors to low separation scores. Figure 5(c) shows subject rated efficiency. Again, scenarios 5&11 performed very poorly. This is readily explained because to meet the time requirement, the subjects had to increase the speed to the maximum as long as possible and then slow down to 250 kts CAS again. Subjects correctly identified this as a reduction in efficiency.

In addition to providing the RSME and performance scores, subjects filled in the open questions indicating that they solved most of the scenarios as planned. For some scenarios subjects indicated that they modified a different trajectory profile than they had originally planned. For example, some scenarios were early and high. Reducing speed earlier would meet the RTO, but it decreased descent performance. A lateral deviation was also required to also meet the altitude requirement at the entry point. For scenarios 5&11, subjects indicated that it was very hard to meet the timing requirement.



(a) How would you judge your performance regarding the time flying over the entry point?

(b) How would you judge your performance regarding maintaining separation with other aircraft?

(c) How would you judge your performance regarding execution of an efficient flight? (For example, consider fuel burn)

Figure 5: Subjective ratings for timing, separation and efficiency.

Results from the rating questions show that subjects were most positive about the conflict overlay. This could be because it provides a direct mapping between the location of waypoints and the goal of separation. The task of 4D replanning while maintaining separation is at the Knowledge Based Behavior (KBB) level of human performance, as it requires making changes to satisfy goals, while keeping the constraints in mind. By maintaining separation using direct manipulation of waypoints, and by showing overlays that reveal the traffic constraints, complexity of the task is reduced, and a lower level of control is possible. Finally, not all scenarios could be solved using only horizontal changes. Because little assistance was provided for the speed and altitude profile modification, the task of maintaining separation overall still remained challenging.

Subjects were only moderately positive about using the trajectory predictions. Some indicated they used it mostly to verify the new plan, while others actively used it while modifying the plan. It was expected that this statement would score more positively, because it helps them visualize how constraints are separated in time. Generally, human beings are better at working in the spatial domain than in the temporal domain(van Marwijk, Mulder, Mulder, van Paassen, & Borst, 2009). Additionally, trends such as large speed differences cannot be observed well by looking at slowly evolving scenarios. Using the trajectory predictions, speed differences could be more easily observed, and conflict aircraft that were initially out of range could also be detected more quickly.

The subjects also provided feedback on what was missing from the interface or what could be improved. Some subjects mentioned they missed the paths of the other traffic on the VSD. Additionally, the subjects indicated that they used the sequence number frequently, as well as the red colors on the path indicating a conflict.

Discussion

In the evaluation experiment, subjects were able to modify their 4D trajectory reasonably well using the interface. From observations during the experiment, run data and from the questionnaire results, several issues regarding the usability of the interface were identified, which are discussed in this section. The scope of this evaluation was to obtain feedback from professional airline pilots on the usability of the display concept. The results obtained are therefore mostly subjective. Future experiments should provide a more objective evaluation of a matured version of the concept.

As the conflict overlay on the ND was positively rated by the subjects, the same principle can be applied to the SPD and VSD. A grid around the selected speed or altitude command can indicate new locations of the command where the conflict will be resolved or not. Additionally, subjects would have liked to see the altitude profiles of other merging aircraft projected onto the own vertical profile. This is expected to provide pilots more information of the altitude plans of other nearby aircraft, and should help them decide on a well-informed course of action. Next to expanding the conflict overlay to the other profiles, an additional overlay can indicate whether the entry point requirements such as time, altitude and speed are met. Multiple overlays together can give pilots a very quick insight in locations that meet most or all requirements.

An important concern raised was cooperation between conflicting aircraft. In this evaluation, a subject had to navigate through a static traffic scenario of fixed 4D trajectories of other aircraft. In a real conflict situation, each of the involved aircraft will notice the conflict and try to resolve it. To prevent that all the affected aircraft will try to solve the conflict individually, a system of rules and priorities could be used, e.g. a modified version of the ‘rules of the air’. These rules would structure how conflicts are resolved by making one aircraft responsible for solving a conflict between two aircraft. Additionally, aircraft resolving a conflict could communicate their intent to the other aircraft in the conflict.

Conclusions and Recommendations

The preliminary evaluation experiment of the transition display concept, proposed in this paper, showed that of the 72 runs, 71 were successfully replanned such that minimum separation was maintained, and where possible timing constraints were met. The one run that resulted in a conflict situation could not be replanned due to resolvable limitations in the current version of the interface. An important feedback is that the subjects responded in a positive way to the features provided by the interface, where the most appreciated feature was the conflict overlay on the ND.

Finally, the planning flexibility offered by the interface resulted in subjects solving scenarios in different ways, but still successfully. An important reccommendation for improvement from pilots was the addition of constraint overlays to the vertical and speed profiles as well. A more complete evaluation of the display concept should also investigate coordination between conflicting aircraft when resolving a mutual conflict, as this is one of the main concerns with self separated airspace, especially when merging towards a common point.

References

AIS Netherlands. (2009). *Aeronautical Information Publication Enroute 1.4.*

Amelink, M. H. J., Mulder, M., van Paassen, M. M., & Flach, J. M. (2005). Theoretical Foundations for a Total Energy-Based Perspective Flight-Path Display. *The International Journal of Aviation Psychology*, *15*, 205–231.

Borst, C., Mulder, M., & van Paassen, M. M. (2010, Sep.). Design and Simulator Evaluation of an Ecological Synthetic Vision Display. *Journal of Guidance, Control and Dynamics*, *33*(5), 1577–1591.

Borst, C., Suijkerbuijk, H. C. H., Mulder, M., & van Paassen, M. M. (2006, Oct.). Ecological Interface Design for Terrain Awareness. *International Journal of Aviation Psychology*, *16*(4), 375–400.

Ellerbroek, J., Visser, M., van Dam, S. B. J., Mulder, M., & van Paassen, M. M. (2011, Sep.). Design of an Airborne Three-Dimensional Separation Assistance Display. *IEEE Transactions on Systems, Man, and Cybernetics, part A: Systems and Humans*, *41*(6).

FAA. (2009). *NextGen Implementation Plan 2009* (Tech. Rep.). Federal Aviation Administration.

Heylen, F. M., van Dam, S. B. J., Mulder, M., & van Paassen, M. M. (2008, Aug.). Design and Evaluation of a Vertical Separation Assistance Display. In *Aiaa guidance, navigation, and control conference and exhibit.*

International Civil Aviation Organization. (1996). *Rules of the Air and Air Traffic Services (Doc 4444)* (Tech. Rep.).

SESAR. (2006). *Air Transport: The Current Situation (D1)* (Tech. Rep.). SESAR Consortium.

van Dam, S. B. J., Mulder, M., & van Paassen, M. M. (2008, Nov.). 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 Marwijk, B. J. A., Mulder, M. A., Mulder, M., van Paassen, M. M., & Borst, C. (2009). A Human-Machine Interface for Replanning of 4D Trajectories. In *International Symposium on Aviation Psychology* (pp. 169–174).

Zijlstra, F. R. H. (1993). *Efficiency in Work Behavior: A Design Approach for Modern Tools*. Unpublished doctoral dissertation, Delft University of Technology.

THE SELECTION OF COMMERCIAL ASTRONAUTS FOR SUBORBITAL SPACEFLIGHT

Brian Kozak, Denver Lopp, John Young, Tim Ropp
Purdue University
West Lafayette, Indiana

When SpaceShipOne won the Ansari X-Prize in 2004, it launched the commercial space tourism industry. In 2007 Burt Rutan said, “We think that 100,000 people will fly by 2020” (Rutan, 2007). This will create a need for qualified crews to operate these spacecraft. The purpose of this qualitative, exploratory study was to investigate the possible selection criteria of these crews. Data was collected from telephone and email interviews with four U.S.-suborbital space tourism companies and Purdue University's astronaut alumni network. Grounded Theory and Truth and Reality Testing were used as the theoretical framework for data analysis. The data gathered suggests that the commercial astronaut should have at least a Bachelor's degree in engineering, have a test pilot background with thousands of hours of pilot-in-command time in high performance jet aircraft, be confident yet humble, and have a fundamental understanding of his/her spacecraft, including spacecraft trajectories, and emergency procedures.

The Ansari Foundation established the Ansari X-Prize which was a prize of $10 million U.S. to spur innovation to create a cheaper, reusable launch system which would allow economies of scale to lower the cost of space access. Scaled Composites' test pilot Mike Melvill became the first commercial astronaut when he flew SpaceShipOne into space at an altitude of 100 km in 2004. Melvill had 23 years of test pilot experience and flown 10 first flights of Burt Rutan's aircraft. Brian Binnie, another Scaled Composites test pilot, flew SpaceShipOne on its historic flight to win the X-Prize in 2004. His flight reached an altitude of 112 km. Binnie flew for the United States Navy for 21 years and graduated from the U.S. Naval Test Pilot School. These two men have been the only current commercial astronauts at the time of this writing (Ansari X-Prize Foundation, 2010).

Review of Government Astronaut Selection

Deke Slayton, one of National Aeronautics and Space Administration's (NASA) first astronauts and head of astronaut selection from 1963 – 1972, said the selection of an astronaut is a complex task that may take up to two years to complete. Slayton wrote, “I had already developed a point system that we used in making the final evaluations on astronaut candidates. There were three parts: academic, pilot performance, and character/motivation” (Slayton, 1994, p. 133). Each has been a separate, independent skill that intertwines with the others to create a well-rounded candidate. These three parts were used to develop the research questions used in this study.

Early astronauts from NASA had to have at least a Bachelor's degree in a related engineering field such as aeronautical or astronautical engineering (Smith, 2005). NASA also wanted professionals who were stable and who had been screened for security because the early space program was a national security interest (Shepard, 1994). For subsequent astronaut groups, the selection criteria was relaxed slightly to a Bachelor's degree in a physical science or engineering field. Currently, NASA requires a minimum of a Bachelor's degree in science or related engineering for astronaut applicants with more emphasis placed on a Master's degree or Doctorate degree (Cernan, 1999).

Similarly, Soviet/Russian Federal Space Agency (RKA) cosmonauts (astronauts) were required to have a technical engineering degree (Gagarin, 2008). However, the RKA was less interested than NASA in the particular type of degree that their astronauts earned. They accepted various types of engineering degrees as long as they could be somehow related to the space program (Linenger, 2000). Currently, the RKA only accepts astronaut applicants who have a technical engineering degree; a science degree does not qualify (Gagarin, 2008). The desire for highly educated people has also been common to the European Space Agency (ESA). Their astronaut candidates should have a minimum of a Bachelor's degree in a technical, scientific or engineering field. A Master's degree with several years of related experience prior to application has also been preferred (Messerschmid et al., 2003).

The spaceflight industry has been highly competitive and as a result the quality of academic preparedness has been important. Gene Cernan, a Purdue University graduate who walked on the Moon during Apollo 17, wrote “Receiving a B instead of a B+ in some college course years earlier might be reason enough to pick the other guy” (Cernan, 1999, p.58).

The RKA was the first agency to launch an astronaut into space in 1961; NASA followed with a suborbital spaceflight a few weeks later (NASA, 1963; RKA, 2009). The early astronauts from both agencies were explorers in the brand new field of spaceflight. Both agencies decided to select military test pilots for the first group of astronauts because of their considerable piloting experiences in experimental vehicles (Clark, 1988). Early spacecraft were highly experimental and fraught with danger because the challenges of flying in space were relativity unknown and the technology to travel in space was in its infancy (Harvey, 2004).

NASA wrote, "The astronauts were first and foremost test pilots, men accustomed to flying along in the newest, most advanced, and most powerful vehicles this civilization had produced. They were talented specialists who loved to fly high-performance aircraft and who had survived the natural selection process in their profession" (NASA, 1963, p. 1). Flying experimental vehicles was dangerous and approximately six test pilots died each year as a result of testing them (Slayton, 1994). The pilots who did survive were ideal for the space program because they were well-rounded pilots with experience related to the interdisciplinary aspects of flying (Borman, 1988).

Unforeseen problems could occur that might require an astronaut to think quickly in unfamiliar situations. Test pilots regularly work and operate in such environments (Freeman, 2000). On the first flight of suborbital flight of SpaceShipOne there were several anomalies. After engine ignition, the vehicle experienced a failure of an electric trim tab which caused it to roll a total of 29 times at speeds ranging from Mach 1 to Mach 3 at altitudes in excess of 200,000ft (Scaled_Composites, 2010). SpaceShipTwo, though built on existing technology, has been a highly experimental vehicle.

The quality and type of person selected as a commercial astronaut has also been very important. The astronauts are flying high performance vehicles in extremely hostile environments with little room for error. The physical and physiological stresses are immense (NASA, 1963). An astronaut needs to be calm under pressure and quick thinking (Freeman, 2000). The qualities of test pilots were also very similar. Shepard wrote, "Esprit de corps, pride, honor, dedication, skill, and courage were all qualities required of the men who would become astronauts" (Shepard, 1994, p. 50).

Methodology

This research project was a qualitative study conducted to contribute to the body of knowledge as it relates to the selection of commercial astronauts for suborbital spaceflight. The study was exploratory because the space tourism industry is less than 10 years old and there are presently only two commercial astronauts. Sekaran wrote, "An exploratory study is undertaken when not much is known about the situation at hand, or no information is available on how similar problems or research issues have been solved in the past" (2003, p. 119).

Wiggins said, "Qualitative research techniques are typically applied in situations where little is known about a particular domain" (1999, p. 164). Furthermore, qualitative research methods are useful for "new fields of study where little work has been done, few definitive hypotheses exist and little is known about the nature of the phenomenon" (Patton, 2002, p.193).

Theoretical Framework

The theoretical framework describes and defines the gathering and interpretation of data in this study as it relates to qualitative research and is a key aspect of such research (Sekaran, 2003). The qualitative research technique Grounded Theory was used as a theoretical framework for the study. It was developed by Barney Glaser and Anselm Strauss as a research method to create a theory from the generalization of recorded data rather than test an established theory (Sekaran, 2003). Grounded Theory allows for opened-ended interviews which are useful to gather as much data about a certain process as possible without being constrained to set answers. A key aspect of Grounded Theory is the bias and credibility of the researcher because the researcher gathered, analyzed and interpreted data based upon his experience and knowledge of the subject. Grounded theory was the fundamental framework used in this study, Truth and Reality Testing was used as a secondary framework to backup data gathered with Grounded Theory. Truth and Reality Testing was useful to determine how data gathered related to what is going on in the real world (Patton, 2002). The researcher used recorded phone interviews and email correspondence to collect data. The researcher was able to establish a rapport with the individual prior to the actual interview. The individuals representing the suborbital space companies were co-founders, vice presidents of engineering/research, former NASA astronauts, or test pilots with actual suborbital spaceflight experience.

Data Analysis

Data analysis of qualitative studies was unique to the study and theoretical framework being used. Patton wrote, "Qualitative analysis transforms data into findings. No formula exists for that transformation" (2002, p. 433).

The analysis of data associated with the study is heavily dependent upon the theoretical framework used within the study and with the bias and credibility of the researcher himself (Patton, 2002). Like quantitative research, there exists a set of tools to allow the researcher to conduct his/her data analysis with credibility (Glaser & Strauss, 1967).

Grounded Theory is one such tool which is based upon the ability of the researcher to fundamentally understand the data gathered and transform the data into meaningful information. The resulting assertions and/or theory generated from the data is the result of the data analysis process. There are several distinct steps in the Grounded Theory process in which the data/interview transcripts were read and analyzed a total of seven times.

During data analysis, the interview transcript was read completely several times with short notes written about key points being made and with questions asked internally such as, 'What is the interviewee really saying? What points are made with this topic?' (Strauss & Corbin, 1990). The interview transcripts were coded via Grounded Theory's open coding which allowed for informative descriptors to be assigned to key phrases in the transcripts. Each of the codes was assigned to a matrix according to the research question asked and the interviewee. There were two different matrices for each research question: one for suborbital space companies and the other for Purdue's astronaut alumni. Once the codes were assigned to their respective matrices, common themes were found among interviewees. The final step of data analysis was to make assertions based upon the data gathered. The assertions made with regard to the research questions were compiled to create a strength of the continuum. The strength of the code was determined by the researcher using the code's location within the transcript, relations to the context in which the code was mentioned, the emphasis placed upon the code itself by the interviewee and via the researcher listening to the audio recording of the interviewee's tone and inflection when the particular code was said (Strauss & Corbin, 1990).

Researcher Bias

I am a graduate of Purdue’s College of Science with Bachelor of Science degrees in Applied Physics (2007), Interdisciplinary Science (2008) and a Master of Science degree in Aviation and Aerospace Management (2010). I am a FAA certified private pilot with approximately 220 hours of Pilot-in-Command time with an additional 200 hours in FAA approved flight training devices. I have read the biographies and autobiographies of most of the 24 astronauts who journeyed to the Moon as well as those key personnel within NASA during the Apollo Program. The most influential book that I have read for this research project was *Deke!*, written by Deke Slayton, who was one of NASA’s first astronauts. During his tenure at NASA, Slayton was responsible for the selection of new astronauts from the mid 1960s to the late 1970s and he selected the crews for all of the Gemini and Apollo missions.

Findings and Assertions

Research Question 1: What kind of educational and/or technical background should a commercial astronaut possess?

Assertion: A commercial astronaut should have an engineering degree at the Bachelor's level or higher.
Definition: an engineering degree is a broad term that could be any sub discipline of engineering as it relates to aerospace such as aeronautical, astronautical, electrical, mechanical or materials engineering. A bachelor's degree is minimal amount of educational knowledge required with preference given to advanced degrees.

> “As it comes to educational requirements, I would say an engineering degree at least as a Bachelor's” (NASA Astronaut A, personal communication, 2010)
>
> “So for educational and technical background whatever the requirements are for military pilots along the lines of engineering training or possibility some classes either undergraduate or master degree level” (NASA Astronaut B, personal communication, 2010)

Research Question 2: What type and how much flight experience should a commercial astronaut have?

Assertion: A commercial astronaut should have a test pilot background.
Definition: A test pilot is someone who flies new and experimental aircraft during the first few flights to verify the design specifications of the vehicle. Furthermore, they 'push the limits' of the vehicle to determine the margins of safety for commercial or military use. A test pilot is also familiar and comfortable with high stress situations associated with testing new, unproven vehicles.

> “Generally those type of vehicles are not going to have enough experience behind them to be routine so people with flight test experience [would] probably be more appropriate for piloting those vehicles” (Suborbital Space Company Representative B, personal communication, 2010).

> "Because of the big unknowns right now, you are going to want test pilots" (NASA Astronaut C, personal communication, 2010)

Assertion: A commercial astronaut should have thousands of hours as Pilot-in-Command (PIC) of high performance jet aircraft (The FAA and the airline industry use the term PIC as time accumulated by a pilot in command of the aircraft even though the autopilot may be used for the majority of the flight. In this definition, PIC time refers to only hand flown time)
Definition: a high performance jet aircraft is a single or multi turbine engine aircraft similar in performance and maneuverability to state-of-the-art military fighters. PIC time is actual pilot control of the aircraft with no autopilot use.

> "I think commercial astronauts, the ones actually flying the vehicle, should have to have the same requirements that NASA astronauts have. And that would be a minimum of 1000 hours pilot in command time" (NASA Astronaut D, personal communication, 2010).
>
> "Somewhere around the order of several thousand hours, probably around two thousand hours in a high performance jet aircraft or a thousand minimum that you would need to have before you had somebody climb into a rocket expecting them to fly passengers to space" (NASA Astronaut B, personal communication, 2010).

Research Question 3: In terms of personality and character, what would make a desirable commercial astronaut?
Assertion: A commercial astronaut should have strong communication skills.
Definition: communication skills are the ability to document and transmit information verbally or via writing. The ability of a pilot to relate a wrong 'feel' of an aircraft to a technician who can fix the problem is extremely important.

> "You've got to be able to work with the engineers and technicians who are developing the spacecraft to be able to, when they find something that needs to be fixed, they need to be able to communicate why it needs to be fixed and work with the people involved to get it fixed" (NASA Astronaut C, personal communication, 2010).
>
> "The ability to communicate phases of flight and the willingness to communicate all of the phases of flight and explain to them what to expect and give them [the passengers] a couple of updates about the way things are going" (NASA Astronaut B, personal communication, 2010).

Assertion: A commercial astronaut should have enormous confidence in his/her vehicle and training and should inspire confidence in his/her passengers.
Definition: the space tourism industry is an emerging market with cutting edge technology and experimental vehicles. In order to successfully operate a spacecraft for passenger revenue flights, a commercial astronaut needs to to be confident in his/her vehicle's design and testing as well as in his/her own abilities to fly the spacecraft safely.

> "[It is a] brand new industry and it is a brand new experience for these folks and so that level of confidence of the flight crew would be critical and so to me that would say being able explain the engineering of the vehicle in the flight, the propulsion, the electrical systems and the ascent environment and the technical perspectives and should be able to explain it and comment everything about it would be an absolutely critical part of the experience that these folks would have" (NASA Astronaut A, personal communication, 2010).
>
> "That is because they are flying people in space who are probably paying per seat in that sense the crew needs to be like a boat captain or cruise director" (NASA Astronaut B, personal communication, 2010).

Assertion: The commercial astronaut should be humble as it pertains to his/her personality and work.
Definition: being humble means having the ability to admit mistakes and learn from them, the ability to take constructive criticism well and the ability to 'check your ego' before piloting a suborbital spacecraft.

> "Throw away the scarf" (NASA Astronaut A, personal communication, 2010).
>
> "I think they ultimately want someone that can check their ego at the door and work effectively in a team environment" (Suborbital Space Company Representative C, personal communication, 2010).

Research Question 4: If training a person to become a commercial astronaut, what are the most important subjects and/or flight areas in which to be familiar?
Assertion: The commercial astronauts should have a fundamental understanding of their spacecraft's engineering, performance characteristics and limitations.
Definition: information about the specific handling characteristics of the spacecraft as it pertains to G limitations, aerial maneuvers, landing speed, maximum bank angle, etc.

> "It would be more of a familiarization with the vehicle. That's important" (NASA Astronaut D, personal communication, 2010).
>
> "Things like engine operation/thrust, flight trajectory (up and down), flight path angle, "zero g" characteristics, heating, flight control response throughout the flight regime, and of course, landing procedures, would be topics of interest" (NASA Astronaut E, personal communication, 2010).

Assertion: The commercial astronaut should be familiar with emergency procedures of the vehicle.
Definition: an emergency procedure is an event that is off-nominal, unplanned or otherwise dangerous to the safety of the vehicle.

> "Also the flight simulation and emergency procedures that would have to be taught and learned" (NASA Astronaut D, personal communication, 2010).
>
> "But for the spacecraft it would be some kind of really dangerous spin or a pressurization, something like that. Which you really wouldn't want the real person to go through because they have to learn. You don't want them to kill themselves" (NASA Astronaut C, personal communication, 2010).

Assertion: The commercial astronaut should be trained in spacecraft trajectories during launch, ascent, cruise, entry and landing.
Definition: suborbital spaceflight is unique when compared to traditional aircraft due to the high speeds, altitudes, maneuvers performed and power of the vehicle. Key parts of understanding the new kind of flight environment are astronautics, physics and aeronautic engineering.

> "So most of the training will have to revolve around the trajectories itself and the attitudes, G loads of those trajectories and then the handling of the flight control systems during ascent and entry and then the off-nominal, potential off-nominal scenarios around all that" (NASA Astronaut A, personal communication, 2010).
>
> "Flying rockets is a completely unique environment" (Suborbital Space Company Representative C, personal communication, 2010).

Research Question 5: When selecting a commercial astronaut, is there anything that you believe is important to consider that we have not discussed?
Assertion: A commercial astronaut is not needed for some spacecraft operations.
Definition: Armadillo Aerospace is designing and building their *Black Armadillo* suborbital spacecraft to operate as a fully automated vehicle. There would be no flight crew members aboard the vehicle during tourist flights.

> "The Armadillo Aerospace vehicle concepts require no crew as such. They are designed to be autonomous in virtually all respects except for the launch controllers and pad ops team. There will be no pilot, commander...or beverage service when we reach cruising altitude :-)" (Suborbital Company Representative A, personal communication, 2010)

Summary of Findings

The assertions made about a commercial astronaut are that he/she: should have at least a Bachelor's degree in engineering, have a test pilot background with thousands of hours of PIC time in high performance jet aircraft, be confident yet humble in personality and have an in-depth knowledge of their spacecraft, including emergency procedures and spacecraft trajectories.

Discussion

The results of this study may be hard to generalize to the commercial space tourism industry due to the uniqueness of each company's spacecraft design. Also, the participants of this study, n=11, represent a small portion of the individuals who have flown in space and who are actively working to develop suborbital spacecraft. In addition, the space tourism industry is yet unproven and highly experimental. The only U.S.-based suborbital space company to operate a suborbital spacecraft was Scaled Composites with three spaceflights of SpaceShipOne in 2004. The spacecraft being developed have been experimental vehicles with little or no previous experience in actual suborbital spaceflight. This could cause the initial selection criteria of suborbital commercial astronauts to be much higher right now as opposed to when the industry becomes more established. Several of the interviewees shared these thoughts.

> "I believe the initial requirements will differ from the later ones when (if) the business becomes established" (Suborbital Space Company Representative C, personal communication, 2010).

> "The qualifications right now will be much higher than they will be in say 10 years...Right now, there are very few slots and the competition is very fierce and there are huge numbers of unknowns. So all of those are going to drive all of these requirements very high right now" (NASA Astronaut C, personal communication, 2010).

Recommendations

Although the results of this exploratory study suggest that the suborbital commercial astronauts should have a strong background in engineering, flight test experience and have strong personable skills, further studies may be conducted to strengthen the results. A larger sample of participants for the next study may be needed. It is the researcher's recommendation to use this study as a basis for further research into the area of commercial astronaut selection. Future research should also investigate a possible training guide for commercial astronauts to prepare themselves at the university level for careers as commercial astronauts. Many of these research topics and questions can only be answered once the space tourism industry is established. There is need for additional research into this emerging industry and this study provides the basis for it with regard to commercial astronaut selection.

Additional information can be found in Brian Kozak's 2010 Thesis 'The Selection of Commercial Astronauts for Suborbital Spaceflight', Purdue University. I can be reached at bkozak@purdue.edu

References

Ansari X-Prize Foundation (2010). *Ansari_x_prize*. Retrieved from http://space.xprize.org/ansari-x-prize

Borman, F. (1988). *Countdown*. New York, NY: Silver Arrow Books.

Cernan, E. (1999). *The last man on the moon*. New York, NY: St. Martin's Press.

Clark, P. (1988). *The Soviet manned space program.* New York, NY: Orion Books.

Federal Aviation Administration Office of Commercial Space Transportation (2010). *2010 U.S. commercial space transportation developments and concepts: vehicles, technologies, and spaceports.* Retrieved from http://www.faa.gov/about/office_org/headquarters_offices/ast/media/ast_developments_concepts_2010.pdf

Freeman, M. (2000). *Challenges of human space exploration.* Cornwall, United Knigdom: Praix Publishing.

Gagarin research & test Cosmonaut training center: Cosmonaut training and testing (2008). Retrieved from http://www.gctc.ru/eng/index.html

Glaser, B. & Strauss, A. (1967). *The discovery of grounded theory*. Chicago, IL: Aldine Publishing Company.

Harvey, B. (2004). *China's space program: from conception to mannedspaceflight*

Linenger, J. (2000). *Off the planet*. New York, NY: McGraw-Hill.

Messerschmid, E., Haignere, JP., Damian & K., Damann, V. (2003). *European astronaut centre prepares of international space station operations*. Science Direct.

NASA. (1963). *Project mercury astronaut selection overview*, Retrieved from http://www.nasa.gov/mission_pagesmercury/missions/astronaut.html

Patton, M.Q. (2002). *Qualtative research and evaluation methods (3^{rd} ed.).* Thousand Oaks, CA: Sage Publications Inc.

RKA. (2009). *Russian federal space agency.* Retrieved from http://www.roscosmos.ru/fsa_naprav.asp?

Rutan, B. (12-January-2007). *Entrepreneurs are the future of space flight*. TED Conference. Long Beach, CA.

Scaled_composites. (2010). Retrieved from http://www.scaled.com/projects/tierone/

Sekaran, U. (2003). *Research methods for business: A skill building approach.* New York John Wiley & Sons Inc.

Shepard, A. (1994). *Moon shot*. Atlanta, GA: Turner Publishing.

Slayton, D. (1994). *Deke!* New York, NY: Forge Books.

Smith, A. (2005). *Moondust.* New York, NY: HarperCollins Publishers.

Strauss, A. & Corbin, J. (1990). *Basics of qualative research.* Newbury Park: Sage Publications Inc.

Wiggins, M.W (1999). *Aviation social science: Research methods in practice.* Aldershot, United Kingdom: Ashgate. New York: Praix Publishing.

KNOWLEDGE TESTS FOR AB-INITIO PILOT SELECTION IN CIVIL AVIATION

Oliver Zierke
German Aerospace Center (DLR)
Hamburg, Germany

A validity study was conducted for two purposes: to evaluate the importance of knowledge tests for pilot selection and to validate the knowledge tests of the German Aerospace Center. The two criteria were a theoretical test given in the Flight Training School and the overall graduation rate. Regression models were calculated to measure the incremental validity of knowledge tests beyond ability tests. A further question addressed the comparison of the predictive validity of knowledge tests and school grades. Knowledge tests contributed 12% incremental validity and thus yielded similar results to school grades (17%). Both regression and dropout analyses revealed knowledge tests to be a better predictor of Flight Training School success than ability tests.

Test batteries often contain knowledge tests, as is the case for the Air Force Officer Qualifying Test (AFOQT, Carretta & Ree, 1996), the Armed Services Vocational Aptitude Battery (ASVAB, Sands, Waters, & McBride, 1997), or the test battery of the German Aerospace Center (DLR, Maschke, 2004). What we refer to as knowledge tests are methods used to reveal the amount of knowledge, skills, and experience of an individual in a given field. Knowledge tests measure aspects of crystallized intelligence in terms of Cattell's theory of intelligence (Cattell, 1987). Meta-analytic research on predictors for job success suggests that knowledge tests show good prognostic validities in general, and particularly for jobs in aviation. Schmidt and Hunter (1998) found a mean correlation of $r = .48$ with job success and an incremental validity beyond cognitive measures of .07 for many jobs. Dye, Reck, and McDaniel (1993) also found a mean correlation of $r = .47$ between knowledge tests and job performance. For the field of aviation, Olea and Ree (1994) reported a mean correlation of $r = .24$ between knowledge tests and practical training success for pilots. In meta-analytical studies, Martinussen (1996) found prognostic validities of $r = .24$ for tests of aviation information and in Martinussen (1998) the same magnitude was reported for tests of mechanical comprehension and aviation information for pilot selection in the Norwegian Air Force. As a result of her meta studies, Martinussen explicitly recommended that pilot selection batteries should include knowledge tests about aviation (Martinussen, 1996, p. 16). The reasons for using knowledge tests for the selection of pilot applicants are different for ready-entry and for ab-initio applicants. Ready-entry applicants are required to have prior job knowledge (Ree, Carretta, & Teachout, 1995), which can be defined as "the cumulation of facts, principles, concepts, and other pieces of information that are considered important in the performance of one's job" (Dye et al., 1993, p. 153). It is assumed that job knowledge has a positive impact on job performance. In contrast, ab-initio applicants are generally tested for more basic knowledge in mathematics, physics, and technical comprehension. This corresponds to the distinction of Caretta and Ree (2000, p. 2): "When selecting applicants for ab-initio training, indicators of ability (i.e. trainability) are emphasized. When selecting from experienced pilots, commercial carriers tend to emphasize indicators of prior experience (…) and flying competence (…)." This study concentrates on the impact of knowledge tests for ab-initio pilots. These tests measure specific knowledge in domains which are important prerequisites for the theoretical training, that is mathematics, physics, technical comprehension and English language.

The following questions will be addressed:

1. Do knowledge tests contribute any unique predictive validity to cognitive ability tests?
2. How valuable are knowledge tests compared to school grades as predictors of success in the flight training school?
3. Which tests best predict training dropouts?

Methods

Subjects

The criterion data were collected from 402 student pilots (88% male, 12% female) in their theoretical ab-initio pilot training at the flight school of a major European airline. All students had graduated from high school with a degree suitable for enrolling at a university. The subjects were about 21 years old on average (*SD* = 2.5, range = 18 to 29).

Measures

The predictor data emerged from the first phase of the DLR selection approximately 8 months before the students began their pilot training. The DLR test battery was administered by computer and included seven cognitive ability tests as well as four knowledge tests and two psychomotor ability tests. The cognitive ability tests covered concentration, memory capacity, quantitative ability, and spatial orientation. The knowledge tests covered the topics of English language, mathematics, technical comprehension, physics, and technical basics. Furthermore, school grades in subjects related to theoretical pilot training were recorded. These subjects are English, mathematics, and physics.

The first criterion is a composite score of two test results from a theoretical training first phase (Starter Course) at the flight school. This Starter Course covers basic mathematical and physical knowledge which is considered to be prerequisite for the following theoretical training for the Air Transport Pilot License (ATPL). The second criterion is the pilots' successful completion of the flight training program. It is a simple pass/fail-criterion. The failed applicants were defined as one group called dropouts.

Analyses

Bivariate correlations between predictors and criteria were calculated. Correlation coefficients were corrected for range restriction and are reported in both forms, corrected and uncorrected. In order to estimate the incremental validity of knowledge tests, multiple regressions were carried out. Furthermore, a dropout analysis with independent-samples t-tests was done in order to find out which tests were most predictive.

Results

1. Knowledge tests and school grades were more strongly related to success in the Flight School Starter Course than the cognitive ability tests (see Table 1, column Correlation).
2. Knowledge tests added 12% to the explained variance of the criterion beyond the cognitive tests. School grades added 17% to the explained variance of the criterion.
3. The most successful tests in predicting the dropouts were the psychomotor tests and school grades followed by the knowledge tests. The cognitive ability tests were least predictive (see Table 1, column Effect Size).

Discussion

The knowledge tests turned out to predict both an early and a late criterion well. This result is in line with most of the meta-analytical results mentioned in the introduction. In fact, their predictions of different criteria were even better than those of cognitive ability tests. However, this was surprising in light of research on ab-initio pilots by Olea and Ree (1994) and Ree, Carretta, and Teachout (1995) who demonstrated the predictive power of general intelligence (*g*) for the AFOQT. These studies did not differentiate between fluid intelligence (*Gf*) and crystallized intelligence (*Gc*). Instead, they defined every test loading on a first factor as *g*, regardless of whether it was a reasoning test or a knowledge test. Actually, about half of the tests they used were knowledge tests in terms of the *Gf-Gc* theory, as for example Mathematical Knowledge, Mechanical Comprehension, and Aviation Information. These tests correlated well with practical training results. It seems that Olea and Ree's *g* includes both *Gc* and *Gf* to the same degree because their test battery, the AFOQT, contains elementary cognitive tasks as well as knowledge tests. Therefore, their results do not contradict the present study.

The present study is consistent with results in the area of admissions to graduate programs. A meta-analysis by Kuncel, Hezlett, and Ones (2001) referring to the Graduate Record Examination included more than 1700 independent samples and demonstrated that domain-specific knowledge tests are better predictors for successful graduation than the ability tests. The same line of results is shown in research on learning and instruction as well as research on expertise. Weinert and Helmke (1995) demonstrated that previous knowledge predicted school grades better than intelligence. Grabner, Stern & Neubauer (2007) showed that the strongest predictor of the attained level of chess experts was deliberate practice and the accumulation of a broad knowledge base rather than general intelligence. Based on this evidence, knowledge tests should receive more attention in aviation psychology.

Table 1

Correlations of the Predictor Variables with the Criterion Starter Course Test and Effect Sizes for Prediction of Dropouts

Variable	*N*	Validity	
		Correlation *r*	Effect Size *d*
Cognitive Tests			
Concentration			
OWT	402	.08	0.03
SKT	402	-.02	0.34
Memory Capacity			
MST	402	.07	0.30
RMS	402	.15**	0.01
Quantitative Ability			
KRN	402	.20**	-0.07
Spatial Orientation			
PPT	402	.19**	0.43
ROT	402	.08	0.08
Psychomotor Ability			
MIC	402	.10*	0.49
SIM	402	.18**	0.69
Knowledge Tests			
ENS	402	.09	-0.10
RAG	402	.32**	0.33
TEC	402	.33**	0.31
TVT	402	.23**	0.69
School Grades			
English	374	.17**	0.21
Mathematics	375	.30**	0.59
Physics	366	.35**	0.60

Note. *r* = observed correlation, ***p* < .01, one-tailed.

References

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., & Ree, M.J. (2000). *Pilot selection methods.* Technical report. Dayton: U.S. Air Force Laboratory.

Cattell, R.B. (1987). *Abilities: Their structure, growth, and action.* Boston: Houghton-Mifflin.

Dye, D.A., Reck, M., & McDaniel, M. (1993). The validity of job knowledge measures. *International Journal of Selection and Measurement*, 1, 153-162.

Grabner, R.H., Stern, E., & Neubauer, A.C. (2007). Individual differences in chess expertise: A psychometric investigation. *Acta Psychologica*, 124, 398-420.

Kuncel, R.K., Hezlett, S.A., & Ones, D.S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations. *Psychological Bulletin*, 127, 162-181.

Martinussen, M. (1996). Psychological measures as predictors of pilot performance: A meta-analysis. *The International Journal of Aviation Psychology*, 6, 1-20.

Martinussen, M. (1998). Pilot selection in the Norwegian Air Force: A validation and meta-analysis of the test battery. *The International Journal of Aviation Psychology*, 8, 33-45.

Maschke, P. (2004). The acceptance of ab initio pilot selection methods. *Human Factors and Aerospace Safety,* 4, 225-232.

Olea, M.M., & Ree, M.J. (1994). Predicting pilot and navigator criteria: Not much more than g. *Journal of Applied Psychology*, 79, 845-851.

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.

Sands, W.A, Waters, B.K., & McBride, J.R. (Eds.) (1997). *Computerized adaptive testing: from inquiry to operation.* Washington: APA.

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.

Weinert F.E. & Helmke, A. (1995). Interclassroom differences in instructional quality and interindividual differences in cognitive development. *Educational Psychologist*, 30, 15-20.

DELTA AIR LINES PILOT SELECTION: 10 YEARS OF VALIDATION RESULTS

Carl C. Hoffmann
Human Capital Management & Performance, L.L.C.
Mebane, NC
Sally H. Spetz
Raleigh, NC
Arianna K. Hoffmann
Human Capital Management & Performance, L.L.C.
Mebane, NC

In the early 1990s Delta Air Lines was faced with a changing pool from which to recruit its pilots, as well as a change in the technology and management of the cockpit. Delta committed to developing a new selection process. After analyzing the results from over a dozen potential tests, three were chosen: CogScreen, the NEO PR-I, and a custom developed Job Knowledge Test. Initial analysis indicated the measures were highly predictive of performance in the cockpit and in training. The cohort of 339 pilots used to validate the selection process has now been followed for 12 years to examine long-term predictive validity. We also present a cross validation of selection measures with a sample of pilots hired in 2007. This study confirms the initial components of the model and indicates that experience and education prior to hire can add significantly to the predictive power of the initial selection model.

In the early 1990's Delta Air Lines developed a new process for the selection and hiring of their pilots. This effort was motivated by several factors, including changes in the pool of available candidates, the move to more highly automated cockpits, and the need for new objective, job-related criteria. Delta was also very interested in finding an effective and efficient set of measures that would minimize the cost and time to screen applicants, while supporting selection of the best candidates.

Guided by an extensive job analysis, reviews of relevant literature, and observations of other global airlines' selection processes, a battery of tests was constructed. These tests tapped four major areas of cognition, knowledge, and ability:

- Standardized Aptitude and Ability Tests: including verbal and numerical ability, spatial ability, mechanical ability, reasoning, non-verbal aptitude
- Cognitive Processing (CogScreen): a computer administered test series designed to measure "the underlying perceptual, cognitive, and information processing abilities associated with flying." (Kay, 1995).
- Personality (NEO PI-R): The Big Five personality factors (Costa and McRae, 1992)
- Job Knowledge Test (JKT): a proprietary test based on a thorough job analysis of the particular demands and activities required of Delta pilots.

In addition to the tests, the selection process included an interview conducted by trained Delta interviewers that assessed behaviors and experience reflecting areas of performance not easily measured by tests. A successful interview outcome is also required for a candidate to be extended a conditional job offer. This paper will focus only on the tests. Future publications will address the interview, which is still in use today, in more detail.

Development of Select-in and Select-out Models

In late 1996 and early 1997, the full battery of tests was administered to a random sample of 200 current Delta pilots hired between 1987 and 1991 (**'87 cohort**), 50 recreational pilots, (**Rec pilots**) and over 600 Delta applicants, of whom 339 were subsequently hired (**'96 cohort**). The job of commercial airplane pilot is a complex and demanding one in which a constellation of knowledge, skills, ability, and traits is required. In developing the hiring process, two models for pilot selection were pursued, a Select-out Model and a Select-in Model. The Select-out Model is directed at finding parameters that distinguish those candidates who are not likely to fulfill the basic requirements expected of a Delta pilot. The Select-in Model is directed at discriminating low performing Delta pilots from the population of successful Delta pilots. Pilot candidates must pass both models in order to be hired.

Select-out Model. The Select-out Model compared test scores of Rec pilots to the ’87 cohort in an effort to differentiate amateur pilots from professionals. A variety of statistical techniques were tried, including multiple regression, factor analysis, cluster analysis, and discriminant function analysis without definitive results. We then undertook a proportionate reduction of error study to identify true positives (Delta pilot scoring above the cut score) and true negatives (Rec pilot scoring below the cut score) on each subtest of the tests included on the battery. Passing scores were set by rank ordering scores and picking a point where those above the passing score included the most Delta pilots and the fewest Rec pilots, while those below the passing score had the most Rec pilots and the fewest Delta pilots.

This process resulted in a reduced set of test measures from the NEO, the CogScreen, and the JKT. The NEO contributed scores from two of the personality measures (Conscientiousness and Neuroticism), the CogScreen contributed scores from seven measures related to accuracy (Math, Symbol Digit Coding, SDC Immediate Recall, SDC Delayed Recall, Shifting Attention Instruction, Pathfinder Combined, Visual Sequence Comparison, and Pathfinder Letter); the JKT contributed a score from each of three subtests, Engineering, Navigation, Aerodynamics, and a separate contribution for the JKT total score. A regression equation was run on these measures that yielded weighted coefficients significant at the .05 level or better.

Because the model is made up of multiple components, it is possible for a very high score on one measure to compensate for a huge deficit in another area. For instance, pilots who score well on job knowledge may need to draw less on multi-tasking ability and speed of processing, because they have ready access to the knowledge they require to do the task at hand. Pilots with a personality profile that allows them to easily draw on the resources of their co-pilots may compensate for a lower score on cognitive processing or job knowledge. Because of the mathematics of the equation, two decision rules were incorporated into the Select-out Model: (1) pilot candidates had to pass at least nine of the 13 measures at or above the 20th percentile of Delta pilot scores; (2) these nine measures had to include at least one JKT measure and at least five CogScreen measures.

Select-in Model. Developing a Select-in Model presented different challenges. We examined performance in the cockpit as well as performance in training for the ’87 cohort. Cockpit performance was measured in flight by Line Check Airmen for 115 of the 200 pilots in the cohort using an evaluation instrument specially developed for Delta (see Hoffmann, 1998). The efficacy of the tests in predicting cockpit performance was presented in workshops at the 1997 and 1999 ISAP meetings and in 1998 at a NATO conference on Collaborative Crew Performance in Complex Operational Systems. While cockpit performance and training measures were highly correlated with test data, valid cut scores could not be determined from this cohort. As a consequence, the Select-in Model was developed using data from the ’96 pilot cohort. We suspect that pilots from the ’87 cohort, being far removed from their hire dates and not under the same pressure in the testing situation as new applicants; may not have had the same level of motivation undergirding their test performance.

The entire test battery was administered to 600 applicants, of whom 339 were hired in the ’96 cohort. Since a cut score was not established on the ’87 cohort data, an interim procedure was adopted. Based on a comparison of test performance in the’96 and ’87 cohorts, three groups were identified: a clear hire group, a clear rejection group, and a borderline group. Decisions on borderline candidates were made on a case by case, qualitative basis. A successful interview was also required for the candidate to be extended a conditional job offer.

New hire training data was collected on the ’96 cohort and used as the criterion measure for determining the Select-in Model. Again, the NEO PI-R, the CogScreen and the JKT contributed significant elements, although the particular subtests varied somewhat from the Select-out Model. For the NEO, the significant factor was Openness. For the CogScreen, the significant subtests were SDC Throughput, Manikin Accuracy, Divided Attention Premature Hits, Shifting Attention Discovery Accuracy, and Pathfinder Number Throughput. For the JKT, the significant subtest was Engineering. Weighting coefficients for these subtests were determined by regression.

Study Purpose

Our initial studies were of necessity short term or cross sectional, and could not address long term outcomes. Using selection criteria to predict long term job performance is of great value in a commercial airline setting since pilots generally stay with the company for 25-30 years. We are now able to report on 12 years of data tracking the ‘96 cohort through a substantial portion of their career development with Delta, allowing us to evaluate the power of the selection criteria to predict pilot performance over time. In addition to testing long term predictive validity, we also had an opportunity to cross validate the continued efficacy of the selection procedures with 264 pilots hired in 2007, a decade after the initial implementation of these procedures. The pass rate on the interview and tests for the 2007 period of hiring was approximately two candidates out of three.

Procedures

For the long term validation study, we followed the ’96 cohort from their new hire training through 2008, tracking training episodes and productivity data over 12 years. Data were collected on the number of times pilots had a problem in any scored segment of training (indoc, initial, recurrent, transition, upgrade and requalification were all considered and each is made up of multiple segments). The complete set of data was available for 330 of the 339 pilots in the’96 cohort. A problem in training was defined as a recorded score of “incomplete,” “not recommended” or “unsatisfactory” on a training segment. It should be noted that such scores do not mean a pilot is immediately disqualified. Instead, the initial follow-up steps are remedial review, re-training, and retesting. Given the significant resources devoted to pilot training, however, such follow-up can be quite costly for the airline. Since the pilots had different numbers of training experiences depending on their career progression and opportunities, a training problem *rate* was also computed. Lost productivity data, measured by the number of scheduled rotation hours per pilot that were not flown, were also collected for each pilot over each of the years since hire.

For the cross validation study with the ’07 cohort, test scores and interview data were collected, as well as the cohort’s new hire training performance measures. In addition, biographical data were collected and hand coded from the applications of the ’96 cohort and the ’07 cohort.

Analysis and Results

To examine long-term predictive validity, the selection tests and background measures were correlated with training problems collected over the 12 years that elapsed from the time of hire. Regression models were produced, using the original components of the Select-in and Select-out models as predictors of training problems. Both models were found to predict performance in training (R-square=0.058). The sample was then divided into quintiles based on each pilot’s predicted training performance. In order to compare across careers that may have caused individual pilots to have more or fewer training episodes, a problem rate was calculated. The rate was calculated per thousand training episodes, since each quintile represents a collection of over 4000 training scored episodes. Table 1 shows the mean number of actual training problems over 12 years and the training problem rate over 12 years. As each group’s predicted performance level decreases, training problems increase, with the quintile with the lowest predicted performance demonstrating the highest level of training problems.

Table 1.

Actual training *outcomes by predicted performance group based on test performance, ’96 cohort*

Predicted Performance Category	Average training problems per pilot, 1996-2008	Problem rate per 1000 training episodes 1996-2007
1 Highest performers (N=66)	.17	3
2 (N=66)	.20	3
3 (N=66)	.35	5
4 (N=66)	.41	8
5 Lowest performers (N=66)	.52	9

The predictive model developed from the test scores of the’96 cohort was applied to the ’07 cohort. Five groups of predicted performance were again defined using the same boundaries established on the ’96 cohort, creating groups that were more equivalent in predicted performance, but not necessarily equivalent in number. Table 2 shows the mean number of training problems and the problem rate for the ’07 cohort in new hire training.

Table 2.

Actual new hire training outcomes by predicted performance group based on test performance, ’07 cohort

Predicted Performance Category	Average problems per pilot, new hire training, 2007	Problem rate, new hire training, 2007
1 Highest performers (N=93)	.07	7
2 (N=53)	.09	8
3 (N=57)	.07	6
4 (N=45)	.11	10
5 Lowest performers (N=30)	.10	9

While the trend of these results is in the desired direction, the effects are not as strong as we would like. This outcome is impacted by multiple factors. First, the N in the lowest predicted performance group has been reduced in comparison to the ’96 cohort. This is due in part to the fact that at the time of selection in 1996, the Select-in Model had not yet been set so some candidates were hired who would have been eliminated by this measure. Additionally, by 2007 a significant amount of content of the JKT had been leaked to candidates, greatly reducing the range and validity of those scores. This problem has since been remedied, fortifying the test against future compromise. Finally, in the decade that passed between our two study cohorts, the characteristics of the pool of available aviator applicants had changed somewhat, perhaps diminishing some of the predictive power of the original model.

In order to explore whether it was possible to enhance the predictive power of the model, background characteristics were collected from the pilot applications and evaluated for potential incorporation into the model. Regression equations were run that combined aspects of education and work history with the original Select-in variables and Select-out variables. The additions strengthened the model’s power for both the ’96 and the ’07 cohort (R-square=0.21). Table 3 and Table 4 show the new model’s training problem predictions for each cohort.

Table 3.

Actual training outcomes by predicted performance group based on new combined model, ’96 cohort

Predicted Performance Category	Average training problems per pilot, 1996-2008	Problem rate per 1000 training episodes, 1996-2008
1 Highest performers (N=66)	.12	2
2 (N=66)	.11	2
3 (N=66)	.23	3
4 (N=66)	.45	7
5 Lowest performers (N=66)	.73	14

Table 4 shows the results of the new model applied to the ’07 cohort, again using the group boundaries established on the ’96 cohort.

Table 4.

Actual new hire training outcomes by predicted performance group based on new model, '07 cohort

Predicted Performance Category		Average problem per pilot, new hire training, 2007	Problem rate, new hire training, 2007
1 Highest performers	(N=122)	.07	6
2 (	N= 63)	.11	11
3 (	N= 43)	.07	6
4 (	N= 33)	.09	8
5 Lowest performers	(N= 17)	.18	17

Results again demonstrate notably higher training problem rates in the groups predicted to be lower performers, providing cross validity evidence for the new model derived from the training experiences combined with background information of the 1996-97 sample. In addition, the new model can be used to predict loss of productivity, as measured by scheduled rotation hours not flown due to sick leave usage. Using the same predicted performance quintiles from the '96 cohort, Table 5 shows the lost productivity hours for each group.

Table 5.

Average lost productivity hours, 1996-2008, by predicted performance group based on new model, '96 cohort

Predicted Performance Category		Average lost productivity hours, 1996-2008
1 Highest performers	(N=66)	195
2 (	N=66)	232
3 (	N=66)	227
4 (	N=66)	223
5 Lowest performers	(N=66)	243

Discussion

This research demonstrates several valuable findings, particularly since it is one of very few studies of pilots at a large commercial airline. The criterion measure in the longitudinal study was based on multiple training evaluations over a twelve year period. Our predictive models were built using subtests from the NEO PI-R, the CogScreen, and a proprietary Job Knowledge Test. Other research has also found the NEO (Campbell 2010a, Campbell 2010b), and the CogScreen (King 1995, Taylor 2000) to be useful in assessments of pilots. Martinussen (1998) conducted a meta-analysis which found the best predictors of pilot performance to be instrument comprehension, mechanical principles, and aviation information. Similar content areas are reflected in the Job Knowledge Test that was built on a thorough analysis of the job of Delta pilots. The research reported here has confirmed that measures of the three major selection components can be used to predict those most likely to succeed at a large commercial air line and that the addition of education and background indicators enhances the predictive power of this model.

At Delta, this combination of factors appears to predict pilot success over more than a decade. In addition to long-term predictive validity, the data shows efficacy in identifying low performers in another cohort of pilots with somewhat different characteristics from the '96 cohort on which the model was built. Those hired in 2007 had much more experience in regional jets and less military experience. As a group they had more total and pilot-in-command hours, but went to less highly ranked colleges and universities. Their profiles on the NEO PR-I were also different, with lower scores on the Conscientiousness, Agreeableness and Extraversion Scales and higher scores on the Neuroticism scale. Finally, their CogScreen profiles indicated that they appeared to focus more on accuracy at the expense of throughput, a combination of speed and accuracy.

The fact that the predictive power of the original model was somewhat less powerful with the 2007 cohort illustrates several issues. Unlike the '96 cohort, the 2007 cohort was selected on the basis of **both** the Select-out and Select-in Models. As Hunter and Burke (1994) make clear, this process results in a restriction of range for the cohort, which is positive for choosing a higher proportion of good pilots, but results in observed validities that are likely to underestimate of real values. The efficacy of an enhanced model with the addition of background variables indicates the need for awareness that any selection model will require monitoring and adjustment over time to ensure sensitivity to changing characteristics of candidate pools and changing workplace conditions.

It should be kept in mind that the predictors in this study are sensitive to the job of pilot as designated by this employer's work and cockpit environment, philosophy, training programs and standards. Delta views the job of pilot as a decision-making position, requiring their pilots to actively engage in the management of their aircraft. The components of the predictive models and their parameters may vary from company to company. Variations may also be expected based on the populations that apply to each airline.

Finally, it is clear that effective selection processes can have a significant impact on operational costs by:

1. Making the selection process more efficient by (a) pre-screening that identifies those most likely to meet selection criteria and (b) focusing the selection process on the best predictors of performance on the job.
2. Reducing training costs for both initial and subsequent training.
3. Reducing the costs associated with lost productivity hours.
4. Hiring pilots that are more committed to their craft and performance of their job.

Acknowledgments

We would like to thank Delta Air Lines for their ongoing support and collaboration on this research.

References

Campbell, J.S., Castaneda, M., Pulos, S. (2010) (a) Meta-analysis of personality assessments as predictors of military aviation training success. *The International Journal of Aviation Psychology, 20,* 92-109

Campbell, J.S. (2010) (b) Five-factor model facet characteristics of non-aeronautically adaptable military aviators. *Aviation, Space, and Environmental Medicine, 81,* 864-868

Costa, P.T., McRae, R.R. (1992) Professional manual: Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO FFI) Psychological Assessment Resources, Inc. Odessa, Florida

Kay, G.G. (1995) Cog Screen: Professional Manual, Psychological Assessment Resources, Inc. Odessa, Florida

King, R.E., Flynn, C. F. (1995) Defining and measuring the "Right Stuff": Neuropsychiatrically Enhanced Flight Screening (N-EFS). *Aviation, Space, and Environmental Medicine, 66,* 951-956

Hoffmann, C.C., Hoffmann, K.P., Kay, G.G. (1998) The role that cognitive ability plays in CRM. Paper presented at NATO Symposium, Human Factors & Medicine Panel on Collaborative Crew Performance in Complex Operational Systems, Edinburgh, Scotland

Hunter, D.R., Burke, E.F., (1994) Predicting Pilot Aircraft Training Success: A Meta-Analysis of Published Research, *International Journal of Aviation Psychology, 4*, 297-313

Martinussen, M., Torjussen, T. (1998) Pilot selection in the Norwegian Air Force: A validation and meta-analysis of the test battery. *The International Journal of Aviation Psychology, 8,* 33-45

Taylor, J.L., O'Hara, R., Mumenthaler, M.S., Yeasavage, J.A. (2000) Relationship of CogScreen-AE to flight simulator performance and pilot age. *Aviation, Space, and Environmental Medicine, 71,* 373-380

USING PHYSIOLOGICAL MEASURES TO IMPROVE TRAINING FOR UAV OPERATORS

Jane H. Barrow[1], Carryl L. Baldwin[1], Daniel M. Roberts[1], Brian A. Taylor[1], Ciara Sibley[2], Joseph T. Coyne[2], Anna Mandulak[3], George Buzzell[1], and Nick Penaranda[1]

[1]George Mason University
Fairfax, VA

[2]Naval Research Laboratory
Washington, DC

[3]Strategic Analysis, Inc.
Arlington, VA

The current research examined use of frontal asymmetry measures of electroencephalography (EEG) activity to assess operator motivational intensity during training of Unmanned Aerial Vehicle (UAV) operators. Participants performed a series of Intelligence, Surveillance, and Reconnaissance (ISR) missions of graduated difficulty. Results demonstrate that frontal asymmetry in conjunction with behavioral measures provides a valuable tool for determining learner workload and motivational intensity. Implications for the use of frontal asymmetry metrics to drive real-time adaptive training are discussed.

Unmanned Aerial Vehicles (UAVs) are currently a key component in many military efforts, including perimeter security, surveillance, reconnaissance and in providing tactical flexibility (Parasuraman, Cosenzo, & De Visser, 2009). A particular area of interest to many researchers looking into UAV operations is that of adaptive automation – the idea that using a metric of operator workload can allow a computer system to automate elements of the UAV task to alleviate workload on the human operator (Wilson & Russell, 2007). A favored method for determining workload is using electroencephalography (EEG) since signals coming from the brain are much quicker to register than the behavioral response. Following this thread of research, our team has looked into using physiological measures like EEG to provide a basis for workload in a real-time adaptive training program.

Cognitive Load Theory (CLT) provides a theoretical basis for the use of a real-time adaptive training program. CLT postulates that keeping an individual at an optimum level of cognitive load will optimize their ability to learn the material at hand (Sweller, 1994). Thus, by using EEG to determine a learner's current level of workload, one could adapt the training to fit the individual's current ability to learn by speeding up training for an individual who is under-loaded or slowing training down (or repeating) for an individual who is overloaded. In addition to cognitive load, one also has to consider the cognitive-affective state of the learner. As used here, the cognitive-affective state is primarily determined by motivation intensity as defined by Brehm & Self (1989) and Wright (2008). Motivational intensity is essentially how much effort a person is "willing" to invest in a given task. It can be distinguished from the person's motivation potential (a relatively enduring trait) by its transient nature. As discussed by Brehm & Self, motivational intensity is a temporary level of motivation determined by the amount of effort a person will expend to satisfy a particular motive.

In general the amount of effort a person expends on a given task increases as the task becomes more difficult. However, if the person appraises the task as being so difficult that expending further effort is unlikely to result in success or, that rewards for achieving a higher level of performance are simply not worth the effort that would need to be expended, then motivation intensity (and effort expenditure) decreases. Fairclough and colleagues (in press) call the point at which a person switches from being willing to expend more effort to a strategy of avoidance or reduced effort expenditure, the "tipping point." This relationship is important to instructional design. Effective learning will take place when learners are challenged to expend effort to stay engaged in the learning task. However, care must be taken not to exceed the "tipping point" of the individual learner, the point where the learner withdraws or disengages from the learning task because it has become either too difficult or is no longer worth the effort.

Frontal asymmetry has gained increased attention over the past two decades for its potential to provide a relatively stable index of individual differences in emotional disorders and its ability to provide index of transient state-dependent types of emotional responding (Coan & Allen, 2004). It is the latter use of frontal asymmetry that is of interest here. Measures of frontal asymmetry may provide an index into the learners' current cognitive-affective approach to the task. In general, greater left hemispheric activation is associated with a an approach-related, goal directed action style characterized by a motivational approach to the task (Davidson, 2004). Conversely, relatively more right hemispheric activation is associated with negative emotions and avoidance response styles. Fairclough et al. (in press) observed greater left hemisphere asymmetry in a high load condition when an incentive was provided, relative to when no incentive was provided. This asymmetry effect for incentive was not observed in low or excessive load conditions. Rather, when no incentive was provided, left frontal asymmetry decreased in a step wise fashion from the low to high to excessive load conditions. This observation supports an interpretation that participants became increasingly disengaged as the task became more and more difficult. The incentive was able to offset the disengagement somewhat in the high load condition; but, incentive was not sufficient when the participant was faced with an excessively high difficulty level. Together, these results indicate that frontal asymmetry scores may be a strong predictor of the learner's motivational state. Frontal asymmetry used in conjunction with other metrics such as, frontal midline theta and parietal alpha may provide a sensitive index of both cognitive effort and cognitive-affective response to the task.

The current study used a part-task UAV simulation in which participants played the part of a mission controller of a Raven in manual mode. This type of task was chosen because it has a variable degree of workload, in which operators spend time looking for a target (low workload), and when the target is detected, they have to give the heading and identification of the target in a rapid manner (higher workload). The heading calculation and vehicle identification tasks were additionally varied from easy to high difficulty. Participants trained on this task over a period of time at each difficulty level such that by the end, they were able to calculate the heading of moving targets relative to the UAV's heading at 30° intervals and distinguish between six different military vehicles from an aerial perspective. It was hypothesized that as participants engaged in the task and learned to calculate the headings and make the identifications, frontal asymmetry would increase over time leading to greater left hemispheric activation as the tasks became more difficult and required further engagement from the participant.

Methodology

Participants

Fifteen participants (10 male) with a mean age of 23.29 years (*SD* = 4.43) participated in this study for partial course credit. All participants had normal or corrected-to-normal vision.

UAV Simulation Task

The UAV task screen was divided into four sections (see Figure 1). The largest portion of the screen displayed UAV flyover videos, created using Virtual Battlespace 2 (VBS2). Above the UAV flyover window was a box displaying the UAV's heading, in degrees. The upper right portion of the screen included a small rectangular box that indicated eye-tracking connectivity status (eye-tracking data and results are not discussed in this paper) and a box that prompted the participant for responses. In the lower right portion of the screen was an image of a compass, demarcated into 30 degree increments.

Prior to the beginning of the task, participants were trained on how to perform the calculation of vehicle heading by mathematically subtracting the target's apparent heading from the UAV's actual heading. Training for vehicle identification occurred at the beginning of each difficulty level and required the memorization of military vehicles by name and appearance. During training, participants were allotted 10 seconds per vehicle to learn the appearance of each via four separate image views (bird's eye, left side, right side, and front-on) presented in a 2x2 grid with the vehicle's name indicated above. Each trial began with the automatic initiation of a UAV flyover video. The simulated UAV flyovers had a viewing angle of 30 degrees between the camera and the ground and the UAV traveled at a speed of about 12.25m/s at an altitude of 150m. During each flyover video, a target vehicle would appear in the distance, with the UAV moving toward and eventually passing over the vehicle. The target vehicle was always moving at a constant speed in a constant direction. For each trial, participants responded by left-clicking the mouse in the UAV flyover window as soon as they

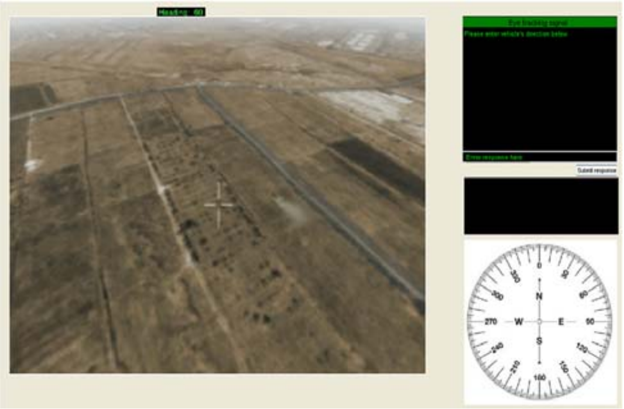


Figure 1. Graphical depiction of the UAV task screen.

located the target vehicle. Participants were then prompted to enter the absolute direction that the target vehicle was traveling (in degrees), which required a mental calculation to account for the UAV heading in relation to the target vehicle's apparent direction. Next, participants were asked to identify the vehicle by name from a multiple-choice list. After submitting this response, participants were then asked to provide a workload rating of mental effort for the heading calculation and vehicle identification on a scale from 1 to 7, with 1 representing the least amount of mental effort and a 7 the most. Participants were then provided feedback as to the correct target vehicle name and direction.

There were three levels of difficulty with 20 trials for each level. Each participant completed all three levels, beginning with the easiest and progressing through the most difficult. In the easiest difficulty level, participants were presented with two target vehicles to learn, and the UAV heading was always set to 0° (North). In medium difficulty level, participants were presented with four target vehicles (the two previous learned plus two more), and UAV headings varied in 90 degree increments (e.g. 0°, 90°, 180°, 270°). In hardest difficulty level, participants were presented with a total of six target vehicles (the four previously learned plus two more), and now UAV & target vehicle headings varied randomly between 0° and 330° degrees in 30°increments.

EEG Recording

EEG was recorded from 12 Ag/AgCl electrodes. Electrode placement followed the standard 10-20 system (Jasper, 1958). Data were recorded from sites Fz, F3, F4, Cz, C3, C4, and Pz. Eye blinks were monitored using electrodes placed above and below the orbit of the left eye to record vertical electro-oculogram (VEOG). A ground electrode was placed in the center of the forehead (10% of the nasion-inion distance, posterior to the nasion), and electrodes were placed at the left and right mastoid processes as reference points for the scalp electrodes. The online recording was referenced to the left mastoid. EEG and EOG were amplified using a NuAmps Neuroscan 40-channel amplifier and recorded using Neuroscan Scan 4.4. Impedance was maintained below 5 kΩ. Filtering was set to a bandpass of 0.1 Hz to 70 Hz. Data was recorded continuously at a rate of 500 Hz and stored on a computer hard disk drive for later analysis.

Experimental Procedure

Participants provided their informed consent and were administered both the Snellen (far) and Rosenbaum (near) eye tests for vision. Participants then completed two questionnaires to collect demographic information and to assess way-finding strategy, and viewed a PowerPoint presentation describing the task. The electrodes were applied to the head of the participant, and then the participant completed the task.

Results

Behavioral Data

Reaction time and accuracy data were analyzed using two separate repeated measures ANOVAs. The data were broken down by difficulty level (easy, medium, or difficult), task type (target search, heading calculation, and identification), and segment (first half or second half of each difficulty condition).

Response Time (RT). The reaction time data were analyzed using a 3 x 3 x 2 repeated measures ANOVA, with difficulty level, task type, and segment as the variables. Mauchly's Test of Sphericity revealed that the difficulty level variable was non-spherical, $\chi^2(2) = 6.224$, $p = .04$, so a Greenhouse-Geisser correction was used for all tests involving that variable. There was a main effect for difficulty level, $F(1.45, 20.28) = 6.027$, $p = .02$, such that RT was significantly faster in the easy condition ($M = 11.56$, $SE = .60$) than in the medium condition ($M = 13.14$, $SE = .54$) or the difficult condition ($M = 13.50$, $SE = .73$). There was also a main effect for task type, $F(2, 28) = 34.51$, $p = .001$, such that RT was significantly faster in the identification phase ($M = 4.80$, $SE = .33$), than in the target search phase ($M = 14.76$, $SE = 1.45$) or the heading calculation phase ($M = 18.64$, $SE = 1.26$). Finally, there was a main effect for segment, $F(1, 14) = 9.38$, $p = .01$, such that RT in the second half ($M = 12.22$, $SE = .50$) was significantly faster than in the first half ($M = 13.25$, $SE = .61$). There was a difficulty level x task type interaction, $F(2.376, 33.27) = 7.42$, $p = .001$, such that RT for heading calculation increased with task difficulty, but RT for target search decreased with task difficulty, and RT for identification peaked at the medium difficulty level but

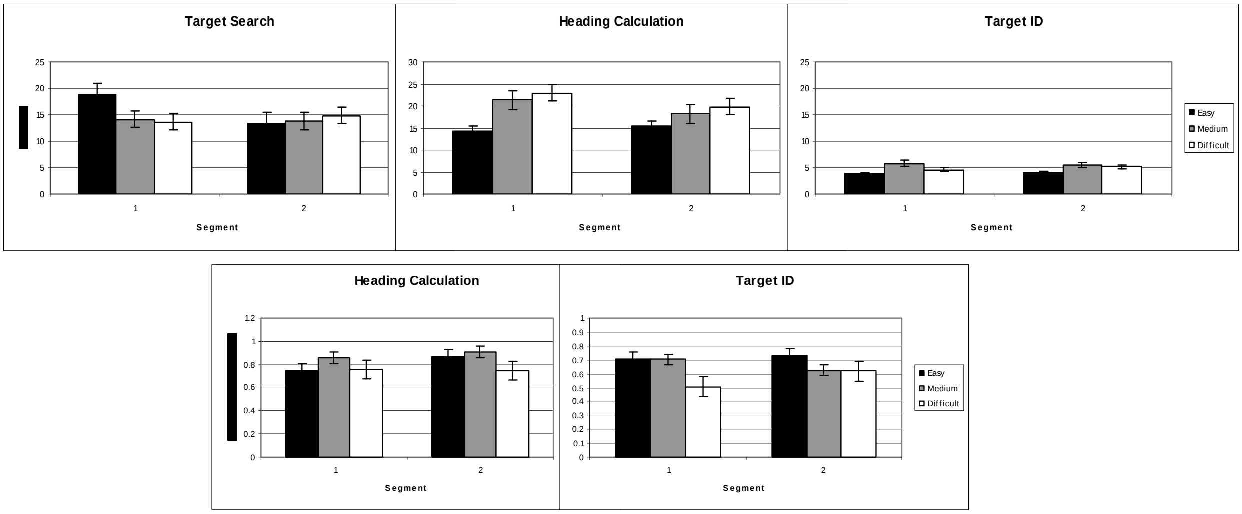


Figure 2. Top three panels show difficulty level as a function of segment (first half or second half) for the three tasks in terms of reaction time. The bottom two panels show difficulty level as a function of segment for heading calculation and target ID in terms of accuracy. Error bars the standard error of the mean.

decreased again in the difficult condition. There was also a three-way interaction, *F*(5, 56) = 7.06, *p* = .001, which is graphed in Figure 2.

Accuracy. The accuracy data were analyzed using a 3 x 2 x 2 repeated measures ANOVA, with difficulty level, task type (minus the target search task, since accuracy was 100%), and segment as the variables. There was a main effect for difficulty level, *F*(2, 28) = 4.98, *p* = .01, such that accuracy was significantly higher in the easy condition (*M* = .76, *SE* = .03) and the medium condition (*M* = .77, *SE* = .03) than in the difficult condition (*M* = .66, *SE* = .06). There was also a main effect for task type, *F*(1, 14) = 9.24, *p* = .01, such that accuracy was significantly higher in the heading calculation phase (*M* = .81, *SE* = .05), than in the identification phase (*M* = .65, *SE* = .04). Finally, there was a significant three-way interaction, *F*(2, 28) = 4.15, *p* = .03, which is also graphed in Figure 2.

EEG Data

The EEGLAB toolbox (Delorme & Makeig, 2004) in conjunction with MATLAB v.2010a (The MathWorks, Natick, MA) was used for analysis of the EEG recordings. The EEG of 3 participants was not included in the analysis due to missing data. EEG was high-pass filtered at 1 Hz to remove linear trends, and re-referenced to the average of the left and right mastoid reference points. For each recording, EEG occurring during three task periods of interest (target search, heading calculation, and identification) was selected. EEG from each of these periods was divided into 1 second, 50% overlapping epochs. Epochs that contained eye blink contamination, as defined as activity exceeding +/- 75 uV on either the upper or lower VEOG electrode, were rejected. Remaining epochs were Hamming windowed and decomposed into frequency spectra with a 512 point fast-fourier transform. The average dB power in each of four frequency bands (theta: 4.2-7.3 Hz, low alpha: 7.3-10.2 Hz, mid alpha: 9.2-11.2 Hz, and high alpha: 10.2-12.2 Hz) was identified for each difficulty condition for electrode sites F3 and F4. The average dB power of F3 was subtracted from F4 to provide the amount of frontal asymmetry. As such, higher numbers indicate greater left-hemispheric activity and lower numbers indicate greater right-hemispheric activity. Additionally, the data was analyzed in terms of the first half of the condition as compared to the last half, to reveal how frontal asymmetry changed over time being exposed to the task.

The frontal asymmetry data were analyzed for the mid-alpha frequency band using a 3 x 3 x 2 repeated measures ANOVA, with difficulty level, task type, and segment half as the variables. For the purposes of this paper, only the mid-alpha data band will be discussed. It is important to note than when discussing the alpha band, higher values indicate less activity – in a sense, when alpha values are high, the brain is idling. There was a main effect for difficulty level, *F*(2, 22) = 3.37, *p* = .05, such that frontal asymmetry demonstrated relatively less left hemispheric activity in the easy condition (*M* = .34, *SE* = .21) and more left hemisphere activity in the medium condition (*M* =

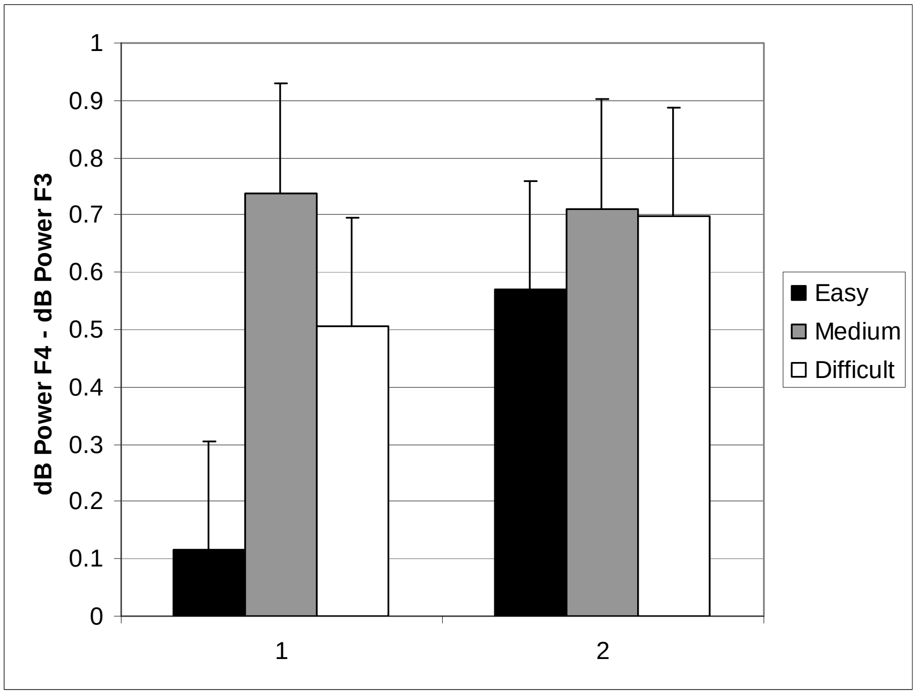


Figure 3. Frontal asymmetry for the first and second half of each difficulty condition for mid-alpha. Error bars are the standard error of the mean. Log power is calculated as

$$(10 \bullet \log_{10}\left(\frac{\mu V^2}{Hz}\right))$$

.72, *SE* = .18) and the difficult condition (*M* = .60, *SE* = .21). This finding replicated previous results indicating greater left hemispheric activity as task difficulty increased. There was also a main effect for task type, *F*(2, 22) = 3.12, *p* = .06, such that there was relatively less left hemisphere activity for the target search (*M* = .49, *SE* = .16) and heading calculation (*M* = .41, *SE* = .19) tasks and greater activity in the left hemisphere for the identification task (*M* = .80, *SE* = .25). Finally, there was an interesting difficulty level x segment interaction, *F*(2, 22) = 4.63, *p* = .02, where left hemispheric activity increased as the task difficulty went up, but in the easy and difficult conditions, there was a large increase in left hemispheric activity in the second half, whereas activity remained approximately the same in the first and second halves of the medium difficulty condition (see Figure 3).

Discussion

The results of this study reveal that as difficulty of the tasks increased, reaction time became slower, accuracy decreased, and there was increasing left hemispheric activity in the brains of participants. This indicates a relatively close coupling between behavioral and frontal asymmetry metrics. Specifically, increases in left hemisphere activity on the frontal asymmetry index were generally associated with more accurate performance. Frontal asymmetry appears to provide an index of a goal-oriented motivational approach. In the current investigation, the difficulty level never appeared to exceed the “tipping point” described by Fairclough and colleages (in press). When broken down by task type, it appears that the heading calculation task was easier than the identification task according to accuracy, though reaction time measures indicate that there may have been a speed-accuracy trade off, such that participants responded very quickly to the identification task, but with poor accuracy. The frontal asymmetry results demonstrate greater left hemispheric activity during the identification task, indicating that the participants continued to try during this task, even with poorer behavioral results. There was relatively less left hemispheric activity for the heading calculation relative to the identification task, indicating that participants found this task somewhat easier. The frontal asymmetry interaction for difficulty level x segment is more difficult to explain. It appears that in the medium difficulty level, participants maintained a motivational approach, as indicated by relatively greater left hemisphere activity. In the easy condition, greater left hemisphere activation in segment two was associated with increased performance in the heading calculation task and maintenance of performance in the vehicle ID task. In the most difficult level, this relationship between increased left hemisphere activation and increased performance may have only been present for the most difficult task – the identification task. It would seem that in the medium difficulty level, participants were trying hard on the task throughout, but in the easy and difficulty conditions they began to try harder as the condition progressed. It is also possible that in the easy condition, which was always presented first, participants found it difficult at first and therefore frustrating (explaining the relatively greater right hemispheric activity during the first half) but with more time became more proficient at the task (explaining greater left hemispheric activity during the second half). In the difficult condition, a similar effect could have taken place, though it is not clear as to why the medium difficult condition would not also show this effect.

The results of this study support the use of frontal asymmetry as a metric for determining learner state during training using EEG. In a real-time adaptive training program, it could provide a means of determining when the learner is engaged independent of performance metrics. This is critical as poor performance could result from either overload and frustration or underload or boredom. In conjunction with spectral analyses, which have been shown to successfully distinguish between different types of working memory usage (Roberts, et al, 2010), it could provide a more accurate assessment of the workload and motivational state of an individual learner which could be used to drive adaptive aiding in a training platform.

References

Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology, 40*, 109-131.

Coan, J. A., & Allen, J. J. B. (2004). Frontal EEG asymmetry as a moderator and mediator of emotion. *Biological Psychology, 67*(1-2), 7-50.

Davidson, R. J. (2004). What does the prefrontal cortex "do" in affect: perspectives on frontal EEG asymmetry research. *Biological Psychology, 67*(1-2), 219-234.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21.

Fairclough, S. H., Gilleade, K., Ewing, K. C., & Roberts, J. (in press). Capturing user engagement via psychophysiology: measures and mechanisms for biocybernetic adaptation. *International Journal of Autonomous and Adaptive Communications Systems*.

Jasper, H.V. (1958). The ten twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10*, 370-375.

Parasuraman, R., Cosenzo, K. A., & De Visser, E. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology, 21*(2), 270-297.

Roberts, D.M., Taylor, B.A., Barrow, J.H., Robertson, G., Buzzell, G., Sibley, C., Cole, A., Coyne, J.T., & Baldwin, C.L. (2010). EEG spectral analysis of workload for a part-task UAV simulation. *Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA: HFES.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295-312.

Wilson, G.F. & Russell, C.A. (2007). Performance enhancement in an uninhabited air vehicle task using physiologically determined adaptive aiding. *Human Factors, 46*(6), 1005-1018.

Wright, R. A. (2008). Refining the prediction of effort: Brehm's distinction between potential motivation and motivation intensity. *Social and Personality Psychology Compass, 2*(2), 682-701.

TEAM VIGILANCE: THE EFFECTS OF CO-ACTION ON WORKLOAD AND STRESS

Andre Garcia, Carryl Baldwin
George Mason University, Fairfax, Virginia

Matthew Funke, Gregory Funke, Victor Finomore, Benjamin A. Knott, Joel S. Warm
Air Force Research Laboratory Wright-Patterson AFB, Dayton, Ohio

Operator vigilance is a vital concern to the Air Force in regard to cockpit monitoring, air-traffic control, and the supervisory control of unmanned aerial vehicles. A key interest is the performance of teams of observers because of the reliance of military operations on good teamwork. Previous literature has examined the efficacy of team vigilance performance by comparing the frequency of target detections by teams in comparison to those obtained by operators working alone. Team performance has consistently exceeded single-operator performance. The present study replicates this effect and provides the initial experimental investigation of the cost of being a team member. Results indicated that team members worked harder but reported less distress than single operators in the performance of a simulated UAV monitoring task.

Vigilance, or sustained attention, refers to the ability of observers to maintain their focus of attention and to detect infrequent and unpredictable targets over prolonged periods of time (Davies & Parasuraman, 1982). The ability of observers to sustain attention and detect these transient signals is of substantial concern to human factors and ergonomic specialists within the Air Force because of the vital role that vigilance plays with regard to enemy surveillance, cockpit monitoring, air-traffic control, and the supervisory control of unmanned aerial vehicles (Warm, Parasuraman, & Matthews, 2008). Accordingly, the Air Force is engaged in studies to further understand the factors that influence vigilance performance and to evaluate the effectiveness of operators who are engaged in vigilance tasks.

Traditionally, vigilance tasks have been considered as tedious but benign assignments that place little demand on operators, and the decrement function, the decline in efficiency over time that typifies performance in vigilance tasks (Davies & Parasuraman, 1982), has been viewed as resulting from task underload and consequent under arousal (Warm et al., 2008). More recent studies have indicated that while they are tedious, vigilance tasks impose a substantial demand upon the information-processing resources of observers and are highly stressful (Warm et al., 2008).

Neurophysiological evidence of high mental workload in vigilance comes from studies examining brain activity in observers using electroencephalography (EEG; e.g., Gevins & Smith, 2007). EEG research suggests that activity in the 4-7 Hz range, known as theta band activity, reflects extant mental work, and more specifically, that theta activity in the frontal midline region varies directly with task demand. In the vigilance domain, a recent experiment by Berka and colleagues (2007) confirms that theta activity increases during performance of a demanding vigilance task, a result that is consistent with other research that links increases in theta activity with increases in mental workload (e.g., Gevins & Smith, 2007)

The stress associated with vigilance task performance has been extensively investigated using the Dundee Stress State questionnaire (DSSQ; Matthews et al., 2002), a multidimensional scale that measures stress experienced in terms of affect, motivation, and cognition. Studies with the DSSQ indicate that participation in vigilance tasks leads to loss of task engagement and increased feelings of distress (Warm, Matthews, & Finomore, 2008).

Of additional interest to the Air Force is the performance of teams of operators because of the reliance of military functions on teamwork for success. Researchers in this area have examined the role of teams in vigilance performance by comparing the frequency of target detections by teams in comparison to those obtained by operators working alone. In most of these studies, if a target was detected by any member of the team, the team received credit for the correct detection or "hit." In terms of correct detections, teams of operators have consistently outperformed their single operator counterparts (Bergum & Lehr, 1962; Hornseth & Davis, 1967; Klinger, 1969; Morgan & Alluisi, 1965; Morrissette, Hornseth, & Shellar, 1975; Pollack & Madans, 1964; Wiener, 1964). However, these studies of team performance have focused solely on performance efficiency and have not examined the costs associated with being a member of a team.

Does working on a team affect the degree of workload and stress associated with vigilance performance? The phenomenon of "social loafing" in which operators exert less effort because they trust their associates to support them would lead to the expectation that being on a team would lower operator workload and stress in comparison to working singly. On the other hand, the phenomenon of "social comparison" would lead to the opposite expectation, because underperforming as a member of a team would make an operator look less competent than her/his associates, and consequently might elevate levels of workload and stress. The goal of this study was to use EEG theta activity and the DSSQ to examine the workload and stress associated with performing a vigilance task as a co-operator relative to performing individually.

Method

Participants assumed the role of either a single UAV controller or a member of a dyadic team of UAV controllers. Participants were assigned at random to the single-operator or co-operator conditions. They were instructed to monitor the clockwise or counterclockwise flight pattern of four UAVs on a simulated air traffic control display. The display was divided into four 90° quadrants, each containing one UAV icon. The task of the controller was to look for cases in which two of the UAVs were on a collision path (the critical signal for detection). In either condition, both the clockwise and counterclockwise flight path directions appeared in a random manner throughout the vigil so that a UAV that was at fault in one flight direction was not at fault in the other. Examples of critical and neutral signals can be seen below in Figure 1.

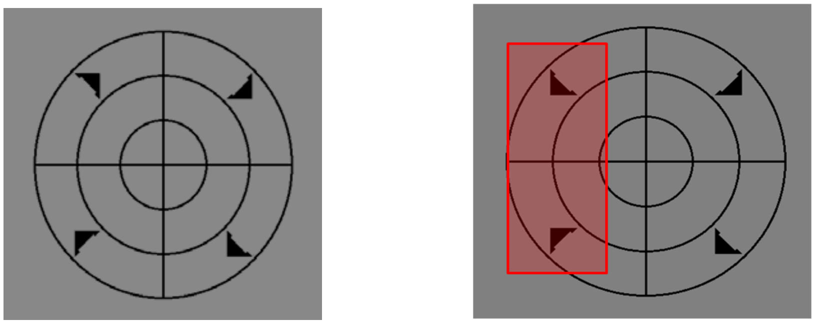


Figure 1. Examples of neutral events and critical signals in the flight path display.

In the single-operator condition, participants performed the vigilance task alone, and were solely responsible for identifying the potential collisions between the UAVs. The co-operator condition used same sex dyads to perform the task. The dyads performed the vigilance task together in an 8 foot × 6 foot room. Although separated by an opaque divider, participants were aware of each other's presence, and were informed that they would be performing the same task. However, they were instructed not to communicate, collaborate, or strategize with each other, as previous research indicates that team communication could negatively influence team performance by distracting team members from the task (e.g., Bergum & Lehr, 1962). Apart from the direction not to communicate with the other member of the dyad, the co-operators were given identical instructions regarding task mechanics to that of the observers in the individual condition.

Fifteen observers (8 women and 7 men) were assigned to the single-operator condition, while 28 observers (14 men and 14 women) were paired to form the co-operator dyads. All participants served in a 40-minute vigil divided into 4 continuous 10-minute periods of watch. In both conditions, the display was updated 30 times/minute with a dwell time of 1000 msec. Sixteen critical signals occurred during each period of watch (four in each display quadrant, two clockwise and two counterclockwise). In both conditions, participants responded by pressing the spacebar on a computer keyboard. In the single-operator condition, participants were credited with a correct detection if they executed a key-press response in the presence of a critical signal, and were charged with an error of commission (i.e., a false alarm) if they made a key-press response to a neutral event. In the co-operator condition, the *dyadic team* was credited with a correct detection if either member of the dyad detected the target correctly, and the *dyadic team* was charged with a commission error if either member made an inappropriate detection response to a neutral event. A CleveMed 8-channel bio-radio was used to record theta activity from sites F3, Fz, Cz, and Pz, as activity at these sites has previously been linked to mental processing and workload (e.g., Gevins & Smith, 2007). Task induced stress was measured by the DSSQ, which was administered prior to and at the conclusion of the vigil.

Results

Performance efficiency. Mean percentages of correct detections in the single-operator and co-operator task conditions are plotted as a function of periods of watch in Figure 2.

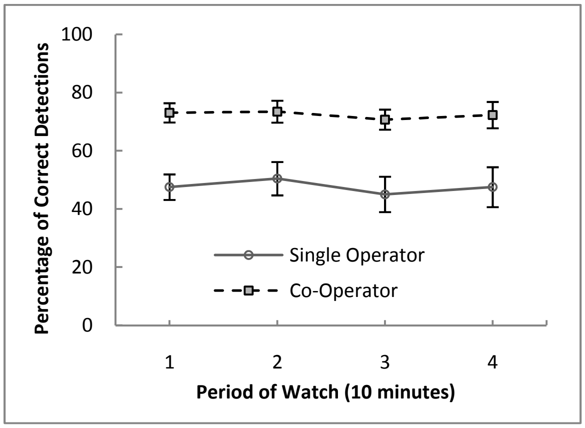


Figure 2. Mean percentages of correct detections in single- and co-operator conditions by periods of watch. Error bars are standard errors.

It is evident in the figure that performance efficiency was greater in the co-operator condition than in the single-operator condition, and that the frequency of signal detections appeared to remain stable over time. These impressions were confirmed by a 2 (conditions) × 4 (periods of watch) mixed-model analysis of variance (ANOVA) of the arcsines of the percentage of correct detection scores (Kirk, 1995), which revealed a significant main effect for conditions, $F(1, 27) = 12.91$, $p < .05$, but not for periods of watch, $F(2.79,75.31) = .71$, $p > .05$. The interaction between these factors was not significant ($p > .05$). In this and all subsequent ANOVAs, the Box correction (Maxwell & Delaney, 2003) was used when appropriate to correct for violations of the sphericity assumption.

An examination of the false alarm scores revealed that errors of commission were rare in this study (i.e., less than 1% of responses). Consequently, these data were not examined further.

Theta activity. Mean recorded theta activity in the co-operator and single-operator conditions by period of watch are represented in Figure 4, below. For purposes of this figure, theta activity data were aggregated across recording sites.

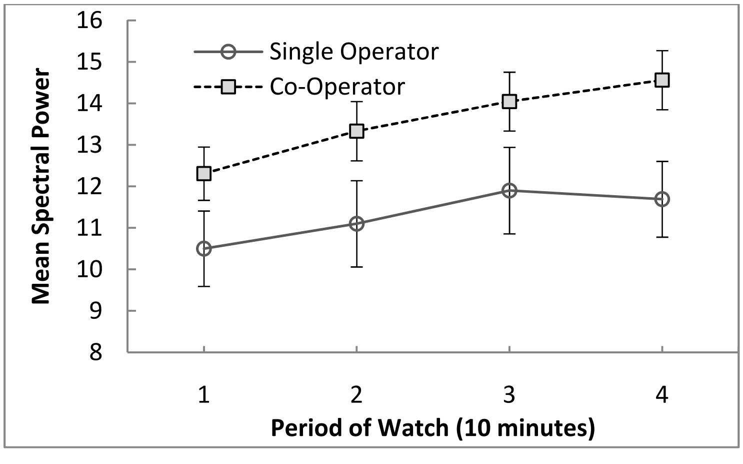


Figure 4. Mean spectral power in the theta band (4-7 Hz), aggregated across recording sites, in the single- and co-operator conditions for each period of watch. Error bars are standard errors.

To analyze theta activity in operators, values were first standardized (*z*-scored) within each individual across monitored sites. A 2 (conditions) × 4 (periods of watch) × 4 (recording site) mixed-model ANOVA revealed a main effect for condition, *F*(1, 41) = 8.23, $p < .05$, and main effects for period, *F*(2.45, 100.30) = 13.89, $p < .05$, and site, *F*(1.89, 77.34) = 228.86, $p < .05$, along with an interaction between period and site, *F*(4.02, 164.73) = 4.73, $p < .05$. Bonferroni corrected *t*-tests revealed that theta activity increased over time at all sites, and that the greatest activity occurred at sites F3 and Fz. Subsequent testing revealed that while theta levels increased over time at all sites, theta levels in these sites were greater than their counterparts by the final period of the vigil. In addition, as is illustrated in Figure 4 above, overall theta levels were greater in the co-operator condition than in the single-operator condition, and theta activity increased over time. All other sources of variance in this analysis were not significant, $p > .05$. In this analysis and the analysis of the DSSQ data to follow, both members of the co-operator dyads were included. Consequently, the co-operator condition had twice as many subjects as the single operator condition. A type III sum of squares was utilized to compensate for the unequal *N* (Field, 2009).

DSSQ stress state. For all observers in the co-operator and single-operator task conditions, pre- and post-vigil DSSQ scores for the worry, task engagement, and distress factors of the DSSQ were standardized against a large normative group with a mean of zero and a standard deviation of one (Matthews et al., 2002). Task-related difference scores were obtained by subtracting the pre-task score from the post-task score. Separate one-way ANOVAs were then computed for each of the three DSSQ factors. A statistically significant difference between conditions was found for the distress dimension, *F*(1, 41) = 9.43, $p < .05$. Analysis of the data for the worry and engagement dimensions revealed no statistically significant differences between conditions, $p > .05$. Mean standardized difference scores (change scores) for all combinations of task condition, period of watch, and DSSQ factors are represented graphically in Figure 3. As is evident in the figure, observers showed little post-vigil change in worry, but they reported themselves as being less task-engaged and more distressed after the vigil than before its start. It is also clear in the figure that participants in the single-operator condition reported a far greater increase in distress (more than 1 standard deviation) than did participants in the co-operator condition.

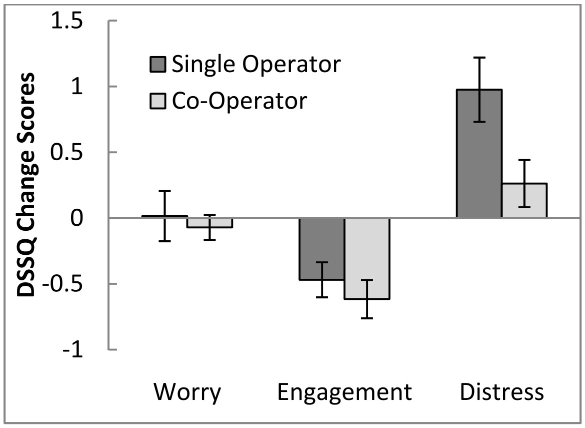


Figure 3. Mean DSSQ change scores for all factors by condition. Error bars are standard errors.

Discussion

As in several previous vigilance studies, co-operator performance in terms of correct detections exceeded that of operators working alone (Bergum & Lehr, 1962; Hornseth & Davis, 1969; Klinger, 1969; Morgan & Alluisi, 1965; Morrissette, Hornseth, & Shellar, 1975; Pollack & Madans, 1964; Wiener, 1964). As noted above, however, those earlier studies made no attempt to determine the cost to observers in terms of workload and stress associated with team membership. The purpose of the present study was to fill that gap in the tapestry of team performance in vigilance. A “social loafing” model led to the expectation that because of task dynamics both workload and stress would be less in the co-operator than in the single-operator condition. Conversely, a “social facilitation” model led to the expectation that workload and stress would be greater in the co-operator condition compared to single operators. The results supported neither model completely.

Consistent with expectations about workload derived from the “social facilitation” model, activity recorded within the theta band signified that participants in the co-operator condition exhibited higher levels of activity, indicating that they devoted more cognitive resources to maintaining their performance levels, and experienced greater mental workload in consequence. This may have been due to a sense of competitiveness, feelings of responsibility to the team, or general feelings of motivation associated with working as a member of a dyad.

With regard to stress, observers in both task conditions demonstrated a loss of task engagement over time, a result that is characteristic of previous vigilance studies using the DSSQ (Warm et al., 2008). However, in addition to the loss of task engagement, observers in the single-operator condition indicated a greater increase in distress after participating in the vigil than did those in the co-operator condition. Rather than accounting for this effect in terms of "social loafing" it is more likely that, given the higher neurophysiological workload scores observed, participants in the co-operator condition experienced less distress because of their knowledge of the "safety net" provided by a teammate.

Overall, these results suggest that military operations which require long periods of sustained attention from operators, such as in UAV surveillance, could be substantially benefitted in terms of increased task performance by the adoption of dyadic teams of operators, and that such benefit will not come at the cost of increased stress to the operators involved. This finding is valuable as increased stress would likely negate the utility of adopting dyadic teams because of the negative effects that stress exerts on the wellbeing and eventual performance of team members (Matthews et al., 2002).

References

Bergum, B.O., Lehr, D.J. (1962). Vigilance performance as a function of paired monitoring. *Journal of Applied Psychology, 46*, 341-343.

Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic,, V.T., Olmstead, R.E., Tremoulet, P.D., & Craven, P.L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine, 78*, B231-B244.

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

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.).Washington, DC: Sage.

Gevins, A., & Smith, M.E. (2007). Electroencephalography (EEG) in neuroergonomics. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The brain at work* (pp. 15-31). New York, NY: Oxford University Press.

Hornseth, J.P., & Davis, J.H. (1967). Individual and two-man team target finding performance. *Human Factors, 41*, 39-43.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove: Brooks/Cole.

Klinger, E. (1969). Feedback effects of social facilitation of vigilance performance: Mere coactions versus potential evaluation. *Psychonomic Science, 14*, 161-162.

Matthews, G., Campbell, S.E., Falconer, S. Joyner, L., Huggins, J., Gilliland, K., Grier, R., & Warm, J.S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress and worry. *Emotion*, *2*, 315-340.

Maxwell, S.E., & Delaney, H.D. (2003). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). New York: Routledge Academic.

Morgan, B.B., & Alluisi, E.A., (1965). On the inferred independence of paired watchkeepers. *Psychonomic Science, 2*, 161-162.

Morrisette, J.O., Hornseth, J.P., & Shellar, K. (1975). Team organization and monitoring performance. *Human Factors, 17*, 296-300.

Pollack, I., & Madans, A.B., (1964). On the performance of a combination of detectors. *Human Factors, 6*, 523-532.

Warm, J.S., Matthews, G., & Finomore, V.S. (2008). Vigilance, workload, and stress. In P.A. Hancock & J.L. Szalma (Eds.), *Performance under stress* (pp. 115-141). Burlington, VT: Ashgate.

Warm, J.S., Parasuraman, R., and Matthews, G. (2008).Vigilance requires hard mental work and is stressful. *Human Factors, 50*, 433–441.

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

Wiener, E.L. (1964). The performance of multi-man monitoring teams. *Human Factors, 6*, 179-184.

THE SART TASK DOES NOT PROMOTE MINDLESSNESS IN VIGILANCE PERFORMANCE

Michael Dillard, David B. Boles

University of Alabama

Matthew Funke, Gregory Funke, Victor Finomore, Allen Dukes, Joel S. Warm, Benjamin A. Knott

Air Force Research Laboratory Wright-Patterson AFB

Gerald Matthews

University of Cincinnati

Raja Parasuraman

George Mason University

Vigilance tasks typically require observers to respond to critical signals on their monitored displays and withhold responding to neutral events. The Sustained Attention to Response Task (SART) features the opposite response requirements which supposedly lead it to promote a mindless, non-thoughtful approach to the vigilance task that lacks attentional focus. To test that possibility, this study compared the SART to the standard vigilance task in terms of perceived mental workload – indexed by the Multiple Resource Questionnaire (MRQ) – and eye tracking activity – reflected via the Nearest Neighbor Index (NNI) – in the performance of a simulated air-traffic control assignment. Observers with both types of tasks identified a subset of identical MRQ dimensions as being highly involved in their monitoring assignment. The NNI scores indicated that observers with both types of tasks experienced higher workload than controls who viewed the display without a work imperative. Evidently, the SART does not promote mindlessness in vigilance performance.

Vigilance or sustained attention tasks require observers to monitor displays for extended periods of time and detect the appearance of critical signals. The signals, which occur infrequently, are embedded in a background of neutral or non-signal events. Observers are typically instructed to make an overt response, such as a button press, to the critical signals and to make no response to the more frequent neutral events. Thus, vigilance tasks can be described as "go/no-go" attentional assignments in which the frequency of "no-go" events outweighs that of "go" events. These assignments are of interest to the aviation community because of the critical role that vigilance plays in military surveillance, supervisory control of unmanned systems, air traffic control, and airport and border security (Vidulich, Wickens, Tsang, & Flach, 2010; Warm, Parasuraman, & Matthews, 2008).

At present, there are two competing models to account for failure of signal detection in vigilance tasks. One of these is the resource model in which the need to make continuous signal/noise discriminations is held to deplete observers' information-processing assets over time, leading to missed signals (Davies & Parasuraman, 1982). As described by Warm et al. (2008), support for the resource model comes from studies indicating that vigilance tasks impose a substantial mental burden on observers as reflected in high scores on the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988), a major instrument for measuring the perceived mental workload associated with performing a task (Wickens & Hollands, 2000), from studies showing that vigilance performance is poorer in tasks that require the use of working memory to distinguish signals from non-signals than in tasks in which signal detection does not involve a working memory component, from neuroimaging studies of resource demand using Transcranial Doppler sonography, and from investigations featuring physiological and subjective report measures indicating that vigilance tasks induce stress in observers that is linked to task demand.

An alternative view of detection failures in vigilance is the mindlessness model proposed by Robertson and associates (Manly, Robertson, Galloway, & Hawkins, 1999; Robertson, Manly, Andrade, Baddley, & Yiend, 1997). The model was prompted by the suggestion that when confronted with repetitive tasks in which signals are separated by long intervals, as in the case of vigilance, a supervisory attention system loses its potency and observers cease to focus their awareness on the task at hand (Shallice, 1988). With this in mind, Robertson and colleagues have asserted that the repetitive nature of vigilance tasks leads to a mindless lack of attentional focus and thence to failures of signal detection. Support for the mindlessness model comes from studies using the Sustained Attention to Response Task (SART; Robertson et al., 1997) which was designed to promote mindlessness in vigilance by inverting the "go/no-go" ratio. With this task, observers are asked to respond to the more frequent neutral events and to withhold responding in the presence the less frequent critical signals. In support of the mindlessness model, research with the SART has shown that failures to detect signals are proceeded by periods of increased routinization and decreased effort, and by findings that absent minded observers do more poorly than non-absent minded observers (Langer, Willmes, Chatterjee, Eickhoff, & Sturm, 2010).

A key issue in dealing with the competing resource and mindlessness models of detection failures in vigilance is the validity of the SART task as a means for promoting mindlessness. That validity has been questioned by Grier and her associates (Grier et al., 2003) who have reported that, as is the case with the traditional vigilance task format (TVF), perceived workload on the SART falls at the upper range of the NASA-TLX, and by Helton and colleagues (Helton et al., 2005) who have shown that with both types of task formats, observers are able to detect subtle patterns in the temporal structure of critical signal appearances. High workload and the detection of subtle changes in task elements do not seem to be consistent with the mindlessness perspective. The present study was designed to investigate the SART-validity issue further in terms of workload and eye scanning behavior. Toward that end, the study featured a vigilance task that required spatial discriminations and the need to visually explore the display in order to locate critical signals (see Figure 1 below).

Recently, Boles and his associates (Boles, Bursk, Phillips, & Perdelwitz, 2007) have introduced a new workload scale, the Multiple Resources Questionnaire (MRQ), which characterizes workload with respect to multiple mental processes based upon a combination of dimensions drawn from Wickens' multiple resource theory (Wickens & Holands, 2000) and factor-analytic studies carried out by Boles and colleagues. The instrument consists of the 17 resource dimensions listed in Table 1. Fifteen of the dimensions reflect encoding and central processing resources; the remaining two are response resources. Using a scale of 0 (no usage) to 100 (extreme usage; Finomore et al., 2008) observers are asked to rate the extent to which a task they just performed utilized each dimension. Research with the MRQ has shown that the instrument is able to uncover different key resource dimensions in tasks involving dissimilar skills such as reading bar graphs, determining the spatial position of a line, word interpretation, medical imaging, and of critical importance for the present study, vigilance (Boles et al., 2007; Finomore et al., 2008). If the SART does indeed promote a mindless, non-thoughtful approach to vigilance performance, one would anticipate that it would engage a more limited subset of resources and employ them at significantly lower level than the standard vigilance format. One goal for present study was to test these possibilities.

Eye-tracking has been used as a tool in aviation psychology since the field's earliest days and a considerable amount of data is available to indicate a close coupling between eye-movements and attention (McCarley & Kramer, 2007; Wright & Ward, 2008). Oculomotoractivity offers an additional medium to assess the degree to which the SART promotes mindlessness in vigilance performance. A recently developed eye-tracking metric known as the Nearest Neighbor Index (NNI) measures the spatial dispersion produced by a pattern of fixations, or more specifically, the ratio of the average minimum distance between observed fixations to the average distance between a hypothetical set of randomly distributed points. Previous research has demonstrated the sensitivity of NNI values to variations in task difficulty such that demanding tasks led to NNI values approaching 1 (i.e., wider fixation distributions), and less demanding tasks led to values approaching zero (i.e., clustered fixation distributions; Di Nocera, Terenzi, & Camilli, 2006). To the extent that the SART promotes mindlessness, one would anticipate less task demand, and therefore lower NNI values compared to the TVF. Moreover, the SART task should result in an NNI that is similar to a control condition with no task imperative.

Method

Observers assumed the role of air traffic controllers monitoring the flight pattern of a squadron of jet fighters on a circular display divided into four quadrants. Within each quadrant was a triangular jet icon. In all conditions, the jet fighter icons would appear to travel on either a clockwise or counterclockwise course (defined by the noses of the plane) throughout the vigil. In the TVF and SART formats, the task of the observers was to look for cases in which one of the jet icons appeared to be flying in an opposite direction relative to the other three aircraft.

Thirty right handed observers (15 men and 15 women) were assigned at random to each format condition. An additional 15 observers served as passive controls who viewed the flight display without an information-processing imperative, and were instructed to simply gaze at the display until the session ended. All participants served in a 40-min vigil divided into four continuous 10-min periods of watch. In all conditions, the display was updated 30 times/min with a dwell time of 1000 msec. Observers in the TVF and SART conditions were allowed 1200 msec from the onset of the signal to indicate a response. Twelve critical signals occurred during each period of watch (three in each display quadrant). In the TVF, observers were instructed to press the spacebar in response to critical signals – which are illustrated in the image on the right in Figure 1 below – and to make no response to non-critical signals – illustrated in the image on the left in Figure 1 below. In the SART condition, opposite instructions were given; participants were directed to press the spacebar for every occurrence of a non-critical signal and to withhold a response upon the occurrence of a critical signal.

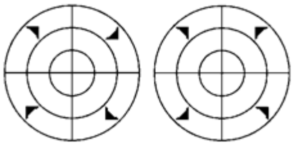


Figure 1. Examples of critical and non-critical signals in the flight path display (after Funke et al. 2010).

Ocular activity data were collected using a Seeing Machines Inc. faceLAB eye tracker, which recorded eye movements at a rate of 60 Hz. Using corneal reflectance to track the movement of the eyes, two desk-mounted infrared cameras recorded information regarding the location of observers' fixations on the visual display throughout the task. Subjective workload was measured using the MRQ immediately following the conclusion of the vigil.

Results

Performance Efficiency. Given the importance of veridical decisions in operational vigilance assignments, the diagnostic accuracy of observers' decisions about the presence or absence of critical signals is vital. Diagnosticity was measured in the present study in terms of Positive Predictive Power (PPP), the proportion of an observer's "signal present" responses that are actually correct and Negative Predictive Power (NPP), the proportion of an observer's "signal absent responses" that are actually correct. As described by Szalma et al. (2006), PPP= (number of correct detections)/(correct detections + false alarms) while NPP = (number of correct rejections)/ **(**correct rejections + misses). PPP and NPP scores of 1.0 indicate a perfectly accurate observer, scores of 0 indicate no correct decisions about signal presence/absence and no diagnosticity. Means and standard errors of the PPP and NPP scores for the TVF and the SART conditions are displayed in Figures 2 and 3, respectively. A 2 (conditions) × 4 (periods of watch) mixed-model ANOVA of the PPP scores showed a significant main effect for condition, $F(1,28) = 17.41$, $p < .05$, $\eta_p^2 = .38$. It is evident in Figure 2 that the mean PPP score in the TVF (89.94) was at the upper level of the PPP range while the mean in the SART format (67.44) was much lower. All other sources of variance in this analysis were not significant, $p > .05$. A similar ANOVA of the NPP scores revealed a significant main effect for periods of watch, $F(2.82, 78.83) = 3.70$, $p < .05$, $\eta_p^2 = .12$. It is evident in the figure that there was an overall temporal decline in the NPP scores but that the degree of decline was quite limited in magnitude. All other sources of variance in this analysis were not significant, $p > .05$. In these and all subsequent ANOVAs, the Box correction was employed to compensate for violations of the sphericity assumption (Field, 2009).

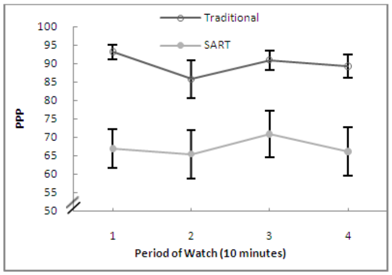


Figure 2. Mean positive predictive power of TVF and SART conditions as a function of periods of watch. Error bars are standard errors.

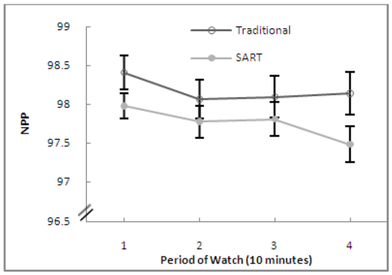


Figure 3. Mean negative predictive power of TVF and SART conditions as a function of periods of watch. Error bars are standard errors.

Workload. Means and standard deviations of the ratings for each of the 17 MRQ dimensions are presented for the TVF and SART format conditions in Table 1. In order to determine a resource profile for each format condition, it was necessary to identify the resources that met a "greater than zero usage" standard within each condition. Toward that end, Bonferroni corrected one-tail *t*-tests with alpha set at .05 were employed in each condition to determine the resource dimensions in which usage ratings were significantly greater than zero. The dimensions that met the usage standard in each format condition are starred in the table. It is evident in the table that the same eight dimensions met the usage standard in each condition. Moreover, the mean levels of engagement for the eight resources were 56.46 and 56.54, for the TVF and SART formats, respectively; values that were above the midpoint of the scale indicating a substantial level of workload. A 2 (conditions) × 8 (dimensions) mixed ANOVA did not reveal a significant main effect for condition or a conditions × dimension interaction, $p > .05$ in each case. In sum, the two format conditions showed identical multidimensional resource profiles and identical high levels of workload in performing the vigilance task.

Table 1. *Means and standard errors that met the inclusion criterion for the TVF and SART conditions.*

Dimensions	**Conditions**	
	TVF	**SART**
	Met Inclusion Criteria	
Auditory Emotional Process	-	-
Auditory Linguistic Process	-	-
Facial Figural Process	-	-
Facial Motive Process	-	-
Manual Process*	29.33 (7.14)	52.67 (8.53)
Short Term Memory Process*	46.00 (10.07)	38.33 (9.36)
Spatial Attentive Process*	91.33 (3.43)	81.00 (6.85)
Spatial Categorical Process*	76.33 (6.73)	83.33 (4.57)
Spatial Concentrative Process*	41.33 (10.64)	25.67 (7.59)
Spatial Emergent Process*	58.33 (9.43)	51.00 (9.07)
Spatial Positional Process*	59.67 (7.21)	57.67 (9.07)
Spatial Quantitative Process	-	-
Tactile Figural Process	-	-
Visual Lexical Process	-	-
Visual Phonetic Process	-	-
Visual Temporal Process*	49.33 (9.31)	62.67 (9.93)
Vocal Process	-	-

Ocular Activity. Fixation point data collected from the TVF, SART, and control conditions were used to calculate NNI scores across the four periods of watch for each format condition and the control condition. Means and standard errors of the NNI scores for each of the three conditions are plotted as a function of periods in Figure 4. It is evident in the figure that the NNI scores for the TVF and SART conditions were higher than those for the control condition, indicating a greater task demand in the TVF and SART conditions in comparison to the control. It is also evident in the figure that the scores for the TVF, SART, and control conditions showed a declining pattern of scanning over time. These impressions were confirmed by A 2 (conditions) × 4 (periods of watch) mixed-ANOVA of the data of Figure 4 which revealed significant main effects for conditions, $F(1, 42) = 6.59$, $p < .05$, $\eta_p^2 = .24$, and periods, $F(2.52, 105.93) = 9.98$, $p < .05$, $\eta_p^2 = .19$. Subsequent Bonferroni corrected *t*-tests with alpha set at .05 indicated that that the mean NNI scores in the TVF (.64) and SART (.68) conditions did not differ significantly from each other ($p > .05$) but they both were significantly greater than the mean for the control condition ($M = .54$), $p < .05$ in each case.

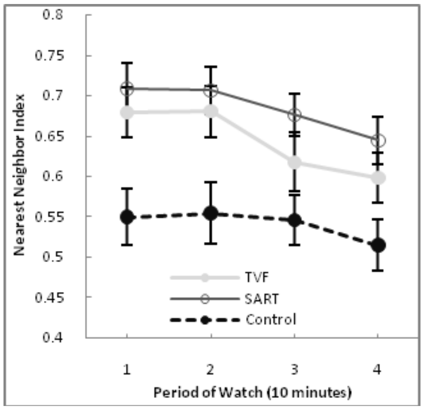


Figure 4. Mean NNI scores for the TVF, SART, and control conditions. Error bars are standard errors.

Discussion

As noted above, a key element in vigilance performance is the diagnostic accuracy of observers' decisions about the presence or absence of critical signals. In terms of signal presence, the PPP data showed that diagnostic acumen was significantly greater when observers performed the vigilance task in the TVF than in the SART format; signal present decisions were 90% accurate in the former condition but only 67% accurate in the latter. A result of this sort could be accounted for in terms of the view that the SART promotes a mindless lack of attentional focus in observers. However, the results of the MRQ workload and ocular workload measures indicate that a conclusion of that sort would be inappropriate.

The principal purpose for the use of the workload and ocular activity measures was to examine the validity of the notion that the SART fosters mindlessness. Toward that end, it was expected that the SART would engage a more limited set of resources on the MRQ than would the TVF and that it would employ them at a significantly lower level than that of the TVF. Neither of these expectations were borne out.

With respect to the MRQ, the two vigilance formats engaged identical ensembles of eight resource dimension many of which involved spatial elements befitting the need of observers to monitor a spatially dynamic visual display. In addition, the mean levels of engagement for the eight resources were above the midpoint of the MRQ scale for both vigilance formats confirming earlier findings by Grier et al. (2003) with the NASA-TLX that both formats induce a substantial level of workload in observers. These outcomes were complimented by the NNI results in which the pattern of eye-scanning was indicative of a higher level of workload in observers across the duration of the vigil than control observers who viewed the display without a work imperative. Rather than depicting the SART as promoting a mindless, non-thoughtful approach to vigilance performance, the results of the present study indicate that under both the TVF and the SART conditions, observers adopt a cognitively active approach in performing a vigilance task.

Given that the SART is not an engine for the promotion of mindlessness, what can account for the finding that PPP is significantly poorer when observers perform a vigilance task in the SART format as compared to the TVF? In seeking an account for this effect, it is helpful to keep four major points in mind. (1) The PPP index reflects the ratio of correct responses to the sum of correct responses and false alarms. Consequently, for any frequency of correct responses, increases in false alarms will suppress the observer's diagnosticity for the presence of critical signals. (2) Grier and her colleagues (2003) have shown that the SART is susceptible to a higher false alarm rate than the TVF, a result also observed in this study in which the mean false alarm rates for the SART and the TVF conditions were 1.15% and 0.26%, respectively. (3) In the case of the SART, a false alarm is defined as an error of omission, i.e., the failure to execute a motor response in the presence of a neutral or non- critical stimulus event. (4) In this study, observers in the SART format were required to make a motor response to neutral events that occurred frequently at the rate of 288 per 10-min period. With these points in mind, the argument advanced by Helton and his associates (Helton, Head, & Russell, in press) that errors of omission in the SART are due to loss of motor control in the form of tactical forced rest stops or "taking a breather" from the need for a high level of continuous responding becomes critical. Instead of lapses of attention, the poor diagnosticity in the SART condition was more likely the result of difficulty in continuously initiating motor responses to an arduous flow of neutral events with a consequent increase in the false alarm rate.

In sum, the results of this study challenge the validity of the proposition that detection failures in the SART emanate from a withdrawal of attentional effort. In so doing, they also challenge the viability of the mindlessness model of vigilance which draws its major support from research with the SART. In addition to implications for theories of vigilance, the present results also have potential meaning at an operational level. Advocates of a mindlessness account of detection failures in vigilance may advocate that steps be taken at the operational level to remove the factors that lead to loss of attentional focus. As Helton et al. (in press) have noted, to the extent to which mindlessness theory represents a misunderstanding of the cause of attentional lapses in vigilance performance, such steps could lead to the adoption of inappropriate solutions to the failure of signal detection in operational vigilance tasks.

References

Boles, D.B., Bursk, J.H., Phillips, J.B., & Perdelwitz, J.R. (2007). Predicting dual-task performance with the Multiple Resources Questionnaire (MRQ). *Human Factors*, *49*, 32-45.

Davies, D.R., & Parasuraman, R. (1982).*The psychology of vigilance.* New York: 1982.

Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A random glance to the flight deck: Pilot's scanning strategies and real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making,* 3*, 271-285.*

Di Nocera, F. Terenzi, M. & Camilli, M. (2006). Another look at scanpath: distance to nearest neighbor as a measure of mental workload. In D. de Waard, K.A. Brookhuis, and A. Toffetti (Eds.) (2006)*, Developments in Human Factors in Transportation, Design, and Evaluation* (pp.1-9). Maastricht, the Netherlands: Shaker Publishing.

Field, A. (2009). *Discovering statistics using SPSS (*3rd ed.). Los Angeles: Sage.

Finomore, V.S., Shaw, T.H., Warm, J.S., Matthews, G., Riley, M.A., Boles, D.B., & Weldon, D. (2008). Measuring the workload of sustained attention: Further evaluation of the Multiple Resources Questionnaire. *Proceedings of the Human Factors and Ergonomics Society, 52,* 1209-1213.

Funke, M., Warm, J.S., Matthews, G., Finomore, V., Vidulich, M., Knott, B., et al. (2010). Static and dynamic discriminations in vigilance: Effect on cerebral hemodynamics and workload. In T. Marek, W. Karwowski, & V. Rice (Eds.), *Advances in understanding human performance:Neuroergonomics,human factors design, and special populations* (pp. 80-90). Boca Raton, FL: CRC Press.

Grier, R.A., Warm, J.S., Dember, W.N., Matthews, G., Galinski, T.L., Szalma, J.A., & Parasuraman, R. (2003). The vigilance decrement reflects limitations in effortful attention, not Mindlessness. *Human Factors,* 45*,* 349-359.

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.

Helton, W.S., Head, J., & Russell, P.N. (in press). Reliable and unreliable warning cues in the Sustained Attention to Response Task (SART). *Experimental Brain Research.*

Helton, W.S., Hollander, T.D., Warm, J.S., Matthews, G., Dember, W.N., Wallart, M., Beauchamp, G., Parasuraman, R., & Hancock, P.A. (2005). Signal regularity and the mindlessness model of vigilance. *British Journal of Psychology*, *96¸* 249-261.

Langner, R., Willmes, K., Chatterjee, A., Eickhoff, S.B., & Sturm, W. (2010). Energetic effects of stimulus intensity on prolonged simple reaction-time performance. *Psychological Research,* 74*, 499-512.*

Manley, T., Robertson, I.H., Galloway, M., & Hawkins, K. (1999). The absent-mind: Further investigations of sustained attention to response. *Neuropsychologica*, 37*,* 661-670.

McCarley, J.S., & Kramer, A.F. (2007). Eye movements as a window on perception and cognition. In R. Parasuraman & M. Rizzo (Eds .), *Neuroergonomics: The brain at work*: New York: Oxford.

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

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, UK: Cambridge University Press.

Szalma, J.L., Hancock, P.A., Warm, J.S., Dember, W.N., & Parsons, K.S. (2006). Training for vigilance: Using predictive power to evaluate feedback effectiveness. *Human Factors*, *48*, 682-692.

Vidulich, M.A., Wickens, C.D., Tsang, P.A., & Flach, J.M. (2010). Information processing in aviation. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (2nd ed., pp. 175-215). New York: Elsevier.

Warm, J.S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors*, 50*, 433-441.*

Wickens, C.D., & Hollands, J.G. (2003). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall,

Wright, R.D., & Ward, L.M., (2008). *Orienting of attention*. New York: Oxford.

SELECTION REQUIREMENTS TO WORK IN FUTURE ATM SYSTEMS

Hinnerk Eißfeldt
DLR German Aerospace Center
Hamburg, Germany

Alexander Heintz
DFS Air Navigation Services Academy
Langen, Germany

Dirk Schulze Kissing
DLR German Aerospace Center
Hamburg, Germany

Different findings concerning staff selection requirements for future Air Traffic Management (ATM) resulting from empirical studies and expert judgement are summarised. The biggest impacts are foreseen for ATCO's, commercial pilots, ATM technical staff and a/c maintenance staff, mainly related to advanced ATM concept features. An empirical study involving ATCOs and pilots encompassing workshops and simulations studying elements of a free flight scenario revealed significant changes in ability requirements for ATCOs and pilots. They indicate higher ability requirements for pilots in future ATM systems and only small changes for air traffic controllers. Pilot and air traffic controller profiles are likely to assimilate with regard to cognitive abilities. 'Operational Monitoring' is expected to become a core requirement for future ATM systems. A definition and a respective requirements analysis scale are proposed.

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. In Europe, the Single European Sky (SES) initiative aims at enhancing capacity and safety as well as reducing cost and impact on environment. Although the European ATM Master Plan resulting from SESAR, which can be seen as the European equivalent to the U.S. NextGen Programme, widely relies on automation, the Human is expected to remain a central part of the future ATM system. The SESAR concept is based on the idea of a business trajectory being carefully planned and finally executed in a complex interaction between all affected partners, with the primary objective to meet the requirements of airspace users. Information processing and decision-making will be strongly supported by automation tools, often integrating data and systems from different partners (e.g. airline operations centres, cockpit, control centres and towers, meteorological services etc.) During the SESAR Definition Phase, the Human Performance implications of the SESAR Concept of Operations were analysed by means of a preliminary Human Factors Case (Eurocontrol, 2007), staffing prediction models and change and transition models (SESAR 2007a).

Results show that many operational improvement steps will require an adaptation or even development of standards and regulations related to operational staff training and competence verification. Far reaching consequences were identified especially for Air Traffic Controllers, pilots, airline operations centre staff and technical staff installing and maintaining air and ground equipment (about 200.000 staff, SESAR, 2007b,c). Generally, the consequences of future ATM systems and procedures for recruitment and training requirements were found to be sometimes under-, sometimes overestimated by operational experts. Since degraded mode operations and unplanned circumstances will still require Human analysis, decision making and action implementation, the basic ability requirements for e.g. ATCOs can be expected to include about today's requirements profile (Eißfeldt & Heintz, 2002). However, for the more advanced operational features larger changes in ability requirements are considered most likely (SESAR, 2007, a, c).

However, it is obvious that impacts of future systems and procedures can only be determined when systematically examined and embedded in simulation studies. The provision of a standard methodology to achieve this will therefore be one element of the future SESAR Human Performance R&D, including structures to ensure its application in the context of development and validation activities in the framework of a European Joint Undertaking, the SESAR JU (SESAR JU, 2009). At the same time, research on impacts on recruitment and training of operational aviation staff has been continued with a relation to SESAR concepts.

The key question of the DLR project 'Aviator 2030' dealt with changes that will concern pilots and air traffic controllers introducing SES. Based on domain experts' point of view, future ATM scenarios were developed. Key aspects key were tested in two simulation studies to identify potential changes in ability requirements for pilots and air traffic controllers prospectively.

Method

A multi method approach was chosen in order to fully cover the sequence of analysis from the operational concept to changes in ability requirements and tests.

Based on domain experts' points of view, anticipated changes in the ATM system were described using a special workshop technique taken from sociological research. The 'Future Workshop' concept was used for the first time in a high-tech environment such as aviation. A set of workshops with pilots and air traffic controllers successfully described scenarios of future ATM, providing a valid basis for further research. A detailed description of the layout and the outcome of the workshops is provided by Bruder, Jörn & Eißfeldt (2008).

A standard tool for job analysis (F-JAS, Fleishman 1992a,b,c) was tailored to aviation-related research by integrating aviation anchors for the current job conditions of air traffic controllers and pilots. In addition, new scales were developed in a similar style to measure requirements not covered in the original material. Applying the F-JAS Aviator 2030 with aviation anchors allowed for an interpretation of whether job incumbents anticipated an increase or a decrease in ability requirements in future ATM systems (Eißfeldt, 2009).

A low-fidelity integrated simulation platform (AviaSim) was developed following a bottom-up approach by combining two off-the-shelf simulators to meet the requirements of high realism, low cost, high adaptability, and full controllability for experimental purposes. It included a short-term conflict alert (STCA), mid-term conflict detection aids, and interactive labels for data ink communication on the ATC side as well as a data link window and a traffic visualization system (Cockpit Display of Traffic Information, CDTI) on the pilot side. In a linked simulation, the transfer of control between air and ground as well as airborne self-separation in Free Flight Airspace, was examined (Eißfeldt et al., 2009; Hörmann et al., 2009).

Results

Exemplary Findings on Changes in Pilot and ATCO Requirements Profiles

Following the Avia Sim Scenario, pilots and ATCOs again rated the amount of the required level of abilities. Figure 1 lists the top ten ability requirements as rated in the AviaSim study for the future scenario. The numbers in brackets refer to the ranking position for the baseline scenario and are well in line with other applications of the F-JAS (e.g., Goeters et al 2004).

On the controllers side the positioning of ‘time sharing’ at the bottom of the top ten list is remarkable as for air traffic controllers this has been the top rated ability requirement in all studies of current ATC so far. For pilots this scale has moved upwards a bit in the free flight scenarios. A common upward trend can be noted for a variety of abilities: ‘perceptual speed’, ‘speech recognition’, ‘stress resistance’, ‘decision making’ and ‘problem sensitivity’ all are becoming more relevant for both professions with the free flight scenario. Pilots and air traffic controllers share 8 of the 10 top future rankings (cf. Figure 1) underlining the notion of profiles assimilating in free flight scenarios. The abilities not in common are: ‘selective attention’ and ‘resilience’ rated high for ATC but not for pilots, whereas ‘spatial orientation’ and ‘auditory attention’ are among the top ten for pilots but not for ATC. If both lists were aggregated into one ‘Aviator Free Flight Profile’ according to their rankings, ‘problem sensitivity’ would come first followed by ‘decision making’ and ‘vigilance’ and ‘visualization’.

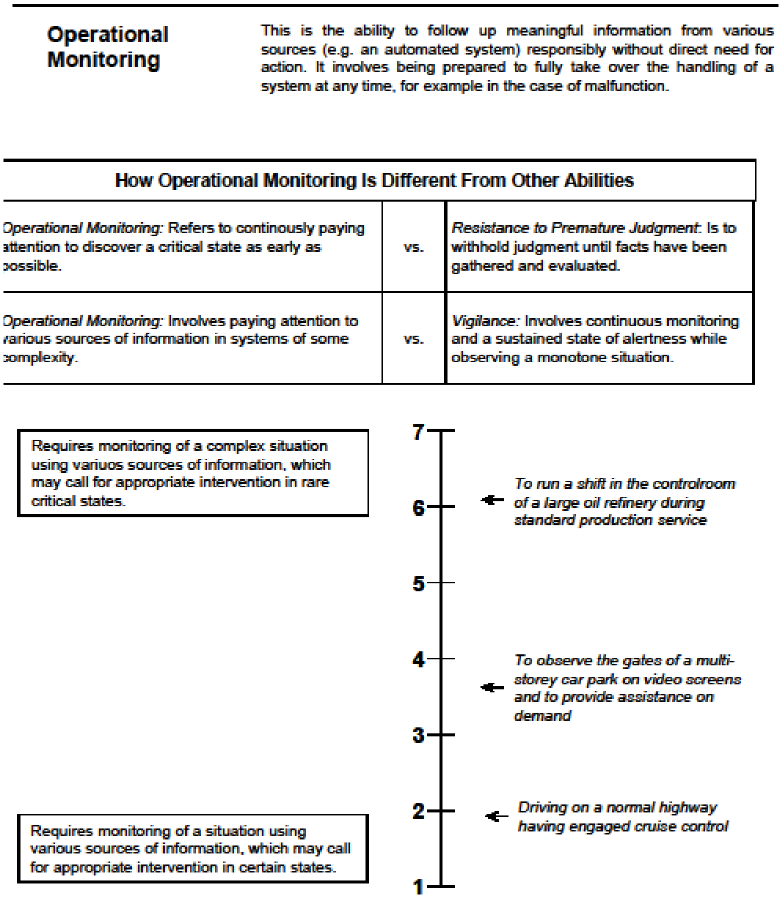
ATC future	Cockpit future
1. Problem Sensitivity (2)	1. Spatial Orientation (5)
2. Decision Making (3)	2. Vigilance (2)
3. Selective Attention (14)	3. Visualization (7)
4. Stress Resistance (9)	4. Problem Sensitivity (12)
5. Speech Recognition (15)	5. Decision Making (9)
6. Resilience (4)	6. Time Sharing (10)
7. Vigilance (5)	7. Speech Recognition (8)
8. Visualization (6)	8. Stress Resistance (11)
9. Perceptual Speed (16)	9. Auditory Attention (1)
10. Time Sharing (1)	10. Perceptual Speed (15)

Figure 1

The top ten ability requirements as rated in the AviaSim Free Flight study for the future scenario. The numbers in brackets refer to the ranking position in the baseline scenario.

Emerging Ability Requirements for Future ATM Systems: Operational Monitoring

Due to the level of automation as envisaged in the future ATM operational concepts, the Aviator 2030 research revealed one specific requirement for pilots and ATCOs to become a critical ability clearly going beyond today's ability requirements. Due to the increase of system based planning and, in consequence, decision making with regard to flight trajectories, the need to effectively monitor the system was found to become much more important. At the same time, the necessity for the operator to take over control under certain circumstances (e.g. handover between ground controlled and free flight mode of operations, non standard operations) is likely to remain critical even in very advanced stages of ATM evolution. Consequently, this ability requirement was further examined and finally defined following the F-JAS format (Fleishman, 1992, a, b, c; Figure 2) in order to enable measuring the ability requirement in future studies.



Operational Monitoring

This is the ability to follow up meaningful information from various sources (e.g. an automated system) responsibly without direct need for action. It involves being prepared to fully take over the handling of a system at any time, for example in the case of malfunction.

How Operational Monitoring Is Different From Other Abilities		
Operational Monitoring: Refers to continously paying attention to discover a critical state as early as possible.	vs.	*Resistance to Premature Judgment:* Is to withhold judgment until facts have been gathered and evaluated.
Operational Monitoring: Involves paying attention to various sources of information in systems of some complexity.	vs.	*Vigilance:* Involves continuous monitoring and a sustained state of alertness while observing a monotone situation.

Requires monitoring of a complex situation using variuos sources of information, which may call for appropriate intervention in rare critical states.

Requires monitoring of a situation using various sources of information, which may call for appropriate intervention in certain states.

Figure 2

Proposed research ability requirement scale on "Operational Monitoring" following the format of the Fleishman Job Analysis Survey F-JAS.

The resulting behavior-anchored rating-scale is named ‘Operational Monitoring’ (Figure 2). It is considered to be marked by some of the ability requirements rated highest by pilots and air traffic controllers for the future scenario: problem sensitivity, situation awareness, decision making and vigilance. Its ends are defined by a description of an extreme degree of the ability. The behavioral anchors are typical task examples for a high/ middle/ low amount of requirement of the ability. Their position on the scale was determined through an expert rating with participants familiar with the format.

Discussion

The findings underline the importance of empirical research involving job incumbents when determining the consequences of future ATM concepts, systems and procedures for selection requirements of operational staff. Structured, facilitated workshops with job holders as well as the inclusion of standardized job analysis tools into validation or other simulation activities promise to be effective means to provide evidence for this frequently discussed issue.

Significant effects were found especially for advanced features. For example, ‘having a picture’ of relevant elements of air traffic is an ability requirement for current ATCOs. During free-flight operations pilots also had to ‘have a picture’ of the surrounding air-traffic. One can consider that the importance of ‘visualization’ capabilities for pilots may change fundamentally with the introduction of airborne-separation procedures. This new ability is not reflected in today’s selection profiles of pilots. It can be assumed that different ability levels concerning ‘visualization’ exist within the present pilot population, as this requirement is not directly tested in many ab-initio pilot selection systems. It will be interesting to see how effective pilot training for self-separation can compensate for these differences in the future.

Pilots and ATCOs in the future will have to take over “manual” control in various non-standard circumstances. The general ability to follow up meaningful information from automated systems without direct need for action will be a prerequisite to show a good performance in these non-standard situations. This shift in job requirements to supervisory control has been identified as a key human-factors issue in advanced ATM systems. One open question here is if this will bring major changes in the ability profiles used for future selection. In this context “Operational Monitoring” may emerge as a useful construct. The value of the new scale will be explored in recent German and European studies on advanced and future ATM systems (e.g. with a newly introduced en route ATM system in Germany).

References

Bruder, C., Jörn, L., & Eißfeldt, H. (2008). When pilots and air-traffic controllers discuss their future. In *Proceedings of the 28th EAAP Conference: The future of aviation psychology* (pp. 354-358). Valencia: EAAP.

Eißfeldt, H. (2009). Aviator 2030 - Ability requirements in future ATM Systems. *In Proceedings of the 15th International Symposium on Aviation Psychology* (pp. 112-117). Dayton, OH: Wright State University.

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.

Eißfeldt, H., Grasshoff, D., Hasse, C., Hoermann, H.-J., Schulze Kissing, D., Stern, C., Wenzel, J. & Zierke, O. (2009).

Aviator 2030 - Ability Requirements in Future ATM Systems II: Simulations and Experiments. In *DLR Forschungsbericht 2009-28*. Köln: DLR.

EUROCONTROL (2007). *The Human Factors Case. Guidance for Human Factors integration*. Ed. 2.0, 29-06-2007. Brussels: Eurocontrol.

Fleishman, E. A. & Reilly, M. (1992a). *Fleishman Job Analysis Survey (F-JAS) Administrator´s Guide*. Potomac, Maryland: Management Research Institute.

Fleishman, E. A. (1992b). *Fleishman Job Analysis Survey (F-JAS) Rating Scale Booklet*. Potomac, Maryland: Management Research Institute, Inc.

Fleishman, E. A. & Reilly, M. (1992c). *Handbook of Human Abilities. Definitions, Measurements, and Job Task Requirements*. Potomac, Maryland: Management Research Institute.

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. 99-109). Aldershot: Ashgate.

Hörmann, H. J., Schulze Kissing, D. & Zierke, O. (2009). Determining Job Requirements for the next Aviator Generation. In *Proceedings of the 15th International Symposium on Aviation Psychology* (pp. 118-123). Dayton, OH: Wright State University.

SESAR (2007a). The ATM Target Concept (Rep. No. DLM-0612-001-02-00a). Brussels: SESAR Consortium.

SESAR (2007b). Human Resources. Task 1.7.1/2/3 – Milestone 3 (Rep. No. DLT-0612-17x-00-12). Brussels: SESAR Consortium.

SESAR (2007c). WP 1.7 - D4. Human Resources. (Rep. No. DLW-0706-017-01-00). Brussels: SESAR Consortium. Retrieved from *http://www.sesar-consortium.aero*.

SESAR (2007d). SESAR Master Plan. (Rep. No. DLM-0710-001-02-00). Brussels: SESAR Consortium. Retrieved from *http://www.sesar-consortium.aero*.

SESAR JU (2009). Statement of Work WP 16 Transversal Aspects. Unpublished. Brussels: SESAR JU.

STUDENT AND INSTRUCTOR PERCEPTIONS OF TRAINING IN A TECHNOLOGICALLY ADVANCED AIRCRAFT

Scott R. Winter and Richard O. Fanjoy
Purdue University
West Lafayette, IN USA

Technologically advanced aircraft (TAA) are defined as those with enhancements such as digital cockpit displays, GPS navigation, moving maps, and autopilots. Many flight-training programs are currently transitioning to such aircraft to more fully prepare their students for commercial flight operations. The technology of modern digitally instrumented training aircraft is similar to that found in advanced commercial aircraft. Students and instructors who operate TAA can provide valuable insight into the process of transitioning from analog to digitally equipped training aircraft. A sample of 216 students and instructors, from a collegiate flight-training program that recently converted from analog instrumented aircraft to the Cirrus SR20, were asked to complete a survey of their perceptions about the transition. Data from these surveys were used to identify perceptual differences between students and instructors for further investigation and to suggest curricular modifications for improved training. This information should provide valuable insights for other flight training programs that are considering or are already operating TAA.

In late 2010, Purdue University transitioned from analog aircraft to a fleet of Cirrus SR20 aircraft equipped with the Perspective avionics package by Garmin. These aircraft are used to complete all private and instrument flight-training courses, and this was the first time technologically advanced aircraft (TAA) were used in the Purdue flight-training curriculum. The Federal Aviation Administration (2006) defines technologically advanced aircraft as equipped with a GPS navigator with a moving map as well as an autopilot, integrated cockpit systems, and ‘glass cockpit’ avionics (p.1). The Aircraft Owner’s and Pilot’s Association’s Air Safety Foundation notes that “new fleet sales to flight schools and university flight departments are almost universally glass cockpit…even for basic trainers” (AOPA Air Safety Foundation, 2007). Previous research into TAA/glass cockpit aircraft investigated pilots’ attitudes and perceptions towards these aircraft (Casner, 2008; Dahlstrom et al, 2006). The current research is focused on how instructors and students who operate TAA perceive training issues. As TAAs become more prevalent in flight training programs, curricular concerns must be addressed quickly to receive maximum benefit from this advanced technology.

Review of the Literature

As TAA have become more common in the national airspace system, training programs have begun assessing the value of converting their fleets from analog to digitally instrumented aircraft. When TAA were first introduced, only 51% of university aviation schools surveyed by Young and Fanjoy (2003) operated newer technology aircraft for student training. Furthermore, it was the view of university flight program administrators that additional training and exposure to advanced avionics was the responsibility of flight operators who would eventually employ student trainees and outside of the objectives of university flight programs (Young & Fanjoy, 2003). Research by Casner (2005) suggested an improved transfer of training for TAA-trained pilots who transitioned to a commercial jet simulator over those who were trained in analog aircraft. Of the subjects who were trained in TAA, 83% were able to successfully complete tasks presented in the jet aircraft simulator compared to 54% of a control group who were not trained in TAA, but did receive assistance from instrument labels in the jet aircraft simulator. When no labeling was present for the control group, the task completion rate dropped to 22%. Although a strong case can be made for training in TAA, a limiting factor for collegiate flight training programs may be the high cost of employing such equipment. TAA may provide a much higher level of preparation and their advanced avionics may be more popular with pilot trainees, but burden sharing associated with more expensive aircraft presents a difficult challenge for flight training operations and student pilots alike.

Casner (2008) completed additional research on advanced aircraft systems with a sample of 134 general aviation pilots from San Francisco area flight schools. Casner surveyed pilots on general attitudes, workload, awareness, learning, retention, error, safety, and pilot preferences. Subjects who participated in this study expressed a clear preference for flying TAA. From the survey sample, 74% indicated they would rather fly glass cockpit aircraft than analog-instrumented aircraft. However, respondents were concerned that pilot reliance on glass cockpit instrumentation may contribute to unsafe flight operations. When asked the question of whether the operation of advanced aircraft systems may lead to stretching the boundaries of safety, 80% of the survey respondents either agreed or strongly agreed with this statement. While Casner's research focused on the perceptions of general aviation pilots, the current study will evaluate perceptions from a sample of university aviation program students and instructors.

Researchers at Middle Tennessee State University (MTSU) investigated the use of TAA in collegiate flight training and the implementation of the FAA Industry Training Standards (FITS) program. The FITS program focuses on aeronautical decision-making and single-pilot resource management skills through the use of scenario-based training. The purpose of the MTSU research was to examine the effectiveness of a FITS training curriculum over a traditional flight-training syllabus for the private pilot certificate and instrument rating training. Students in this study, using the FITS syllabus, completed an instrument rating with an average of 88.66 flight hours compared to 134.3 flight hours for those participants who used a traditional syllabus (Craig, Bertrand, Dornan, Gossett, & Thorsby, 2005). Although the MTSU researchers used TAA in their study, the findings reflect a focus on the FITS program rather than aspects of the equipment used. In the current study a traditional syllabus was used with Cirrus aircraft to limit the survey focus rather than adding additional variables associated with different training curricula.

Dahlstrom, Dekker, and Nahlinder (2006) studied the implementation of TAA in an ab initio flight training school. Instructors interviewed prior to training in the Cirrus SR-20 anticipated 5 problem areas during the transition: the use of displays, aircraft speed, use of the side yoke, work environment, and flight safety. During interviews with the same instructors, after the implementation of the TAA, researchers found that the implementation had been less problematic than anticipated. “The planning and preparation of the training material (particularly the hints for instructors) were unanimously seen as the main reason for the successful implementation” (Dahlstrom, Dekker, & Nahlinder, 2006, p. 140). It is anticipated that some of the concerns from that study may be further reinforced by the current research.

Methodology

Participants

The sample population for the current research included 216 students and flight instructors from a university flight-training program. Survey respondents included 50 student pilots and 40 part- and full-time flight instructors for a response rate of 41%. Four surveys were not considered in the analysis due to incomplete responses. Participants were told that their participation was voluntary and their responses would be confidential. All participants were required to be at least 18 years or older to complete the survey and to have flown the Cirrus aircraft during the 2010 fall semester. Participants were advised that they could only complete the survey once.

Instrumentation

Data for this research was gathered with an electronic survey created using Qualtrics, an online survey development software program. The survey consisted of 32 questions that addressed biographical data and possible concern areas for completing training in a technologically advanced Cirrus SR20 aircraft. Institutional Review Board (IRB) approval was obtained for the survey and anonymity of the respondents was maintained by the survey website. Participant's e-mail addresses were accessed from flight training records and made available by the Director for Flight Training for initial contact purposes. An introductory e-mail was sent to eligible participants that provided information on the survey, stated the eligibility requirements, and requested completion of the survey instrument. The survey window was open for the first month of the spring 2011 semester. Participants were sent two follow-up e-mails reminding them of the survey completion deadline.

Results and Discussion

Participant Demographics

Of the 216-targeted subjects, 90 responded to the survey, producing a response rate of 41%. Although four surveys were incomplete, there were 48 usable responses from student pilots and 38 useable responses from flight instructors. All of the students reported their age as between 18 and 21 years old, and 80% had less than 200 hours total flight time. The majority of flight instructors were between 19 and 23 (82%), and 63% had between 200 and 500 hours total flight time. However, 25% of the instructors had more than 700 total flight hours. Most students had very little TAA experience prior to beginning the fall semester with 90% reporting less than 5 hours of G1000 avionics use and 90% reporting zero hours of Cirrus flight time. Flight instructors reported slightly higher levels of experience than students with both G1000 avionics and Cirrus aircraft. Most students (72%) reported no experience with the G1000, compared with 26% of instructors who reported no G1000 experience. It was unclear from survey responses whether the prior G1000 experience was completed in an actual aircraft or a G1000 trainer available in the university's simulator facility. Regarding actual in-flight experience, 43% of instructors reported some prior Cirrus flight time, compared to only 10% of students.

Perceptions of the SR20 for Training

Three survey questions focused on student and instructor perceptions of using an SR20 for training. Participants were asked to comment, given their experience, on whether or not the SR20 was a good primary training aircraft, if they preferred to train in an SR20 rather than a conventional aircraft such as a Piper Warrior, and if there was too much information available in the SR20 cockpit for primary training. The survey provided a Likert scale for answers with response options of strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree. The results are shown in Table 1.

Table 1
Perceptions of the SR20 Aircraft for Training

Students			Instructors	
Questions	*n*	*Agree/Strongly Agree*	*n*	*Agree/Strongly Agree*
SR20 is a good primary trainer	48	45%	38	26%
Prefer SR20 over conventional aircraft	48	49%	38	26%
Too much information in SR20 for primary training	48	31%	38	39%

A larger percentage of student pilot respondents than instructors felt the SR20 was a good primary trainer and preferred it to conventional aircraft. Of the student respondents, 45% agreed or strongly agreed that the SR20 was a good primary training aircraft and 49% agreed or strongly agreed to a preference for training in it rather than in a more traditional aircraft such as a Piper Warrior. When instructors were asked if the SR20 was a good primary trainer and if they preferred training in an SR20 over a more conventional aircraft, 26% agreed or strongly agreed to both questions. The differences in response levels might reflect a flight instructor bias associated with safety concerns, flight schedules, and program completion deadlines that become more pertinent in an aircraft with which they were less familiar. Although further study is needed in this area, it may be that these additional pressures forced instructors to take a more reserved view of the SR20's utility as a primary trainer. In addition, slightly more instructors (39%) than students (31%) agreed or strongly agreed that there was too much information in the SR20 cockpit for primary training. This difference might reflect an instructor awareness of available information and concern that students did not fully utilize such information.

Instructor concerns about the suitability of the SR20 as a primary training aircraft may be further impacted by the limited experience of instructors in the SR20 with 57% having no prior Cirrus experience before the fall 2010 semester. While full time flight instructors completed a full transition training course for the SR20 at Cirrus Aircraft in Duluth, MN, part-time flight instructors who were only allocated some ground training and 3 hours of flight time to transition to the aircraft before being assigned to student pilots. Several instructors expressed concerns about this issue with comments such as, "we were thrown into the airplane with very little knowledge of the airplane or G1000 operations," or "lack of instructor knowledge of G1000 and Cirrus systems – more training necessary." Of the additional survey comments expressed by flight instructors, 30% referenced a lack of training or the desire to have had more training before beginning to teach students. This concern with instructor training experience may well have been influential in responses to questions about the suitability of the SR20 as a primary training aircraft and suggest a key consideration for flight operations with an impending transition to TAA.

Workload, Situational Awareness, and Safety

Participants were asked how they perceived the primary flight display (PFD), multi-function display (MFD), traffic advisory system, and terrain awareness warning system affected their workload during flight. The majority of student and instructor respondents reported that the PFD, MFD, traffic advisory system, and terrain awareness system either reduced or had no effect on workload. However, 14 (37%) of instructors in the survey felt the MFD, in particular, increased their workload. Of the student respondents, 10 (21%) felt that the traffic advisory system, in particular, contributed to an increased cockpit workload. Workload perceptions are depicted in Table 2.

Table 2
Workload Perception of Avionics Components

	Students				Instructors			
Questions	*n*	Increase	Decrease	No Effect	*n*	Increase	Decrease	No Effect
PFD	48	1	34	13	38	3	23	12
MFD	48	3	42	3	38	14	21	3
Traffic Advisory System	48	10	32	6	38	5	22	11
Terrain Awareness Warning System	48	2	24	22	38	2	14	21

In follow-up questions, all respondents were asked if the traffic advisory system caused them to spend more time looking inside the aircraft and less time scanning for traffic. The results were similar for instructors and students in response to this statement with 59% of the flight instructors and 54% of the students agreeing or strongly agreeing with this statement. When asked if situational awareness was better in an SR20 over a conventional aircraft such as a Piper Warrior, 69% of instructors and 53% of students agreed or strongly agreed. These numbers were slightly less than the results reported in Casner's study of San Francisco area pilots. In that study, 85% of pilots agreed or strongly agreed that their situational awareness was better in an advanced cockpit (Casner, 2008). A possible explanation for the difference may be the sample demographics and different flight experience levels of participants in the two studies. Pilots in the Casner study had a mean total flight time of about 1,500 hours with perhaps much greater situational awareness than the much lower experienced participants in the current study.

In a final focus of the current study, participants were asked if they felt the SR20's cockpit would cause pilots to continue into deteriorating weather conditions. Findings suggest 40% of student respondents and 40% instructors agreed or strongly agreed with this statement. This statement was compared to a similar statement from Casner's 2008 study. Pilots from Casner's study were asked if advanced cockpit systems would cause pilots to push the boundaries of safety. In that study, 80% of the participants agreed or strongly agreed with that statement (Casner, 2008). The controlled environment of the university aviation program and strict weather and dispatch criteria clearly limits a pilot's ability to opt for flying into deteriorating conditions and may have biased the current study participant responses to this question.

Summary and Conclusions

Findings from the current study suggest that instructors may be less supportive than students of the SR20 as a primary training aircraft over conventional aircraft such as the Piper Warrior. Particular concerns from instructors about MFD operations and from students about distractions associated with the traffic advisory system were noted in survey responses. These findings may have been biased by the brief exposure instructors and students in the sample had to SR20 flight operations and training activity before the study survey was completed. Findings from the current study do provide support for earlier research by Casner (2008) that suggests pilots view situational awareness with TAA as improved over conventional aircraft. However, the Casner study contention that pilots believe TAAs may lead them to fly into deteriorating weather conditions received less support from the current study. Follow-on research should be conducted with a pilot training cohort of students and instructors who have more experience in a TAA to ascertain the validity of the current study findings. In addition, further study is needed to determine the minimum preparation appropriate for instructors before they are assigned students to teach in a new fleet of aircraft. The findings of the current study as well as growing prominence of TAA in primary flight training operations underscores the need for further investigation into appropriate curricular and instructor preparation for the current and future generations of professional and general aviation pilots.

References

AOPA Air Safety Foundation. (2007). *Technologically advanced aircraft safety and training.* Aircraft Owner's and Pilot's Association. Federick: AOPA Air Safety Foundation.

Casner, S. M. (2008). General aviation pilots' attitudes toward advanced cockpit systems. *International Journal of Applied Aviation Studies , 8* (1), 88-112.

Casner, S. M. (2005). Transfer of learning between a small technically advanced aircraft and a commercial jet transport simulator. *International Journal of Applied Aviation Studies , 5* (2), 307-319.

Cirrus Aircraft. (December 31, 2010). *Inside Cirrus.* Retrieved from http//:news.cirrusaircraft.com/?tng=production

Craig, P. A., Bertrand, J. E., Dornan, W., Gossett, S., & Thorsby, K. K. (2005). Ab initio training in the glass cockpit era: new technology meets new pilots. *International Symposium on Aviation Psychology 2005.* Dayton: International Symposium on Aviation Psychology.

Dahlstrom, N., Dekker, S., & Nahlinder, S. (2006). Introduction of technically advanced aircraft in ab-initio flight training. *International Journal of Applied Aviation Studies , 6* (1), 131-144.

Federal Aviation Administration. (2006). FAA/Industry Training Standards (FITS) questions and answers. Retrieved from http//:www.faa.gov/training_testing/training/fits/media/fits-qa.pdf.

Young, J. P., & Fanjoy, R. O. (2003). Advanced collegiate flight automation training: what is needed and at what cost? *International Journal of Applied Aviation Studies , 3* (2), 215-225.

Improving Human Factors Training: Perceptions of Retired Aircraft Maintenance Technicians

Kim Wayne Robinson
Oklahoma State University
Broken Arrow, Oklahoma
Frederick D. Hansen Ph.D.
Oklahoma State University
Tulsa, Oklahoma

Human Factors training is a major concern for airlines and Maintenance Repair Organizations. This study taps into the reservoir of knowledge and experience of retired Aircraft Maintenance Technicians (AMTs) with respect to aviation maintenance human factors. A survey was given to retired AMTs with 20-30 years of maintenance experience to rank the relative importance of each of the FAA's Dirty Dozen of human factors in maintenance accidents. Follow-up focus groups were conducted to further explore recommendations on how to prevent human factor accidents related to maintenance. This paper discusses the findings from the original survey and preliminary findings from the first focus groups.

As aircraft maintenance technicians conduct either scheduled or unscheduled maintenance, errors and mistakes occur. According to the Federal Aviation Administration (FAA), about 80 percent of maintenance mistakes involve human factors (Adams, 2009). The risk of experiencing more human error is a definite byproduct of a demanding maintenance work schedule. How to prevent these airline maintenance related human factor induced errors from occurring is a continual challenge. Gordon Dupont developed the Dirty Dozen in his work with Transport Canada. The Dirty Dozen (Table 1) has been published in a series of safety posters along with discussions of available safety nets to provide a tool for training, for situational awareness, and as a checklist of human factors issues in aviation maintenance (Dupont, 1997). These posters continue to be used in aviation maintenance training and human factors awareness.

Table 1**.**

Gorden Dupont's Dirty Dozen (Dupont, 1997)

Most Common Causes of Maintenance Accidents

- Lack of Communications
- Complacency
- Pressure
- Stress
- Lack of Teamwork
- Lack of Assertiveness
- Norms
- Lack of Knowledge
- Lack of Resources
- Fatigue
- Distraction
- Lack of Awareness

In 2006 and 2007, the FAA conducted two major surveys among aviation safety inspectors and the international maintenance community to determine human factors that required immediate attention. The survey focused on issues such as fatigue, training, leadership, error reporting, and documentation. The Dirty Dozen were presented to the safety inspectors to rank as challenges. "The top three were pressure, complacency, and norms" (Johnson & Hackworth, p.34)". The FAA has a three day employee training course (Human Factors in Aviation Maintenance) that is now required.

Maintenance Resource Management (MRM) is another recent conceptual phenomenon that goes after the human side of accidents and has bled over into the various facets of the global aviation system (Taylor and Christensen, 1998). Its many applications are poignant and delve into the various aviation environments that are unique to the aircraft maintenance technician.

MRM is part of the soft skills that have come to the forefront in the aviation and especially the airline maintenance business. This concept allows employees to better understand their function within a company's operation and its efforts to obtain a safe and error-free maintenance product. One factor that increased attention on MRM was the ValuJet accident in 1997 that resulted in more surveillance and sharpened focus on maintenance program requirements, as well as airworthiness responsibilities between operators and the third-party repair stations with which they outsource maintenance (Al-Amoudi, 1998).

MRM, one of the best-kept industry secrets, is the newest tool in the airline maintenance management arsenal. The demand for this type of training became apparent following the Aloha Airlines flight on April 28, 1988, in which approximately 18 square feet of the upper cabin separated from the aircraft. Initial investigation of the accident revealed fatigue damage and disbanding of the metal structure that was in part caused by improper maintenance, inspection, and supervision. It was found that maintenance human factors may have played an important role in this accident (Gauntt, 2000).

Airline maintenance complexes have very large and diverse personnel requirements that a MRM program can positively impact. Airlines have used MRM in training maintenance and ground personnel on how to recognize their own limitations, which, in turn, may affect or influence individual and team performance. The successful MRM programs have upper management support, training for all maintenance departments, continuous clear communication feedback, and a uniquely tailored MRM implementation plan.

Several successful in-house programs were established with large commercial carriers such as Continental Airlines, US Airways, and United Airlines. Alteon (a Boeing Company affiliate) conducts free 2-day MRM seminars to airlines and any organizations who desire error-reductions (Maynes, 2005). If in-house MRM programs are not available to airlines, outside MRM contractors such as Grey Owl and Flight Safety International are two suppliers of such programs (Gauntt, 2000).

A need exists to develop a more effective human factors training program based on real causes of aircraft maintenance accidents. In order to do this, the real-world views of experienced AMTs are necessary. We are losing a wealth of knowledge from AMT retirees that could be used to influence the next generation of human factor training programs. Steps need to be taken to capture this knowledge so that it can be incorporated into future training. If this is not done immediately, it will be lost forever, and future training will lack the valuable insight that the experience from this group could provide.

Purpose of the Study

The purpose of this study was to describe the perceptions of retired aircraft maintenance workers concerning human factor causes of aircraft accidents (maintenance-related specifically). This was accomplished by first surveying retired AMTs concerning their views of the influence of various Human Factors (specifically the Dirty Dozen used by the FAA) in causing aircraft maintenance accidents. Two follow-up focus groups were conducted with selected AMTs to discuss the survey results and to describe in detail their views related to training on these issues.

Methodology

This is a descriptive study to provide a profile of the human factors perceptions of retired aircraft maintenance technicians. The population for this study was retired American Airlines Aircraft Maintenance Technicians who meet monthly at a Transport Workers Union (TWU) union hall. The group of TWU retiree association members normally has a 25-150 membership and they are on an average, age 55 and above, normally white, male, ex-worker with 5-30 years maintenance experience at American Airlines. Retirees normally meet monthly to keep up on issues affecting their retirement. It was at one of these normal monthly meetings that the initial survey was passed out.

Initial data was collected through a short survey. Fifty surveys were handed out during a meeting of retired mechanics with 29 surveys completed. The goals of the survey were to identify participants for focus groups, and to discover how the AMTs would rank the importance of each of the Dirty Dozen based on their years of experience on the hangar floor.

Two follow-up focus groups were chosen from among the survey participants who volunteered. Questions presented to the focus group centered on the top four human factor causes. The focus group was asked about what these factors look like in the work environment and what strategies are or could be used to overcome these causes. Participants were also asked about human factor training and the good and bad techniques of the trainers. Results from the focus groups will be used to develop the next level of human factors training for maintenance workers at American Airlines.

Research Questions

The primary research questions for this study focused on how retired AMTs rank human factor causes of accidents using the FAA Dirty Dozen. Specifically, what strategies/safety nets do retirees believe should be used at the hangar floor to help prevent human factor causes of accidents? How do retirees feel about the development of the next phase of human factor training? What topics are important and relevant to the next generation of AMT training?

Findings and Analysis

Preliminary results

Survey results. Survey taken May 2010 showed that 90% of the respondents had 20 years or more of aircraft maintenance experience. Participants were asked to choose the top four human factors problems from the Dirty Dozen list and place those selections in the High box, and then the least significant four in the Low box, and then the remaining four were then entered in the Medium Box. All responses were considered however only the four top responses in the High, Medium, and Low categories were noted in Table 2. The numbers in each box indicate the total number of responses for each factor considered High, Medium, or Low. Lack of Assertiveness made the top four for each the Medium and Low categories while Fatigue was in the top four for each the High and Medium categories. It is interesting to note that Lack of Communication and Complacency failed to make the top four in any category even though Complacency was in the top three items from the major FAA surveys mentioned earlier.

Table 2.

Categorize the "Dirty Dozen" human factor causes of maintenance-related accidents in terms of High, Medium and Low.

High box	Medium Box	Low box
Presssure-18	Lack of Resources-16	Lack of Teamwork-17
Lack of knowledge-16	Lack of Assertiveness-13	Norms-16
Distraction-15 Fatigue	-12	Stress-12
Fatigue- 14	Lack of Awareness-11	Lack of Assertiveness-12

Note. Only the four highest factors in each category were included which explains why the total don't reflect all choices from the 29 participants.

Next, respondents ranked the Dirty Dozen inside the three categories of High, Medium, and Low (see Table 3). It is important to note that there are numerous differences between the results of the two questions. Ranking any set of items from 1 to 12 will inevitably result in a great deal of diversity based on personal experiences. Complacency was not considered a major influence for either ranking and that Fatigue, while in the top four in both the High and Medium box in Table 2, it did not score high enough to be included in any of the boxes from Table 3.

Table 3.

Rank from highest to lowest the human factor cause inside the High, Medium, and Low Boxes.

High Box	Medium Box	Low Box
Lack of Awareness-8	Lack of Assertiveness-9	Lack of Communication-5
Pressure -7	Lack of Resources-7	Norms-7
Distraction-7 Pressure	-9	Lack of Knowledge-4
Lack of Teamwork-5	Lack of Teamwork-5	Stress-5

Focus Groups. On June 11, 2010, the first Focus Group met. Four members were present. All had twenty years or more experience with one individual the recipient of a golden wrench award (prestigious lifetime mechanic award winner). Other members of the group included an extensive variety of mechanic experience, consisting of three Quality Assurance inspectors, two line maintenance mechanics and one shop level mechanic.

On June 17, 2010, the second Focus Group met. This focus group was based on the knowledge of investigator- awareness of recent past retirees. Four members were present with one member unable to participate due to medical problems. All participants had twenty years or more experience with one individual having being recently retired after working with the investigator for over fifteen years. Maintenance experience of the group consisted of two Quality Assurance inspectors, two line maintenance mechanics, and one shop level mechanic.

Focus Group 1 comments included.

- Pressure is external and can come from maintenance supervisors (foreman, your boss) who allow the aircraft maintenance schedule to override safety sometimes.
- Fatigue caused by long hours, side jobs, inadequate sleep (off shift hours circadian system upset), overtime (needed and manufactured) and the extra effort of field trips. Mechanics are "burning the candle at both ends".
- Foremen suggest shortcuts to get the plane off and this sets the environment up for lack of knowledge

Focus Group 2 comments included.

- Pressure is sometimes self-induced. Stress includes you wanting to do the best job and not wanting to mess it up.
- Fatigue is the effect of not managing overtime rightly- "They worked 20 hour [shifts]…nobody twisted anybody's arm [forced you to take overtime]."
- Pressure might bring a lack of awareness. "I remember a supervisor [management] walking off the end of a wing."

Acknowledgement

The complete results of this study will be included in the dissertation of Kim Robinson. Investigator would like to thank Transport Workers Union 514 in Tulsa for allowing the survey and focus groups events to occur at their union hall. The views of the research report do not reflect the views of Transport Workers Union, American Airlines, and the City of Tulsa.

References

Adams, C. (2009). Human factors: Beyond the "Dirty Dozen". *Aviation Today*. Retrieved 3 March, 2010 from http://www.aviationtoday.com/am/repairstations/Human-Factors-Beyond-the-&quotDirty-Dozen%22_33730.html

Al-Amoudi, H. (1998). Humans Factors in Aircraft Maintenance. Bell & Howell Information Company). *California State University Dominguez Hills, UMI 1389319* (Master of Science in Quality Assurance), 1-64. (Digital Dissertations UMI)

Dupont, G. (1997). The dirty dozen errors in aviation maintenance. In Meeting Proceedings Eleventh Federal Aviation Administration Meeting on Human Factors Issues in Aircraft Maintenance and Inspection: Human error in aviation maintenance (pp. 45-49). Washington, DC: Federal Aviation Administration/Office of Aviation Medicine.

Gauntt, T. (2006). Maintenance Resource Management if you haven't been to an MRM class, why not? Retrieved 01/28/2006 from http://www.aviationtoday.com/cgi/am/show_mag.cgi?pub=am&mon=0800&file=08mainte...

Johnson. W.B & Hackworth, C. (2008). *Human Factors in Maintenance*. Flight Safety Foundation; AeroSafetyWorld.

Maynes, L. (2005). A look at Crew Resource Management and Maintenance Resource Management. *Mobility Forum, 14*(2), 4-6.

Taylor, J., & Christensen, T. (1998). *Airline Maintenance Resource Management.* Society of Automotive Engineers, INC.

EVALUATION OF A SEPARATION ASSISTANCE DISPLAY IN A MULTI-ACTOR EXPERIMENT

J. Ellerbroek, M. M. van Paassen, M. Mulder
Faculty of Aerospace Engineering, Delft University of Technology,
Kluyverweg 1, 2629 HS Delft, The Netherlands.

In the past, several display concepts have been developed, as aids in the task of airborne self-separation. In several of these display concepts, the interface helps the pilot solve the conflict, as opposed to automation providing an explicit resolution. Especially in this case of manual problem solving, (implicit) interaction between the actors in a conflict becomes an important factor. An experiment was conducted to evaluate an EID-inspired, constraint-based separation assistance display, where all aircraft in each conflict were controlled by pilot subjects. In the experiment, several conflict scenario's have been evaluated, where coordination between pilots could either follow implicitely from the conflict geometry presented by the interface, or, require additional, explicit rules ("rules of the air") to be solved in a coordinated fashion.

In the current ATM concepts for unmanaged airspace, aircraft will fly completely predetermined 4D trajectories, where automation will provide resolution advisories for traffic (or other) conflicts that may result from uncertainties that arise during flight (RTCA, 2002; SESAR Consortium, 2007). In this situation, the pilot's task will be one of monitoring separation, and selecting and applying resolution advisories, provided by the automation. He should, however, be able to judge the fidelity of a proposed resolution, and be able to intervene in case the automation fails.

Furthermore, because conflicts will be resolved in a decentralized fashion, determining the resolution to a conflict will require coordination between the actors in that conflict. This means that for automated, as well as manual conflict resolution, predictability of decisions will be essential to guarantee an acceptable level of safety. In situations where there is not enough time for negotiation, implicit coordination will be required, e.g., by following a predetermined set of rules that dictate which aircraft should maneuver, and how it should maneuver. In worst-case scenarios, pilots will have to manually determine resolution maneuvers, for instance when the automation has failed, or other reasons why a pilot decides to resolve the conflict manually. This poses limits on the complexity of the coordination rules. For automated resolution advisories, high rule complexity can make it difficult for pilots to understand the rationale behind resolution advisories, potentially resulting in non-conformance and distrust of the system (Schild & Kuchar, 2000; Lee & Moray, 1992; Parasuraman & Riley, 1997).

For adequate situation awareness, and proper interaction with automated systems, and between actors in a conflict, it is therefore necessary for regulation and automation to be transparent and understandable to the operator. The work presented in this paper is part of an ongoing study on the design of a separation assistance interface that can fulfil this role (van Dam, Mulder, & van Paassen, 2008; Heylen, van Dam, Mulder, & van Paassen, 2008; Ellerbroek, Visser, van Dam, Mulder, & van Paassen, 2011). The display concepts developed in this study try to realize proper support, by showing the implications of other traffic for the affordances of locomotion, and how they relate to constraints that result from ownship performance limits. By going beyond visualizations that relate only to the automation logic, these displays help pilots gain deeper knowledge of the functions and relations within the work domain. These displays should provide support in routine as well as unforeseen situations, where the pilot may have to rely on his own skills to resolve a conflict.

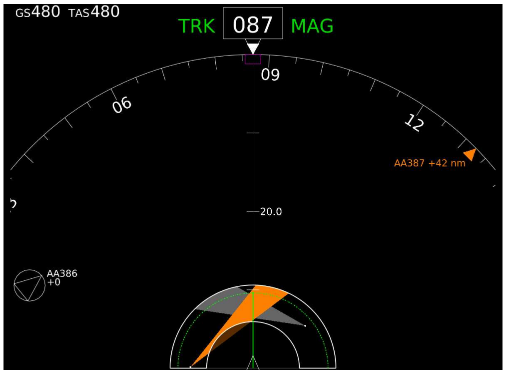


Figure 1: The horizontal separation assistance display is based on a classical Boeing navigation display, with an added separation assistance overlay. The overlay provides a functional presentation of the affordances for aircraft airspeed and track angle using a horizontal projection of the three-dimensional velocity-vector affordance space.

The work presented in this paper will focus on the coordination rules that can be used with these display concepts, in multi-actor resolution of traffic conflicts. An experiment was defined to evaluate coordination behavior in worst-case scenarios, in which pilots have to resort to manual determination of conflict resolutions. In the experiment, a horizontal, constraint-based separation assistance display was available to the pilots to evaluate conflicts, and to determine resolution maneuvers, see Figure 1. The following two sections will present a set of coordination rules that can be used with the display, and describe the experiment and discuss the results, respectively. The paper concludes with a summary of these findings, and plans for future work.

Implicit coordination for manual control

For implicit coordination between actors in a conflict to function consistently well, a set of rules must be defined that keeps pilots from selecting opposing resolutions. These rules may be based on extensions of the visual flight rules (International Civil Aviation Organization, 1996), but in most cases, a cooperative resolution can also be derived from the conflict geometry, see Figure 2. This type of coordination is related to the conflict solution that results in *minimum path deviation*. Consider the nominal aircraft position at time t:

$$\mathbf{x}\left(t\right) = \mathbf{x}_0 + \int_{t_0}^{t} \mathbf{V}_{orig}\left(t\right) \mathrm{d}t \tag{1}$$

The path deviation for a maneuver can be derived from difference between maneuver and original velocities:

$$\Delta x = \int_{t_0}^{t_1} \left|\mathbf{V}_{sol}\left(t\right) - \mathbf{V}_{orig}\left(t\right)\right| \mathrm{d}t = \int_{t_0}^{t_1} \left|\Delta\mathbf{V}_{sol}\left(t\right)\right| \mathrm{d}t \tag{2}$$

Using (2), it can be shown that the path deviation is minimized by minimizing $\Delta\mathbf{V}_{sol}$.

Figure 2 shows a traffic conflict with two aircraft, and the derivation of their velocities relative to each other. The circles visualize the horizontal separation margin around each aircraft, and the areas between the triangle lines tangent to each circle show the conflicting values for each relative velocity vector. In this figure, $\Delta\mathbf{V}_{sol}$ is the vector distance between $\mathbf{V}_{rel}$ and the nearest constraint zone leg. The shortest distance is found when $\Delta\mathbf{V}_{sol}$ is taken perpendicular to the constraint zone leg (van Dam et al., 2008; Bilimoria, 2000). Figure 2 also illustrates that, as long as $\mathbf{V}_{rel}$ is closer to one leg than to the other, a single optimum for $\Delta\mathbf{V}_{sol}$ can be found, and that both aircraft share this optimum. Therefore, implicit coordination is guaranteed when the optimum is selected as a resolution.

For situations where there is no unique geometrically optimal solution, an additional set of rules is required. For the experiment, the following 'rules of the air' were used: aircraft being overtaken have the right of way and overtaking aircraft must remain clear by altering heading to the right. When two aircraft are approaching each other head on they must both alter heading to the right.

Because the separation assistance display presents the pilot with a velocity action space that is based on the conflict geometry, it can support both coordination strategies. Geometrically optimal solutions can be selected using the display, by changing speed and heading to move the speed vector to the nearest conflict zone leg. Also, selecting a velocity vector to the left, or to the right of a conflict area is analogous to passing the intruder aircraft to the left or the right. Ownship will pass in front of the intruder when the velocity vector crosses the respective constraint area on the display.

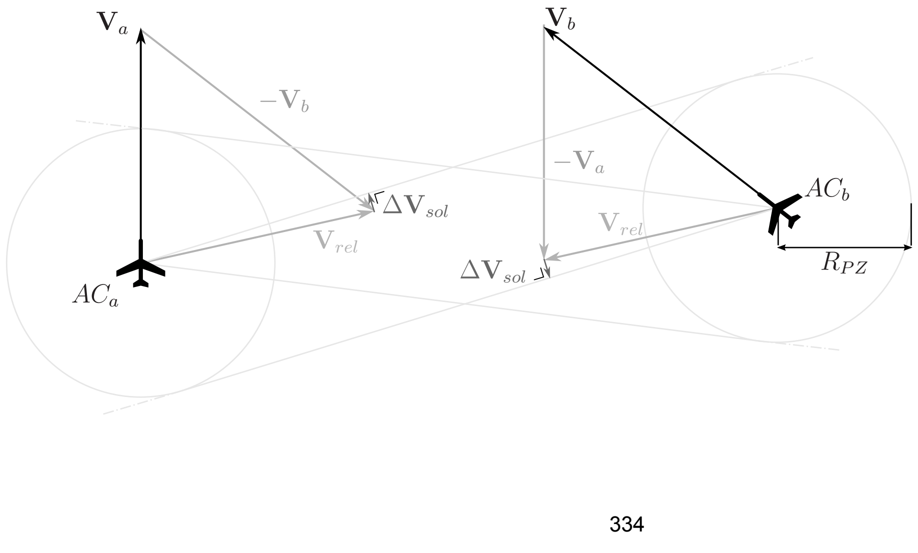


Figure 2: Geometrically optimal solutions guarantee implicit coordination, for all conflict geometries with the exception of collision courses. Because of the rotational symmetry of constraint zones of both aircraft, selecting the optimal solution for aircraft AC_a will always be complementary to the geometrically optimal solution for aircraft AC_b.

Experiment

To evaluate the coordination of manual resolution maneuvers between actors in traffic conflicts in unmanaged airspace, a multi-actor, traffic separation experiment was performed. To obtain analyzable pilot responses, as well as the interactions between those responses, pairs of pilots were placed in two-aircraft traffic conflict situations, with a loss of separation in the near to short term future.

Method

Each session consisted of a continuous presentation of five consecutive conflict scenarios, that needed to be resolved manually, with the aid of a constraint-based separation assistance display. Traffic conflicts were always between two human actors, and were designed using parameters *conflict angle*, *time to first loss of separation*, and *CPA distance*.

Apparatus and Aircraft Model: The experiment was performed on two physically separated, fixed-base pilot stations. Each setup featured two LCD screens: one showing a Primary Flight Display, the other showing a Navigation Display with separation assistance overlays. Participants could control display settings and auto-pilot heading and speed modes through physical EFIS selector and Mode Control panels.

The aircraft models employed in the simulation were low-order, quasi-linear models of a Boeing 707-300, and an Airbus A330, see table Table 1. The model coefficients were obtained from EUROCONTROL's BADA aircraft database. The simulation was run in realtime, at an update rate of 100 Hz. The experiment was conducted with zero wind, and no turbulence.

Experiment Design and Procedure: The experiment was designed as a within-subjects repeated-measures, where factors *aircraft model* and *conflict geometry* were varied. The *aircraft model* factor was introduced to illustrate the effect of a reduced speed margin on the availability of (optimal) resolution options. Because every aircraft type suffers from reduced speed margins at increasing altitude (stall speed and critical mach number converge with increasing altitude), speed margins are an important factor for conflict resolutions at cruise altitude.

The conflict geometry was designed based on three factors: conflict angle, time to first loss of separation, and the distance between the two aircraft at the closest point of approach. Here, the conflict angle determines the shape and orientation of the constraint zone, and the magnitude of the closing speed between the two aircraft. Varying conflict angle between scenarios, therefore, is a way to minimize memorizing/learning effects between scenarios. The distance at the CPA, d_{CPA}, determines whether a unique optimal solution to the conflict can be found ($d_{CPA} \neq 0$), or whether coordination based on an additional set of rules is required ($d_{CPA} = 0$). The time to first loss of separation varied between 3 - 5 minutes, which meant a medium to high level of urgency for each conflict scenario. Each of the conflicts was designed with two participating aircraft.

Each experiment session required two subjects. These subjects were invited and briefed separately, and were not informed that the conflicting aircraft were controlled by a second participating pilot. After a briefing on the experiment and the functioning of the separation display, subjects performed approximately one hour of training. To avoid learning effects, but still reach a stable level of performance and sufficient understanding of the information presented by the separation assistance interface, separate example scenarios were used for training.

The measurement phase consisted of 10 conditions, presented in a randomized block design. Subjects performed each scenario in both aircraft, resulting in 20 measurement trials per subject. The trials were combined in blocks of five continuous conflict scenarios. This meant that for each set of five scenarios, all participating aircraft were present in the same simulated airspace, during the course of the five trials. Aircraft that do not participate in the current conflict were placed at different flight levels, to avoid previous and future conflicts having an effect on the affordance space of the current trial. After each trial, subjects were asked to fill in a short questionnaire concerning their resolution decision.

		Boeing 707-300	Airbus A330
TAS_{min}	[kts]	282.4	331.1
TAS_{max}	[kts]	530.1	471.5
TAS_{cruise}	[kts]	485.0	432.0

Table 1: *Relevant data for the aircraft models in the experiment. The difference in cruise speed influences conflict geometry, and the reduced speed margin for the Airbus can limit the resolution possibilities.*

Table 2: *Rules and strategies for conflict resolution.*

1. Safety has the main priority: Ensure sufficient separation at all times.
2. Avoid resolutions that result in parallel tracks.
3. If available, apply the geometrically optimal solution.
4. When a unique optimal solution is not available, apply rules of the air.
    - 4a. An aircraft being overtaken has the right of way and the overtaking aircraft must remain clear by altering heading to the right.
    - 4b. When two aircraft are approaching each other head on they must both alter heading to the right.
    - 4c. Aircraft from the right have the right of way. Remain clear by passing behind that aircraft.

Subjects and Instructions to Subjects: Sixteen experienced glass-cockpit pilots participated in sets of two, 15 male, and one female. Experience in terms of flight hours per pilot ranged from $2,000$ to $16,700$ hours. Subjects were asked to perform an experiment, where they should resolve traffic conflicts in unmanaged airspace. They were informed that the results would be used to evaluate a concept for a separation assistance interface. To avoid "gaming" effects, (e.g., pilots creating, or prolonging conflicts on purpose), pilots were not informed that there was a second participant, and that they were, in fact, flying against a human "opponent". Instead, they were told that during the measurements, intruder aircraft could participate in the resolution of a conflict, by using certain automated logic.

Prior to the experiment, pilots received a short briefing on the geometrical concepts behind the display, how to use the display, and on the experimental setup. An important aspect of this briefing was to instruct the pilot on the rules and strategies for conflict resolution, see Table 2

Dependent Measures: Dependent measures for this experiment consisted of several objective and several subjective measures. Objective measures were the *solution choice per pilot* in terms of vector change dimensions (heading and/or speed), and applied tactic (optimal state change vs. rule of the air), and the level of cooperation between pilots. *Safety* was measured in terms of minimum separation, and the initial reaction time was used as a measure of *performance*. These measures were constructed from recorded parameters position, heading, and selected speed and heading. Subjective measures consisted of online SA questions, and a post-experiment questionnaire.

Experiment Hypotheses: Several studies involving manual (horizontal) conflict resolutions found that pilots prefer to keep velocity constant (Hoekstra, 2001; Steens, van Dam, van Paassen, & Mulder, 2008). It was therefore hypothesised that the majority of the maneuvers would be heading-only. It was also hypothesised that conflict geometries with a small, non-zero expected CPA distance result in the largest amount of opposing resolutions, as the choice between the optimal solution and applying rules of the air is less clear for such conflicts. Conflict geometries where $d_{CPA} = 0$ will show more coordination based on the rules of the air, whereas conflict geometries with large expected CPA distances will mostly be solved implicitely, where pilots use the shortest-way-out principle.

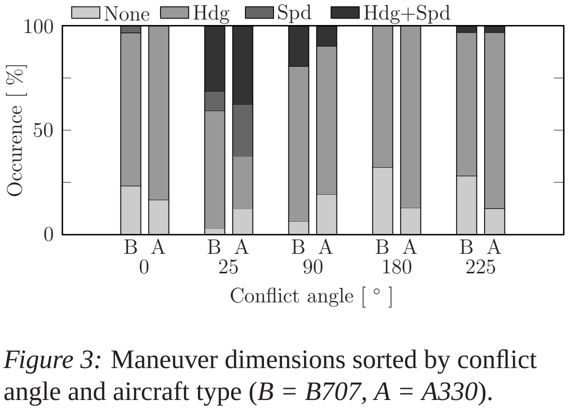


Figure 3: Maneuver dimensions sorted by conflict angle and aircraft type (*B = B707, A = A330*).

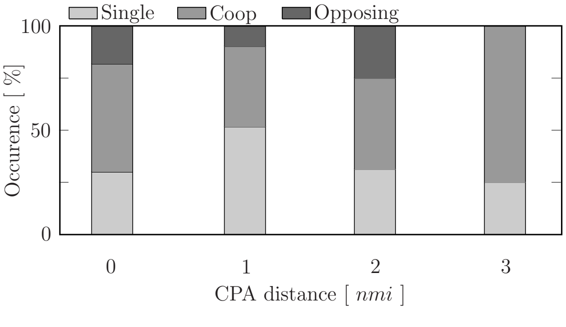


Figure 4: Level of cooperation between pilots sorted by conflict CPA distance.

Results

The resolution maneuvers in the experiment can be grouped by the flight parameters that were changed to resolve each conflict. For horizontal conflict resolution the available maneuver options are heading and speed changes. Therefore, solution choice is a categorical measure with four levels: *no action*, *heading only*, *speed only*, and *combined heading and speed*. The selection of a maneuver will depend on conflict geometry, aircraft performance limitations, phase of flight, and personal or airline preference. Table 3 shows the maneuver choice average for the entire experiment. As was hypothesised, the majority of the resolution maneuvers was heading only (almost 70 %), which can be attributed to personal or airline preference (Hoekstra, 2001). Figure 3 shows the maneuver choice sorted by conflict angle and aircraft type. For conflict angles $0°$, $180°$, and $225°$ this figure shows that (nearly) no speed maneuvers were used. These conflict angles result in (near) head-on or parallel (take-over) courses. In these situations, speed changes have no effect other than speeding or delaying a loss of separation, and only heading changes can be used to resolve such conflicts. A notable exception to the preference for heading resolutions is found in the $25°$ conflict angle scenarios, especially for the A330 (61% of the resolutions involved a speed change). In this situation, large heading changes are required to resolve the conflict. Also, for the A330, the max. speed line hides the tip of the intruder triangle, making it difficult to detect intruder intent, and impossible to determine the correct coordination rule. Giving way to the intruder by slowing down might then be considered the safest course of action.

Figure 4 and Table 4 show the level of cooperation between pilots, by CPA distance, and on average, respectively. Table 4 shows that pilots selected opposing solutions in 16% of the measured trials. This can be a matter of insufficient training, but it can also be an indication of a weakness of the interface. The most prominent cause for the opposing solutions was that the two pilots applied different rules: 64% for scenarios where $d_{CPA} = 0$, and 45% for scenarios where $d_{CPA} \neq 0$. These are errors where the wrong rule is applied. For all values of d_{CPA} this can be an indication that pilots could not reliably retrieve the required information from the display. In other cases, the correct rule was applied, but an error was made while evaluating the rule (7% for $d_{CPA} = 0$, and 55% for $d_{CPA} \neq 0$). In scenarios where $d_{CPA} = 0$, the direction of the maneuver depends on a previously stored 'rule of the air'. Therefore, when a wrong maneuver is made in such a scenario, it is because the pilot did not remember the applicable rule correctly. For scenarios where $d_{CPA} \neq 0$, the rule requires the direction of the maneuver to be derived from the display. In such scenarios, an erroneously applied rule can also be an indication that pilots could not retrieve the required information from the display.

50% of the measured trials were solved cooperatively. Figure 4 shows that this occured most frequently for scenarios with the largest conflict CPA distance. In situations where d_{CPA} is large, the velocity vector of ownship is close to the edge of the constraint zone belonging to the conflict. The optimal solution (the shortest way out of the triangle) is clearly visible on the interface, and guarantees implicit coordination when both parties strive for minimum path deviation. The scenarios with the smallest, non-zero CPA distance showed the lowest percentage of cooperation. In these situations, the optimal solution is less evident, and the choice between applying the optimal solution or applying the rules of the air becomes less clear.

The minimum separation was used as a measure of safety, by comparing the measured value to the defined separation minimum. The separation minimum was violated in 3 out of 160 measured trials. In all three cases, this occurred during a premature return to nominal heading and speed, and in all cases, the incursion was minimal. Reaction time was used as a measure of performance, but showed no significant variation across conditions.

As a subjective measure, pilots were quizzed randomly from a set of traffic awareness questions during the experiment. Although most questions were answered correctly, there are two notable exceptions. When asked whether the other aircraft was slower or faster than the own aircraft, pilots gave more unsure and wrong answers in conflict scenarios where the tip of the conflict zone (which also indicates the tip of the intruder velocity vector) was not visible on the display. Another question that was often answered wrongly was whether or not the other aircraft participated in the resolution maneuver. This cue is visible from the movement of the conflict zone on the display, which can be difficult for pilots to see without extra visual cues. Results from the post-experiment questionnaire also identify this as the most important issue with the display.

Table 3: *Percentages maneuver choice.*

None	16.67 %	Heading only	68.91 %
Speed only	3.85 %	Heading and speed	10.58 %

Table 4: *Percentages pilot cooperation.*

Single pilot solution	33.97 %
Cooperative solutions	50.00 %
Opposing solutions	16.03 %

Conclusions

A multi-actor, traffic separation experiment was performed, to evaluate the coordination of manual resolution maneuvers between actors in traffic conflicts in unmanaged airspace. Similar to previous studies, results from the experiment showed a considerable preference for heading-only maneuvers. As expected, difficulties with implicit coordination between actors in a conflict occurred for conflict geometries that do not clearly fall into a single category of coordination rules.

In the experiment, sixteen pilots participated, in pairs of two. For practical reasons, this is already a considerable amount of subjects. However, because of the nature of most of the measurements (i.e., categorical data with uneven expectations for the outcome per category), sufficient statistical power in the data requires a sample size closer to 50 groups, or 100 pilots, possibly even more. Because an experiment of this magnitude is difficult to realize, a follow-up study has been initialized that employs pilot decision models in a Monte-Carlo simulation, in an effort to identify the influence of behavioral characteristics on separation coordination and safety.

Acknowledgements

The authors gratefully acknowledge the pilots that participated in this study, and would like to thank MSc students W. M. G. Pas and R. W. Jutte for their help in carrying out the experiment. 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

Bilimoria, K. D. (2000). A geometric optimization approach to aircraft conflict resolution. In *Aiaa guidance, navigation, and control conference and exhibit.*

Ellerbroek, J., Visser, M., van Dam, S. B. J., Mulder, M., & van Paassen, M. M. (2011, Sep.). Design of an Airborne Three-Dimensional Separation Assistance Display. *IEEE Transactions on Systems, Man, and Cybernetics, part A: Systems and Humans*, *41*(6).

Heylen, F. M., van Dam, S. B. J., Mulder, M., & van Paassen, M. M. (2008, Aug.). 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.

International Civil Aviation Organization. (1996). *Rules of the Air and Air Traffic Services (Doc 4444)* (Tech. Rep.).

Lee, J. D., & Moray, N. (1992). Trust, Control Strategies and Allocation of Function in Human-Machine Systems. *Ergonomics*, *31*(10), 1243–1270.

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, *39*(2).

RTCA. (2002, Mar.). *Airborne Conflict Management: Application Description V2.5* (Tech. Rep. No. RTCA SC-186). Federal Aviation Authorities.

Schild, R., & Kuchar, J. K. (2000). Operational Efficiency of Maneuver Coordination Rules for Airborne Separation Assurance System. In *3rd USA/Europe Air Traffic Management R&D Seminar Napoli.*

SESAR Consortium. (2007, Sep.). *SESAR Definition Phase D3: The ATM Target Concept* (Tech. Rep. No. DLM-0612-001-02-00). Eurocontrol.

Steens, C. L. A., van Dam, S. B. J., van Paassen, M. M., & Mulder, M. (2008). Comparing Situation Awareness for two Airborne Separation Assistance Interfaces. In *AIAA Guidance, Navigation and Control Conference and Exhibit.*

van Dam, S. B. J., Mulder, M., & van Paassen, M. M. (2008, Nov.). 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.

HUMAN PERCEPTION OF FLASHING LIGHT EMITTING DIODES FOR AIRCRAFT ANTI-COLLISION LIGHTING

Chris Yakopcic, John Puttmann, Benjamin R. Kunz, Mark Holleran, Brandon Wingeier, Ali Hashemi, Kenneth Stapp
University of Dayton
Dayton, OH

Light Emitting Diodes (LEDs) have the potential to replace incandescent bulbs currently used in aircraft anti-collision lighting. LEDs require less power to operate, and possess the ability to flash without the addition of moving parts. Compared to incandescent bulbs, however, LEDs yield a slightly different spectral output and a different intensity profile when flashing. The impact of these differences on the viewer's ability to detect the light was examined to determine if LEDs can successfully replace incandescent bulbs on aircrafts and runways. Using an automated system to drive an LED with variable intensity and duration, the light source was displayed to naïve participants to establish visibility thresholds for solid and pulsed LEDs. Participants were asked to immediately respond as to indicate if a light was present. The data collected were examined and applied to the different models currently available for determining the effective intensity of a pulsed light.

In recent years, designs for aircraft anti-collision lighting have incorporated Light Emitting Diodes (LEDs) due to their low power consumption compared to incandescent bulbs. LEDs also have a longer operating life and do not require moving parts to flash. As the technology for anti-collision lighting is updated, the old standards for intensity and pulse width must be verified to ensure that LEDs are detected by the human eye equivalently to existing alternatives. LEDs differ from incandescent bulbs for two main reasons. The spectral output of an LED is significantly different than that of an incandescent bulb. A white incandescent bulb produces a yellowish output and most white LEDs produce a white that has a much stronger blue component. Second, the intensity profile, or pulse shape of a flashing LED is rectangular, and the rotation of an incandescent bulb (or the mechanism surrounding it) produces a rounded pulse shape. The purpose of this experiment was to measure the effective intensity of LEDs, and to compare the results with the values produced by the analytical models for determining effective intensity.

Effective intensity was defined by Ohno and Couzin (2002) as the luminous intensity of a steady light source that has the same visual range as the pulsed light in question. The current standard for measuring the effective intensity of flashing lights was proposed by A. Blondel and J. Rey in 1911. This experiment involved human subjects viewing a lamp housed in a contraption with a rotating disc that provided for the flashing of the light. Since then, the technology used to develop aircraft anti-collision lighting has changed considerably, yet the equation proposed by Blondel and Rey is still used as the standard.

The Blondel-Rey equation has been evaluated both experimentally and analytically for use with LEDs. In addition, other models have been proposed as alternatives for measuring the effective intensity of a pulsed light source known as the Allard method, and the form-factor method (Ohno and Couzin 2002). Ohno and Couzin (2002) conducted a theoretical study of the models used for determining effective intensity and proposed the modified Allard method as a more accurate alternative for multi-pulse flashing lights. J. D. Bullough et al at Rensselaer Polytechnic Institute conducted a study where different colored LEDs were compared to incandescent lights to see if participants detected a difference in perceived intensity between the two types of light sources (2006). In an experiment conducted by Nurmi (2004), a flashing and a steady light were presented individually to participants. The absolute threshold for detection was established for participants as a function of the size of the aperture through which the light was displayed and the amount of ambient lighting.

The experiment presented in this paper involved displaying both a steady and a pulsed LED at intensities that are believed to be in a range above and below the absolute threshold of detection for a dark adapted human subject. As opposed to a flashing light, a single pulse was presented to the subject to allow for a better comparison to the Blondel-Rey and Allard models for effective intensities, and provided a greater amount of control during the experiment. This experiment was unique because it was completed using two different methods for measuring detection thresholds, and the results for each of the methods were compared. The work presented is the result of a project sponsored by the FAA with a student team in the Engineering Innovation Center at the University of Dayton.

Methods for Modeling Effective Intensity

The models for effective intensity that were studied in this paper were the Blondel-Rey equation and the Allard method. When the LED pulse shape is square, the form-factor method exactly matches the results produced by the Blondel-Rey equation. Also, the modified Allard method was meant to approximate the Blondel-Rey method for a square pulse shape. Since the pulse shape of the LED in this experiment was square, the form-factor and modified Allard methods were not considered as they will produce the same results as the Blondel-Rey equation.

The Blondel-Rey equation can be seen in (1), where *I(t)* is the intensity profile of the pulse output from the LED, and *a* is the visual time constant that was experimentally determined to be 0.2s by Blondel and Rey. For a square pulse, *t1* and *t2* are the rising and falling edge of the pulse respectively. The value I_{eff} is the effective intensity of the light pulse in question.

$$I_{eff} = \frac{\int_{t_1}^{t_2} I(t)dt}{a+(t_2-t_1)} \tag{1}$$

The Allard method is based on a convolution of the intensity profile of the pulsed LED, and visual impulse response as defined by Allard. The equations for the Allard method can be seen below in (2) and (3), where *I(t)* again represents the intensity profile of the light pulse. The function *i(t)* is the convolution of *I(t)* and the visual impulse response *q(t)*, which is defined in equation (3). The visual time constant is also set to 0.2s in this equation, and the effective intensity is defined as the maximum of *i(t)*.

$$i(t) = I(t) * q(t) \tag{2}$$

$$q(t) = \frac{1}{a}e^{-\frac{t}{a}} \tag{3}$$

Experimental Method

Materials and Apparatus

For testing, participants were seated 50 feet away from the LED apparatus. The LED was white with a value of (0.301, 0.293) on the CIE chromaticity diagram. Curtains were hung to remove the possibility of reflection from the walls. The LED was housed in a matte black box with circular baffles to reduce the scattering of the observed light. A chinrest was used to ensure the subjects were looking in the direction of the LED. To obtain intensity values that encompassed a test subject's threshold, neutral density filters were used to reduce the light output of the LED to the order of micro-candelas (µcd). The intensity of the LED was varied in a wide range thought to encompass each subject's threshold intensity, and the subject responses were used to precisely determine each subject's threshold for the steady and pulsed LED.

A MATLAB program was developed that automatically controlled the LED that was observed by the test subjects. The MATLAB script was capable of controlling the pulse width to a resolution of 2 milliseconds. The intensity of the LED was determined by the voltage output of a data acquisition unit that was controlled by the MATLAB script. The MATLAB script fully automated the testing procedure for displaying the LED and collecting the user responses using input obtained from a "yes" button and a "no" button placed below the subject's right and left had respectively. The script for this experiment was capable of generating LED signals at two different pulse widths 5 seconds, and 226 milliseconds. The 5 second pulse width was decided to be large enough for the light to be considered steady, as the effective intensity of a 5 second pulse with normalized intensity was determined to be 0.962 using the Blondel-Rey equation, and .994 using the Allard method.

Procedure

Participants first completed a visual acuity test using the standard Snellen chart. Only participants with 20/30 or better vision (with corrective lenses, if necessary) were included in subsequent analysis. Following the task instructions, participants were seated, the lights were shut off and dark adaption began, lasting for 10 minutes. At this point the pretest was administered so that the subject could practice using the yes and no buttons and observe the LED before testing.

After the pretest, the experiment was divided into four parts. The first part used the method of limits (MOL) to determine the subject's threshold. On the first trial, the LED was presented at the lowest intensity in the test range. On each subsequent trial, the LED intensity increased incrementally until the maximum intensity was reached. On the next trial, the highest intensity was displayed with LED intensity decreasing incrementally on each subsequent trial to complete a cycle. On each trial, participants pressed the “yes” button if they detected the lights and the “no” button if they could not detect the light. Each trial consisted of a 5 second display of the LED followed by a 1.5 second break period. The subject had the ability to respond during the 5 seconds when the LED was on, but not during the break period. In the MOL test there were three complete cycles for a total of 114 trials, (19 unique intensities each presented 6 times).

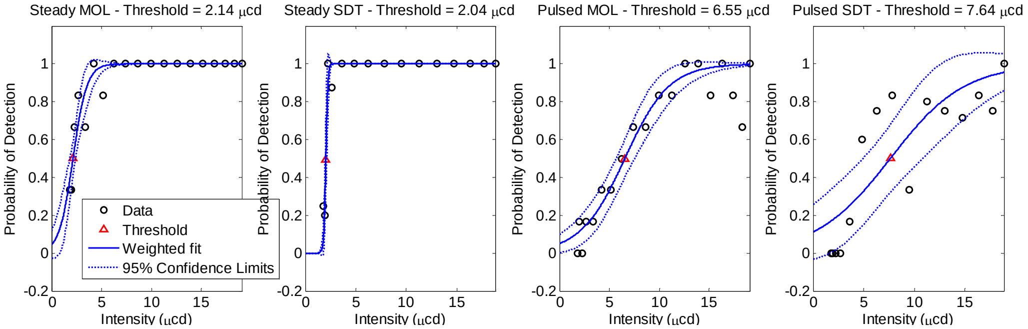
The second part of the test was based on experimental methods used in signal detection theory (known as the SDT test from here on), and presented 16 different intensities in a random order. Each intensity value was presented a maximum of 10 times but with a probability of 0.5 that a blank trial (0 intensity) would occur. Although the order of the intensities was random, the ten trials associated with that intensity were presented in a group so that the blank trials could be specifically associated with each of the intensities. This provided data that would determine the subject bias for each of the intensities, via calculation of ROC curves.

Parts three and four had the same procedure as the first two, except that the LED had a pulse width of 226 milliseconds. The subject still had 5 seconds to respond as to whether they could see the light, followed by a 1.5 second break. After this data was collected the pulse data was compared with the steady light data to determine the effective intensity.

Results

The threshold for each subject was calculated individually using a logistic regression, and then the overall data was compared to the Blondel-Rey and Allard equations. Figure 1 shows an example of the data that was collected for a single subject. The results for each of the four tests are displayed as the probability of correctly detecting the LED as a function of LED intensity. The plot in Figure 1(a) shows the results for the steady LED for the MOL test, and the plot in Figure 1(b) shows the data for the steady SDT test. The plot in Figure 1(c) shows the results for the MOL using the short pulsed LED and the plot in Figure 1(d) shows the results short pulse SDT test. In each of the plots, the raw data is presented along with the logistic regression and the 95% confidence limits. The threshold point is interpolated as the point where the subject correctly detects the light 50% of the time.

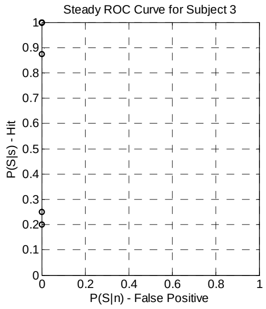
The data in Figure 1, depicting results from a single participant, represents desirable results as both of the testing methods (MOL and SDT) produced similar threshold values for a steady LED. Likewise, the threshold obtained from the pulse tests was also similar when comparing the methods. The shorter pulse test yielded a higher threshold than the steady LED, which is the result expected when comparing to the theoretical methods for calculating effective intensity in equations (1 - 3).



(a) (b) (c) (d)

Figure 1. Example of data collected for a single test subject (known as participant 3) that provided a desirable result. The plots show the subjects responses for (a) the steady light using MOL test, (b) the steady light using the SDT test, (c) the pulsed light using the MOL test, and (d) the pulsed light using the SDT test.

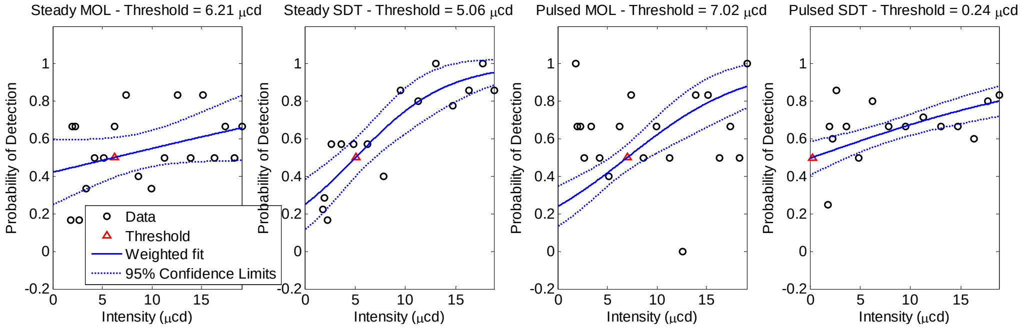
The receiver operator characteristic (ROC) curves were also plotted for each of the subjects. ROC curves for participant 3 are shown in Figure 2, where each data point represents a different intensity. Although Green and Swets (1988) caution against generating ROC curves when signal intensity is the variable, insight into the subject’s response bias that could not be obtained when looking at the data in Figure 1. Figure 2(a) shows the ROC curve for the steady LED and Figure 2(b) shows the curve for the pulsed LED. The data in Figure 2 shows that this subject completed the test with no false positives, as each of the data points resides in the region of zero false positive probability. This means that the subject did not respond yes to any of the blank trials, which further validated the clean threshold data displayed in Figure 1.



(a) (b)

Figure 2. ROC curves generated for the SDT data from test participant 3. The plots display the ROC data for (a) the steady SDT test and (b) the pulse SDT test.

Figures 3 and 4 display the test results for a different test subject to show how the results different greatly between subjects. It can be seen that the separation in the threshold between methods differs greatly for the pulsed LED because there was not enough data below a 50% probability to generate an accurate threshold. The data is also much more variable when compared to the previous subject in Figure 1. As indicated by the ROC curve in Figure 4, there were a notable number of false positives (where a participant reported detecting a light when none was presented). This participant comparison shows the variability between subjects encountered during data collection.

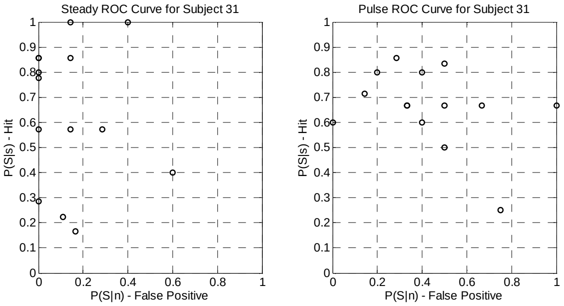


(a) (b) (c) (d)

Figure 3. Example of data collected for a single test subject that provided an undesirable result. The plots show the subjects responses for (a) the steady light using MOL test, (b) the steady light using the SDT test, (c) the pulsed light using the MOL test, and (d) the pulsed light using the SDT test.

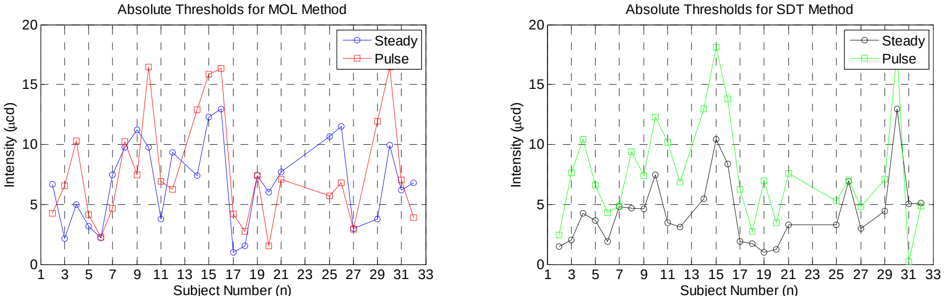
A 2 (test method: MOL or SDT method) x 2 (LED pulse width; steady or pulse) repeated measures ANOVA performed on threshold values revealed significant main effects of test method, $F(1, 25) = 10.13$, $p = .004$, LED pulse width, $F(1, 25) = 13.52$, $p = .001$ and an interaction between test method and LED pulse width, $F(1, 25) = 12.86$, $p = .001$. Planned paired-sample t-tests indicated significant differences between thresholds for pulse LEDs and steady LEDs obtained using SDT method, $t(25) = -5.95$, $p < .001$ and between thresholds for steady LEDs determined using SDT and steady LEDs determined using MOL, $t(25) = 4.40$, $p < .001$. Notably, there was no significant difference between thresholds for pulse and steady LEDs determined using MOL ($p = .23$).

The expected difference between thresholds for pulsed vs. steady LEDs was only apparent for thresholds using the SDT method. The plot in Figure 5(a) shows the threshold value that was determined using the MOL test for both the steady and pulsed LED, and the plot in Figure 5(b) shows the thresholds obtained using the SDT test. Each of the thresholds is plotted in micro-candelas as a function of the subject number. Of the 32 participants tested, 6 were omitted because threshold data could not be obtained from their responses. The threshold for the pulsed LED ($M = 7.74$, $SD = 4.40$), was nearly always higher than the constant LED ($M = 4.45$, $SD = 2.87$). The lone exception, subject 31, was described in Figures 3 and 4 in which data indicated a large amount of false positives. The difference between pulse (*M* = 7.79, *SD* = 4.64) and steady LED thresholds (*M* = 6.87, *SD* = 3.54) was much smaller for MOL thresholds. A one-sample t-test suggests that the difference between pulse LED thresholds and steady LED thresholds was significant only for SDT data (*t*((25, test value = 0) = -5.95, *p* < .001).



(a) (b)

Figure 4. ROC curves generated for the SDT data from test participant 31. The plots display the ROC data for (a) the steady SDT test and (b) the pulse SDT test.



(a) (b)

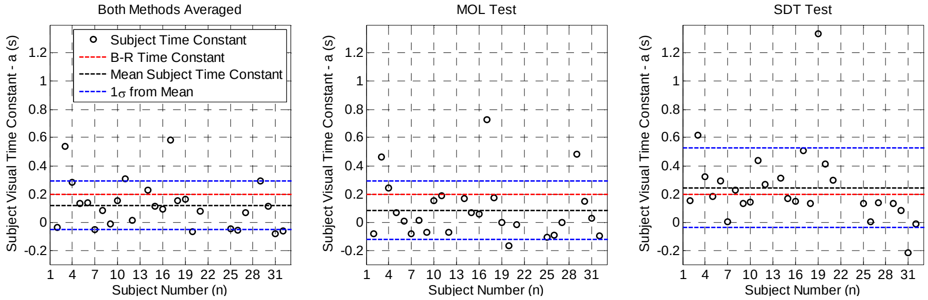
Figure 5. Threshold data determined for each subject using (a) the MOL test and (b) the SDT test.

Calculation of Effective Intensity and Model Comparison

The data was analyzed in relation to both the Blondel-Rey and the Allard method for determining effective intensity. First, the threshold data collected was taken as a ratio of the threshold for the steady light to the threshold for the pulsed light. The ratio was taken using just the MOL data and just the SDT data. Then a third ratio was calculated that was the average of the thresholds of the two methods for each subject. The ratios calculated are as follows: R_{MOL}=0.882, R_{SDT}=0.575, R_{TOTAL}=0.730. The ratios were also calculated by using the Blondel-Rey and Allard equations using a pulse width of 226ms. The intensity of the pulse is not important as the ratio is independent of pulse intensity in either case because effective intensity scales linearly with the peak intensity of the pulse. The ratios were calculated as follows: R_{BR}=0.531, R_{AL}=0.677. It can be seen that the SDT test provided data that is closest to the ratio produced by the Blondel-Rey equation, and the total data was a closer match to the data provided by the Allard method.

Since the pulse shape in question was a square wave, the integral of the light pulse could be simplified to a multiplication of the pulse width and the amplitude. Then, the equation was rearranged to calculate the visual time constant based on the experimental data collected. The peak of the pulse *I(t)* was assumed to be the threshold for the pulsed light and I_{eff} was assumed to be the threshold calculated for the steady light. Figure 6 shows the visual time constants that were calculated for each subject when the thresholds from each method were averaged in Figure 6(a), when just using the MOL test thresholds in Figure 6(b) and when just using the SDT test thresholds in Figure 6(c). The new visual time constants were calculated as follows: a_{TOTAL}=0.119, a_{MOL}=0.085, a_{SDT}=0.245. A t-test analysis indicated no statistically significant difference between the value obtained using the SDT data and the visual time constant proposed by Blondel and Rey, *t*(25, test value = 0.2) = .80, *p* = .43. Data from the MOL test provided a time constant that was statistically different from 0.2 seconds, *t*(25, test value = 0.2) = -2.82, *p* = .009.

$$I_{eff} = \frac{\int_0^T I(t)dt}{a+\Delta T} \rightarrow I_{eff} = \frac{\max(I(t))\Delta T}{a+\Delta T} \rightarrow a = \frac{\max(I(t))\Delta T - I_{eff}\Delta T}{I_{eff}} \qquad (4)$$



(a) (b) (c)

Figure 6. Visual time constant determined for each subject using the result in equation (4), where the plots display the result for (a) the average of both methods, (b) the result for the MOL test, and (c) the result for the SDT method.

Conclusion and Future Work

This paper compared visual thresholds for both a steady and a pulsed light source using two different methods for establishing absolute thresholds. We conclude that the SDT method provided results more consistent with the analytical models for effective intensity, as the threshold for the pulsed light was consistently greater than threshold for the steady light. When examining the results of the MOL test, there was no clear relationship between the threshold intensity and type of signal presented to the participants. Moreover, the time-constants derived using data collected using the SDT test more closely resembled the values predicted by the Blondel-Rey equation than those obtained using the MOL test.

More work must be completed before these adjustments to the effective intensity modeling equations can be accepted. Since only one pulse width was presented to subjects in this experiment, the comparison to the effective intensity modeling equations can only be related using one data point. In the future multiple pulse widths will be tested to verify that the modeling equations match the experimental data as a function of the LED pulse width.

Acknowledgements

Research reported in this paper was conducted under the Flight Deck Program Directive / Level of Effort Agreement between the Human Factors Research and Engineering Group (AJP-61), FAA Headquarters and the Aerospace Human Factors Division (AAM-500) of the Civil Aerospace Medical Institute. The authors of this paper would like to acknowledge Dr. Nazih Khaouly, Dr. Andy Mead, and Dr. Nelda Milburn of the FAA for their support in this project. Also the authors would like to thank Professors John Hageman, Dr. Phil Doepker, Dr. Bill Moroney, and Dr. Joe Haus for their insightful discussions and academic support.

References

Blondel, A., & Rey , J. (1911). Sur la perception des lumieres breves a la limite de leur portee. *Journal de physique thtorique et appliqute., 1*, 530-550.

Bullough, J. D., Yuan, Z., & Rea, M. S. (2006). *Brightness/Luminous Intensity Values for Blue, White and Green Aviation Signal Lights Using Light Emitting Diodes*. Troy, NY: Lighting Research Center, Rensselaer Polytechnic Institute.

Green, D. M., & Swets, J. A. (1988). *Signal Detection Theory and Psychophysics*. California: Peninsula Publishing.

Nurmi, T. (2004). The Determination of the Effective Light Intensity of LEDs. In *Physical Measurement and Visual and Photobiological Assessment*. Tokyo, Japan: CIE Expert Symposium on LED Light Sources.

Ohno, Y., & Couzin, D. (2002). Modified Allard Method for Effective Intensity of Flashing Lights. In *Temporal and Spatial Aspects of Light and Coulour Perception and Measurement, Expert Symposium*. Veszprem, Hungary: Commission Internationale de L'Eclairage.

ALTITUDE-EXTENDED SOLUTION SPACE DIAGRAM FOR AIR TRAFFIC CONTROLLERS

J. Lodder, J. Comans, M.M. van Paassen, M. Mulder
Delft University of Technology, Faculty of Aerospace Engineering
Kluyverweg 1, 2629HS Delft, The Netherlands

The solution space diagram was developed to assist air traffic controllers and pilots in dealing with traffic. Up until now, it has been used to solve conflicts in the horizontal plane. Especially in the context of Air Traffic Control, it is important to also include the vertical dimension. This paper describes an approach to incorporate this vertical dimension in a two dimensional display. The altitude extended solution space diagram will be calculated taking into account the Altitude Relevance Bands of all aircraft involved. In this way, the algorithm can discard conflict zones that can never lead to a conflict. Based on this algorithm, a display prototype has been developed that is able to show the effect of altitude changes to the controller. This display will be used to perform an evaluation experiment to assess the benefits of including altitude information.

The solution space diagram (Figure 1) has been introduced to assist pilots and air traffic controllers (ATCos) to deal with handling traffic situations (Dam, Abeloos, Mulder, & Paassen, 2004; Velasco, Mulder, & van Paassen, 2009). The diagram presents a visualization of the heading and speed constraints imposed by traffic surrounding traffic. Such a constraint-based approach to interface design was inspired by the Ecological Interface Design (EID) framework (Vincente & Rasmussen, 1992). By showing the constraints, instead of showing a predefined solution, an operator can see all the boundaries of his operational envelope. Based on this the operator can make an informed decision on how to handle a particular situation.

A key task in the Air Traffic Control (ATC) domain is merging a number of aircraft at a specific waypoint (Hermes, Mulder, van Paassen, Boering, & Huisman, 2009). A number of aircraft enter an ATCos sector, and have to leave at a specific waypoint without getting into conflict with each other. In an ATC context, a conflict is defined as a situation that will lead to loss of separation. In other words, a conflict occurs when an aircraft is on a trajectory that brings it within a specified minimum distance from an other aircraft. The distance requirement can be split into a horizontal and a vertical requirement. In the vertical plane, aircraft must be spaced by at least $1000\ ft$. In the horizontal plane the minimum distance is between 3 and $5\ Nm$. Both requirements can be combined to define the protected zone (PZ). The PZ is a volume of airspace surrounding an aircraft in the shape of a hockey puck with a radius of 3 to $5\ Nm$ and a thickness of $2000\ ft$.

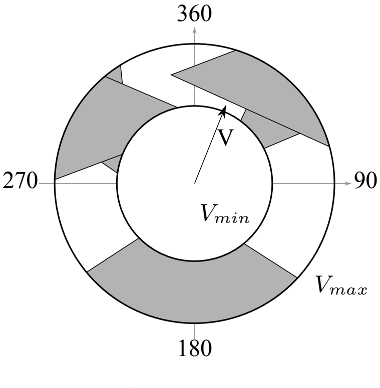


Figure 1: The Solution Space Diagram

The solution space presents the constraints imposed on an aircrafts velocity by the horizontal part of the conflict zone in a velocity diagram as shown in Figure 1. It shows which combinations of speed and heading will eventually lead to a loss of separation. The diagram is constructed by first calculating the velocities that will lead to a conflict with each surrounding aircraft, called the intruding aircraft. The calculated conflict zones are then clipped by an annular section that has an internal radius equal to the minimum velocity, V_{min}, and an outer radius equal to the maximum velocity, V_{max}, of the controlled aircraft. The annular section represents the full performance envelope in the horizontal plane, the gray areas represent the subset of this envelope that leads to a conflict. In this way, an ATCo can see how how the traffic surrounding an aircraft under observation affect the instructions that can be given.

A drawback of the solution space is that it is only presenting conflicts in the horizontal plane. Flying, on the other hand, is a three dimensional activity. This vertical component becomes especially important in climb and descent maneuvers. When only aircraft on the same altitude are shown on the display, it cannot be used during climb and descent maneuvers. When all aircraft are shown, regardless of altitude, the display will provide false conflicts.

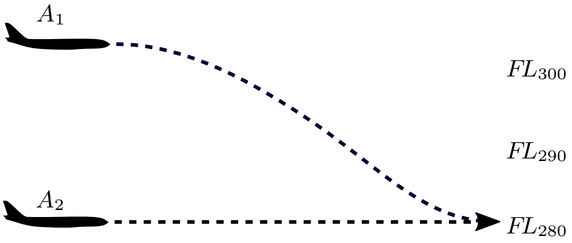


Figure 2: Two aircraft involved in a descent

Consider the situation in Figure 2. When an SSD would only be calculated taking into account traffic at the same altitude, A_1 would start a descent without knowing about A_2. If the speed of both A_1 and A_2 are approximately equal, A_1 would fly straight into the protected zone of A_2. This would only show up once A_1 has actually entered the protected zone.

Treating all aircraft as if they were on the same altitude would put A_2 in the SSD of A_1, but it would also indicate a conflict because both A_1 and A_2 are in the same horizontal position. Wether or not the situation remains a conflict depends on the velocities of A_1 and A_2. If the difference in velocity is large enough, A_1 would end up either in front or behind A_2 without loss of separation. To avoid this, the conflict zones will need to be calculated taking this into account. This procedure will be explained in the Estimated Overlap Section.

The goal of this research is to develop a display that incorporates information in the vertical plane on a solution space display in the context of ATC. This paper will first introduce the procedure to decide which surrounding aircraft are relevant during a vertical maneuver. Next, a technique to determine the time interval during which the conflict zone for a specific intruder is valid will be discussed. The final section describes the resulting display that will be used to evaluate the altitude extended SSD.

Altitude Based Filtering

In order to make sure an aircraft can be allowed to climb or descend, the ATCo has to verify that there will be no conflicting traffic interfering with the maneuver. The aircraft that could potentially interfere can be determined by a technique called Altitude Based Filtering.

The first step in altitude based filtering is to compute an Altitude Relevance Band (ARB). The ARB is the altitude interval in which an aircraft will move during a vertical maneuver as shown in Figure 3. One side of the interval will be determined by the current altitude of the aircraft, the other end of the interval is defined by the altitude the aircraft is climbing or descending to. When an aircraft is not performing a vertical maneuver, the ARB has no thickness and is equal to the current altitude.

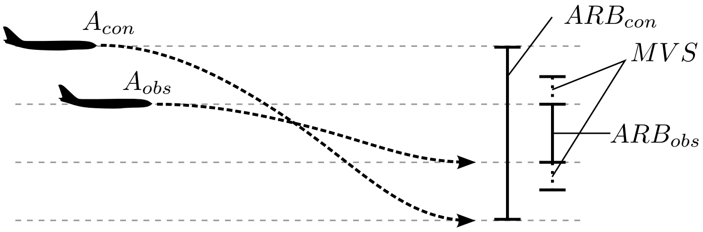
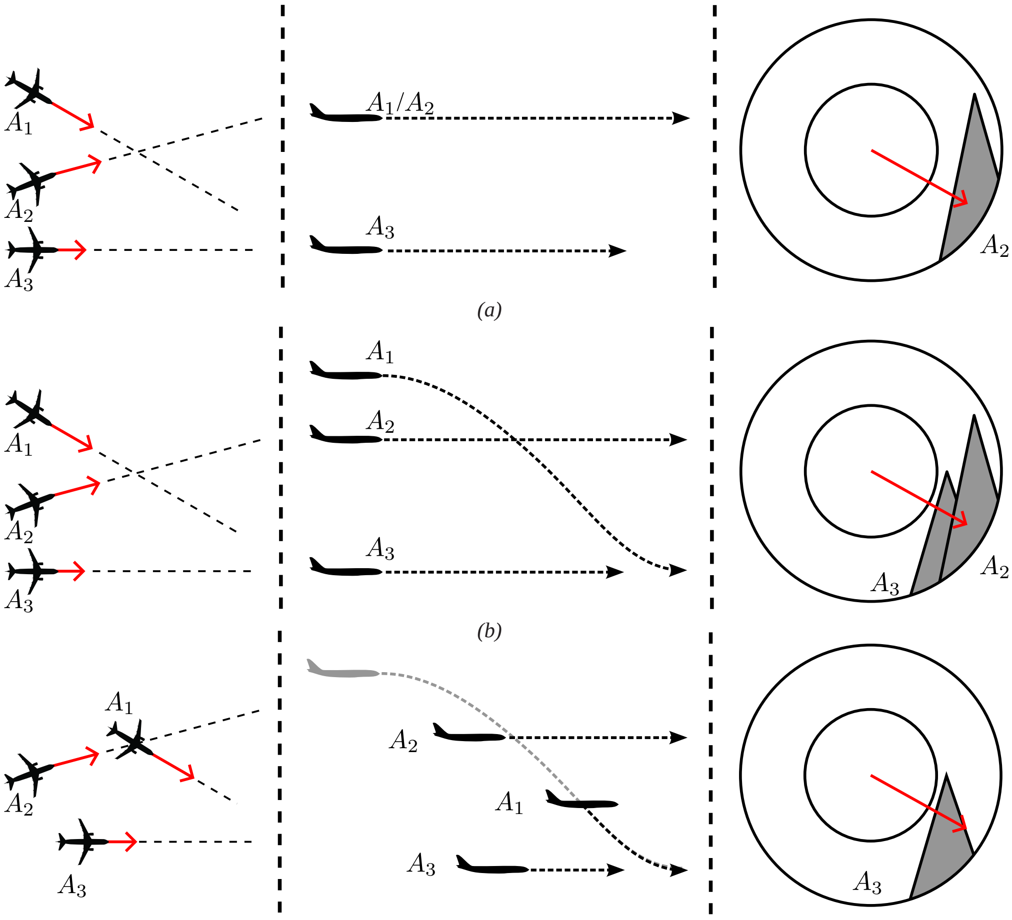


Figure 3: Definition of the Altitude Relevance Band

The second step is to add the minimum vertical separation to the ARBs of the observed aircraft. In the context of the solution space, the observed aircraft are the aircraft surrounding the aircraft for which the solution space is being calculated.

The final step is to determine which aircraft have overlapping ARBs. These will be the aircraft that could potentially get into a conflict. Figure 4 shows an example of using altitude based filtering to draw the solution space. Figure 4 (a) shows the horizontal and vertical situation of two aircraft on the same altitude and one on a different altitude. The right column shows the solution space calculated for aircraft A_1. Since no vertical maneuvers are performed, only A_1 and A_2 can be in conflict. In the solution space diagram, only the conflict zone of A_2 shows up.

Figure 4 (b) shows the situation when A_1 would start a descent. In this case, A_1 is crossing altitudes of A_2 and A_3. This results in both conflict zones being drawn in the solution space diagram. When the situation progresses, A_1 will have descended below A_2. At this moment, A_1 will not be able to get into conflict with A_2 anymore and only the conflict zone of A_3 will be drawn on the solution space.



(c)

Figure 4: Altitude based filtering example

Estimated Overlap Time

Altitude based filtering only looks at the ARBs of aircraft to determine if there could be conflicts. This effectively gets rid of all aircraft which will never be on the same altitude and can never be in conflict. After this procedure, there can still be aircraft left which will never get into a conflict with the controlled aircraft. Consider a controlled aircraft at 30000 ft that needs to descend to 15000 ft. There could be an observed aircraft that will be at the exact same location in 60 s, but at 15000 ft. Since it is physically impossible for any commercial aircraft to descend 15000 ft in 60 s it will never be possible to get in conflict even though the ARBs are overlapping.

When performing vertical maneuvers, an aircraft crosses all intermediary altitudes between its current altitude and its required altitude. The time at which an aircraft crosses a certain altitude depends on two main factors. The vertical speed and the time at which the aircraft will start its descent. This time is mainly driven by ATC. When the ATCo gives a command, the pilot will initiate his maneuver. There might be some delay between receiving a clearance and executing the maneuver. In the best case scenario, the delay can be close to zero, in the worst case, it might be in the order of a few minutes. Next to this unknown time delay, the actual rate of climb or descent is also unknown. As with the time delay, it should be possible to make assumptions about the fastest and slowest maneuvers for a specific situation.

Based on these time delay and vertical speed intervals, a time versus altitude diagram can be plotted as shown in Figure 5. This diagram is created by computing the fastest and slowest descent. The diagram shows the evolution of altitude with time. At t_{0_f} a controller issues a command. The fastest descent, which has no time delay and maximum vertical speed, starts immediately and can be seen as the left line in the diagram. After the maximum delay, at t_{0_s}, the slowest maneuver with the lowest vertical velocity is initiated. This is represented by the right line.

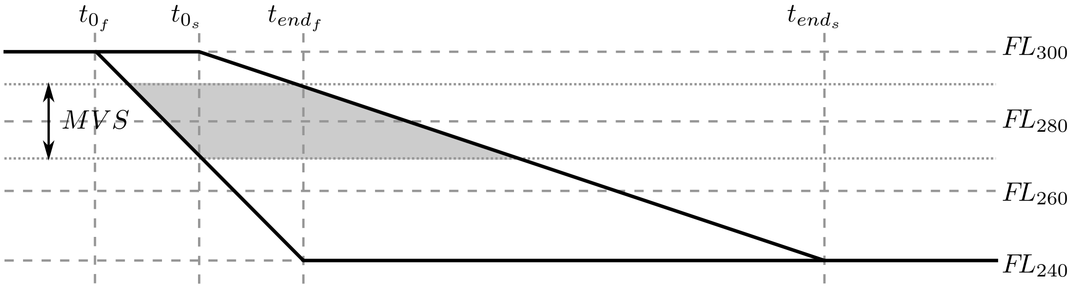


Figure 5: Time-altitude diagram for a single aircraft

The time-altitude diagram immediately shows the estimated time interval during which an aircraft will be on a given altitude. When taking into account the minimum vertical separation discussed before, a prediction of the time interval during which a conflict can occur can be estimated. An example of this is shown in Figure 5 by the gray area. The gray area represents the relevant combination of time and altitude for an aircraft flying at 28000 ft taking into account a minimum vertical separation of 1000 ft. The lowest and highest time value of the gray area determine the relevant interval when crossing an aircraft flying at constant altitude.

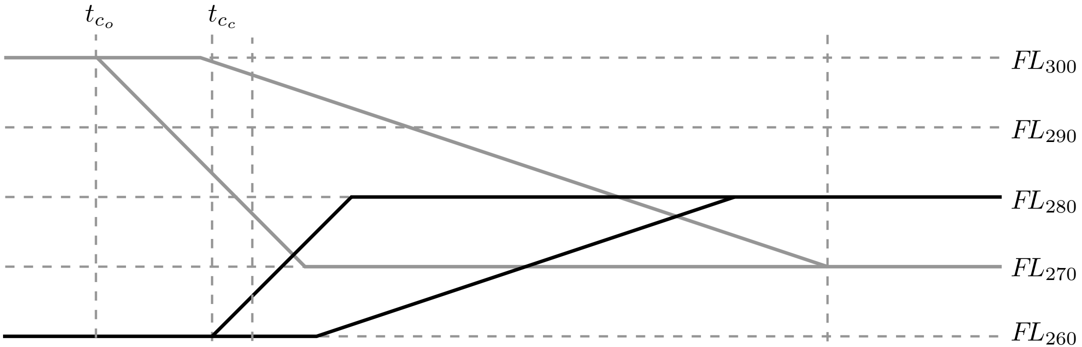
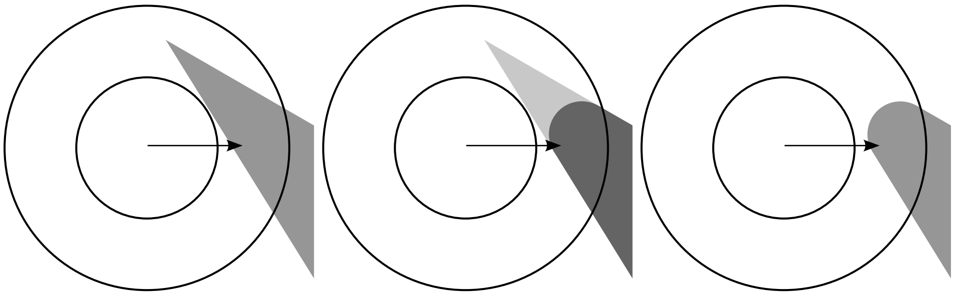


Figure 6: Time-altitude diagram for a climbing and descending aircraft

Figure 6 shows an example of a time altitude diagram for a situation where traffic is not maintaining altitude. While the higher observed aircraft is descending, the controlled aircraft is climbing. The earliest possible conflict time, t_e, occurs when the distance between the fastest descent line and the fastest climb line become smaller than the minimum vertical separation. The latest possible conflict time, t_l, is determined by the point at which the vertical distance of the slowest profile becomes larger than the minimum vertical separation.

Based on the predicted conflict time interval, the solution space can be truncated. In its most basic form, a conflict zone shows all conflicts that can occur in a time interval from 0 s to ∞ s. The time it will take until loss of separation takes place depends on the position of the velocity vector within the conflict zone. The shorter the distance to the tip, the longer it will take. If, for example, the velocity of the controlled aircraft is exactly at the tip of the solution space, it will fly with exactly the same velocity as the observed aircraft. Therefore, both aircraft are flying in parallel and will never move closer. In other words, it will take an infinite time to get a loss of separation. Moving the velocity just a little into the conflict zone will result in a small relative velocity which will gradually bring the two aircraft closer. The further the velocity is moved into the conflict zone away from the tip, the higher the relative velocity becomes and the faster the aircraft will enter each others protected zone. Based on this principle, it is possible to truncate the conflict zone based on a time interval. In this way, only the relevant part of the conflict zone will show up in the solution space.

Figure 7 shows an example of the truncation process. Figure 7a shows the full conflict zone for a conflict ranging up to infinity. In this case, the conclict zone is a sharp triangle. As explained before, the tip of the triangle corresponds to the velocity of the observed aircraft and represents a conflict at infinity. Decreasing the range of the conflict time interval will result in a situation shown in Figure 7b. The original conflict zone is shown in light gray while the remaining part is shown in darker gray. Because of the circular nature of the protected zone, the endcap of the truncated conflict zone will also be circular. The end result of the truncation process is shown in Figure 7c.



(a) Complete Conflict Zone *(b)* Part to be truncated *(c)* Resulting truncated Conflict Zone

Figure 7: Truncation of the conflict zone

Interface prototype and proposed experimental evaluation

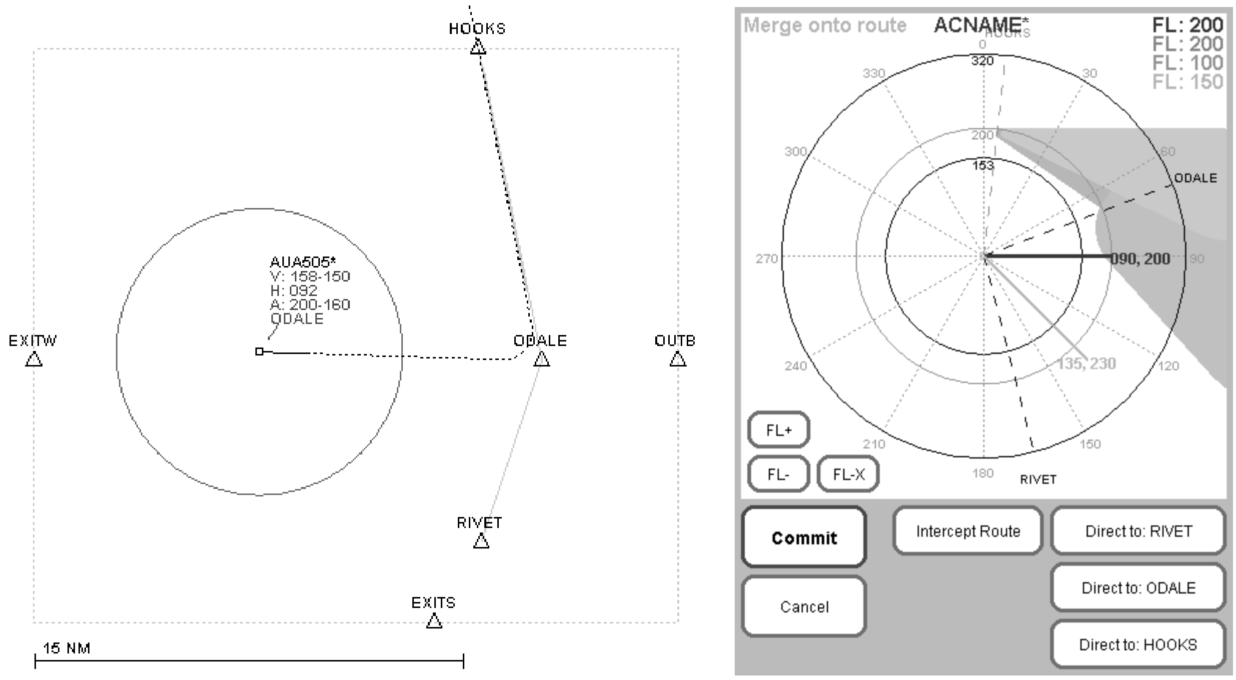
To evaluate the altitude based filtering method, an ATC simulation has been developed that incorporates the altitude extended solution space display. The simulation consists of a standard plan view display, Figure 8 (a) and a solution space diagram, Figure 8. The controller can select an aircraft in the plan view display and give the selected aircraft heading and speed commands in the solution space display like in previous experiments (?, ?). The controller can press the FL- and FL+ buttons to inspect the effect of issuing a vertical command. Once the controller is satisfied he can commit his commands and the aircraft will change its trajectory.

An evaluation experiment will be conducted to investigate the effect of including altitude information in the solution space diagram. Six subjects will take part in the experiment. They will control four different scenarios, two with a low traffic level, two with a high traffic level. The scenarios will be flown with the solution space visible or not visible. This will result in four combinations of high & low traffic and solution space on & off. At one minute intervals the test subjects will be prompted to rate their experience workload level on a scale of 1 to 5.

After each experiment condition the test subjects will be asked to fill in a questionnaire. This questionnaire aims to investigate what elements of the simulator aids in alleviating workload and generating a mental picture of the traffic situation.

Several performance metrics will be calculated from the gathered data. These metrics are for example number of separation losses during a run, number of aircraft delivered at their requested exit condition, distance of aircraft traveled through the sector versus optimal travel path and number of commands given during an experiment run.

The combination of the workload measurements, questionnaire results and performance metrics will be used to investigate the areas where the display can be improved.



(a) Plan View Display

(b) SSD Display

Figure 8: Experiment display

Conclusions

This paper presented a technique to improve the solution space diagram to assist Air Traffic Controllers in planning vertical maneuvers. Altitude based filtering was used to determine which aircraft will never be able to get in conflict with each other. By calculating an estimate of the earliest and latest time an aircraft can reach an altitude, the conflict zones can be truncated to remove even more irrelevant information.

These techniques were used to develop a simulation which will be used to conduct an evaluation experiment. This experiment will use scenarios with varying complexity to assess the benefits of an altitude extended solution space display.

References

Dam, S. Van, Abeloos, A., Mulder, M., & Paassen, M. van. (2004). Functional presentation of travel opportunities in flexible use airspace: an EID of an airborne conflict support tool. In *Systems, man and cybernetics, 2004 ieee international conference on* (Vol. 1, p. 802 -808).

Hermes, P., Mulder, M., van Paassen, M. M., Boering, J. H. L., & Huisman, H. (2009, November). Solution-Space-Based Analysis of the Difficulty of Aircraft Merging Tasks. *Journal of Aircraft*, *46*(6), 1995-2015.

Velasco, G. Mercado, Mulder, M., & van Paassen, M. (2009, August 2-5). Analysis of Air Traffic Controller Workload Reduction Based on the Solution Space for the Merging Task. In *Proceedings of the AIAA Guidance, Navigation and Control Conference.* Toronto, Canada.

Vincente, K. J., & Rasmussen, J. (1992). Ecological Interface Design: Theoretical Foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, *22*(4), 589–606.

PORTRAYING THE PASSAGE OF TIME IN A TIMELINE INTERFACE FOR SUPERVISORY CONTROL

Terry Stanard, Patrick Dudenhofer, Sarah Spriggs, Gloria Calhoun, Lamar Warfield
Air Force Research Laboratory, 711 HPW/RHCI
Wright Patterson AFB, Ohio

Heath Ruff
Ball Aerospace & Technologies Corp.
Fairborn, Ohio

Timeline displays are a promising user interface concept for supervising multiple unmanned vehicles. The essentials of a timeline display include a timescale, current time indicator, and lines overlaid with symbology to represent the duration and timing of events such as refueling, transit to destinations, and mission durations. Little research has investigated alternatives to portraying time and events. The present study investigated two ways to portray the passage of time: a moving timeline with a stationary current-time indicator, and a stationary timeline with a moving current-time indicator. Participants viewed videos of dynamic timelines portraying the passage of time these two ways, and also at two speeds of movement. While watching the videos they answered questions requiring information retrieval and temporal problem solving. Participants answered more quickly and accurately with the moving timeline condition at the faster movement speed, particularly on questions requiring planning decisions.

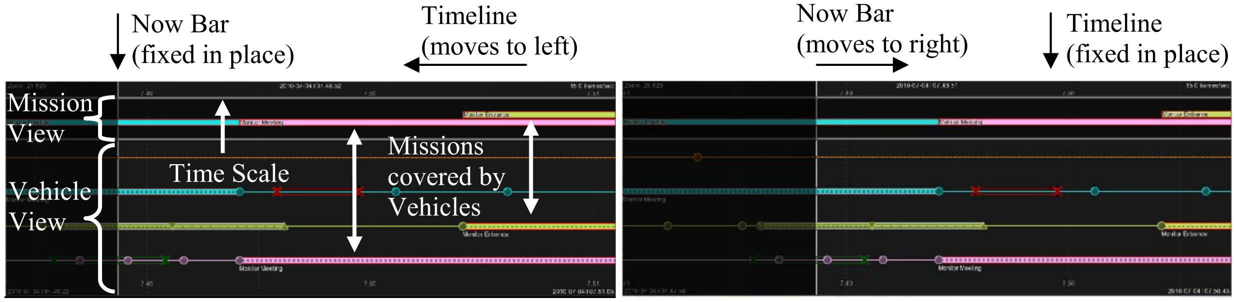
Recent advances in automation have enabled unmanned vehicles to operate semi-autonomously. In this context, vehicle operators will shift away from actively controlling the vehicles to supervising multiple vehicles that essentially pilot themselves. This new work setting will be challenging, requiring operators to continually monitor missions as they are performed, problem solve, and dynamically re-plan missions in response to changing conditions in the environment and situation (Scott, Mercier, Cummings, &Wang, 2006).

To effectively manage multiple unmanned vehicles, operators will need interface technology providing effective visualizations of vehicle operations. Command and control displays have traditionally included map-based displays that allow operators to monitor manned or unmanned vehicles in geospace. However, situation awareness and proactive management of unmanned assets may also be improved through interfaces that portray the temporal dimension. Limited research supports this view. A relatively simple temporal display was found to facilitate scheduling performance better than a typical geospatial display in a simulated weapon-target scheduling task for a manned, single-ship naval warfare scenario (Rousseau, Tremblay, Lafond, Vachon, & Breton, 2007). A coordinated suite of timeline visualizations was found to improve the speed and quality of dynamic replanning solutions in an Air Force airlift operations center (Scott, Roth, Truxler, & Wampler, 2009). Temporal displays specific for supervisory control of multiple unmanned assets have also been designed. Hanson, Roth, Hopkins and Mancuso (2004) describe a task-based temporal interface for interacting unmanned vehicles teams that supports synchronization and analysis of mission plan changes. Cummings and Mitchell (2005) designed a decision support display that included timelines for each of four vehicles tasked with destroying time-sensitive targets. Color-coded blocks arranged along a timeline depict scheduled tasks (e.g., battle damage assessment, mission events, waypoints, windows of opportunity).

The Air Force Research Laboratory (AFRL) is designing a temporal interface that includes a timeline-based representation of missions and resource constraints for multiple unmanned vehicles (see Figure 1). The timeline window features a time scale at the top and a vertical “now bar” marking the current time. In the mission view, each horizontal bar represents the start and end times for a mission and the bar color denotes vehicle assignment. In the vehicle view, a horizontal line is devoted to each vehicle with symbology overlays pertaining to the tasking and status of vehicle (e.g., refueling needed, ready to deploy, performing a mission).

Unfortunately, there are few research results and design guidelines for timeline displays to direct our efforts. As a start in addressing specific design issues related to our temporal interface development, we investigated the orientation of a timeline display, finding an advantage for a horizontal over a vertical orientation for providing temporal situation awareness (Spriggs, Warfield, Calhoun, & Ruff, 2010). The study reported here addresses another design decision: how to portray the passage of time in a timeline display.

The majority of timeline designs to date use a moving timeline with a fixed now bar to represent the current time (Figure 1A). The mission and event symbology move from right to left, passing through the now bar as the events actually occur. The fixed now bar creates a stationary frame of reference for the current time, which means the user can always glance at the same place on the display to see what events are imminent. However, there is also rationale for a stationary timeline with a now bar that moves from left to right (Figure 1B). A stationary timeline may make it easier to retrieve information and engage in mission planning, because the symbology remain fixed and only the now bar moves. Decisions such as which vehicle to assign to a new mission might be easier for an operator if the timeline and symbology representing planned events are fixed in place on the display. Of course, the now bar would eventually move off the screen, and a mechanism would be needed to reset the placement of the now bar so it remains continuously visible.



(A) Timeline Moving/Now Bar Stationary (B) Timeline Stationary/Now Bar Moves

Figure 1. Two ways of showing the passage of time in the AFRL timeline display. Picture A illustrates a moving timeline with a stationary now bar, while Picture B shows a stationary timeline with a moving now bar.

The utility of a moving versus a stationary timeline may be a function of how fast the displayed symbology moves. With a display area of fixed size, movement speed can be manipulated by varying the time span represented on the time scale: the beginning of the time period (left side of the format) and end of the time period (right side of the format). When the displayed time span is larger (for example, six minutes or greater), changes in the location of displayed symbology from moment to moment are relatively small. When the time span is shorter (for example, three minutes or lower), changes in the location from moment to moment of displayed symbology, including text, occurs more quickly. Time spans under 10 minutes are relevant for supervising small unmanned vehicles.

When the time span is three minutes and motion is more rapid, we predict speed and accuracy in judgments concerning temporal events will be poorer with a moving timeline compared with a stationary timeline. The quicker movement of the symbology could make it more difficult to cull information of interest, compared with the stationary timeline, where the timeline symbology is fixed in place on the display and only the now bar moves. When the time span is six minutes, we predict there will be little difference between the stationary and moving timeline in how quickly and accurately participants judge temporal events. This is because the perceived speed of symbology movement is predicted to be negligible.

Method

Experimental Design

Sixteen volunteer Air Force employees participated in this study (10 men, 6 women, mean age = 36.6 years). All participants reported having vision correctable to 20/20 and normal color vision. Trials were blocked by movement type (moving timeline, stationary timeline). Within a movement type, participants experienced both movement speed conditions before experiencing the other movement type. The order of movement type was counterbalanced across participants and the order of movement speed was counterbalanced within movement type. Within each combination of movement type and movement speed, participants completed 8 trials, each with a different temporal format drawn from Spriggs et al. (2010). During each trial they answered 4 types of questions. Question order was not counterbalanced but each type of question appeared an equal number of times in each ordered position across trials. This resulted in a 2 (Movement Types) x 2 (Movement Speed) x 4 (Question Types) x 8 (Replication/Format) within-subjects design. Participants answered a total of 128 questions within 32 trials.

Movement Type. The passage of time was represented by either a moving timeline and timescale (with a stationary now bar), or a stationary timeline and time scale (with a moving now bar). Because symbology would normally emerge onto the format in the moving timeline condition as time elapsed, special measures were taken to try to equate the amount of visible symbology between the movement type conditions over the duration of a trial. The format appeared the same in either condition at the start of the trial. In the moving timeline condition, no new symbols appeared at the far right portion of the format as the timeline moved to the left. Missions that didn't end within the trial period were displayed as a single, thick line without overlaid symbology.

Movement Speed. Speed was manipulated by changing the duration of the visible span of time, with the shorter time scale (164 s or 2:44 min) producing faster movement of symbology (2.63 mm/s or 9.73 pixels/s) compared with the longer time scale (368 s or 6:08 min) which produced a slower speed (1.17 mm/s or 4.34 pixel/s). A format appeared the same at the start of the trial in both speed conditions, the only difference being the span of the timescale represented.

Question Types. Each of the eight formats provided sufficient information to answer four different types of questions (see Table 1). Each question involved one or more of three levels of situation awareness described by Endsley (1995). The types of questions are representative of some but not all envisioned operations in supervisory control of multiple unmanned vehicles. During presentation of a format, participants answered four questions (one question per question type). Questions appeared at varying intervals from the start of the trial (either 15, 20, 25 s since trial start). The period of time between subsequent questions was varied (20, 25, 30 s). Participants had 20 s to answer each question before it was removed and counted as a miss. Only one question was visible at a time. The duration of the trials/videos varied between formats (110, 115, 120 s) depending on the time when questions appeared and the period of time between questions.

Table 1

Question Types Posed with Temporal Formats, with Examples of Each Question Type.

Mission Planning: Given a situation, what vehicle would be best to perform the mission? "Which vehicle would you send to cover the 'Escort Governor' mission?", "If vehicle 4 lost comm., which vehicle would you send to cover its missions?"
Counting: How many missions, events, vehicles are planned? "How many vehicles are currently performing missions?", "How many vehicles are currently deployed?"
Next Event: What event will occur next? "What vehicle will land next?", "Which vehicle will next begin a new mission?"
Event Time: What time will an event occur? "What time will vehicle 3 land?", "What time will vehicle 2 enter restricted space?"

Procedure

After completion of their informed consent to participate, participants reported background information and received a briefing on the temporal format symbology and experimental conditions. Next, two training trials were completed with the first assigned combination of movement type and speed. Participants were then shown their accuracy scores, as well as a briefing which explained, for all trials, why each candidate response was either correct or incorrect. Following this training, the first block of 8 experimental trials was conducted for the assigned movement x speed combination. These procedures were repeated for the three other blocks of movement type x movement speed trials.

At the start of a trial, the participant observed a video showing a dynamic temporal format presented on a 24 in diagonal flat screen liquid crystal display (1920 x 1200 pixels). During the playback of the video, the participant answered each of the four questions as they appeared by typing a number on the keyboard keypad. For three of the four question types, the correct answer was always a single digit. Event Time category questions required entry of a clock time, and the participant typed a six digit number conforming to hours, minutes, seconds (example: 120345 for 12:03:45). Response time for each trial was measured from the time the question appeared until the answer was entered. Participants did not receive feedback on their performance during the experimental trials. After the end of the trial, the experimenter configured the software for the next trial.

To reduce the likelihood that participants retained knowledge of format features, a fifteen-minute session was conducted between the second and third trial blocks that consisted of training and trials involving a different temporal display task on a touch screen monitor (not reported in this paper). After completion of all experimental trials, a questionnaire was administered asking participants to compare the two timeline movement conditions and movement speeds in terms of their ability to retrieve information and maintain situation awareness. Total session time per participant, including the fifteen minute alternate temporal display task, was approximately 2.5 hr.

Results

Accuracy and response time in answering questions during the presentation of dynamic timeline formats were analyzed as a function of movement type, movement speed, and question type. Overall, subjects were able to answer questions before the 20 s timeout nearly 100% of the time in both timeline movement conditions. The replication variable was found to have a high measure of internal consistency. For example, response time replication data had a Cronbach's alpha = .980. As a result, data were collapsed across replications. The resulting performance data sets (accuracy and response time) were analyzed with a repeated measures Analysis of Variance (ANOVA) model.

Accuracy

Data for the percentage of questions answered correctly failed tests of normality and were subsequently transformed using root-arcsine. The data plotted are untransformed means. Average accuracy differed as a function of timeline movement type (*F*(1, 15) = 8.53, *p* = .011). More questions were answered correctly when the timeline moved (93.6%) compared with a stationary timeline (91.0%). Accuracy also differed as a function of movement speed (*F*(1, 15) = 12.88, *p* = .003), with questions answered more accurately when the timeline symbology moved at the faster speed (94.1%) compared with the slower speed (90.4%). The effect of question type was also significant (*F*(3, 45) = 30.29, *p* = .000). The post hoc tests showed that Mission Planning and Event Time questions were answered correctly more often on average than Counting and Next Event questions.

There was a significant movement speed and question type interaction, *F*(3, 45) = 18.97, *p* = .000 (see Figure 2). The post hoc test showed that Next Event questions were answered correctly more often on average in the fast speed than the slow speed condition. The predicted interaction of timeline movement type and movement speed for the accuracy measure was not significant (*F*(1, 15) = 0.271, *p* = .610).

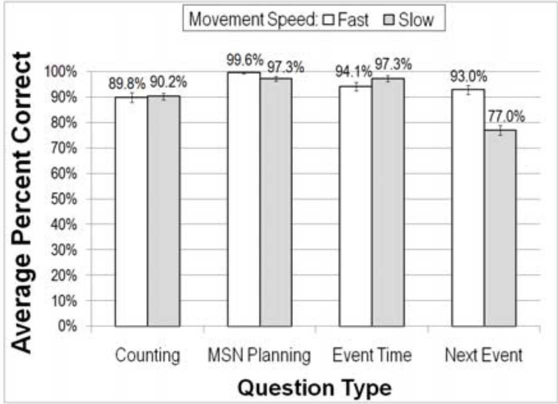


Figure 2. Average percent correct responses for each question type as a function of movement speed. Error bars are the standard error of the means. MSN is an abbreviation for Mission.

Response Time

Response time data also failed the normality test and were transformed using natural log. The data plotted are untransformed means. Average time to answer questions differed significantly based on question type, *F*(3, 45)

= 69.91, *p* = .000. Post hoc tests showed that response time was quickest with Mission Planning and Next Event questions, followed by Counting questions, and the slowest responses were with Event Time questions.

There was a significant interaction between movement speed and question type, *F*(3, 45) = 3.74, *p* = .018 (see Figure 3, left). Average response times were quicker on three of the four question types in the fast movement condition, notably Counting questions. There was also a significant interaction between movement type and question type, *F*(3, 45) = 2.91, *p* = .045 (Figure 3, right). In general, average response times to answer questions were quicker with the moving timeline. However, none of the post hoc tests for these interactions comparing the differences between question type and movement speed or movement type were statistically significant.

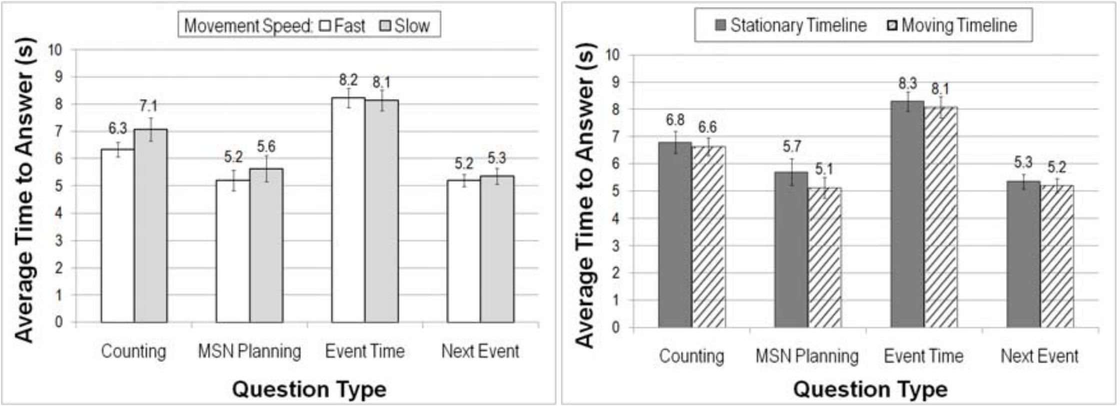


Figure 3. Response time to answer question types as a function of movement speed (left), and movement type (right). MSN is an abbreviation for Mission.

Discussion

In this study we predicted the type of timeline movement portraying the passage of time would impact speed and accuracy in answering questions about unmanned vehicle replanning situations and upcoming events. We evaluated two movement types. In one condition the timescale, vehicle events and mission events depicted on the timeline moved from right to left as time passed, and a vertical line denoting the current time (now bar) remained in a fixed place on the timeline format. In the other movement type, the timescale and events remained stationary in the format and the now bar moved from left to right as time passed. We predicted when the visual speed of movement on the timeline was relatively high, a moving timeline format would make it harder to retrieve information and interpret the timeline compared with a stationary timeline, where events remained in the same place and only the now bar moved.

Contrary to expectations, performance was superior with the moving timeline condition. Accuracy in answering questions was higher with a moving timeline. Response time was faster with the moving timeline condition, particularly on Mission Planning questions. One explanation for these results may be found in basic visual processes. The human visual system is very sensitive to motion and the interposition of objects may be easier to apprehend when there is apparent motion of objects against a fixed background (Dr. Scott Watamaniuk, personal communication, February 16, 2011). Watamaniuk noted there were multiple symbols in motion in the moving timeline condition and also multiple cues to a stationary reference. One intended reference was the now bar, but the surrounding interface and the Liquid Crystal Display panel itself also represented fixed, stationary frames. As a result, judgments such as the number and identity of missions and events, which ones will occur next, and which vehicles are available for new tasking, may be easier when there is an apparent visual flow of the timeline symbology against a stable background. In regards to the subjective data, participants' impressions were also aligned with the performance data. For instance, participants were asked to rank their preferences for each combination of movement type and speed of motion conditions. Inspection of the rankings suggests that movement type influenced rankings more than movement speed. Twelve of 16 participants indicated that having the timeline move was preferred over a stationary timeline.

Although there was no significant interaction of movement type and movement speeds, faster movement speeds produced superior performance. There did not appear to be a speed-accuracy tradeoff. Figure 2 shows accuracy was higher at the faster speeds, particularly with Next Event questions which were correctly answered 16% more often in the fast movement condition. Figure 3 shows response time was also quicker in the faster timeline speed compared with the slower speed, notably for Counting questions (delta=0.8 s) and Mission Planning questions (delta=0.4 s). We attribute this performance improvement to increased urgency and task engagement afforded by faster movement in both the moving timeline and stationary timeline conditions.

Other interesting findings are performance differences between question types. Mission Planning questions were answered significantly more quickly and accurately than others. This trend is opposite to what was found in a previous study looking at orientation of a timeline (Spriggs et al., 2010). The same format content and many of the same questions were used in both studies, and yet Spriggs et al. found the planning questions were answered least accurately and most slowly. We postulate three possible reasons for this. First, participants in the previous study had an average of 52 s exposure to each format across trials. In the current study, participants had approximately 8 min exposure to a particular format across trials. We believe Mission Planning questions to be relatively difficult to answer when exposure time is low, and the much greater exposure time in the present study assisted with diagnosing vehicle availability in order to answer questions. Secondly, the earlier study only used static formats without any visible movement, and so the presence of movement may improve ability to diagnose vehicle availability. Finally, the present study used a different question-answer paradigm. The timeline format was continuously visible during the question-answer process, while in Spriggs et al., the format was removed when participants indicated they were ready to answer the question.

A possible future study to explore these issues could replicate the current study, but vary exposure time and include a third timeline speed condition where there is very little visible movement in the timeline format (manipulated by a long time scale in the timeline). This study would help us to better understand whether it is the presence of movement, longer exposure time, or a combination of both which aid Mission Planning decisions. Results from such a study should provide insights into which method for portraying the passage of time is better, depending upon whether the supervisory control application uses a timeline as a primary display versus a secondary display. In the latter, information is retrieved via short glimpses which may mitigate any value of temporal flow.

Acknowledgments

We would like to thank Dr. Scott Watamaniuk for lending his expertise in interpreting results of this study.

References

Cummings, M.L. & Mitchell, P.J. (2005). Managing multiple UAVs through a timeline display. *Proceedings of the Information Technology @Airspace*, AIAA-2005-7060.

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

Hanson, M.L., Roth, E., Hopkins, C.M., & Mancuso, V. (2004). Designing human system interfaces for supervising multiple UAV teams. In *Proceedings of the 42nd AIAA Aerospace Sciences Meeting*, AIAA-200401305.

Rousseau, R., Tremblay, S., Lafond, D., Vachon, F., & Breton, R. (2007). Assessing temporal support for dynamic decision making in C2. In *Proceedings of the Human Factors and Ergonomics Society,* 1259-1262.

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 Society*, 699-703.

Scott, R., Roth, E., Truxler, R., & Wampler, J. (2009). Techniques for effective collaborative automation in dynamic air mission replanning. In *Proceedings of the Human Factors and Ergonomics Society*, 202-206.

Spriggs, S., Warfield, L., Calhoun, G., & Ruff, H. (2010). Orientation of a temporal display for multi-unmanned aerial vehicle supervisory control. *International Conference on Human-Computer Interaction in Aerospace (HCI-Aero)*, Cape Canaveral, FL.

VIGILANCE DECREMENTS IN A SUSTAINED ATTENTION TASK: EXAMINATION OF A MITIGATION STRATEGY

Guy A. French   Thomas R. Carretta
Air Force Research Laboratory
Wright-Patterson AFB, OH 45433

John K. Flach
InfoSciTex
Wright-Patterson AFB, OH 45433

A study was conducted to examine the effectiveness of perceptual and cognitive intervention tasks on mitigating vigilance decrements commonly observed in sustained attention tasks. Sixteen participants were randomly assigned to one of two experimental intervention conditions (perceptual or cognitive). Half of the participants completed a 45-minute "No Intervention" Control trial first, followed by one of the Intervention trials, also 45 minutes. The other half completed one of the Intervention trials first, followed by the No intervention trial. Following each trial, participants completed the SSSQ and the NASA TLX. As expected, a general decrease in objective performance over time was observed. However, contrary to expectations and prior research (St. John & Risser, 2007), the experimental intervention tasks did not reduce decrements in target acquisition performance over time, nor did they reduce subjective workload. The authors discuss methodological differences between the St. John and Risser study and the current study that may have contributed to differences in the effectiveness of the vigilance mitigation interventions in the two studies.

The capability to capture video and/or still imagery through sensors on remotely piloted aircraft (RPA) and fixed sensors has increased dramatically in recent years. The use of RPAs to support military operations is becoming increasingly widespread. RPAs are particularly well suited for performing missions requiring persistence due to their unrivalled endurance and low visual, radar and acoustic signatures. Their ability to search for, loiter over, and track ground targets for extended periods of time make them highly valued assets. For many long endurance RPAs, the duration of missions is such that operator fatigue becomes a critical factor (Wilson, Russell, & Caldwell, 2006). Operator fatigue also plays a role in the ability of sensor operators/image analysts when the task is to monitor multiple video feeds from fixed sensors for suspicious activities and potential targets.

Operator fatigue has been addressed by applying work shifts, changing out the crew mid mission at scheduled intervals (typically 8 hours). However, vigilance decrements begin to affect human performance long before physical fatigue begins to affect operators. This is especially true for sensor operators and image analysts performing long missions. These missions require personnel to monitor sensor video imagery for long periods of time for targets and suspicious activity. The low rate of occurrence of such activities over long periods leads to performance decrements in terms of reduced target acquisition rates and longer response times. By building a better understanding of the factors that contribute to these vigilance decrements, we can identify and evaluate technologies designed to address and mitigate factors that contribute to vigilance decrements and augment human performance.

Perceptual vigilance (Davies & Parasuraman, 1982; Molloy & Parasuraman, 1996)) has been studied for several years, resulting in the identification and characterization of performance decrements in sustained attention tasks. Theorists have attempted to explain vigilance decrements as a function of arousal/motivation (Vroom, 1964; Yerkes & Dodson, 1908), workload/multiple resource theory (Wickens, 2002), and other factors. However few effective mitigation technologies have been implemented to address the issue (Schroeder, Touchstone, Stern, Stoliarov, & Thackray, 1994; St John & Risser, 2007).

St. John and Risser (2007) examined the utility of perceptual and cognitive interventions for mitigating the vigilance decrement. Their approach combined aspects of arousal and resource theories (Arousal-Resource model). They introduced an intervention in the form of a secondary task designed to draw upon resources separate from those required by the primary visual vigilance task. The interventions were an auditory alarm “ring tone” that required sensory perception only and two auditory cognitive tasks that required participants to mentally reorder strings of 3 or 4 spoken digits. They hypothesized that the two cognitive digit task interventions would arouse participants, replenish depleted resources, and re-engage them in the vigilance task. They also hypothesized that the cognitive tasks would be more effective than the simple alarm because they were more demanding and engaging. Participants performed a 45 minute laboratory vigilance task twice, once in a control condition without any intervention and once with one of the three interventions. In the intervention conditions, participants received the intervention whenever they missed a target. This intervention method served as a proxy for a closed-loop system in which operators would receive interventions whenever low attention was detected by psychophysiological measures, prior to an actual miss. All three interventions significantly reduced misses by approximately 30%. Participants who showed greater vigilance decrements in the baseline (no intervention) condition showed more improvement from all interventions. That is, more vulnerable participants benefited most. The cognitive interventions performed as well as, but no better than, the simple alarm. The cognitive tasks also interfered with target detection performance on occasions when the interventions occurred while a target was viewable. However, the alarm was rated as more frustrating and less appropriate than either of the cognitive intervention.

The current study examined the utility of St John and Risser’s (2007) Arousal-Resource approach for mitigating the effects of vigilance decrements in a task where a single operator was required to monitor sensor feeds from two fixed remote sources.

Methods

Participants

Participants were 16 civilian and military employees (13 males, 3 females) stationed at Wright-Patterson AFB, OH. They ranged in age from 21 to 44, with a mean of 27.75 years. All participants reported being in good health, with normal visual acuity after correction and no problems with peripheral vision or color blindness. Most reported some prior simulator experience (56%) and video game experience (81%). All participants were volunteers and no compensation was provided for their participation.

Measures

Several types of data were collected. These included objective measures of perceptual vigilance ability and target acquisition performance (hits, misses, false alarms), demographic/background data, and subjective measures of mood/stress and workload.

Objective Experimental Task Performance. Objective measures of the experimental task performance were hits, misses, and false alarms.

Perceptual Vigilance Task (PVT). In this task (Temple et al, 2000) participants monitored the presentation of 8- by 6-mm light grey capital letters consisting of ‘**O**’, ‘**D**’, and a backwards ‘**D**’ centered on a video display screen. The letters were constructed in 24-point type using an AvantGarde font and were exposed for 40 milliseconds against a visual mask that consisted of unfilled circles on a white background. The participants task was to use the mouse to indicate when the target letter ‘**O**’ was presented. Responses were scored as hits, misses, and false alarms.

Biographical Questionnaire. This questionnaire collected information in order to characterize the sample and assist in interpretation of participants’ performance on the target detection task. Items elicited information about participants’ sex, age, general health, wellbeing, previous experience with simulator-type environments, previous experience with video games, and whether they had vision correctable to 20/20 acuity and normal peripheral and color vision.

Short Stress State Questionnaire (SSSQ). The SSSQ (Helton, 2004) is an abbreviated version of the Dundee Stress State Questionnaire (DSSQ; Matthews et at., 2002). It consists of four subscales: mood state, motivation, thinking style, and thinking content. Scores for three factors are derived from the subscales: task engagement, distress, and worry. Additionally, one component of the SSSQ assesses participants' perceptions of their physical and mental workload. One version of the SSSQ is administered before task performance and the other is administered after task performance.

NASA Task Load Index (NASA TLX). The NASA TLX (Hart & Staveland, 1988) is a subjective workload assessment measure that allows users to evaluate their interactions with various human-machine systems. A paper-and-pencil version which is part of the SSSQ was used to assess operator workload during the experiment. Participants rated their subjective workload on a scale that ranged from 0 (low) to 100 (High) for each of 6 subscales: Mental, Physical, Temporal, Effort, Performance, and Frustration An overall workload score based on a weighted average of ratings on the 6 subscales also was computed.

Procedures

Participants remained seated throughout the experiment, except during breaks. The study began with a pre-briefing regarding research objectives, procedures and informed consent. This was followed by administration of the Demographic Data Questionnaire. Next, participants completed an abbreviated version of the Dundee Stress State Questionnaire (DSSQ) called the Short Stress State Questionnaire (SSSQ) to establish a baseline of their stress level. Next, the 12-minute Perceptual Vigilance Task (PVT) was administered to estimate participants' perceptual vigilance ability. Scores from the baseline PVT were correlated with objective measures of performance on the experimental task to examine relations between the tasks. Following the short vigilance task, participants completed the SSSQ to assess changes from their baseline (pre-vigilance task) level. They also completed the NASA TLX to assess subjective workload. There was a 10-minute break following completion of the PVT and questionnaires.

Next, participants were randomly assigned to one of the two experimental intervention conditions (perceptual or cognitive). Half of them completed the "No Intervention" trial first, followed by one of the Intervention trials. The other half completed one of the Intervention trials first, followed by the No intervention trial. Following each target acquisition block, participants completed the SSSQ and the NASA TLX.

Those doing the No Intervention trial first completed a 3-minute practice session during which the target was 3 times more frequent than during the actual experiment. Task performance feedback was provided (i.e., hits, misses, and false alarms) during the practice trial. After a short break, the 45-minute experimental trial followed during which no performance feedback was given. Following the experimental trial, participants completed the post-trial SSSQ and NASA TLX to assess their subjective stress and workload, then took a 10-minute break before beginning the second experimental session. The procedures were similar for the Intervention conditions. For these conditions, participants began with a 3-minute practice session during which the targets occurred 3 times more frequently than in the experimental session. During the practice session, participants received feedback on target acquisition performance (i.e., hits, misses, and false alarms) and the intervention occurred 6 times. After a short break, participants completed a 45-minute experimental trial during which no performance feedback was given. The post-test SSSQ and NASA TLX were completed following the experimental session to assess participants' subjective stress (mood state) and workload.

The experimental task was to monitor simulated video feeds from 2 remote fixed sensors positioned to monitor traffic intersections in an urban setting and to designate targets/suspicious behaviors as they were detected. Participants were instructed to designate targets and respond to interventions, depending on the experimental condition. Figures 1 illustrates the displayed imagery as viewed by study participants. Each screen (left and right) displayed information for one of the two remote sensors.

Figure 1. Displayed imagery as viewed by study participants.

Analyses

In order to examine trends over time, the Perceptual Vigilance Task (PVT) and Experimental Task (ET) each were divided into 3 equal time intervals and objective performance (hits, false alarms) scores were calculated for each interval. Repeated measures analyses of variance were used to examine trends in performance over the 3 intervals. Correlational analyses examined the relations between the PVT and ET measures. Pre-PVT SSSQ scores served as a baseline by which to evaluate changes in subjective mood state following each of the Experimental Tasks (Control and Intervention conditions) using related samples t-tests. NASA TLX means were examined to compare subjective workload levels for the Control and Intervention conditions of the Experimental Task using related samples t-tests. All analyses used a .05 Type I error rate.

Results

Perceptual Vigilance Task

As expected, over time there was a trend toward a lower target acquisition (hit) rate (67.75%, 61.84%, and 62.28%) and a higher number of false alarms (7.50, 7.81, and 7.93) on the Perceptual Vigilance Task (PVT). However, neither of these trends was statistically significant. Further, contrary to expectations, performance on the PVT was not related to performance on the experimental task. The strongest relations were between the number of false alarms on the PVT and the number of false alarms on the control ($r = .287$, $p = .14$) and intervention ($r = .384$, $p = .07$) tasks. Examination of the pre- and post-PVT SSSQ scales revealed an increase on Distress (6.38 vs. 14.31, $t(15) = 14.17$, $p < .001$), but no difference for Engagement (24.13 vs. 22.63, $t(15) = 1.33$, ns) or Worry (6.38 vs. 5.75, ($t(15) = 0.74$, ns). Mean post-PVT workload scores were elevated for Mental (85.6), Temporal (71.9), and Effort (80.6) scales relative to the scale midpoint of 50. Mean Overall workload was 62.5.

Experimental Task

Results on the ET were mixed. Comparisons between the Control condition and each of the intervention conditions were not statistically significant for either the number of hits or false alarms. There was a significant overall effect for time interval for both hits ($F(2, 30) = 8.33$, $p < .001$) and false alarms ($F(2,30) = 3.82$, $p < .05$). However, neither trend showed a consistent decline in performance as was expected. The number of hits declined from the first to the second time interval, but recovered somewhat in the third interval. See Figure 1. The number of false alarms decreased from the first to the second interval (i.e., performance improved), but increased in the third interval.

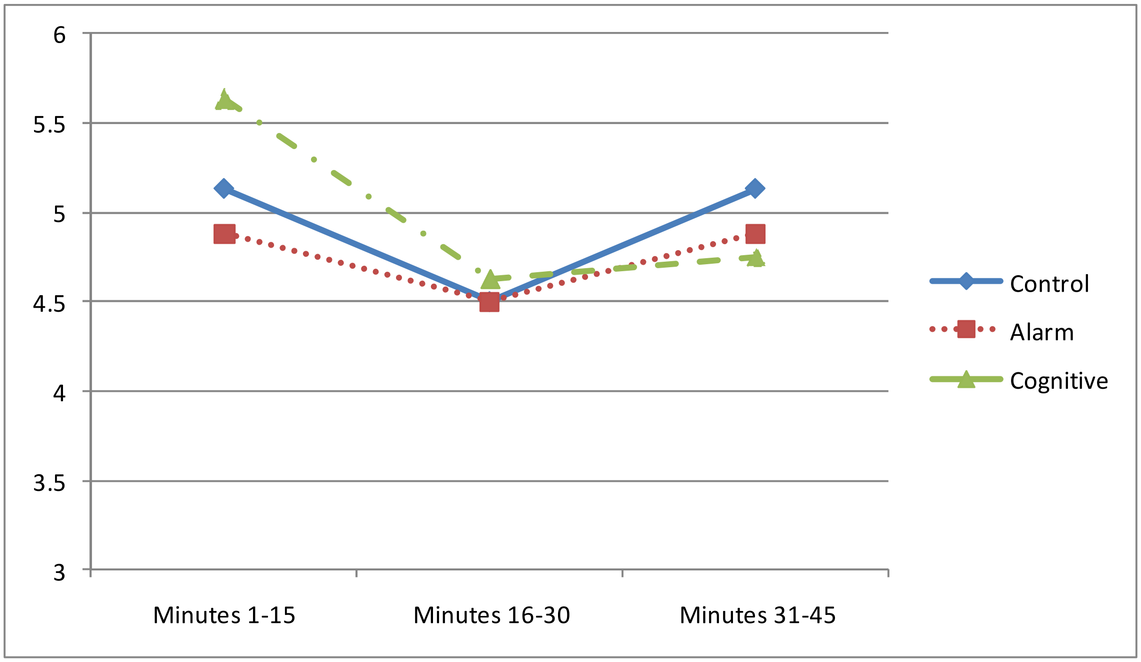


Figure 2. Number of Hits on the Experimental Task by Intervention Condition and Time Interval.

Examination of the SSSQ scores revealed mixed results. Using the pre-PVT scores as a baseline, SSSQ Engagement decreased after both the ET Control condition (24.13 vs. 20.25, t(15) = 2.90, p < .01) and the Intervention condition (24.13 vs. 20.38, t(15) = 4.34, p < .001). Worry increased following the Control condition (6.38 vs. 8.63, t(15) = -2.12, p < .05), but not the Intervention condition (6.38 vs. 7.63, t(15) = -0.99, ns.). There were no significant changes in Distress scores following the experimental tasks. Comparisons between the post-ET Control and Intervention conditions showed no significant effects.

Comparisons between the NASA TLX subjective workload scores for the post-PVT and each of the post-ET conditions revealed that subjective workload for most of the scales and for Overall workload was higher following the PVT (62.50) than after either of the ET tasks (Control: 62.50 vs. 34.27, t(15) = 6.36, p < .001; Intervention: 62.50 vs. 32.39, t (15) = 7.39, p < .001). The Control and Intervention tasks did not differ on Overall workload (34.27 vs. 31.92, t(15) = 0.53, ns).

Discussion

As expected, there appeared to be a vigilance decrement across the 45 minute task, with the larger decline from the first 15 minute period to the second 15 minute period. What was somewhat surprising was the lack of a difference between the baseline and intervention conditions. St. John and Risser (2007) reported fairly stable performance for each of the three intervention conditions. Methodological differences between the two studies may have contributed to differences in results. To begin, there was a major difference in the primary tasks in the two studies. In St. John and Risser, participants performed a sensor monitoring task that involved detecting a single critical signal repeatedly over 45 minutes. The event exposure duration was 400 ms and the event rate was 1 per 2 seconds with a critical signal (target) rate of 3 per minute. The critical signal was a truck icon that was slightly larger than the non-critical signal. All signals occurred in one of 6 fixed screen locations.

The primary task in the current study was a step closer to a common sensor operators task. Rather than a single critical signal occurring at fixed locations, there were 11 suspicious behaviors used as critical signals that could occur anywhere in the two sensor streams. Further, the non-critical signals were “daily life” activities in an urban scene continuously presented. Thus, the non-critical signals overlapped with the targets. Finally, the critical signal rate in the current study was 1 per 2.5 minutes across the two scenes with a median exposure time of 11.0 seconds. Despite these differences in event rate, exposure time, number of displays monitored, and observer uncertainty about spatial location of the critical signal, a vigilance decrement was observed. The difference between the studies occurred in the effectiveness of the interventions.

One possible explanation for the difference between the findings of the two studies is the intervention schedule. St. John and Risser implemented their interventions with simulated psychophysiological monitoring where the intervention was triggered by a missed target. In the present study, a simpler approach was taken by implementing a constrained randomized schedule. Such an approach, if effective, would eliminate the need for physiological monitoring to achieve performance benefits. Alas, the current study indicated that such hopes were unjustified. Whether the reason for different results lies in the differences in the nature of the experimental tasks, the intervention schedule, or some interaction of the two, it appears at this point that physiological monitoring may be necessary to achieve the performance benefits of a system based mitigation strategy.

References

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

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). NY Elsevier.

Helton, W. S. (2004). Validation of a short stress state questionnaire. *Proceedings of the 48th Meeting of the Human Factors and Ergonomics Society.*

Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliland, K., Grier, R., &Warm, J. S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion, 2,* 315-340.

Molloy R, & Parasuraman, R. (1996). Monitoring an automated system for a single failure: vigilance and task complexity effects. *Human Factors, 38*, ,311-322.

Schroeder, D. J., Touchstone, R. M., Stern, J, A., Stoliarov, N., & Thackray, R. (1994). *Maintaining vigilance on a simulated ATC monitoring task across repeated sessions,* DOT/FAA/AM-94/6. Washington, DC: U.S. Department of Transportation, Federal Aviation Administration, Office of Aviation Medicine.

St. John, M., & Risser, M. R. (2007). Mitigating the vigilance decrement: Perceptual versus cognitive interventions. *Proceedings of the 4th International Conference on Augmented Cognition,* Baltimore, MD.

Temple, J. G., Warm. J. S., Dember, W. N., Jones, K. S., LeGrange, C. M., & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors, 42,* 183-194

Vroom, V. H. (1964). *Work and motivation*. NY: Wiley.

Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*, 150-177.

Yerkes, E. 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.

SLEEP DEPRIVATION EFFECTS ON COGNITIVE PERFORMANCE

Hoermann, H.-J.
Steinmann, M., Uken, T. & Elmenhorst, E.-M.
German Aerospace Center (DLR)
Institute of Aerospace Medicine
Hamburg and Cologne, Germany

An experimental sleep deprivation study has been launched at the German Aerospace Center (DLR) in order to determine effects of varying degrees of sleepiness and alcohol on cognitive performance. A total of 48 subjects in cohorts of eight subjects each will stay for twelve consecutive days and nights in the AMSAN sleep laboratory in Cologne. During their stay in the laboratory, subjects are deprived of sleep in a successive manner totally and partially. In addition, on one day they are exposed to moderate alcohol levels. In between the interventions two recovery days are provided per design. A short test-battery of pilot's and air traffic controller's aptitudes is administered to the subjects including spatial orientation, perceptual speed and control of attention tests. In addition, self-concept of mental fitness is measured via a questionnaire prior to and after the cognitive tasks in order to examine whether potential performance decrements due to fatigue are recognized by the subjects. The study design and some preliminary data of the first two cohorts (N = 16) are presented in the paper.

After a continuous time of wakefulness or insufficient rest within 24 hours, human performance can be impaired significantly. Scientific research showing adverse effects of acute fatigue on different cognitive processes is well established and summarized for example in a recent meta-analysis by Lim & Dinges (2010). However, many laboratory studies did administer primarily simple reaction time tasks and therefore it is less apparent whether and how more complex cognitive functions are affected by sleep deprivation. Although, fatigue has remained in aviation among the top 10 safety concerns of the NTSB for over 20 twenty years, still it appears to be a severe issue in accident and incident investigations (National Academy of Sciences, 2011). In the US, crew fatigue has been linked to at least 10 major accidents and even more incidents since 1990. FAA has recently announced to address the fatigue issue among commercial pilots by proposing new flight time, duty and rest requirements based on latest scientific findings in the field.

Based on earlier research at the DLR sleep laboratory of combined effects of sleep deprivation and alcohol on performance (Elmenhorst et al., 2008, 2009) a new study is currently ongoing with extended periods of wakefulness up to 38 hours and a wider scale of performance tasks including cognitive, psychomotor tests and a number of self-assessments. This research aims at identifying potential combined and differential effects of alcohol and sleep deprivation on a variety of psycho-physiological measurements. Some initial thoughts and findings are described below.

Method

Sample

A total of 16 healthy subjects (8 female, 8 male) with an average age of 26 years participated in the study until the time this paper has been written. By end of this year 48 subjects will have been gathered. All subjects are mentally and physically in good conditions with no recent history of sleep disorders, medication and drug or alcohol abuse. In advance of the study all subjects received a comprehensive briefing and training in all performance measures. They signed an informed consent form and were paid for participation. The study was approved by the Ethics Committee of the North Rhine Medical Board.

Procedure

Two cohorts (out of finally six), each with eight subjects stayed for twelve consecutive days and nights in the DLR sleep laboratory AMSAN (Samel et al., 1997). All subjects adjusted their sleep-wake cycle to the laboratory conditions one week in advance of the study. Additionally, they were trained intensively on all performance tests with at least 20 trials per test distributed across the two preceding weeks of the experiment.

The first one and a half experimental days and the first night were used for adaptation to the study environment (Adapt). All measurements throughout the following 48 hours period (i.e. until day four) served as a baseline (Base) for the psycho-physiological parameters. On night four subjects were deprived of sleep totally (TSD, no sleep) or partially (PSD, four hours of sleep). Night five and the following days six and seven served as recovery period (Reco). During night seven again all subjects were deprived of sleep totally or partially with two more days and nights for recovery. In the evening of day ten all subjects consumed alcohol up to a calculated blood alcohol concentration of 0.1 % (ALC). During the following night ten, additionally they were deprived of four hours of sleep (PSA, four hours of sleep). Two more days and nights provided sufficient time for recovery before the end of the study on day thirteen. The experimental interventions were rotated systematically from cohort to cohort to avoid sequential effects. However, at the time of writing this paper only two cohorts had completed the experiment.

Support personnel were present in the facility day and night. Medical doctors were available throughout the experiment to provide medical care if needed. Subjects did not leave the facility for the entire time of the experiment. In order to avoid potential influences on alertness, they were asked not to engage in physical exercise, thrilling games or to take nicotine or caffeine.

As per experimental protocol performance testing took place every three hours except during sleeping periods. All performance measures were administered by computers in the private rooms of the subjects and supervised remotely by the investigating team. Each testing session lasted about 45 minutes. Altogether each subject completed 63 testing sessions during the twelve days in the laboratory. No explicit performance feedback was provided.

Performance Measurement

Three tests from the DLR test battery of air traffic controller and pilot selection tests (Goeters, 2004) were chosen to assess the impact of TSD, PSD, PSA, and ALC on important performance factors of aviators.

Aircraft-Position Test (FPT): The Aircraft-Position Test is a test of spatial orientation. Each item consists of two equal aircraft silhouettes with different headings displayed on a computer screen. The task is to mentally rotate one of the aircraft silhouettes according to a number of heading and direction instructions (e.g. 90°R, 270°L, 180°, 90°R). The letters “L” (left) or “R” (right) indicate the direction of the rotation. At the end of these rotations, subjects have to compare the two silhouettes in order to determine a final heading change needed to reach a complete match between the two aircraft. The FPT has a maximum of 112 items in 5 minutes. The error rate and the number of correct answers are counted.

Attention Control Test (KBT): The Mental Concentration Test involves a combination of different cognitive functions such as visual search, working memory, decoding speed, and simple arithmetic under time pressure. On the screen a search area is presented with a set of 19 symbols connected to numbers. For each task, two symbols are displayed which have to be “translated” into the corresponding digits by reference to the search area. As soon as the sum of the two digits has been entered via a numeric keypad the next task appears. After every ten tasks a new search area is presented. In ten minutes subjects can solve a maximum of 200 tasks. The error rate and the number of correct answers are counted.

Visual Perceptual Speed (OWT): The Visual Perceptual Speed test measures the ability to quickly grasp certain details of visually presented information. Images are presented for less than two seconds depicting two instruments and two sets of objects. Subjects have to enter the perceived two instrument readings and afterwards to respond to the question how many objects of either set A or B they have seen. The error rate and the number of correct answers is counted. The maximum score is 63 points.

In addition to these three aviation tests, one test from the AGARD STRES battery, the Unstable Tracking Task (Santucci et al., 1989), the Psychomotor Vigilance Task (Dinges & Powell, 1985), and the Lane Change Task (Mattes, 2003) for measuring driving performance were administered. However, data of these tests are not included in this paper.

Subjective Measurement

Subjective Fatigue Checkcard (FAT): Subjective fatigue levels (FAT) were assessed before and after each performance testing session using the Samn-Perelli Subjective Fatigue Checkcard (Samn & Perelli, 1982). Subjects reported their current levels of fatigue by ratings of ten different mental states. FAT results in total scores ranging from 0 to 20, which can be categorized into four different fatigue levels related to performance capabilities: class I - severe fatigue; class II – moderate to severe fatigue; class III – mild fatigue; class IV – sufficiently alert. A lower FAT total score indicates a higher level of subjective fatigue. Perceived quality of sleep was rated once per day (except for TSD) by using a sleep diary.

Self-concept of Performance (SOP): Subjects were asked to assess their perceived level of performance on anchored 6-point Likert scales ranging from 1 = Min to 6 = Max. Each self-rating scale was directly related to one of the administered performance tests as described above. Subjects assessed their expected performance level immediately before each testing session and also retrospectively after each testing.

Additional Data

A number of other psycho-physiological parameters were also collected at nights and during some of the testing sessions. The results of these psycho-physiological data will be described elsewhere.

Results

In general, the motivation of the subjects to participate in the study was very high. They arrived well prepared with their training and sleep logs for the adaptation day. Though they shared the facility the entire time together with people they had not seen before, we did not encounter controversies or any tense situations in the first cohorts. Just in a few cases medication was needed against headache. A few data had been lost because one subject left a day earlier due to family reasons and one subject arrived four days later because of a prior illness.

The treatment effects were clearly visible in the acute subjective fatigue scores (FAT). According to the Fatigue Checkcard scoring procedure, subjects remained on average in the highest class IV (= alert) during the majority of measurements on baseline and recovery days. During PSD/PSA and TSD they dropped down to class II (= moderate to severe fatigue). It remains to be shown if cumulative sleep debt was present. In fact, the recovery periods for the second and third treatment seemed to be slightly longer than for the first. The mean course of subjective fatigue is depicted in Figure 1 for the first eight subjects. The scores for the second cohort appeared to be equivalent but cannot be included in the chart because of the different sequence of treatments. In summary, these data show that the treatment had substantial effects as expected on the subjective levels of fatigue of the participants.

The mean performance course for the KBT for cohort 1 is shown in Figure 2. In spite of the intensive training before the experiment the data revealed a positive performance gradient especially during the adaptation and baseline days. This is confirmed by the data of the second cohort and also for the FPT and OWT that subjects obviously continued improving their average performance by about 5% to 10% over the 63 repeated measurements. The magnitude of this effect has to be taken into account for the evaluation of the degree of performance impairment due to sleep loss and alcohol. Therefore, we have decided to use the average of the baseline measurement (Base2) and the measurement on the recovery day following the treatment (second recovery day for TSD) at the corresponding point of time as baseline scores in order to provide a preview of effects for the purpose of this paper.

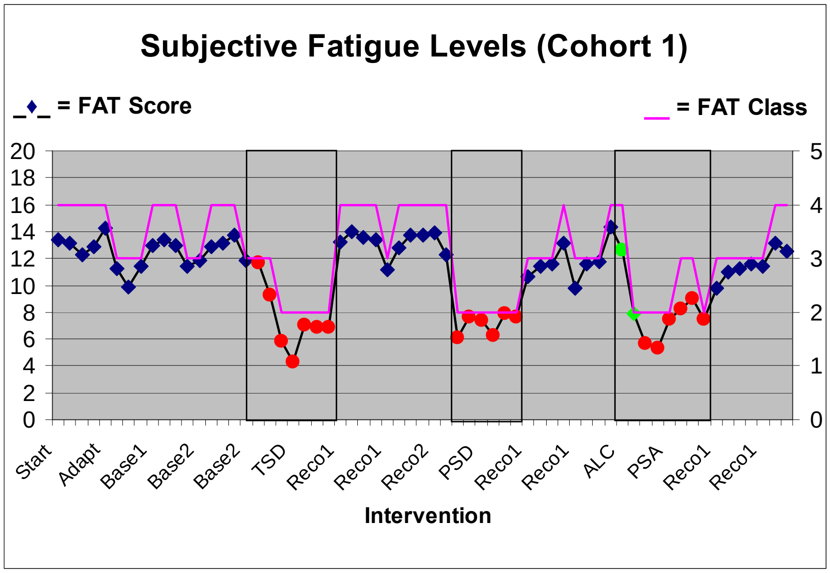


Figure 1. Subjective levels of fatigue throughout the experiment (Cohort 1, N=8). Colour and boxes indicate periods of treatment (blue = no treatment, red = TSD/PSD/PSA, green = ALC)

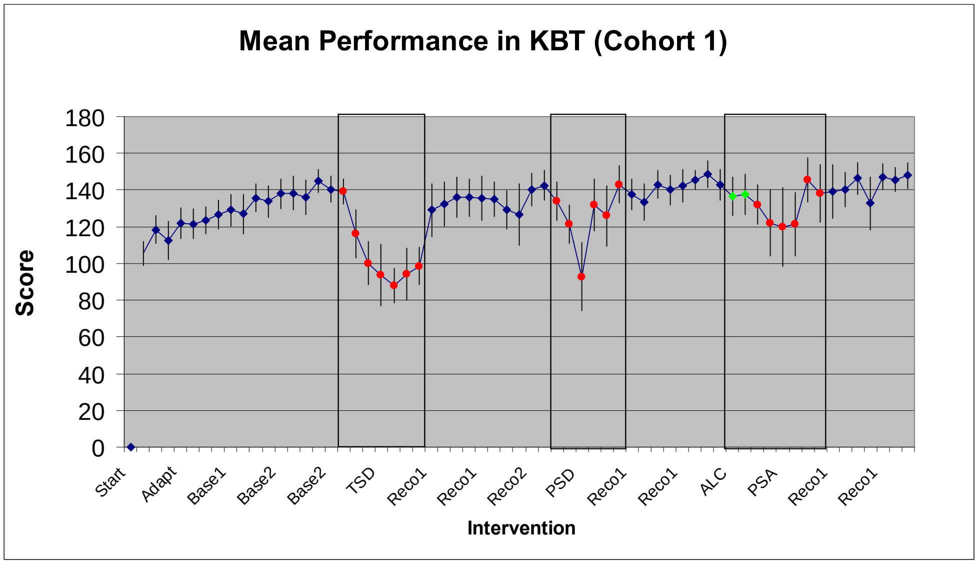


Figure 2. Mean performance scores in KBT (Cohort 1, N=8). The error bars indicate the corresponding standard deviations.

At the first glance the effects of sleep deprivation on levels of performance as shown in Figure 2 seem to be massive especially for TSD. Statistical significance of systematic treatment effects was tested with t-tests with dependent samples. The results of the 9am measurements for TSD, PSD, and PSA and the 6pm measurement for ALC are shown in Table 1. Each of the three tests offered total scores for aspects of performance quantity and accuracy (error rate). TSD at 9am was equivalent to 26 hours of wakefulness. For PSD and PSA subjects had 4 hours of sleep in the preceding 26 hours. Alcohol intake was scheduled 2 hours in advance of the 6pm test session on day ten, which means a calculated blood alcohol concentration of 0.07% to 0.08% at the time of the respective performance measurement.

Though our results are only preliminary and shown here primarily for the purpose of illustrating the effectiveness of the experimental design, a rather consistent pattern of effects seemed worth to be noted. While sleep deprivation impaired performance quantity significantly for most of the tests, alcohol had significant effects primarily on performance accuracy but less on performance quantity. Under the influence of TSD for example, performance was slowed down by about 40%, while the error rates remained rather unaffected. Statistical effect sizes varied from -0.6 standard deviations of performance decrements in the partial sleep deprivation and alcohol conditions (moderate effects) up to -1.5 standard deviations of performance decrements for the condition of total sleep deprivation (large effects).

Table 1. *Effects of sleep deprivation and alcohol on performance. Results of t-tests with two-tailed significance levels. Significant effects (p < 0.05) are in boldface characters. The effect size d was calculated according to Morris & DeShon (2002).*

Performance Test	Quantity	Error Rate
TSD		
FPT	**t(14) = 5.8, p = 0.00, d = -1.5**	t(14) = -1.3, p = 0.22
KBT	**t(14) = 4.8, p = 0.00, d = -1.4**	**t(14) = -3.1, p = 0.01, d = 0.8**
OWT	**t(14) = 3.0, p = 0.01, d = -0.8**	t(14) = -1.9, p = 0.08
PSD		
FPT	**t(15) = 2.2, p = 0.04, d = -0.6**	t(15) = 1.1, p = 0.30
KBT	**t(15) = 3.2, p = 0.01, d = -0.8**	t(15) = -2.0, p = 0.07
OWT	t(15) = 0.3, p = 0.80	t(15) = 0.5, p = 0.61
PSA		
FPT	**t(13) = 2.4, p = 0.03, d = -0.6**	t(13) = -0.9, p = 0.39
KBT	**t(13) = 2.1, p = 0.05, d = -0.7**	t(13) = -1.4, p = 0.18
OWT	t(13) = -0.9, p = 0.39	t(13) = 2.0, p = 0.07
ALC		
FPT	t(15) = 1.6, p = 0.13	**t(15) = -2.4, p = 0.03, d = 0.6**
KBT	t(15) = 1.2, p = 0.23	**t(15) = -2.6, p = 0.02, d = 0.6**
OWT	t(15) = -1.1, p = 0.28	t(15) = 1.4, p = 0.18

Discussion

According to the initial analyses massive performance decrements were identified for all three tests after a period of 26 hours of wakefulness. So far our data have shown that sleep deprivation can affect performance on spatial orientation and complex attention tasks by a reduction of work quantity. The amount of deterioration is significantly higher for TSD compared to PSD and PSA. Performance in the perceptual speed test was also impaired but to a lesser extent. When being debriefed subjects reported having had more difficulties with the more complex tasks because of "interruptions" of their cognitive processes due to fatigue. When loosing the thread of thought they had to go back and restart the task from the beginning, which cost them time. Compared to FPT and KBT performance in the OWT seemed to have a higher degree of automaticity, which was less likely to be interrupted.

It is interesting to compare the potential influences of alcohol and fatigue for the current subsample of 16 subjects. The data indicated that while fatigue seemed to reduce the level of productivity in general, under the influence of alcohol work quantity remained unchanged. However, working speed seemed to be on cost of accuracy because the error rate was deteriorating significantly after alcohol consumption. Since our study is still ongoing, what is reported here are only preliminary findings, which need to be verified when all subjects have completed the experiment.

With respect to operational crewmembers, lower performance does not necessarily imply less safety. Final outcomes depend for example on how pilots decide on their course of action ahead. It was shown in other research that fatigued pilots tend to avoid riskier options and require more time to collect necessary information before finalising a decision. However, this can lead to higher time pressure during high workload phases as well. And also the freedom of choice can be limited by contextual factors. In further analyses of this study we will analyse how the measures of self-awareness of performance are related to differential effects on the performance tests. Self-efficacy is an important factor in decision making. If confidence is not realistic decisions can lead to undesired outcomes.

References

Dinges D.F., Powell J.W. (1985). Microcomputer analysis of performance on a portable, simple visual RT task during sustained operations. Behavior Research Methods. Instruments and Computers, 17, 652–655.

Elmenhorst D. , Elmenhorst E.-M., Luks N., Maass H., Mueller E.W., Vejvoda M., Wenzel J., Samel A. (2009). Performance impairment during four days partial sleep deprivation compared with the acute effects of alcohol and hypoxia. Sleep Medicine, 10, 189–197.

Elmenhorst, E.-M., Elmenhorst D., Luks N., Maass H., Vejvoda M., Samel A. (2008). Partial sleep deprivation: Impact on the architecture and quality of sleep. Sleep Medicine, 9, 840–850.

Goeters, K.M. (Ed.) (2004). Aviation psychology: Practice and research. Ashgate, Aldershot.

Lim, J. & Dinges, D.F. (2010). A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. Psychological Bulletin, 136, 375-389.

Mattes, S. (2003). The lane change task a tool for driver distraction evaluation. In H. Strasser & H. Bubb (Eds.), Quality of work and products in enterprises of the future. Ergonomia, Stuttgart.

Morris, S.B. & DeShon, R.P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychological Methods, 7, 105-125.

National Research Council of the National Academies (2011). Issues in commuting and pilot fatigue: Interim report. The National Academies Press, Washington, D.C., www.nap.edu .

Samel, A., Diedrich, A., Drescher, J. Lorenz, B., Plath, G., Vejvoda, M. & Wenzel, J. (1997). Langzeitmonitoring psychophysiologischer Größen in der Flugphysiologie. Internist, 38(8), 755-769.

Samn, S.W., Perelli, L.P., 1982. Estimating aircrew fatigue: a technique with application to airlift operations. Brooks AFB, USAF School of Aerospace Medicine. Technical report SAM-TR-82-21.

Santucci, G., Boer, L., Farmer, E., Goeters, K.M., Grisset, J., Schwartz, E., Wetherall, A., Wilson, G. & Yates, R. (1989). Human performance assessment methods. AGARDograph N° 309. NATO-AGARD, Neuilly-sur-Seine.

A FACTOR-ANALYTICAL PERSPECTIVE OF SOPITE SYNDROME ASSESSMENT IN AEROSPACE SYSTEMS

J. Christopher Brill & Brittany N. Neilson
Old Dominion University
Norfolk, Virginia, U.S.A.

Aerospace systems require pilots to perform complex tasks under demanding conditions. There is an unrecognized component, which has deleterious effects on human performance, called sopite syndrome. Sopite syndrome is characterized by intense drowsiness despite receiving an adequate night's rest, difficulty concentrating, and lack of motivation. Currently, sopite syndrome is measured exclusively by a 39-item self-report questionnaire called the Mild Motion Questionnaire (MMQ). The purpose of the present research is to develop a short-form for the MMQ that can be used for quick assessments in applied settings, while maintaining internal consistency. Participants (N = 422) completed the MMQ by indicating how they feel following exposure to mild, non-sickening motion. Principal-axis factor analysis with oblimin rotation identified a two-factor solution comprised of 25-items with 2 dimensions: adverse effects and positive affect. Internal consistency was .86. Discussion of efforts to validate the short-form MMQ and the multidimensionality of sopite syndrome is included.

There has been a large effort to improve pilot performance through the development of monitoring and assistance programs (Russo, Stetz & Thomas, 2005; Lehrer, Karavidas, Lu, Vaschillo, Vaschillo & Cheng, 2010; Liang, Lin, Hwang, Wang & Patterson, 2010). Specifically, these programs focus on performance as it relates to higher cortical functioning, such as vigilance, skill acquisition, and workload. However, a largely unrecognized factor intrinsic to flight can contribute to performance decrements: sopite syndrome. Graybiel and Knepton (1976) identified sopite syndrome as a manifestation of motion sickness that presents itself with a unique set of symptoms, namely, drowsiness, fatigue, and lack of motivation. Research has documented the presence of sopite syndrome among pilots undergoing operational flight training (Flaherty, 1998), as well as passengers and operators in common transportation systems (e.g., commercial planes, trains, boats; Lawson & Mead, 1998). Sopite-like symptoms have also been observed during space flight (Kanas & Manzey, 2008). Nevertheless, there has been little implementation of assessing the degree to which sopite syndrome moderates pilot performance.

The purpose of this study is to construct a valid, yet practical, short measure of sopite syndrome that can be used alongside other pilot performance assessments. In doing so, we hope to further explore the multifaceted nature of sopite syndrome. At present, the only measure specifically designed for assessing sopite syndrome is a 39-item self-report measure, the Mild Motion Questionnaire (MMQ; Lawson, Kass, Muth, Sommers & Guzy, 2001). Due to its length, the MMQ may be cumbersome for rapid assessments conducted in-flight or in the laboratory. Our aim is to reduce its length to include only items that capture the most explained variance of sopite syndrome.

Method

Participants

Four hundred and twenty-two undergraduate psychology students at Old Dominion University participated in a larger study that included the Mild Motion Questionnaire (MMQ). There were 570 total participants (401 women, 164 men) in the larger dataset, but 148 contained missing data and were eliminated from data analysis. Participants had to be 18 years of age or older. The median age of the adult group was 19 (*SD* = 4.13).

Materials and Procedure

The Mild Motion Questionnaire (MMQ) was completed anonymously and administered simultaneously with several other questionnaires as a part of a larger study. Participants were asked to retrospect on feelings following mild, non-sickening motion using a 5-point scale (*not at all* to *very strongly*). Optional course credit was granted for those who completed the entire survey.

Results

The factor extraction method performed was principal-axis factor analysis with direct oblimin rotation (delta = 0). The extraction technique was chosen based on the non-normality of the dataset and our a priori theory that at least some covariance among sopite-related variables exists (Fabrigar, Wegener, MacCallum & Strahan, 1999; Costello & Osborne, 2005). An oblique rotation method (i.e., direct oblimin rotation) was used, as orthogonality of variables is not representative of the multidimensional nature of sopite syndrome (Costello & Osborne, 2005). Initially, a 6-factor solution was identified by retaining all factors with Eigenvalues greater than 1.0. Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .946, above the standard criterion of .6 (Norusis, 1999), and the Bartlett's test of sphericity was significant at $p < .001$ indicating a strong relationship among the variables. While this model accounted for 54.5% of the variance, three of the six factors accounted for less than 1% of the variance. It should be noted that many of the factors did not contain at least 3 item loadings greater than .3, which is a standard cutoff for factor loading (Costello & Osborne, 2005).

Previous research using Monte Carlo analyses suggest that too many factors are retained using the Eignenvalue above 1.0 criterion (Velicer & Jackson, 1990). The Scree plot, which is an alternative method for retaining factors, indicated two distinct factors above the elbow. The analyses were run again using principal-axis factor analysis with direct oblimin rotation while forcing a 2-factor setting. The 2-factor solution accounted for 44.4% of the total variance, with Factor I accounting for 33.5% and Factor II accounting for 10.9% of the variance, respectively. The 2-factor model maintained good sampling adequacy and strength among variables (KMO = .946, Bartlett's test of sphericity $p < .001$).

Since the purpose of this study was to create a shorter version of the MMQ, rotated factors with item loadings above a cutoff of .6 were included, as compared to the standard .3 cutoff (Costello & Osborne, 2005). This procedure resulted in a 25-item solution with two factors, as represented in Table 1. Factor I (Cronbach's alpha = .94) appeared to represent the adverse effects from mild motion exposure, and Factor II (Cronbach's alpha = .91) was composed of items relating to positive affect, a unique effect experienced from exposure to mild motion. Cronbach's alpha for the total 25-item MMQ-short form was .858, indicating that good internal reliability was maintained.

Discussion

The purpose of the present study was to develop a short-form for the Mild Motion Questionnaire (MMQ) that can facilitate quick and accurate assessments in applied settings, while maintaining good psychometric properties. We used a more stringent cutoff of .6 for item loadings and retained two factors, which broadly represent adverse effects and positive affect resulting from mild motion exposure. The magnitude and pattern of factor loadings for our sample group appeared relatively consistent with the two previous investigations (Lawson, Kass, Muth, Sommers & Guzy, 2001; Lawson, Kass, McGrath & Campbell, 2006). Lawson et al. (2001) developed the MMQ and cited 7 factors with Eigenvalues above 1, similar to our initial 6-factor solution. However, they used a cutoff of item loadings above 0.43 and found a four-factor solution, which included head/body symptoms (Cronbach's alpha = .92), relaxed/content (Cronbach's alpha = .90), drowsy/fatigued (Cronbach's alpha = .87), and poor concentration/motivation (Cronbach's alpha = .87). A more recent study by Lawson and colleagues (2006) established a 3-factor solution comprised of somatic, affect, and sopite dimensions using a .4 cutoff. Their 3-factor solution collapsed together the original "poor concentration/amotivational" and "drowsy/fatigued" factors to create a new "sopite" dimension. The head/body symptoms reflected the somatic dimension and relaxed/content reflected the affect dimension. To our knowledge, this 3-factor solution is not currently being adopted for use but provides evidence that convergence of factors may yield a simpler model.

A frequently cited concern with factor analysis is the decision to use of a more complex model to explain the greatest amount of variance or the simplest model that explains less of the variance (Thompson, 2004). We wanted to maintain levels of variance explained comparable to previous studies while making our model less convoluted and smaller. Lawson et al.'s (2006) 3-factor model accounted for 48% of the variance; similarly, our 2-factor solution accounted for 44.4% of total variance with a more stringent cutoff of .6 and less items. While theoretically sopite syndrome is multidimensional and represents more than our 2-factor solution, a short-form should serve as a screener for potential cognitive decrements. Thus, dividing the factors of the scale into negative and positive symptomology is functionally appropriate for assessing human performance. The outcome of our analyses further reinforces the multidimensional nature of sopite syndrome. Nevertheless, the convergence of items into fewer factors suggests there may be limitations to our current methods for assessing sopite syndrome. Results from experiments (e.g., Graybiel & Knepton, 1976) and anecdotal evidence frequently identify cognitive impairment as a component of sopite syndrome; however, items representing this aspect were affected by factor convergence. One potential explanation for this is participants' retrospections of feelings following exposure to mild motion may not reflect actual symptoms. Moreover, our sample was mostly college freshman, who were most likely reflecting on common travels (e.g., driving to school or work), which may not tax cognition sufficiently for impairment to manifest in an obvious fashion. Moreover, our participants may have lacked insight into the extent to which they experience cognitive impairment with motion – perhaps as a function of the impairment itself. People whose lifestyles or careers involve frequent and prolonged exposure to motion (e.g., pilots, astronauts, commercial truck drivers) may offer greater insight into typical sopite symptoms. This will provide a more accurate representation of the population of interest. Moreover, the MMQ should (and will) be implemented in the laboratory for mild motion studies. By using it as a repeated-measures "state" assessment tool, we will be able to assess symptoms development following exposure to mild motion. Future studies will concentrate on the use of the MMQ as a state measure for rapid on-site assessments.

The present study sought to explore the multidimensionality of sopite syndrome while developing a short-form questionnaire for assessing it. Our prospective solution is a two-factor 25-item self-report questionnaire, which accounts for 44% of the variance and offers and internal consistency of .86. We will continue our efforts to develop and validate a short-form measure for sopite syndrome for use in both field settings and the laboratory.

Table 1

Factor Loadings for Fixed 2-Factor Principal Axis Factoring with Oblimin Rotation of Mild Motion Questionnaire

Descriptor	Factor I	Factor II
Lethargic/ Sluggish	**.772**	-.189
Weak	**.770**	-.281
Fuzzy-headed/ Foggy-headed	**.761**	-.354
Uncoordinated	**.745**	-.286
Dizzy	**.741**	-.448
Light-headed	**.713**	-.462
Disoriented	**.702**	-.357
Tired	**.695**	.043
Fatigued	**.691**	-.070
Drowsy	**.675**	.067
Off-balance/ Wobbly	**.660**	-.317
Shaky/ Jittery	**.658**	-.258
Disconnected/ Detached	**.656**	-.130
Hard to keep eyes open	**.653**	.139
As if drugged	**.628**	-.073
Headache	**.626**	-.381
Lazy/ Unmotivated	**.623**	.045
Quiet/ Not communicative	**.613**	-.180
Irritable/ Annoyed	**.612**	-.263
Stomach awareness	.592	-.370
Confused .5	72	-.175
Sleepy .570		.216
Want to be alone	.570	-.147
As if in a trance/ Hypnotized	.567	.009
Floating .5	59	-.028

Apathetic .5	26	.046
Blurred vision	.449	-.136
Hard to concentrate	.443	-.235
Distant .4	09	-.004
Distracted/ Preoccupied	.357	-.053
Yawning .354		.342
Peaceful -.238		**.823**
Comfortable -.	309	**.820**
Relaxed -	.233	**.817**
Pleasurable -.242		**.730**
Content/ Happy	-.333	**.698**
Calm -	.190	**.649**
Soothed -.	104	.567
Bored .267		.363

Note. Factor loadings > 0.6 were retained and are printed in boldface.

Acknowledgements

This research was funded by a grant from the NASA Virginia Space Grant Consortium, contract #104766. Dr. J. Christopher Brill is the Principal Investigator.

References

Costello, A. B. & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation. 10*(7), 1-9.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299.

Flaherty, D. E. (1998). *Sopite syndrome in operational flight training* (Master's thesis). Retrieved from Defense Technical Information Center (DTIC). (NTIS No. AD-A354942).

Graybiel, A. & Knepton, J. (1976). Sopite syndrome: A sometimes sole manifestation of motion sickness. *Aviation, Space, and Environmental Medicine, 47*, 873-882.

Kanas, N., & Manzey, D. (2008). *Space psychology and psychiatry*. Segundo, CA and Dordrecth, The Netherlands: Microcosm Press and Springer.

Lawson, B. D., Kass, S., McGrath, C., & Campbell, J. (2006). *Development of the mild motion scale.* Unpublished manuscript. Naval Aerospace Medical Research Laboratory, Pensacola, Florida.

Lawson, B. D., Kass, S., Muth, E., Sommers, J., & Guzy, L. (2001). Development of a scale to assess signs and symptoms of sopite syndrome in response to mild or nonsickening motion stimuli. *Abstracts of the Aerospace Medical Association 72nd Annual Scientific Meeting*, Reno, NV.

Lawson, B. D., & Mead, A. M. (1998). The sopite syndrome revisited: Drowsiness and mood changes during real or apparent motion. *Acta Astronautica, 43*, 181-192.

Liang, G., Lin, J., Hwang, S., Wang, E. M., & Patterson, P. (2010). Preventing human errors in aviation maintenance using an on-line maintenance assistance platform. *International Journal of Industrial Ergonomics, 4,* 356-367.

Lehrer, P., Karavidas, M., Lu, S., Vaschillo, E., Vaschillo, B., & Cheng, A. (2010). Cardiac data increase association between self-report and both expert ratings of task load and task performance in flight simulator tasks: An exploratory study. *International Journal of Psychophysiology, 76*, 80-87.

Norusis, M. (1999). *SPSS 9.0 Guide to data analysis*. Upper Saddle River, New Jersey: Prentice Hall.

Russo, M. B., Stetz, M. C., & Thomas, M. L. (2005). Monitoring and predicting cognitive state and performance via physiological correlates of neuronal signals. *Aviation, Space, and Environmental Medicine, 76*(7), C59-C63.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Washington, DC: American Psychological Association.

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor-analysis- Some further observations. *Multivariate Behavioral Research, 25*(1), 97-114.

A LARGE SCALE VALIDATION STUDY ON AIR TRAFFIC CONTROLLER SELECTION AND TRAINING – DESIGN, CHALLENGES AND RESULTS

Kristin Conzelmann
DLR German Aerospace Center
Hamburg, Germany

Alexander Heintz
DFS Air Navigation Services Academy
Langen, Germany

Hinnerk Eißfeldt
DLR German Aerospace Center
Hamburg, Germany

A validation study with 476 air traffic controller trainees of DFS German Air Navigation Services has been set up, encompassing the complete data from the selection of candidates to the completion of their training. The design includes a detailed coding of interview contents, questionnaire data, and results of the reference sample of 13,716 applicants. Data analysis involves the prediction of training success, training performance, and trainees' personal evaluation of the selection and training process. The success rate of 81 % was satisfactory. Selection measures were adequate to predict pass-fail and performance criteria from institutional training (i.e., theoretical exams). Basic ability measures and a semi-structured interview predicted success in training best. Performance in early training courses was positively related to successive stages and to overall training success. Repetition of exams and switching of training courses/units increased the likelihood of failing. The limits of a correlation-based approach to examine validity are discussed.

Due to the low base rate of suitable applicants, the high safety relevance of the job, and the high costs of ATC training, the selection of ab in itio air traffic controller (ATCO) trainees requires a valid and efficient approach to recruitment. The selection procedure for DFS German Air Navigation Services ATC trainees has been developed and conducted in close cooperation with DLR German Aerospace Center. The continuous and intense quality assurance of the process is based on four major pillars, i.e., job requirements analyses (e.g., Bruder, Jörn, & Eißfeldt, 2008; Eißfeldt, & Heintz, 2002), careful selection and regular training of involved psychological and operational staff (Seidel, Pecena, & Eschen-Léguedé, 2009), intense and regular cost benefit analyses (Heintz, 2004), and regular psychometric validation. The selection process involves multiple stages, starting with a paper sift based on published application criteria (e.g., age, education, language training) and a pre-selection based on a biographical questionnaire. The first stage of testing covers a comprehensive set of mental abilities (i.e., memory, concentration, attention, English), followed by a stage of work sample testing (two multiple-task performance tests). The third stage focuses on teamwork abilities (two exercises); the final phase involves an oral English test and a biographical interview including problem-solving tasks. Successful trainees undergo a medical examination according to Eurocontrol class 3 requirements. For each stage and the final selection decision, clearly defined cut-off values and compensation mechanisms are applied.

The validity of the selection program had already been proved in two former large scale validation studies (Damitz, Eißfeldt, Grasshoff, Lorenz, Pecena, & Schwert, 2000; Eißfeldt & Maschke, 1991). These former studies resulted in well-directed adaptations and further developments of the test program (e.g., Pecena, 2003). In international scientific literature, several other validation studies in the field of air traffic controller selection have been reported (for an overview see Broach & Manning, 1998). The current validation study aims primarily at analyzing the complete selection procedure applied during 1997 and 2006 and the reliability and validity of the training process. The major objective of aptitude selection for ATCOs is to increase the probability of success in a safety relevant and costly training. The study at hand was designed to fulfil professional standards of quality assurance and to further enhance the efficiency and validity of the selection and training process.

Method

The sample involves N=476 ATCO trainees (mean age 20.52, S=1.74; 66% male) who were selected between 1997 and 2006. As predictors complete testing data from all selection stages including detailed coding of the interview content were available. Particularly in the first selection stage, tests scores of the same performance domain can be partly compensated (e.g., a low score in one attention test can be compensated by a higher score of another attention test). Therefore, in addition to single test scores, composite scores of performance domains used for selection decisions were created and used for analysis. In order to apply correction for range restriction, testing data of the corresponding reference sample of N=13,716 test takers was available.

The following criteria were used for analysis: intermediate and final result in terms of pass vs. failure; data of all theoretical and practical exams during institutional training (IT), exam repetition, total duration of training and duration of on-the-job training (OT), and questionnaire data involving self-reports from all phases of selection and training. Compound scores were created of performance assessments and results of written exams. The following composite scores were established: IT overall theoretical and overall practical exam score, composites for each training stage (e.g. initial stages such as “Basic” and “ATC” Course), composite scores of the trainers’ ratings in practical exams on detailed performance criteria (e.g. communication, strip handling). See table 1 for a summary of predictors and criteria.

The data set at hand made the data analysis methodologically challenging. Since data was collected over a course of nine years, selection tests were further developed, i.e., item material, count of items, and test duration changed, stanine scores were applied for the major analyses instead of raw scores. The trainees of the validation sample were trained in two different training systems (DATS DFS Air Traffic Controller Training System, DATS 1, N= 430 vs. DATS 2 N=46; DFS, 2010), and for different combinations of licenses (Aerodrome N=91 vs. En route and Approach N=385). Thus, aside from the total sample, several subsamples had to be analyzed taking into account a sufficient sample size. The possibility to cross-check the results across subsamples, was considered as an advantage and yielded additional proof or disproof of the findings. Criterion data was on the whole not normally distributed; trainees were evaluated within a small spectrum of possible grades and got mostly positive ratings. Consequently, non-parametric tests such as Mann-Whitney U-Tests, Spearman correlations and the more complex logistic and ordinal regression analyses were applied to analyze ordinal and categorical data. Logistic regression analysis enabled predicting the dichotomous pass/fail criterion. Structure equation modeling was used to predict the results of later selection stages out of the preceding stage. In additional comprehensive analyses stepwise logistic regression analyses were applied. In the case of data allowing for parametric statistical analysis the corresponding methods were carried out (i.e., Pearson correlations, t-Tests, linear regression, discriminant analysis). Multivariate correction for range restriction was done with the Range J software (Johnson & Ree, 1994) for normally distributed interval-scaled criteria.

Analyses were performed for each selection measure separately. In addition, comprehensive (e.g. multivariate) analyses were calculated in order to include the relationships among the measures. Focus on analysis was the DATS 1 sample since only one trainee of the DATS 2 sample failed. Results were cross-checked with the DATS 1 En route and approach controller sample (=DATS 1, ACC, N=363).

Table 1

Summary of predictor and criterion variables.

Predictor data	**Criterion data**		
			Composite scores
Demographic data (age, sex, education, A-level grade-point-average etc.) Test data • Basic ability tests (concentration, attention, memory, English) • Personality questionnaire • Work sample tests • Team Exercises (group and dyadic) • Oral English exam • (Coded) interview content • Final aptitude level ratings	Institutional Training (IT)	Detailed theory performance assessments in initial training	• Training stages • Overall theory score
		Detailed simulation performance in initial training	• 12 Overall scores on detailed performance ratings
		Detailed results of student license examination	• Overall practical score
		License assignment (Aerodrome vs. en route)	
	Operational Training(OT)	OT and total training duration	
Corresponding test data for reference sample (except for interview content)		Pass/fail in intermediate and final check out	
Questionnaire on perception of selection and training (only for subsample)			

Main findings

Success rate and training performance. With respect to the success rate, all trainees who failed to complete the training in their initial license assignment (Aerodrome vs. En route) were counted as failure including drop out due to medical reasons or cancellations of training for personal reasons. Overall, 80.7% of the selected trainees of the validation sample successfully validated as ATCOs, corresponding to the regular quality assurance analyses of DFS. The success rate among trainees for aerodrome control towers was higher than for en route and approach controllers (89% vs. 79%). Most failures occurred during operational training (64.8% of all failures). A significantly higher success rate was observed for female trainees, (88.9% compared to 76.4% for male trainees). There were also remarkable differences between the success rates of the DFS units ranging from only 69 % (in one unit) to 100% in some Aerodrome units. While the success rate for 18-19 years old (86.7 % of N=243) and 20 years old trainees (86.0 % of N=157) exceeded the overall success rate, 39% of the trainees with an age of 23 or higher (N=59) failed to validate. Successful trainees had a grade point average (GPA) close to 2 (ranged from 1=very good to 6=insufficient) whereas unsuccessful trainees' GPA approached 3 ($r_{age-GPA}$=.21; p<.01).

Predictive validity of the selection procedure. Correlations between ability domains of the first selection stage and pass-fail criteria were all positive and significant with respect to concentration, attention and English (r=.08, p<.05 - r=.16, p<.01). All ability domains including the English test correlated consistently with the first training stages (r=.09, p<.05 to r=.29, p<.01). The correlation pattern between the pre-selection tests and the training subjects was positive and mainly significant (r=.09, p<.05 to r=.42, p<.01), too. Concentration and attention correlated significantly with the number of repeated exams (r=.12, p<.05 - r=.13, p<.05, DATS 1, ACC).

Analyses of work sample tests revealed singular significant relationships with subjects of theoretical training (r=.11 - r=.16, p<.01) and with the overall scores on detailed performance ratings (i.e., Traffic planning, Strip Handling, Situational Awareness, Theory, r=.11 - r=.19, p<.05). Team exercise performance revealed some singular significant relationships, for example, with training duration being shorter with a better result in the decision making rating (DATS 1, ACC, r=-.11, p<.05). Neither work sample tests nor team exercises contributed significantly to the prediction of overall training success. This result was surprising since both selection stages proved their validity in the last validation study (Damitz et al., 2000; Höft & Pecena, 2004). A closer look into the data revealed differential predictive validity of work sample test sub scores for male and female applicants, resulting in a differential importance of sub scores for the criteria. These differences are, however, compensated by the overall test score and the significance of the results disappears. Performance in the oral English exam (selection) explained the variance of the English examination result (first training stages) up to 30%. The better the English was judged by the experts, the fewer exams were repeated by the trainees.

In the semi-structured interview the selection board rates an applicant on several dimensions: general motivation, job motivation, cooperation, stress resistance and interactive proficiency. These so-called risk ratings were related to the training criteria. Failure in training was significantly related to a high risk in general motivation (r=-.13, p<.05, particularly in IT), job motivation (r=-.11, p<.05) and cooperation (r=-.13, p<.01, particularly in OT, and student license examinations). A high risk in general motivation increased the probability of exam repetition. The higher the risk was expected to be in interactive proficiency, the longer the training lasted. During the interview, psychologists rate the applicant on specific variables such as parental support, hobbies, efficiency of studies / school. These and additional variables that were hand-coded out of the interview minutes were also related to training success. Correlations were mostly positive and in the expected direction. Particularly, questions about social (e.g., experiences with teachers, relationship to superiors, group membership, self-evaluation) and motivational issues (e.g., efficiency of career, course of studies, hobbies, job motivation) proved to be important predictors of training success in IT and OT (r=.11, p<.05 – r=.23, p<.01).

As a part of comprehensive analyses, stepwise logistic regression analysis was performed with the selection tests on the pass/fail criterion. Concerning the DATS 1 model, 77.4 % of the trainees were assigned correctly to pass and fail (Chi^2=23.78, df=4; Nagelkerke's R^2=.29). Cross checking with the DATS 1, ACC sample revealed an even better classification rate of 82.5 % correct (Chi^2=31.09, df=5; Nagelkerke's R^2=.40). Almost every selection stage contributed to the model fit with at least one significant test score. Including the aviation specific personality scale that is administered in the context of basic ability testing, the model even improved with respect to the DATS 1, ACC sample (classification correct: 88%; Chi^2=43.94, df=6; Nagelkerke's R^2=.54). The personality scale is only used for interview preparation instead of being applied as a hard criterion in selection. Thus, variance of the personality scale is not yet utilized and has a greater chance to result in significant validation findings. Multiple correlation for the prediction of the overall IT theory score was R=.49 (R=.40 uncorrected, DATS 1, ACC sample). Pre-selection tests predicted the IT theory score best. The overall practical IT score could not be predicted as well as the theoretical score (R=.34; uncorrected: R=.29, DATS 1, ACC sample).

Validity of the training. Training and OT duration were affected by the working position (en-route controllers took longer than aerodrome controllers), change of sector group (e.g., OT duration with change: 21.22 months, S=7.26 compared to 17.54 months without change, S=6.09; Z=-2.66, p<.01), change of training course (e.g., total training duration change excluded: 27.55 months, S=5.53; change included: 32.55, S=5.34; Z=4.37; p<.01). Success rate without changing a training course was 87.9% compared to 64.5% including a switch of training course (N=20; Chi^2 = 13.297, p<.01, w=.17). Results showed a significant tetrachoric correlation between the failure of exams and the pass/fail ratio (r=.41, p<.01). However, 51.7% of the trainees failed in at least one exam during IT, indicating that failing an exam does not necessarily imply total failure. However, without repeating an exam, the success rate was 90% (N=207) compared to 72.2% with resit of exams (Chi^2 = 24.15, p<.01, w=.23). Comparably, fewer repeated practical exams resulted in a higher chance of succeeding in OT and total training (r=-.27, p<.01). Better results (test exams, trainers' evaluations) in IT stages increased the likelihood of OT and total training success (correlations between r=.09, p<.05 and r=.35, p<.01). Within the training stages, there were consistent positive correlations of performance in theoretical test subjects with the pass/fail criterion (with few exceptions, i.e., Aircraft principles of flight, Navigation and English). The trainers' evaluations in the Center course (En route/Approach controller) predicted success in overall training and OT significantly (r=.11- r=.19, p<.05). The better the trainees scored on the twelve DATS criteria (i.e., communication, strip handling) in the practical IT exams, the more likely they finished training successfully.

Perception of selection and training among trainees. The questionnaire reflecting the trainee's perception of various aspects of the selection and training process pointed out relevant insights. The trainees felt they were sufficiently informed throughout the process, they were supported individually to achieve their optimal performance, and both selection and training were appropriate to achieve their objectives in training and in the job. IT and the selection phase received better evaluations compared to the OT phase. Women stated that they received more feedback in IT (M=3.44, S=.66) and OT (M=3.62, S=.56) compared to men (M=3.07, S=.81 and M=3.39, S=.66, all differences p<.05). Male trainees, however, found the transition to simulation training easier than women (M=3.41, S=.71 compared to M=3.09, S=.88, p<.05). Ratings differed also according to the unit. Failed trainees evaluated OT worse than trainees who validated.

Discussion

The overall training success rate was sufficient and increased compared to former evaluations. However, remarkable differences concerning the gender of the trainees have to be explored further. Evidence for factors beyond the ability level is under examination in order to identify ways to enhance the success rate of male trainees. The educational level and age are significantly related to training success, which confirm the relatively strict application of criteria for ATCO applicants at DFS, i.e., accepting only applicants with A-level exam who should not be older than 24 years. Operational units can be objectively informed on the impact of failed exams during initial training and their limited relationship to failure in the operational training. This helps to prevent a "Pygmalion" effect negatively impacting the attitude towards trainees in the OT phase (Rosenthal & Jacobson, 1992).

Concerning the selection process, as a main consequence of the study and to further increase the success rate, one should focus on tests which are not already used as hard or explicit criteria in the selection procedure in order to make better use of additional sources of variance. For example, some personality scales and specific categories of questions within the semi-structured interview yielded additional gain in predicting training success. Despite comparatively low correlation effects of test adaptations based on former studies, the increase of the training success rate confirms the success of these adaptations. Correlation-based validation approaches in selection processes with a very low variance in predictor and criterion data are limited. The selection cut-off values (selection rate of 5-6%) and the limits for high safety related training performance are comparably strict; usually "false positive" selection decisions are avoided. Without considering the increase of the training success rate, there would be a risk of underestimating selection tests and training exams that did not prove to be valid in terms of correlation-based analyses within the selected group but probably affect the success rate in a positive way. Moreover, there is a strong risk of overestimating tests which reveal high correlations with training performance irrespective of the training success rate of trainees selected based on these tests.

References

Broach, D. & Manning, C. (1998). Issues in the selection of air traffic controllers (pp. 237-271). In M. W. Smolensky & E. S. Stein (Eds.), *Human factors in air traffic control*. San Diego, CA: Academic Press.

Bruder, C., Jörn, L. & Eißfeldt, H. (2008). Aviator 2030: *When pilots and air traffic controllers discuss their future*. In A. Droog & T. D'Oliveira (Eds.) Proceedings of the 28th EAAP Conference, Valencia, Spain (Vol 2 pp.354-384). Valencia: European Association for Aviation Psychology.

Damitz, M., Eißfeldt, H., Grasshoff, D., Lorenz, B., Pecena, Y. & Schwert, T. (2000). *Validierung des DLR-Auswahlverfahrens für Nachwuchsfluglotsen der DFS Deutsche Flugsicherung GmbH: Ergebnisse des Projektes Qualitätssicherung*. DLR FB 2000-45. Hamburg: DLR.

DFS German Air Navigation Services (2010). *DFS Initial Training for ATCO Syllabus and Training Event Plans. Version 1.1.* Langen: DFS.

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.

Eißfeldt, H. & Maschke, P. (1991). *Bewährungskontrolle eines psychologischen Auswahlverfahrens für den Flugverkehrskontrolldienst anhand von Kriterien der Berufsausbildung*. DLR-FB 91-11. Hamburg: DLR.

Heintz, A. (2004). *Cost-benefit analysis in the selection of Air Traffic Controllers.* In Proceedings of the 26th EAAP Conference - Aviation Psychology: Costs and Benefits. Sesimbra: European Association for Aviation Psychology.

Höft, S. & Pecena, Y. (2004). Behaviour-oriented evaluation of aviation personnel: an assessment center approach. In Goeters, K.-M. (Ed.), *Aviation Psychology: Practice and Research* (pp. 153-170). Hampshire, England: Ashgate Publishing Ltd.

Johnson, J.T. & Ree, M.J. (1994). Range J: A Pascal program to compute the multivariate correction for range restriction. *Educational and Psychological Measurement, 54,* 693 - 695.

Pecena, Y. (2003). *An Assessment-Center Approach to ATCO Selection – An Evaluation Study.* In Proceedings of the second EUROCONTROL Selection Seminar (pp. 201-208). Luxembourg, 16.-18.Nov.2003.

Rosenthal, Robert & Jacobson, Lenore (1992). *Pygmalion in the classroom.* Expanded edition. New York: Irvington.

Seidel, K., Pecena, Y., & Eschen-Léguedé, S. (2009). *Computer-based assessor training - a possibility to solve the dilemma of economical efficiency and quality?* In Proceedings of the 14th European Congress of Work and Organizational Psychology, Santiago de Compostela, Spain, 13.-16. May 2009.

REDUCING PATHOLOGICAL STRESS EFFECTS AND INCREASING PILOT PERFORMANCE DURING UNEXPECTED IN-FLIGHT EVENTS.

Wayne L. Martin; Patrick S. Murray; Paul R. Bates

Griffith University Aerospace Safety Centre, Brisbane, Australia

The inherent reliability of the modern aircraft means pilots rarely experience actual emergencies, or novel, unexpected events. When events do occur, increased arousal levels may have pathological effects on pilots' abilities to deal optimally with the situation, leading to increased likelihood of undesired aircraft states. Amygdala based appraisals of unexpected events may cause over-arousal through lack of expectation, lack of previous experience of such events (either directly or vicariously), and through poor individual perceptions of the ability to handle such events. Routine discussion of novel or emergency events widens pilots' event knowledge database and raises expectation of event occurrence. Individual perception of efficacy in such events is heightened through increased and more readily accessible knowledge, allowing more positive appraisals, which reduces arousal level and improves performance. A pilot study using scenario based discussion at a New Zealand Airline showed very positive perceptions of utility and efficacy and will be discussed.

The ubiquitous reliability of the modern aircraft has added substantial improvements to air safety. However, as a result the average pilot rarely experiences real emergencies, or novel, unexpected events. While simulator training allows exposure to such events, this happens rarely, perhaps only four days per year, with prolonged periods of routine operations out on the line, the norm. While older pilots may remember the days when engine failures were not uncommon, the airline pilot of today could go through the remainder of their career with some statistical surety of never experiencing a major powerplant failure. Engine reliability is such that the prevalence of other systems failures or automation related issues are more commonly cited in modern aircraft incident and accident statistics. Accidents such as the Air France Flight 447 loss of control over the Atlantic (BEA, 2009), the Turkish Airlines Flight 1951 loss of control on approach at Amsterdam (The Dutch Safety Board, 2010), and the Qantas A330 incident off Western Australia involving a loss of control inflight (ATSB, 2008) are typical recent examples where unexpected events resulted in accidents or undesired aircraft states.

When actual events do occur, increased arousal levels may have pathological effects on pilots' abilities to deal optimally with the situation, leading to an increased likelihood of undesired aircraft states. Over-arousal or acute stress has been strongly associated with reduced performance (Stokes and Kite, 1994) with Amygdala based appraisals of unexpected events at times causing over-arousal through lack of expectation, lack of previous experience of such events (either directly or vicariously), and through poor individual perceptions of the ability to handle such events. Anything which can be done therefore to reduce stress levels during critical events is likely to engender positive effects on performance.

Routine discussion in the briefing room or the aircraft of novel or emergency events has several benefits during subsequent critical incidents. As well as widening pilots' event knowledge database, it raises expectation of event occurrence. Individual perception of self-efficacy in such events is heightened through increased and more readily accessible knowledge, allowing more challenging, rather than threatening appraisals, and therefore reducing arousal level. A pilot study using scenario based discussion at a New Zealand Airline in 2010 showed very positive perceptions of utility and self-efficacy amongst the Pilots who participated.

Discussion

Substantial data exists (eg. Boeing, 2010) which suggests that the reliability of modern aircraft, coupled with increased flight path awareness tools such as Enhanced Ground Proximity Warning Systems (EGPWS), Vertical Situation Displays (VSD), Head Up Displays (HUD), and Electronic Flight Bags (EFB), have greatly reduced the number of Controlled Flight Into Terrain (CFIT) accidents over the last few decades. Coupled with ubiquitous reliability in modern aircraft engines and improved systems, the trend in flight safety is continuing to improve. Regardless of these tools and equipment, aircraft still continue to have accidents, with the predominance in the statistics suggesting that in-flight loss of control is now the most common cause (Boeing, 2010).

Of these accidents, humans have continued over the evolution of aviation, to contribute around 70-80% to contributory causes (O'Hare, Wiggins, Batt, & Morrison, 1994; Wiegmann and Shappell, 1999; Yacavone, 1993) and while the technology has morphed over the generations, so too has the nature of the human contribution. Typical areas of concern in modern aircraft generally include automation management, loss of situational awareness, poor judgement and decision making, vigilance issues, complacency, spatial disorientation, and physical factors. While these issues have always been contributory, the very nature of the modern airliner, with its automation, reliability, and endurance, has made some of these more of an issue than was previously the case.

The level of pilot performance during unexpected, novel, or emergency events varies widely. In some, well documented cases, such as those on American Airlines Flight 1592 which ditched in the Hudson river in New York (NTSB, 2009), the British Airways Flight 009 which lost all four engines in a volcanic ash encounter in Indonesia (UK AAIB, 1982), or those on United Airlines Flight 811 which made an emergency landing following a cargo door failure (NTSB, 1992), the pilots performed exceptionally well, both individually and as a crew. In other cases however, pilot performance has been badly affected by startle and/or acute stress effects resulting in undesirable aircraft states or even accidents. Recent examples include the Turkish Airlines Flight 1951 accident at Amsterdam (The Dutch Safety Board, 2009), and the Air France Flight 447 which crashed in the Atlantic, where the pilots failed to recover from what appear to be fairly recoverable situations following unexpected events.

Stress and under-performance are likely where pilots have never considered how to deal with a situation before (Hancock & Szalma, 2008). Generating a solution to a problem and putting into effect a strategy for dealing with it are much harder from scratch, particularly under the effects of acute stress. Having a stored plan for dealing with a range of novel events, allows individuals to simply recall these strategies and apply them to the situation at hand; a much simpler task under stress. These "cognitive pre-plans" may simply be management strategies which can be applied to a number of conceivable events.

Startle is a phenomenon which affects all humans and most animals, and varies in its intensity depending on individual susceptibility and expectation levels (Muto and Wierville, 1982; Warrick, Kibler & Topmiller,1965). Research has shown that startle can affect information processing for up to 30 seconds (Thackray, 1988; Vlasak, 1969; Woodhead 1959, 1969), which in a dynamic, complex situation such as an unexpected novel or emergency event, can have substantial effects on situation outcome. It is likely that these effects are generated not from the initial involuntary startle reflex effects (such as eyeblink and aversive movement), but rather from an amygdala based activation of the sympathetic nervous system which may accompany it, which is epitomised by the “fight or flight” reaction (Canon, 1929). Unfortunately there is little that can be done to overcome these phenomena other than to raise expectation levels through conditioning (Roberts, 2003), and greater exposure to such events which creates “previous experience” and a sense of self-efficacy which will reduce sympathetic nervous system arousal levels.

Acute stress caused by over-arousal can have significant effects on situation outcome, with emotion-focussed coping mechanisms such as freezing or denial having potentially disastrous consequences. Stress occurs when individuals appraise a situation as potentially harmful and that they are insufficiently equipped to deal with it. Primary appraisal is the amygdala based assessment process which determines whether a stimulus is benign/positive or involves loss, harm, threat or challenge. Secondary appraisal determines the best method of coping with threats and is generally dealt with in two ways: problem-focussed coping, in which the individual deals with the problem, and emotion-focussed coping which simply changes the individual’s relationship to the problem (Lazarus & Folkman, 1984; Monat & Lazarus, 1991). Of these, emotion-focussed coping may use problematic methods such as freezing or denial, and will likely therefore have negative implications for situation outcome. The following diagram is proposed to show the relationship between appraisal and information processing:

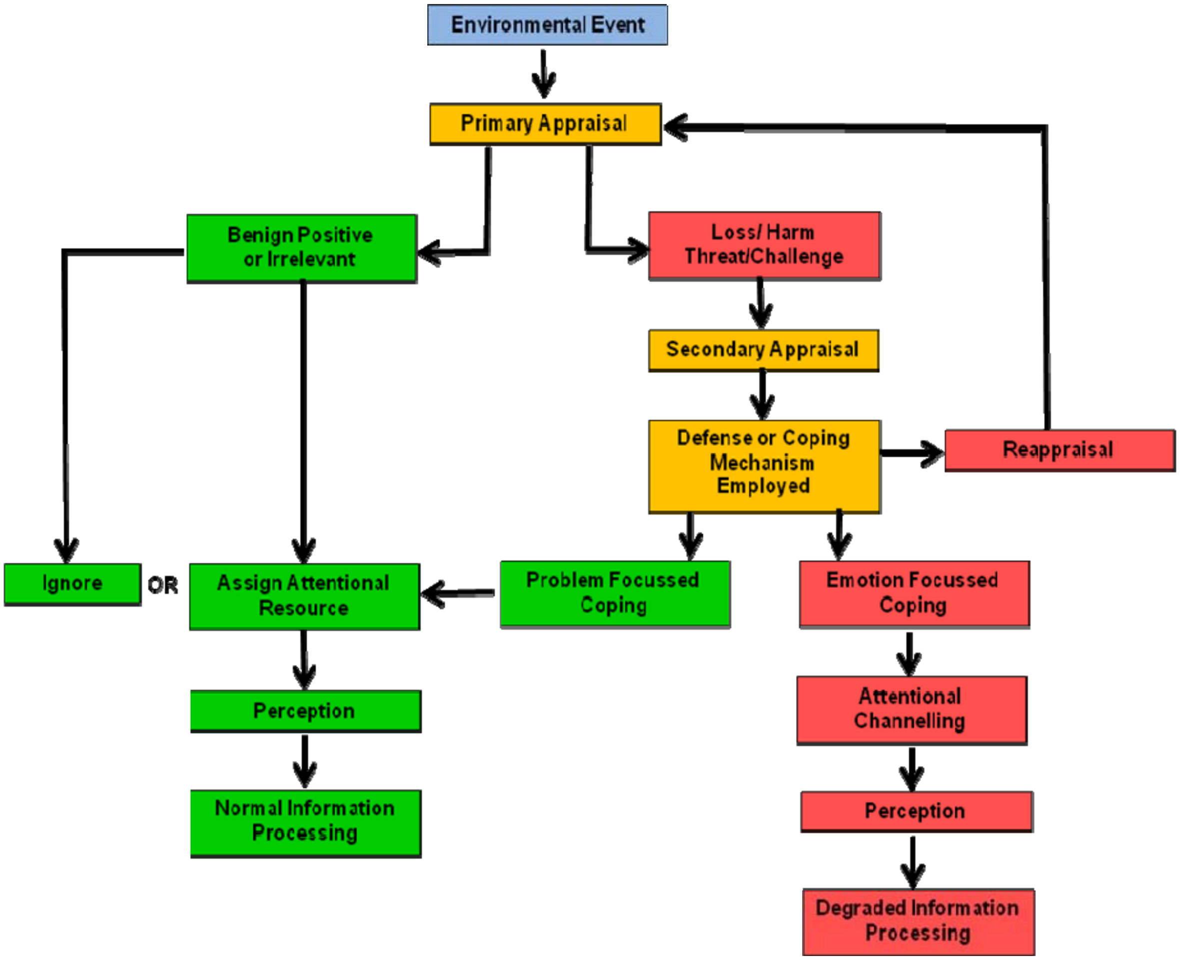
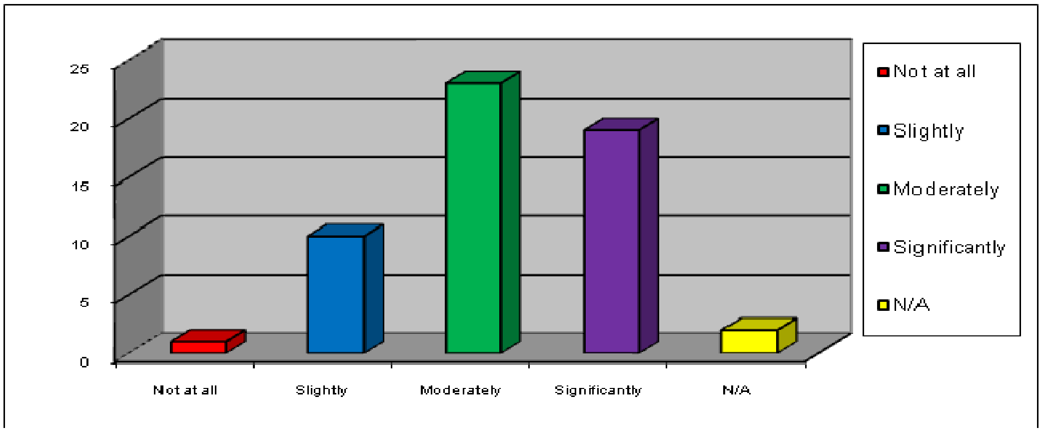


Figure 1. A Conceptual Model of Appraisal and Information Processing

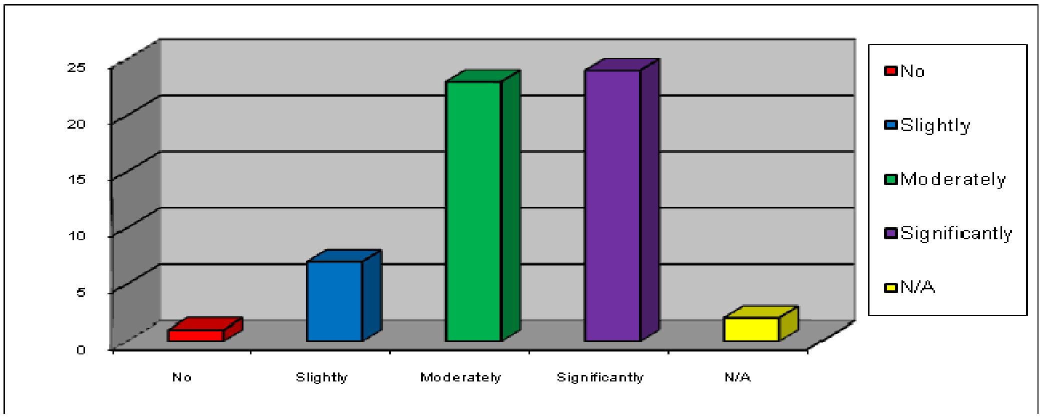
(Martin, Murray & Bates, 2010)

A pilot study was carried out in 2010 at a New Zealand based international airline (Martin, Murray & Bates, in press). The project was entitled "*What would you do if....?*" and involved ten weeks of trial followed by a short survey. During the trial pilots were encouraged to spend a few minutes enroute each day discussing novel events and emergencies, considering how they would handle the aircraft, what they would do in terms of diversions, checklists and communications, and what best resources to utilise to deal with the problem. Following the trial a survey was conducted which analysed the sense of utility and self-efficacy for novel events which pilots developed during the trial. The following graphs show some of the significant results:

Question 4: Do you think that these discussions have raised your expectation level for surprise events?



Question 6: As a result of these discussions do you think that you would be better prepared to handle one of these novel or emergency events if it happened unexpectedly?



(Martin, Murray & Bates, 2011)

Conclusion

Reduced expectation of failure and conditioned over-confidence in aircraft reliability are realistic consequences of ubiquitous normality in the sophisticated and failure-tolerant aircraft becoming more and more prevalent in airline operations. This lack of expectation, coupled with an enduring emphasis on traditional failures during training and checking, are having negative effects on situation outcomes during unexpected novel or emergency events.

Pathological stress effects such as denial and freezing, coupled with over-arousal associated with startle or surprise, can be mitigated to some extent by greater levels of expectation, and greater pilot self-efficacy for the handling of such events. This expectation and efficacy can be improved by organisational and personal interventions which would encourage pilots to discuss novel and emergency events during quiet periods enroute, as a means of developing "cognitive pre-plans" for dealing with such events. The traditional "What would you do if .... happened?" has commonly been used in military transport operations and in airline command training, but has been under-utilised in normal line operations. Discussion of novel events allows pilots to form a plan for what they would do in a given situation, free of stress and startle effects, which they can then store as a series of related processes and strategies for some time in the future. Regular revisiting of these strategies allows for consolidation in long-term memory and associations with a large range of situations. This in turn allows these strategies to be utilised from memory in the event of a novel incident, either directly, or through association with some previously considered and similar event. Strong memories in the long-term memory are comparatively more resilient to stress effects, and the greater the depth and breadth of previous event "pre-plans" therefore, the more seamless utilisation of effective processes and strategies for dealing with such events or similar events.

A pilot study at a New Zealand based international airline found that pilots who participated in scenario discussions of novel events generally had raised levels of expectation and a greater sense of self-efficacy for dealing with such events. A willingness to continue with scenario discussions beyond the study indicates the sense of utility of the discussions which was evident amongst participants.

As evidence based training starts to become more widespread and an acknowledgement of changing needs to deal with real world failures in aircraft becomes recognised, the use of scenario based discussions to bolster the capabilities of pilots in dealing with unexpected novel and emergency events would seem to be complementary. Further research is warranted.

References

ATSB (2008). *In-flight upset, VH-QPA, Airbus A330-303, 154 km west of Learmonth, Western Australia, 7 October 2008.*. Retrieved from: http://www.atsb.gov.au/publications/investigation_reports/2008/aair/ao-2008-070.aspx

BEA (2009). *Interim Report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight 447 Rio de Janiero – Paris.* Retrieved from http://www.bea.aero/docspa/2009/f-cp090601e1.en/pdf/f-cp090601e1.en.pdf

Boeing, (2010). *Statistical Summary of Commercial Jet Airplane Accidents: Worldwide Operations 1959-2009* Retrieved from: http://www.boeing.com/news/techissues/pdf/statsum.pdf

Cannon, W. B. (1929). *Bodily Changes in Pain, Hunger, Fear and Rage: An Account of Recent Research Into the Function of Emotional Excitement*, (2nd ed.). New York: Appleton-Century-Crofts

Hancock, P. A. & Szalma, J. L., (2008). *Performance under stress.* Aldershot, UK: Ashgate

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping.* New York: Springer.

Martin, W.L., Murray, P.S., & Bates, P.R. (2010). The effects of stress on pilot reactions to unexpected, novel, and emergency events. In *Proceedings of the 9th International Australian Aviation Psychology Association Symposium,* Sydney, April 18-22, 2010.

Martin, W.L., Murray, P.S., & Bates, P.R. (2011). What would you do if.? *Aeronautica 1,* (1), 1-16.

Monat, A. & Lazarus, R. S. (1991). *Stress and coping.* New York: Columbia University Press

Muto, W. H., & Wierville, W. W. (1982). The effect of repeated emergency response trials on performance during extended-duration simulated driving. *Human Factors, 24,* 693-698.

NTSB (1992). *Aircraft Accident Report* Number AAR-92/02: United Airlines Flight 811, Boeing 747-122, N4713U, Honolulu, Hawaii, February 24, 1989. Retrieved from: http://www.ntsb.gov/publictn/1992/aar9202.htm

O'Hare, D., Wiggins, M., Batt, R., and Morrison, D. (1994). Cognitive failure analysis for aircraft accident investigation. *Ergonomics*, 37, 1855-69.

Roberts, R. C. (2003). *Emotions: An essay in aid of moral psychology.* Cambridge, UK: Cambridge University Press.

Stokes, A. F. & Kite, K. (1994). *Flight stress: Stress, fatigue, and performance in aviation.* Aldershot, UK: Ashgate.

Thackray, R. I. (1988). Performance recovery following startle: A laboratory approach to the study of behavioural response to sudden aircraft emergencies. *FAA Technical Report No. DOT/FAA/AM-88/4, Civil Aeromedical Institute, Federal Aviation Administration,* Oklahoma City, USA.

The Dutch Safety Board (2010). *Crashed during approach, Boeing 737-800 near Amsterdam Schipol Airport, 25 February, 2009.* Retrieved from: http://www.onderzoeksraad.nl/docs/rapporten/Rapport_TA_ENG_web.pdf

Vlasak, M. (1969). Effect of startle stimuli on performance. *Aerospace Medicine, 40,* 124-128

Warrick, M. J., Kibler, A. W., & Topmiller, D. A. (1965). Response time to unexpected stimuli. *Human Factors, 7,* 81-86.

Wiegmann, D. & Shappell, S. (1999). Human error and crew resource management failures in Naval aviation mishaps: A review of U.S. Naval Safety Center data, 1990-96. *Aviation, Space, and Environmental Medicine,* 70, pp. 1147-51.

Woodhead, M. M. (1959). Effect of brief noise on decision making. *Journal of The Accoustic Society of America, 31,* 1329-1331.

Woodhead, M. M. (1969). Performing a visual task in the vicinity of reproduced sonic bangs. *Journal of Sound Vibration, 9,* 121-125.

Yacavone, D. W. (1993). Mishap trends and cause factors in Naval aviation: A review of Naval Safety Center data, 1986-90. *Aviation, Space and Environmental Medicine,* 64, 392-5.

SIMILAR ACCIDENTS: LESSONS LEARNED? COUNTERMEASURES!

Ronald John Lofaro, PhD
Embry-Riddle Aeronautical University Worldwide, adjunct Associate Professor; (FAA ret.)
Orange Beach, AL 36561
Captain Kevin M. Smith
United Air Lines Captain, (ret.), USN Captain (ret.)
Mesquite, NV 89027

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.

All-too-many accidents in the past decade have major similarities. Yet, we continue to have these accidents. Are there countermeasures to reduce the risks of similar (or non-similar) accidents ? We propose a differing view of accident causation that can lend itself to preventing future accidents and uncovering more generalizable causes. Many accident reports fall back on the standard panacea for accident prevention: training. Training is a large part of the answer in accident prevention**, but** only very specialized, focused decision performance training with rigorous evaluation. We will show a template for such training and evaluation as another countermeasure, one that provides a uniform approach and guidelines for flight simulator scenarios (LOFT). However, we have concluded that even these efforts at a optimum solution still may fall short. We look towards an automation-based, threat identification/risk management cockpit display, incorporating all of the above to prevent runway excursions and other CFIT.

Aviation safety demands accident prevention. Today, we have the National Transportation Safety Board (NTSB) tasked with investigating major accidents, and each military service has a Safety Center investigating their own aviation accidents. All produce reports seeking to identify causation and recommend changes to preclude future accidents. Yet, we have hit a 40 year plateau in the accident rate, which may indeed be slightly rising. Why? The focus here will be on what is now called runway excursions - usually overruns after landing - which continue to be by far the most common type of aircraft accident, normally leading to aircraft damage, and, on occasion involving fatalities. To date, no reports seem to have impacted runway excursions. The level of concern with excursions has reached a stage where the FAA is doing a study, led by Dr. Kathy Abbott, on such excursions: "Operational Use of Flightpath Management Systems." While not yet finished, preliminary findings suggest flightcrew have never been properly trained to operate highly automated aircraft. Perhaps they cannot, at present, be effectively trained, the FAA report implies, because there are no checklists for many of the automation-related problems that pilots frequently encounter. This leaves them having to manage using ingenuity. The authors believe that a decision-making in the operational environment is a lacuna in training; such training and evaluation in specially-designed LOFT scenarios may overcome the need for instantaneous ingenuity. The authors' have written often, beginning in 1993, Lofaro and Smith (1993), and continuing even now, that the overarching issue in safety was actually that of decision making; that risk identification and management is the primary role of the captain/flight crew; that decision making is the functional aspect of risk management. To resume: the FAA report's message is that regular systemic failings in training, identified in real operations, show that airline operations today contain an identified potential for hundreds of latent accidents and incidents unless changes are made. (the authors heartily concur). Recent runway excursions will be quickly examined for similarities in conditions, actions, consequences. This paper will discuss accident causality models and generalizability of findings. Next, using the literature, indicate generic but effective training modalities of accident prevention for this spate of similar accidents and for any accident with similar conditions. These modalities are embedded in a framework for generic accident prevention training and evaluation based on decision performance, and refer to a template for flight simulator (FS) LOFT scenario design that allows for both flight control and cognitive (CRM/ human factors) skills with evaluation. Finally, the paper will attempt a step into the future as we now believe that training, practice, evaluation may need another, more direct, boost. This paper proposes a mission adaptive, real-time, collaborative dispatch-flightcrew display; a display that enables both threat identification and risk management (TIRM), and is an accurate decision-making aid. We will lay out a template and protocol for developing an automation-based, threat identification/risk management cockpit display and its placement an existing cockpit display. We would be remiss if we did not give several *caveats*: The paper is limited to 6 pages. This will preclude detailed presentation, analyses and argumentation, and will also preclude presentation of most material from NTSB (Reports are available at www.ntsb.gov).

Several Similar Accidents

Accident One

Southwest Airlines flight #1248 attempted a landing at Chicago's Midway International airport in adverse conditions, rolled through a blast fence, an airport perimeter fence, across a crowded adjacent roadway striking several passenger automobiles. One bystander's life was lost; he was parked on a street across from the airport perimeter fence, several people on-board seriously injured, and property destroyed. The conditions, at the time of the approach and attempted landing were low visibility and falling snow, compounded by the fact that the runway was slippery and braking action advisories were in effect. The Captain had about a 200 foot ceiling and ½ mile visibility, close to CAT 1 minimums, and unfavorable wind conditions, with a short runway. When the factors are combined (cumulative risk), the Captain had an aircraft that was outside normal conditions and into non-normal/emergency operating conditions: high risk. The touch-down point was at 4,500 feet down the 6,502 foot runway, when (with the plane's weight and speed) it would have needed 5,300 to safely stop. At touch- down, the plane was slightly over the speed for landing, and had there not been some dryer runway conditions on portions of the runway, the plane would have had an excursion that may have ended by crashed into the building across the street with resultant loss of lives and injuries.

Accident Two

American Airlines (AA 331) attempting to land in Kingston, Jamaica, in heavy rain, departed the runway, blasted through a fence, skidded across a road. It broke apart and halted a scant 40 feet from the Caribbean Sea. This flight did result in four serious injuries, all to passengers. A quick aside: at the writing of this Chapter, there is no NTSB report, preliminary or final, available. The data presented comes from various news services. The crew had contacted Jamaica Air Traffic Control (ATC) to request the Instrument Landing System (ILS) approach for Runway 12. ATC, however, advised them of tailwind conditions on Runway 12 and offered a circling approach for landing on Runway 30. The crew repeated their request for Runway 12 and were subsequently cleared to land on that runway with the controller further advising the crew that the runway was "contaminated" (wet). When the risk factors are combined, the Captain had an aircraft that was outside normal conditions and into non-normal/emergency operating conditions and, therefore at high risk. Because of the 14 knot tailwind, groundspeed at landing was 162 knots (186 mph), far above the recommended landing speed for this aircraft. The aircraft, with a heavy fuel load, touched down some 4,100 feet down the 8,910 feet long runway. Normal touchdown would be between 1,000 feet and 1,500 feet. The flight crew, with an over-the-limit tailwind, "decided" to landed "long and hot" on a contaminated runway with no overrun.

Other accidents

Yet, the above are far from the total runway excursions in the past decade or so. American Airlines (AA) flight 1420 had a very similar disaster in 1999, at Little Rock, AK. The Captain continued an approach when severe thunderstorms were in/over the airport, the crosswind was over AA limits for a crosswind landing, the spoilers were not deployed on touchdown. The plane overran the runway, crashed into light poles and ILS stanchions and eleven persons died, with score of injuries. Southwest Airlines already had an runway overrun accident (SW 1455) in the year 2000 as did American (AA flight 2253) in 2010. There are many others: One-Two-Thee flight 269; Kingfisher flight 4124 and more. All had similar approach and runway conditions. None attempted a missed approach/go-round; diversion to alternate.

Similarities

Given the conditions, these landing should never have been attempted. But, there was no attempt to abandon the approach/ landing and then proceed to the assigned alternate. As a last resort, they could have pulled up and declared a missed approach. Why did not they divert enroute, or abandon the approach, or reject the landing and proceed to the alternate? What occurred in these accidents is series of either non-decisions or poor decisions, culminating in the attempt to shoot the landing. The plane/crew, as per our ODM and risking risk paradigms, Lofaro and Smith, (1993, 1998); Smith and Lofaro (2001b, 2003), were at high risk. But, they were seemingly unaware of, or unconcerned with, risk. They were dealing with one condition at a time and seeing if they could legally attempt a landing. But, rising risk is additive/cumulative. While we tend to think of one risk factor at a time, like wind or visibility, the reality is a cumulative effect that occurs when the real impact of the conditions are taken together, resulting in a much higher risk level than the conditions taken as discrete events. This was not done and the price was paid each time.

Accident Causation

These accidents could be a result of a cognitive-based pair of causes: "plan continuation bias" and workload/overload. In times of stress, the decision maker will choose from among all the information available only those facts which support a preconceived solution, "tunnel vision," i.e., functional fixation and decision bias. As overload occurs, the first action is to prioritize, then shed tasks to reduce overload. These two cognitive (dis)functions can work together, especially in the absence of any decision-making model and training. But, the lack of operational decision making skills and training are the root causes. Why the emphasis on decision-making? It is patently clear that before any action, or non-action, occurs a decision is made or a decision point is missed. An *Aviation Week & Space Technology* article by Taverna and Gallagher (2010) clearly states that, in future cockpits, there are needs for "collaborative decision-making," "aids to decision making," and "cognitive resource management." We later will propose an attempt to meet these needs. The current paradigm in use for finding accident causes is a linear path, often referred to as a causal chain. There are two problems with this: the first being that any accident is a compendium of, almost entirely one-off conditions, personnel and actions. Finding out, even if accurately, what caused airplane X to crash is a snapshot in time; a snapshot that can never be replicated. Minimal, if any, generalizable knowledge inheres in the accident findings. Before we get bombarded with critiques, we do not deny some generalizable knowledge may occur, the operative words being "some " and "may." The second problem we see with accident investigations refers back to the linear pathway viewpoint. We posit that an accident is more on the order of a mosaic, whose pieces are composed of personnel, actions, conditions; a mosaic that can have many ways of fitting the pieces together, with no real negative consequences. But, there exists one or more ways where the "fit" results in an accident. Put another way, there can prior poor decisions and behaviors that seemingly “worked” in the past (blind luck? the “not your day to die” phenomenon?), but, at some point in time, with changed conditions, the fit of these previous decisions and actions no longer make a viable mosaic resulting in an accident. Is there a causation schema that can be more generalizable-at the very least for excursions?

Operational Decision Making and Accident Analysis

This will be a short section; please see full-up exposition at Lofaro and Smith (2008). Since 1993, we have been presenting/publishing on the concepts of a rising risk continuum, the critical mission factors that would make risk rise, and the pilot/crew's primary function of mission of risk identification and risk management. Since approximately the early 2000's, the concept of risk management as an integral part of aviation safety finally seems to have emerged. Over the past 18 years, we have developed an operational decision-making paradigm (ODM) as the way for a flightcrew, in-flight, to identify/manage risk and thus complete the mission safely. ODM is the process, often under time pressure and often with little or no margin for error, that is the functional aspect of risk management by the pilot. ODM involves integrating SA with the recognition and assessment of those factors that are critical to safe flight in order to identify and respond to the risk. If the level of risk rises, effective risk reduction strategies need to be employed to keep that risk within manageable limits.

Components of the ODM Model

The ODM components are: 1.The Operational Envelope (Mission Space). 2. Situational Knowledge/Risk Location (Location within the Ops Envelope). 3. The Critical Mission Impact Areas and the Critical Mission Factors which comprise the Impact Areas/Risk Location (Hostile Agents invading the Mission Space). 4. The Rising Risk Continuum/Risk Location and 5. Cumulative effect (a concept whose use is embedded throughout all the above components). *It must be strongly noted that all of these components are so intertwined that any separation or sequencing of them is artificial.* Integrated throughout the model is the pilot's role as risk manager. We refer the reader to Smith and Lofaro (2001); Lofaro and Smith (2008). We have used ODM as an accident causation tool and shown it to be effective across almost all types of accidents; all involve a set of decisions, see Smith and Lofaro (2009 Workbook). However, we have come to believe that even ODM training and specially designed LOFT scenarios and evaluation are not the whole answer. Having said that, we still go to...

LOFT Design and Components

Again, this will be a short section; again, please see Lofaro and Smith,(2008). We need to get to the crux of this paper: a mission adaptive, real-time, collaborative dispatch-flightcrew display; a display that enables threat identification, risk management, and is an accurate decision-making aid. Flying is an integrated, mission-oriented activity and must be evaluated as such. In the early 1990's, Captain Smith created the framework for a model that demonstrated the CRM human factors skills and the flight control skills are interrelated, interdependent, and often simultaneous in execution. Flight proficiency skills and knowledge are interwoven, interdependent, and necessarily interact with the CRM skills/knowledge differentially across tasks and conditions. These interactions can be identified/specified by a matrix-type crew mission performance model (MPM) using the tasks, which comprise a mission/flight leg; this we term Integrated CRM, (Lofaro, 1992a). Added to Integrated CRM is the Mission Performance Model (MPM) which relates directly to LOFT design and flightcrew evaluation while "flying" a LOF. By using the behavioral markers (behaviorally-anchored descriptors) that make up both the CRM and flight control skills clusters, a LOFT can be designed where the MPM and the ODM are used as structural base. Such LOFT's can be evaluated by using the behaviorally-anchored descriptors mentioned above. Yet, as with the ODM, we again came to believe that even such specialized LOFT training is not the whole answer; on to...

An Automation-Based, Threat Identification/Risk Management (TIRM) Cockpit Display

We may be accused of attempting to find that fool's gold: an error-proof (idiot-proof?) system for safety. Be that as it may, our proposal is to incorporate within the Dispatch-Cockpit system a software tool, with cockpit display, that is designed to quickly and easily communicate mission risk. Such a display tool could be used interactively enroute to bring to the Captain's situational awareness the risk-laden factors that could have a detrimental effect to mission success. Space precludes all but an outline for such an onboard risk management display which we would call a threat display, albeit threat displays for the air transport mission may sound bizarre. However, if one takes a look at a modern cockpit, there are already some sophisticated threat displays. As one example, TCAS II. Adverse winds are, in the ODM model, listed as a hostile agent, i.e., a threat. In fact, windshear is adverse winds and there also currently is a cockpit display for it.

The What and the Where

During the mission planning phase, preload and make visible all current and known mission critical factors/hostile agents. This risk data can be uploaded to the cockpit along with the flight planning forecast (FPF). Next, update risk relevant mission critical factors/hostile agents en route. Dispatchers maintain contact with flight crews after they are airborne to keep them advised of weather conditions, alternate landing plans, and necessary changes in altitude. Dispatchers have many sources of information that are unique, such as holding times, discussions with a crew that just landed ahead of your flight, etc. The Dispatcher can feed the Captain updated, risk relevant information into his and the cockpit's risk displays, as well as legal parameters. As a corollary: The Captain would also, as they are encountered en-route, downlink any critical mission factors to Dispatch. Thus, both the flight crews and Dispatch must identify various intervening factors as hostile agents from the critical mission factors/hostile agents.
An in-flight risk management display could be integrated into the lower engine indicator and crew alerting system (EICAS) of the cockpit display system and be connected to a similar display utilized by the Dispatcher. This display would show a series of three lights: green yellow and red.

Green light----- Risk is Low; normal operations prevail; continue mission as planned.

Yellow Light---Risk has risen to Moderate; one or more mission critical factors are in play and have, or will impact flight safety (e.g. windshear advisories at destination airport). Some modification to flight plan needed.

Red Light----Risk level has risen to High. At least three mission critical factors are in play. Significant alterations to the mission flight plan are urgently required. Example: At any point in time, the dispatcher doing flight following would uplink the illuminated light display, say the yellow light. The Captain could then query the illuminated light to see what has changed and what factors are involved. Say severe turbulence over Iceland was reported by Air France, with resultant on-board injuries. The Captain would either agree to the yellow, or red light, by pushing on the light. If a collaborative modification is developed and agreed-on, the dispatcher would then uplink a reroute away from the severe turbulence, with new fuel calculations. But, if the Captain did not agree to the color-related risk factors, he would not push any light. Thus, there is no degradation of what is called Captain's authority, as per Federal Aviation Regulation (FAR) 91.3.

The How

Begin with joint and interactive training of flight crew and Dispatch of a shared mental model that flight crew and dispatch would use; the ODM. It can be used in the mission planning stage and, in the en-route stage of a flight. Part of this training would be on how to use model. The difficult part is developing a software-driven, cockpit display; the TIRM. ODM-trained Captains and Dispatch personnel could be used here along with software engineers in a set of workshops using a knowledge engineering technique called Small-Group Delphi Paradigm, (Lofaro, 1992b) in this effort. Next would come usability testing involving both dispatch and flight crew, (Maliko-Abraham and Lofaro, 2001a). The final step is implementation. The first air carrier to develop and test this new approach would both serve as a trial case and as a proof of concept. Such dispatch/cockpit crew integration would be entirely at the discretion of each air carrier.

Summary

The pilot/flightcrew's prime function is risk identification, assessment and action. This occurs in an time-compressed, unforgiving, decision making environment. We began with a decision making model (ODM) for training, then went to a specialized LOFT design and crew evaluation template where accident analysis can provides data of use in LOFT design. We went a step beyond by proposing the design and develop of an automated cockpit display based on the ODM. A cockpit display warning system that up-links with Dispatch enabling a collaborative ground-air decision making process. Such an automated cockpit display (TIRM) can enable the pilot/flightcrew to make decisions based on accurate identification and assessment of risks. We see the end result (the TIRM display) as a decision aid in identifying and managing risk, thereby reducing risk to the aircraft and all onboard. Aviation safety and the flying public are the beneficiaries.

References

Lofaro, R.J. (1992a). A Small Group Delphi Paradigm. In The Human Factors Society *Bulletin,* Vol. 35, No. 2.*(* pp.1-6).

Lofaro, R.J. (writer/editor), (1992b). Workshop on Integrated Crew Resource Management (CRM). DOT/FAA/RD-92/5. NTIS: Springfield, VA.

Lofaro, R.J. and Smith, K.M. (1993). The Role of LOFT in CRM Integration. In *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).* University of Denver Press: Denver CO.

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). Taylor and Francis: Philadelphia, PA.

Maliko-Abraham, H. and Lofaro, R.J. (2001a). Air Traffic Delays: New ATC Equipment Coupled With New ATC Candidate Selection Methods May Be One Answer. In *Proceedings* of the 2001 National Air Traffic Controller's Association (NATCA). (pp 132-136). ATCA, Arlington, VA.

Smith, K.M. and Lofaro, R.J. (2001b). A Paradigm for Developing Operational Decision-making (ODM). In *Proceedings* of 2001 SAE World Aviation Congress (WAC) Conference. Seattle, WA: 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.

Smith, K.M. and Lofaro, R.J. (2009). Professionalism in Airline Operations…and Accident Investigations? In *Proceedings of the 15th International Symposium on Aviation Psychology.* (pp. 154-159). Dayton, OH: ISAP. *Note: S*mith/Lofaro did a full-day Workshop with accompanying Workbook.

Taverna, M.A. and Gallagher, B. (2010). Dematerialized. In *Aviation Week &Space Technology,* Vol.172, No.89 (pp.52-53). AW&ST: NY, NY.

TOWARDS INTEGRATING TRAFFIC AND TERRAIN CONSTRAINTS INTO A VERTICAL SITUATION DISPLAY

C. Borst, P. Rijneveld, M. Mulder, M. M. (René) van Paassen
Delft University of Technology, Faculty of Aerospace Engineering
Kluijverweg 1, 2629HS Delft, The Netherlands

Future airspace operations will allow flight crews to plan and fly their own preferred route and time of arrival without much intervention from air traffic control. Thereby, pilots will become more responsible for planning their own route while maintaining safe separations from traffic and/or terrain. This demands for strategic and tactical planning tools that supports pilots in these tasks. The work in this paper focuses on supporting the airborne separation assurance task in the vertical plane by means of portraying traffic and terrain conflict zones onto an enhanced Vertical Situation Display. In a simulator evaluation the experimental display was compared to a baseline display that only showed a terrain profile and intruder aircraft location relative to ownship. The experiment results revealed that although the overlays decreased pilot workload, and resulted in slightly less traffic conflicts, decision-making and conflict awareness did not significantly improve.

To meet the future demand of air transport, in terms of flight safety and its environmental impact, airspace operations are changing. New concepts for Air Traffic Management, such as NextGen and SESAR, permit a more flexible use of airspace (SESAR Consortium, 2007). That is, air traffic operations will be based on trajectory optimization instead of procedural control and each aircraft operating within this new environment will be responsible to adhere to its planned trajectory while maintaining safe separations from traffic, terrain, and adverse weather cells. In other words, more and more Air Traffic Control (ATC) tasks will shift towards the flight deck. To prevent an increase in pilot workload and undesirable side effects, new tools on the flight deck are required to support pilots in these tasks.

One approach may be to delegate airborne separation assurance tasks to the automation. This may be a valid choice, because the employment of automation in aircraft has shown to significantly increase flight technical performance, decrease workload, and increase flight safety over the last three decades. An example of a successful piece of automation is the aircraft's flight management system, that is able to optimize the flight and minimize the fuel use much better than any human could ever perform by means of consulting aircraft operating manuals.

Despite numerous other benefits, automation has also led to a host of human performance problems, including "out-of-the-loop" situation awareness and vigilance problems, transient workload peaks, skill degradation, difficulties in reassuming manual control, and decreased job satisfaction. Billings extensively discussed the pitfalls of current aviation automation and advocated a more human-centered approach to flight deck design since the flight crew still has the final authority and responsibility to ensure safety(Billings, 1997). A human-centered approach starts with the recognition that humans, unlike (current) technology, can adapt their behavior under new circumstances. When human operators are well-informed, this variance in behavior can provide a positive outcome from unexpected events where automation would have dramatically failed. Supporting such adaptive behavior requires a paradigm shift in flight deck design.

A requirement for promoting effective cooperation between humans and automation is that the automation is transparent enough to allow for observing its performance and for comprehending its functionalities well enough. In this paper we explore a constraint-based approach to interface design to accomplish a synarchy between humans and automation. This approach is inspired by the Ecological Interface Design (EID) framework that recognizes the tight inter-connection between humans and technology(Flach, Vicente, Tanabe, Monta, & Rasmussen, 1998). Previous studies in the horizontal plane showed that such a constraint-based approach can be successful(Van Dam, Mulder, & Van Paassen, 2008). The focus of this paper is on the operational context of airborne self-separation tasks in the vertical plane. Therefore, a Vertical Situation Display (VSD) will be used and enhanced to support such tasks.

Towards a Tactical Planning Tool

A VSD is standard in most modern civil aircraft such as the B737-800, the A380, and the B787. A common layout of a VSD, shown in Figure 1, portrays the aircraft's vertical flight status by using the along-track distance on

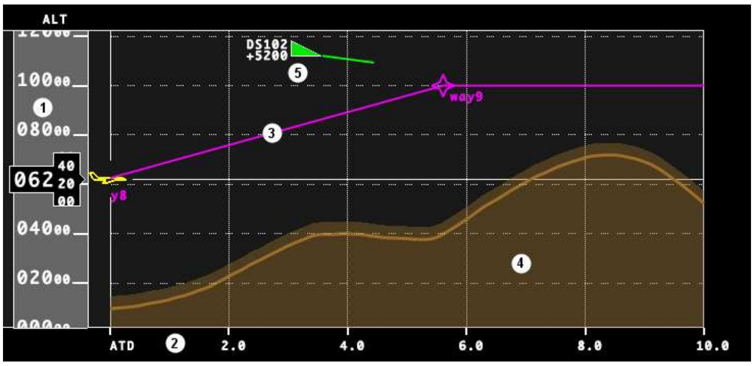


Figure 1: A typical VSD layout showing: 1) altitude, 2) along-track distance, 3) planned trajectory, 4) terrain profile, and 5) position and flight direction of another aircraft.

the horizontal axis and the altitude on the vertical axis. This status picture allows pilots to monitor the flown trajectory, observe traffic, and preview the terrain profile along the planned trajectory.

In a nominal situation, such a status view allows pilots to monitor the aircraft's path following performance relative to the aircraft's intentions. However, in case another aircraft is going to cross the intended flight path, tactical deconfliction may be required to ensure safe separations. It may be clear from Figure 1 that such a status view does not indicate potential conflicts, safe fields of travel, and hint what can be done to circumvent a potential conflict. For example, a deviation from the planned trajectory to mitigate a conflict could very well result in another conflict with other traffic and/or terrain. Of course on the short-term, a Traffic Collision Avoidance System (TCAS) or Terrain Awareness Warning System (TAWS) may issue an alert and command a resolution advisory to prevent a direct collision with another aircraft or terrain, respectively. However, these last-resort warning systems are not well integrated, are known to issue false alarms, and they do not allow pilots to anticipate on and evaluate conflicts and their resolutions(Billings, 1997).

To allow for tactical deconfliction (in the vertical plane) from traffic as well as terrain on a medium-term time scale (such as 5 minutes prior to impact), an extension to the VSD would be necessary. Instead of designing a new type of command display and associated automated algorithms, the automation can supply the visualizations that pilots can use to make their own decisions. Previous studies in the horizontal plane showed that meaningful mappings of the work domain constraints can be visualized in such a way that they allow pilots to *directly* perceive the nature of the conflict and the actions that can be undertaken to ensure safe separations(Van Dam et al., 2008). Inspired by that study, it is hypothesized that similar visual mappings can be useful for the vertical flight situation as well.

Extended Vertical Situation Display

A constraint-based approach to interface design starts with a work domain analysis to identify the constraints governing the work domain. Once a representation of the work domain has been composed, EID continues by finding a visualization for the constraints thus discovered. The goal of EID is to transform a *cognitive* task into a *perceptual* task by providing meaningful information about the work domain that humans can directly perceive and act on accordingly. Previous studies in the application of EID in aviation demonstrated that making the internal (e.g., aircraft maneuvering performance) and external constraints to flight (e.g., terrain and traffic) – and their relationships – perceptually evident on the interface can promote sound decision-making(Van Dam et al., 2008; Borst, Mulder, & Paassen, 2010).

Internal Constraints

The internal constraints are formed by the limitations of the aircraft itself. The internal constraints are divided into maneuvering constraints and energy constraints. The aircraft's vertical maneuver space is defined by its minimum speed, its maximum speed and its climbing capabilities. The minimum speed is defined by the stall characteristics, the maximum speed by the structural limitations, and the climbing capabilities by the maximum thrust of the engines.

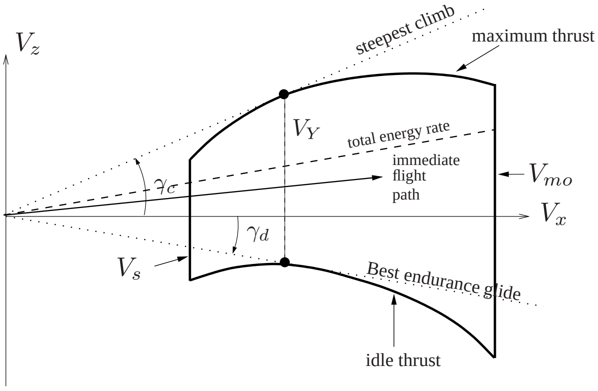


Figure 2: Aircraft vertical maneuvering envelope.

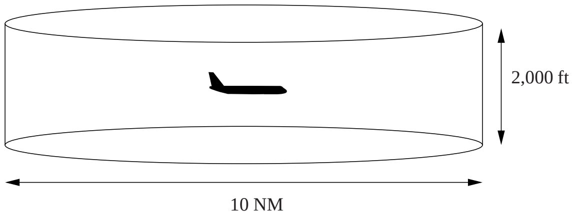


Figure 3: Aircraft protected zone (PZ).

Also the gliding capabilities without thrust may be of interest, for example, when an engine failure has occurred. Altogether these constraints form a boundary for stationary flight conditions, called the performance envelope, see Figure 2. To accomplish a stationary (climbing or descending) flight, it is important that pilots use the throttle and elevator to control the energy state of the aircraft in such a way that the kinetic energy (speed) does not increase or decrease during the climb/descent. Failing to perform a proper energy management strategy can result in stall or overspeed conditions. The aircraft's instantaneous flight path vector and total energy state can also be mapped in the performance envelope (Figure 2).

External Constraints

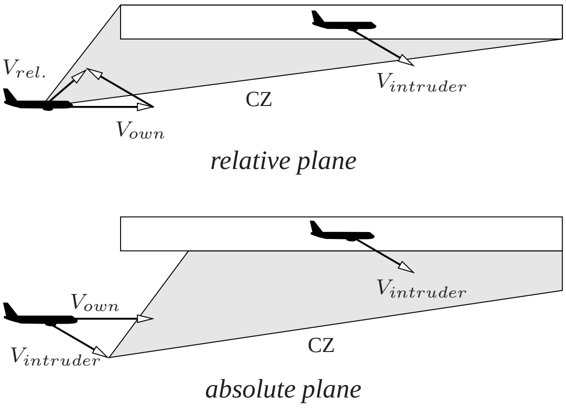
In this paper, only traffic and terrain are considered as external constraints to flight. Regarding traffic, minimum safe separation with respect to other aircraft can be defined using a virtual coin-shaped area around each aircraft, known as the Protected Zone (PZ). The dimensions of this area are the current separation minima: a height of 1,000 ft above and below the aircraft and a radius of 5 NM, see Figure 3. When an aircraft enters the PZ of another aircraft, the separation criteria are violated. The external terrain constraints are formed by the shape of the terrain and (man-made) obstacles. In order to have a safe obstacle clearance a Minimum Safe Altitude (MSA) of 1,000 ft is generally adopted.

Interface Mappings

Enabling pilots to evaluate potential conflicts and the opportunities for resolutions, the external constraints to flight should be made observable relative to the internal constraints. Regarding traffic constraints, basic vector calculus can be applied to construct a conflict geometry that indicates whether or not two or more aircraft are on a collision course. Figure 4 indicates how this is done: from the ownship position two lines are drawn to the left and right most points of the intruder aircraft's PZ. This triangular shape is called the Conflict Zone (CZ). As long as the relative flight path vector stays outside of the CZ, there will be no loss of separation. However, steering the relative flight path vector is a rather difficult and not intuitive task. By applying basic vector geometry the CZ can be translated to the absolute plane such that pilots only have to steer their own flight path vector out of the CZ. The CZ can be directly mapped within the performance envelope: the CZ then marks the area within the performance envelope where the tip of the flight path vector may not be located. Regarding terrain constraints, the highest peak along the route demands the aircraft's minimum climb rate (or potential energy rate) to safely clear the peak. Extending a line from the mountain peak towards the ownship altitude, and within the maneuvering envelope, then defines the safe field of travel (Figure 5). To ensure the opportunity to clear the peak, the terrain peak line should be below the upper boundary of the envelope.

Working with the Cues

Figure 6 shows a screen capture of the extended VSD (EVSD) with an identical situation as displayed in Figure 1. Contrary to Figure 1, it is now clear that there is a traffic conflict and the pilot can resolve the conflict by reducing the airspeed while continuing on the planned trajectory. Thus the status view of object positions has now been extended to a status view of conflicts as well as the opportunities for resolution. The constraint overlays are



relative plane

absolute plane

Figure 4: Conflict Zone (CZ) construction.

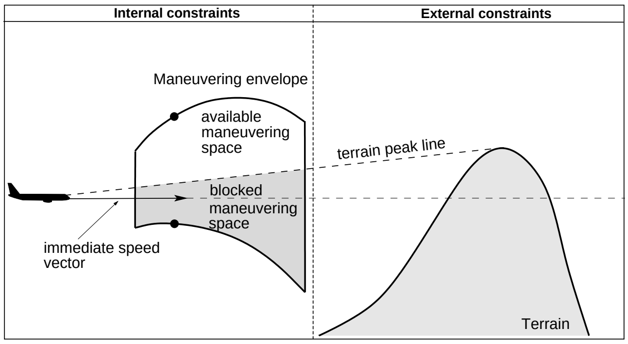


Figure 5: Terrain conflict mapped in envelope.

Figure 6: Layout of the extended VSD (EVSD).

continuously presented, even when there is no immediate threat to safety. As such, pilots can early detect possible threats to safety and avoid them by choosing an efficient and 'economic' maneuver.

Mapping all constraints into one display was challenging as they apply to different domains: traffic and internal maneuvering constraints are primarily defined in the speed and altitude domain, whereas terrain constraints are defined in the distance domain. Combining these two domains into a single display was done by adding a horizontal speed and vertical speed tape along the axes of the EVSD, whereby the speed and altitude domains and linked by a look-ahead time. Therefore, the flight path vector not only indicates the airspeed (length) and flight path angle (direction), but also the horizontal distance and altitude that can be reached within the look-ahead time.

To evaluate the severity (or proximity) of a traffic conflict, pilots can observe the geometry of the CZ. That is, the angle between the legs of the CZ gives pilots an idea of the distance of the intruder aircraft. A large angle between the legs indicates that a conflict is close, whereas a small angle indicates that the intruder aircraft is far away. To resolve a conflict, pilots should simply aim flight path vector outside the CZ by adjusting speed and/or flight direction. The means to aim the flight path vector outside CZ are the throttle (to manipulate the vector's length) and/or the elevator (to manipulate the vector's direction).

For terrain conflicts, the steepness of the terrain peak line indicates the severity or proximity of the conflict. That is, a steep peak line consuming a large portion of the maneuvering envelope indicates a high mountain peak far ahead, or a relatively small peak very nearby. Which one of the two possibilities applies can easily be detected from the shown terrain profile in the distance domain. To avoid a terrain conflict, while performing a steady symmetric climb, the flight path vector and the energy line should be aligned and aimed above the terrain peak line.

Experimental Evaluation

To evaluate the EVSD as a tactical planning and decision-support tool, a pilot-in-the-loop evaluation (in a fixed-base flight simulator) was done using 12 professional glass-cockpit airline pilots in a mixed within- and between-subjects setup. The pilots were instructed to follow a reference trajectory with associated speed commands and avoid any traffic and terrain conflict by solely adjusting either their airspeed or altitude, or both. They were not given specific strategies to solve the conflicts, because the display was meant to support decision-making, rather than command it. Therefore, they were told to solve conflicts in a safe way, with minimum deviations from the intended flight path and airspeed. The trade-off between safety and flight path efficiency was to be made by the pilots themselves.

The *independent* variables of the experiment were the conflict scenario (SCENE, a within-subjects variable) and the display configuration (DISP, a between-subjects variable). SCENE had 12 levels: 4 traffic conflicts, 3 terrain conflicts, 4 mixed conflicts, and one conflict-free scenario. DISP had two levels: a baseline VSD (as shown in Figure 1) and the EVSD. The pilots were divided into two groups. Six pilots only operated with the EVSD while the remaining six operated with the baseline VSD.

The *dependent* measures in the experiment were: 1) the performance, evaluated in terms of target speed deviations, flight-path deviations, and instantaneous load factor, 2) the pilot conflict awareness, subjectively measured by means of a verbal questionnaire during the runs, 3) the pilot workload by means of a NASA TLX rating scale, and 4) the safety in terms of crashes, PZ incursions, and MSA incursions.

Results

Analysis of the performance measure showed that only SCENE had a significant effect on the performance $(F(11, 110) = 4.036, p < 0.01)$. The results, grouped per conflict type, can be seen in Figure 7. Post-hoc analysis (SNK, $\alpha = 0.05$) did not reveal significant differences between these groups. From the figure it can be seen that in the traffic conflict scenarios the average performance scores were higher for the EVSD, meaning that pilots adopted a more efficient strategy to resolve a traffic conflict. When using the VSD, pilots opted for a safe solution by executing the steepest possible climb more frequently. Although safe, these maneuvers were less efficient in most cases. Terrain conflicts, however, were less efficiently resolved by the pilots when using the EVSD. When avoiding traffic, pilots sometimes flew too close to the terrain, causing an MSA incursion and performance penalty.

Regarding conflict awareness, no significant effects were found for neither DISP nor SCENE. Although pilots were inclined to be more ‘conflict aware’ and confident about their answers when operating with the EVSD, no hard conclusions can be drawn.

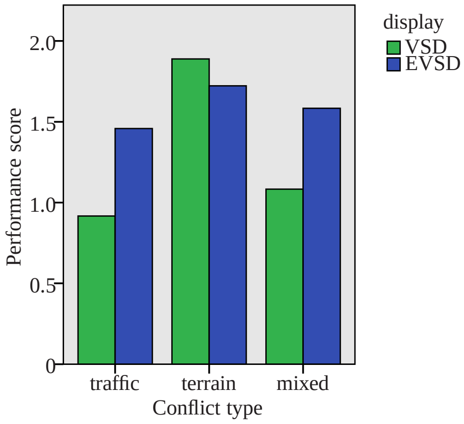
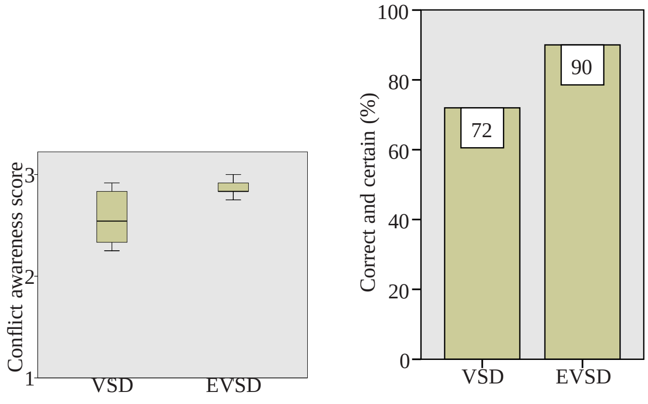


Figure 7: Average performance scores per conflict type.



(a) Boxplot of the conflict awareness scores

(b) Percentage of correct and certain answers

Figure 8: Boxplot of the inflight traffic awareness answers with the percentages of correct and certain answers.

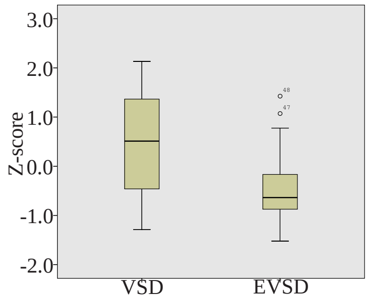


Figure 9: Boxplot of the workload measurements.

	VSD	EVSD
PZ intrusions (66)	5	3
MSA intrusions (42)	0	4
crashes & collisions (42)	0	0

Figure 10: Safety in terms of PZ incursions, MSA incursions, and crashes.

Since workload was only measured at the end of each run and not for each individual scenario, a separate ANOVA (only for DISP) was done. This result was significant ($F(1,47) = 11.542, p < 0.01$). Figure 9 shows a boxplot of the measurements that indicate a higher workload for VSD compared to the EVSD. Investigation of the TLX subscales revealed that the factors 'mental demand' and 'temporal demand' were much higher for the VSD pilots.

Regarding safety, the number of PZ intrusions, MSA incursions and crashes are shown in Figure 10. For EVSD all traffic incursions occurred in one and the same scenario, where the pilots initially solved the conflict by decreasing speed, but they did not anticipate on a later climb, where they needed an additional speed decrease. The MSA incursions for EVSD pilots was because pilots tried to fly as close as possible to the terrain to stay as close as possible to the intended flight path. Additionally, presenting constraint information also invites pilots to fly close the limits of safe system performance(Borst et al., 2010). Despite these results, no crashes have been recorded.

Conclusions and Recommendations

The extended Vertical Situation Display (EVSD), designed using a constraint-based approach to interface design, should allow pilots to diagnose traffic and terrain conflicts in the vertical plane and perceive the opportunities for resolution. An experimental evaluation showed, however, that the EVSD did not prove to be a significant improvement over a baseline VSD in terms of conflict awareness, resolution efficiency, and safety.

For further development of the EVSD, it is recommended to explore scenarios with multiple intruders, changing aircraft configurations (flaps, speedbrakes, gear), and malfunctions. Also, intent information should be included in the conflict zone visualizations to allow for a strategic planning tool. Finally, the effect of wind conditions on the performance envelope and solution space should be investigated since the EVSD relates the aircraft performance envelope, defined in the velocity domain, to distances relative to the ground.

References

Billings, C. E. (1997). *Aviation Automation: The Search for a Human-Centered Approach*. Lawrence Erlbaum Associates.

Borst, C., Mulder, M., & Paassen, M. M. van. (2010). Design and Simulator Evaluation of an Ecological Synthetic Vision Display. *Journal of Guidance, Control, and Dynamics*, *33*(5), 1577–1591.

Flach, J. M., Vicente, K. J., Tanabe, F., Monta, K., & Rasmussen, J. (1998). An Ecological Approach to Interface Design. In *Proceedings of the human factors and ergonomics society 42nd annual meeting* (pp. 295–299).

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. (2008). Ecological interface design of a tactical airborne separation assistance tool. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, *38*(6), 1221–1233.

A VIDEO PROTOTYPING METHODOLOGY FOR EVALUATING NOVEL INTERFACE CONCEPTS IN COCKPIT DISPLAYS

Paul McKay, Stacey D. Scott, Jonathan M. Histon
Department of Systems Design Engineering, University of Waterloo
Waterloo, Ontario, Canada

Gerard L. Torenvliet
Esterline|CMC Electronics Inc.
Ottawa, Ontario, Canada

Modern cockpit displays contain a multitude of complex information sources. Integrating new interface concepts into an existing cockpit display to produce a high-fidelity prototype suitable for user testing can be extremely time-consuming. Discount prototyping methodologies are needed to enable user testing at earlier stages of the design cycle to ensure appropriate changes occur and high quality interfaces result. Video prototyping can provide a useful step between low-fidelity, static prototypes and higher-fidelity software prototypes. However, existing video prototyping methods are designed to elicit user feedback on design concepts. While user feedback is important to the adoption of aviation interfaces, it is also desirable to examine performance using more complex metrics, which have traditionally required the development of a fully interactive software prototype. We propose a new scenario-driven video prototyping methodology that allows designers to apply complex metrics during early-stage user evaluations.

Developing and assessing display concepts for complex, dynamic task environments, such as modern aircraft cockpits, can be both time consuming and resource intensive. The development of a prototype cockpit environment of sufficient fidelity for use in human participant evaluation can require as much as several months of effort from one or more software development experts. In an effort to reduce the work required to develop or modify testable dynamic interfaces, we examined the concept of video prototyping, which has been common in the human-computer interaction (HCI) domain for the past two decades (Mackay, Ratzer, & Janecek, 2000; Vertelney, 1989).

Video prototyping can provide a useful step between lower-fidelity static prototypes, such as paper or still image prototypes, and higher-fidelity software prototypes. However, use of these video prototypes has traditionally been limited to demonstrating interface concepts as a way of obtaining user feedback (Bardram, Bossen, Lykke-Olesen, Nielsen, & Madsen, 2002; Bardzell et al., 2006; Halskov & Nielsen, 2008; Mackay, et al., 2000; Muller, 1991; Tognazzini, 1994; Vertelney, 1989; Young & Greenlee, 1992). While obtaining user feedback is useful and important in the early stages of the user centered design process, the low fidelity and lack of interactivity in these non-software prototypes typically prevents them from being used in later stage performance-based interface evaluations. In the past, this has meant that examining the performance of a design concept based on complex concepts such as situation awareness required the development of a fully interactive software prototype. We propose that a new form of scenario-driven video prototyping, using video authoring techniques to show proposed display concepts layered atop existing interfaces, can provide several advantages over traditional prototyping techniques.

The primary advantage of the proposed technique is that it enables the development or modification of testable dynamic interfaces with reduced time and effort compared to traditional software prototyping. This reduction is possible because the work required is similar to adding special effects to films, meaning that it shifts the type of tools and expertise required from software development to traditionally lower-fidelity techniques like video editing and graphic design. Additional advantages associated with the proposed new technique include:

- the reduction in development time and cost enables user testing to be carried out earlier in the design process;
- the ease of creating overlays on an existing interface allows for rapid, low-cost user interface testing; and
- the scenario-driven methodology allows for performance testing based on complex concepts such as situation awareness.

In the following sections, we overview existing video prototyping techniques and identify limitations that led to the development of our proposed technique, describe the methodology for creating and evaluating a video prototype, and present a case study showing our use of this methodology to develop and evaluate a new display concept for the aviation context.

Video Prototyping

The use of video in the prototyping process began in the late 1980s with initial work primarily using video to record physical prototypes being manipulated by the designers to show their concept of use (Muller, 1991; Vertelney, 1989; Young & Greenlee, 1992). Further work using video as a prototyping tool has developed other ways of recording the design concepts, including using performers to show interaction with mock-ups of an interface (Tognazzini, 1994), using software rendering tools to generate 'virtual' video prototypes (Bardram, et al., 2002; Halskov & Nielsen, 2008), and using video game characters as virtual performers in prototype videos (Bardzell, et al., 2006). While all of these techniques are useful for communicating a design idea, they share two main limitations that prevent them from being used to test the utility of an interface design.

First, the scenarios on which these video prototypes are based are typically created by the designer to showcase their interface rather than on representative use cases. In some cases, such as Bardram (2002) and Halskov's (2008) studies of ubiquitous computing in the medical domain, scenarios capture an envisioned world and so may not be an accurate representation of the way the interface will function in an actual implementation. While these 'mock-up' scenarios can work very effectively as a method of demonstrating a design concept for the purpose of soliciting user feedback, they are not sufficiently realistic to be used for formal evaluations.

Second, many video prototypes are intended to function as storyboards, and so portray users interacting with an interface. As a consequence, the interfaces under design are typically not shown in detail, or are in detail for only a limited amount of time. Tognazzini (1994) talks specifically about some of the filmmaking techniques and directing decisions made in creating a video prototype to limit the amount of 'full-resolution' screen time needed. This works well for storyboarding a design concept as it limits the level of interface development required, but it is less useful for formal interface evaluation as it limits the level of detail available to be studied.

Methodology

To address the limitations with existing video prototypes, we developed a methodology for creating high resolution video prototypes based on realistic scenario data. The methodology has been developed for use at an intermediate stage of the user-centered design process (Preece, Rogers, & Sharp, 2002). The process can be applied to either new interfaces, or modifications of existing interfaces. It is assumed that several iterations of lower fidelity prototypes have been previously developed and evaluated to arrive at a relatively mature design concept. The key elements of the methodology being proposed can be divided into three phases.

Phase 1: Scenario Development and Data Collection

The first step is to develop an appropriate scenario and collect relevant data for use in creating a video prototype. The scenario should be a representative use case for the system being studied, so it will generally be necessary to consult with subject-matter experts to ensure that the scenario is realistic. There are many analysis tools available to assist in selecting and developing a representative scenario for evaluation, such as task analysis (Crandall, Klein, & Hoffman, 2006; Diaper & Stanton, 2003) or cognitive work analysis (Vicente, 1999). In the development and initial case study of this methodology, the analysis method we used was a form of operational sequence modeling (Chapanis, 1996).

Once a representative scenario is selected, data must be collected so that the scenario can be developed into a set of videos showing the proposed interface design in representative task conditions. The relevant data will vary depending on the scenario and the system being designed, but the primary component will be high-resolution video of the interface design showing the state of the display throughout the scenario. This video can be created using a screen-capture tool, such as Camtasia Studio[1], which records live interface graphics and user interactions during system usage. Other potentially relevant data to be recorded could include audio, interaction (cursor input, keystrokes, etc), or gaze-tracking information, among others.

---

[1] http://www.techsmith.com

Phase 2: Video Prototype Creation

Using the data recorded in the first phase of the video prototype methodology, the captured video can be synchronized with the other captured data (audio, interaction, etc) to produce a first set of control videos. These videos are then edited to produce a treatment set of videos showing the proposed interface design. This can be accomplished by using a standard video editing suite to create additional video channels with graphical overlays to show the modifications to the existing interface or the dynamic elements of a new interface.

Phase 3: Video Prototype Evaluation

The final phase of the methodology is the evaluation of the video prototype with representative users to examine the effects of the interface modification or new interface design. This evaluation takes a similar form to a typical usability evaluation with the exception that the participants are not able to interact with the prototype, and therefore need to be given tasks that are appropriate for the scenario context but still allow for assessment based on the desired metrics. Examples of such tasks could include: recording a log of scenario events, evaluating the performance of participants in the video prototype, or additional external tasks such as manual control.

Case Study – Supporting Collaboration in Modern Cockpits

The methodology described above has been used to evaluate an innovative interface design for advanced cockpits. New cockpit avionics architectures are emerging that use cursor control devices and keyboards for pilot interaction with individual and shared displays. This form of architecture has a number of advantages compared to a conventional glass cockpit, but brings some challenges as well. One of these challenges, resulting from the concentration of avionics controls into a keyboard and cursor control device, is crewmembers' potential loss of peripheral awareness cues of each other's actions. The design we developed in this case study aimed to restore some of this lost information by augmenting an existing interface design with information about operator usage history, including both input (keyboard or mouse) and visual (gaze) activity. To ensure that the design concept for visualizing this usage history information was sufficiently mature for performance testing, it was developed through several iterations of low fidelity prototypes before moving onto the video prototyping evaluation.

Phase 1: Capturing a Flight Scenario

An existing high-fidelity software cockpit interface prototype of an advanced two-pilot cockpit that enabled virtual flight simulation was used in phase 1 to develop a representative flight scenario and to collect data of in-flight cockpit display interactions. Three participants (two pilots and one air traffic controller) were recruited to act out the scenario. Display usage of the primary flight and navigation displays were recorded using the FRAPS[2] screen capture tool, while usage of the flight information display was captured using Camtasia Studio. A digital camcorder with lavalier microphones was used to record the prototype setup, including the radio and intercom conversation from the three participants. A gaze tracking system was used to capture the visual interaction data of the co-pilot, while the pilot's visual interaction data were approximated based on a post-scenario interview, screen capture videos, and the wide angle video of the prototype setup.

Phase 2: Creating Control and Treatment Video Prototypes

In phase 2, the collected data were then used to create a control and a treatment set of video prototypes. The treatment videos showed visual traces of operator usage history by placing color-coded borders around the interface components that were viewed or edited by the pilots. The opacity of these borders were adjusted to indicate the recency of use (i.e., when an interface component was viewed or edited, the border for that component was set to full opacity, and would fade away over time when the component was not being used). The Adobe Premiere[3] video editing tool was used to create the video prototypes. Each prototype contained the interface sequences and interface usage history (treatment condition) of 30-minutes of flight scenario captured in Phase 1. Figure 1 (left) shows an un-treated snapshot from one of our screen capture videos, and Figure 1 (right) shows the same snapshot after editing.

---

[2] http://www.fraps.com

[3] http://www.adobe.com/products/premiere

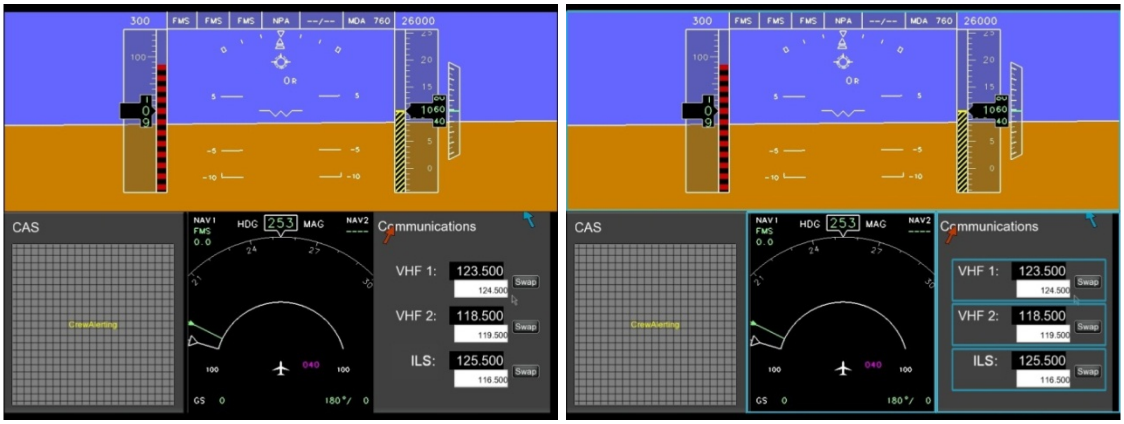


Figure 1. Un-treated screen capture snapshot (left) and proposed interface treatment (right).

Phase 3: Video Prototype Evaluation

In the final phase, a formal user evaluation of the developed video prototype was then conducted with representative users. An experimental display setup was assembled with a form similar to the prototype cockpit that was used as the basis for the design (Figure 2). Eleven trained pilots with a minimum of 15 flight hours participated in the study. A between-subjects experimental design was used, in which the participants watched the videos for either the control condition (five participants) or treatment condition (six participants).

Figure 2. Prototype cockpit used in video prototyping evaluation.

Participants were given two main tasks to perform while watching the scenario videos. Their first task was to take the role of an evaluator. This involved paying attention to the events of the scenario and the actions of the flight crew and, after the scenario, rating their individual and group performance. The second task involved completing a scenario log sheet by recording information about a variety of relevant flight information (such as radio frequency changes and ATC clearances), and the timing of flight events. The primary evaluation task was selected because its continuous cognitive aspect would make the secondary task challenging; instead of simply watching and listening for cues about information relevant to the second task, they needed to pay attention to the scenario events and integrate them into an overall understanding of the scenario and the performance of the flight crew.

Requiring participants to complete a log sheet provided a measure of how well they understood the scenario events and the actions of the flight crew (i.e., a basic measure of situation awareness). A measure of situation awareness was obtained by comparing each participant’s log sheet to a master log sheet that included all possible events. The reliability of the log sheet information depended on participants' cooperation with the data recording process (i.e., participants may understand more information than they record on the log sheet). Events were considered to have been recorded correctly if the participants recorded the correct information (e.g., a new radio frequency) at approximately the correct time (within one minute before or after the actual time).

After the scenario, participants completed a brief questionnaire that included three questions asking them to rate the crew’s performance and one question asking them to rate their confidence that their scenario log sheet captured all the relevant scenario information. Each question used a 7-point Likert-style rating scale. Finally, participants were interviewed using a semi-structured process that elicited additional details on three general topics: the post-scenario questionnaire, the information on the scenario log sheet, and the cockpit interface. Participants in the treatment condition were asked additional questions dealing specifically with the interface augmentation.

Case Study Results and Implications

The video evaluation provided unique and valuable insights into the strengths and weaknesses of the proposed interface treatment. The most interesting results arose from the participant debriefing interviews. Five of the six treatment condition participants reported using the usage history information to maintain awareness of important scenario information. One of these participants, who missed a radio frequency change while looking at a chart, noticed and recorded the change when he looked back up at the display; the interface treatment helped the participant gain awareness of the radio frequency change and, thus, supported their awareness of the situation.

The case study also highlighted an important challenge in performing the video evaluation: selection of the task for observers, and developing relevant quantitative measures of variables of interest. In the case study, the quantitative measure of participant awareness using the log sheet information did not show any statistically significant differences between the treatment and control groups; however, this was likely a result of two main limitations with the evaluation process. First, the use of a between-subjects design with a small sample size made it unlikely that any potential differences between groups would be detected. This limitation could be relatively easily addressed in future studies by increasing the sample size or by using a within-subjects experimental design.

The second limitation of the evaluation process was the use of the scenario log sheets as the primary measure of participant awareness. The self-reported nature of the log sheets made it difficult to ensure that performance was measured consistently across participants. More highly experienced pilots, for example, tended to report much less information, yet demonstrated a clear understanding of the scenario events during the post-scenario interviews. It is possible that experienced pilots decided to focus on watching and evaluating the actions of the flight crew for the purposes of the flight evaluator task, knowing that they could recall the flight event details from memory if needed, while participants with less experience were not as confident in their ability to evaluate the flight crew and instead focused on the log sheet task. Applying a standardized situation awareness measurement technique such as SAGAT (Endsley, 1990) may help reduce such variations in recorded awareness data in the future.

Discussion

The goals for the video prototyping methodology discussed in this paper were to allow user testing at a lower time and cost compared to conventional software prototyping techniques, and to allow this user testing to examine the performance of a design concept based on complex metrics such as situation awareness. The results from our initial use of the methodology indicate that it does have the potential to succeed at both of these goals. The use of video for prototyping allowed a single researcher with little or no software development experience to collect the necessary data, create a prototype, and conduct a performance evaluation with a similar level of effort as would have been required for a software expert to develop an interactive prototype. Additionally, the evaluation using the video prototypes generated results that demonstrated both the utility and the limitations of the proposed interface design concept for supporting awareness.

Methodology Considerations

In addition to addressing the limitations discussed above, several other considerations are relevant for future use of this video prototype evaluation method. Perhaps the most important consideration is the task participants are asked to perform during the evaluation. While the flight crew evaluator task worked well for our scenario as it approximated the task of monitoring a highly automated aircraft, such an evaluation task may not be suitable to other domains; the task must be tailored to the domain and scenario being studied.

Another consideration relates to the use of approximate data for the pilot's point-of-gaze. It was initially unclear whether participants would easily notice a difference in activity between the pilot and co-pilot visual borders, possibly leading participants to distrust the interface treatment and begin to ignore it. However, of the six participants in the treatment condition, only one participant mentioned a difference, observing that the co-pilot's eye movements seemed to dart around more than the pilot's. Obtaining real gaze data for use in the prototype did have some benefits in that it made the process of prototyping the visual borders somewhat faster and ensured that they were accurate representations of the co-pilot's eye activity, but these benefits came at a cost of the time required to set up and calibrate the gaze tracker and analyze the point-of-gaze data. We suspect that using only the approximate method would have greatly accelerated phase 1 with little cost to the realism of the video prototype.

Based on this successful use of approximate gaze tracking data, it is possible that creation of the video prototypes could be further accelerated by using approximate data for other aspects of the prototype. For example, in our evaluation, it is possible that phase 1 could have been accomplished using the automation of the existing cockpit prototype and simulating the input interaction and audio stream for the two pilots and air traffic controller. In this way, the video prototype could have been created by consulting a single expert pilot to confirm that the prototypes were realistic, instead of using three pilots to act out the scenario. It is also possible that the video prototype method could be used in cases where the interface design is not based on an existing prototype by creating a video based on a still image interface design and realistic approximations of a usage scenario.

Conclusion

This paper has presented a new discount prototyping methodology suitable for testing complex and dynamic interface concepts, such as advanced aviation interfaces, at fairly early stages in the design process. The proposed methodology uses a new form of video prototyping that adapts film-like special effects applications to digital video screen captures of existing interfaces. The proposed method enables designers of complex interfaces to begin performance testing of novel display concepts much earlier in the overall design process, and make appropriate modifications, before extensive and costly software development is needed.

Acknowledgements

This research was supported by the Natural Sciences and Engineering Council of Canada (NSERC). We thank members of the University of Waterloo Human Factors Research Group for feedback on the methodology and case study interfaces. We also thank participants of the data collection and user evaluation phases of this research.

References

Bardram, J., Bossen, C., Lykke-Olesen, A., Nielsen, R., & Madsen, K. H. (2002). *Virtual video prototyping of pervasive healthcare systems*. Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques, London, England.

Bardzell, J., Bardzell, S., Briggs, C., Makice, K., Ryan, W., & Weldon, M. (2006). *Machinima prototyping: an approach to evaluation*. Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles, Oslo, Norway.

Chapanis, A. (1996). *Human Factors in Systems Engineering*. New York, New York, USA: John Wiley & Sons, Inc.

Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: a practitioner's guide to cognitive task analysis*. Cambridge, MA: The MIT Press.

Diaper, D., & Stanton, N. (Eds.). (2003). *The Handbook of Task Analysis for Human-Computer Interaction*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Endsley, M. R. (1990). A methodology for the measurement of situation awareness. *Situational Awareness in Aerospace Operations,* 1/1-1/9.

Halskov, K., & Nielsen, R. (2008). Virtual video prototyping. *Human-Computer Interaction, 21*(2), 199-233.

Mackay, W. E., Ratzer, A. V., & Janecek, P. (2000). *Video artifacts for design: bridging the Gap between abstraction and detail*. Proceedings of the conference on Designing interactive systems, New York City, New York, United States.

Muller, M. J. (1991). *PICTIVE- an exploration in participatory design*. Proceedings of the SIGCHI conference on Human factors in computing systems, New Orleans, Louisiana, United States.

Preece, J., Rogers, Y., & Sharp, H. (Eds.). (2002). *Interaction Design: Beyond Human-Computer Interaction*: John Wiley & Sons, Inc.

Tognazzini, B. (1994). *The "Starfire" video prototype project: a case history*. Proceedings of the SIGCHI conference on Human factors in computing systems, Boston, Massachusetts, United States.

Vertelney, L. (1989). Using video to prototype user interfaces. *SIGCHI Bulletin, 21*(2), 57-61.

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work.* Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Young, E., & Greenlee, R. (1992). *Participatory video prototyping*. Extended abstracts of the SIGCHI conference on Human factors in computing systems, Monterey, California.

APPLYING PRINCIPLES OF MUSIC SCENE ANALYSIS TO SYNCHRONOUS AUDITORY WARNING SIGNALS

Matthew J. Davis
The Ohio State University, School of Music
Columbus, Ohio

In emergency situations pilots are often presented with the difficult task of distinguishing between simultaneous auditory warning signals, each with varying levels of importance. This inability to effectively discriminate between synchronous warnings can lead the pilot to ignore certain signals, to misinterpret them, or to be simply unaware of their presence. The creation of signals that could be easily construed and distinguished from other simultaneous signals would not only be desirable but would also contribute to increased situational awareness and better decision-making during emergencies.

The focus of this study centers on creating a system of rules or methods for designing auditory warnings based on stream segregation principles common in music. These methods were derived from a study by David Huron which examined the role of voice-leading rules in music stream analysis (Huron 2001). "Voice leading" refers to a set of compositional options that is taught to university students of western classical music. These rules help students compose music that is consistent with a particular classical style. For instance, one of the common rules is the avoidance of parallel octaves. A typical university music student might be told to avoid "parallel octaves" because it "sounds bad". However, in Huron's study, a parallel octave might be explained as having poor streaming potential. When the goal is to maximize the number of musical streams ("voices"), the compositional technique must take into account (however indirectly) a system of rules or preferences that maximizes the listener's ability to distinguish between voices.

While much research has been conducted examining the appropriate properties of warning signals, (Puyer 2005, Selcon 1995, Stanton 1994, Stanton 1999), it would be prudent to examine the contributions of auditory streaming research in relation to music. This can be accomplished by manipulating the onset synchrony, timbrel differences, binaural qualities, rhythmic identities, amplitude modulation, temporal continuity, and pitch proximity, to name a few. By applying these principles to warning signals, this study has sought to create a system of auditory warnings that contains more efficient differentiating properties in addition to conforming to a more unified stylistic entity.

Method

Using Huron's musical perceptual principles as a guideline, this study proposes a set of rules for maximizing the distinguishing properties of auditory warnings.

Table 1

Auditory Warning Rules Derived from Common Voice Leading Principles

---

1. Maintain a minimum frequency separation of 30-50 Hz between signals
2. Avoid concurrent pitches that share musical unisons, octaves, or fifths
3. Intensity Contour (gradual onset, brief peak, gradual reduction to normal intensity)
4. Urgency (or primacy) Separation (High, Medium, Low)
5. Multi-Signal Rhythmic Synchrony/Asynchrony
6. Harmonicity/Non-harmonicity
7. Optimal Pitch Range for Warnings (600-1200 Hz)
8. Optimal Pitch Range for Cautions (175 Hz - 600 Hz)
9. Overcome Cabin Noise (60-88dB in jet aircraft, 70-90 dB in GA)
10. Timbrel Differentiation (sharp timbres for warnings, mid timbres for cautions, soft timbres for advisories)
11. Timbrel Identities (create timbrel "families" linked with certain systems)
12. Rhythmic Identities (create rhythmic "families" linked with certain functions)
13. Tempo Modulation (increase tempo as urgency increases)
14. Mix tones and speech for high mental load tasks (i.e. whooop whooop, "pull up")
15. Melodic Motion (i.e. upward for "pulling up", downward for "pushing down", etc.)
16. Pitch Urgency Scale (most urgent warnings contain highest pitches)
17. Signal Component Limitation (signals should contain no more than three successive pitches)
18. Signal Familiarization (familiar signals should be used when possible)
19. Automatic Intensity Adjustment adjusts to allow for radio communication. (Peryer 2005)
20. Signal Localization (signals sound as though emitted from their cooperating warning lights

---

Once these methods were created and a demonstration set of warnings was derived, it seemed necessary to compare the new signals with warnings currently employed in aircraft. Both the new and old signals were employed in a correlational pilot study that made use of a simple task observing a participant's ability to distinguish between various synchronous signals. The task was employed using Max/MSP, a visual programming language designed for multimedia. In this case, it was used both for data collection and for designing the new warning signals. The audio samples were divided into two groups: *"Old"* referring to the signals recorded from the aircraft, and *"New"* referring to the signals created expressly for this study. The groups were further subdivided into 1-6 groups of synchronous signals, that is, how many signals would be playing at the same time. (Refer to Table 2).

Table 2

Division of Groups and Synchronous Signals

Old Signals	New Signals
1 Signal at a time	1 Signal at a time
2 Signals at a time	2 Signals at a time
3 Signals at a time	3 Signals at a time
4 Signals at a time	4 Signals at a time
5 Signals at a time	5 Signals at a time
6 Signals at a time	6 Signals at a time

The experiment was initiated when a participant clicked on the *"Play/Next"* button, which caused the program to randomly select between the "old" or "new" groups before randomly selecting between 1-6 synchronous patterns. Thus, when "*Play/Next"* was selected, a random group of synchronous signals would play. The participant would then select how many signals he or she could identify. Once sure of their answer, they would then select the "*Submit Answer"* button, which would stop the auditory stimulus and save the participant's answer. Out of the six groups of synchrony, only six combinations of synchronous signal groups would be selected (randomly) while never repeating. This meant that the "Old/New" groups would be individually heard 36 times, and the experiment (with both groups together) would last exactly 72 trials.

The subject pool was comprised of university music students, both undergraduate and graduate, from The Ohio State University School of Music. While an ideal pool of participants would consist of pilots, trained musicians were deemed a suitable substitute for this pilot study. The experiments took place in the Cognitive and Systematic Musicology Lab at the School of Music.

Old Warning Signals

Warning signals were recorded from a Lear 35 and a Lear 31 provided by Spectra Jet, Inc. In both aircraft, the signals were recorded using stereo microphones placed in the pilot's seat. While every effort was made to record these sounds in the most realistic manner possible, it is necessary to keep in mind that these aircraft were not airborne and did not have an accurate representation of the ambient noise that would occur in flight. To account for this, an audio track of aircraft cabin noise from Microsoft Simulator X used was as ambient noise for both groups of signals.

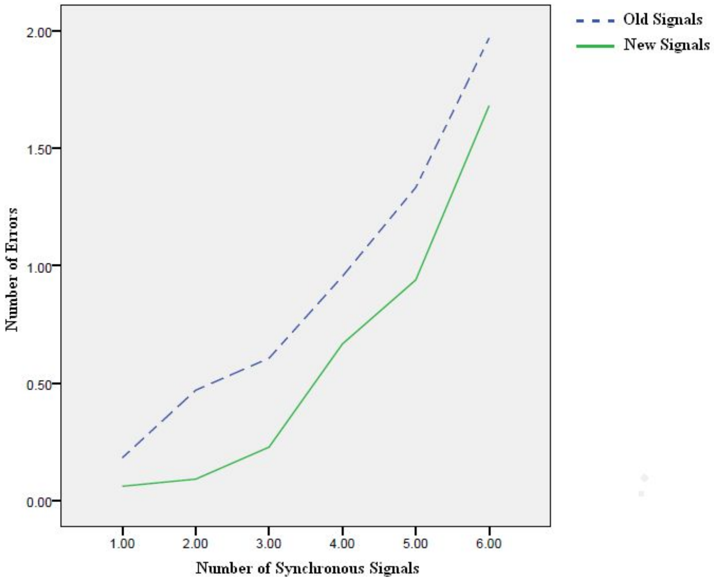
New Warning Signals

These warnings were designed almost exclusively to maximize the detectability of numerous auditory streams. The six signals were designed based on the twenty principles derived from musical voice leading rules. (See Table 1).

Results

The results of this pilot study, as seen in Figure 1, show the correlations of the "Old Signals" and "New Signals". These results display an expected trend of errors increasing as the number of synchronous signals increases. The graph also provides evidence that is consistent with the hypothesis that warning signals derived from principles of music scene analysis are better able to remain distinguished from other signals playing in synchrony.

Figure 1



Note: (n = 11)
Old Signals: Person Correlation 0.680. Correlation is significant at the 0.01 level (2-tailed).
New Signals: Person Correlation 0.663. Correlation is significant at the 0.01 level (2-tailed).

Discussion and Conclusion

This pilot study is merely the beginning of the possibilities that exist in applying musical principles to auditory warning signals. While being able to distinguish between four, five, or even six synchronous warnings is an impressive academic feat, it is not very applicable compared to coupling auditory warnings with visual and vibro-tactile warnings. The purpose in demonstrating the increased streaming capability of these new warnings is to introduce a new method for designing warning signals that are less likely to be confused with others.

Acknowledgments

This study was made possible through the help of Spectra Jet, Inc., who allowed me to obtain recordings of auditory warnings currently used in a Lear 35 and Lear 31a. (This study does not necessarily reflect the opinion of Spectra Jet, Inc.). I would also like to thank my colleague, Brandon Paul, as well as other members of the Cognitive and Systematic Musicology Laboratory at The Ohio State University School of Music for providing essential critiques and further ideas.

References

Huron**,** David. (2001). Tone and Voice: A Derivation of the Rules of Voice-leading from Perceptual Principles. Music Perception, Vol. 19, No. 1 (2001) pp. 1-64.

Peryer , G., Noyes , J., Pleydell-Pearce , K., & Lieven , N. (2005). Auditory Alert Characteristics: A Survey of Pilot Views. International Journal of Aviation Psychology 15(3), 233-250.

Selcon, S.J., Taylor, R.M., & McKenna, F.P. (1995). Integrating Multiple Information Sources: Using redundancy in the design of warnings. Ergonomics, 38(11), 2362-2370.

Stanton, N.A. (1994) Human Factors of Alarm Design. Taylor and Francis Publishers: London.

Stanton, N.A. & Edworthy, J. (1999) Human Factors in Auditory Warnings. Ashgate Publishers: Aldershot.

PERIPHERALLY-LOCATED VIRTUAL INSTRUMENT LANDING DISPLAYS

Nathan Bulkley
Zachary Spielman
Brian P. Dyre
University of Idaho
Moscow, Idaho

We examined how the location and spatial extent of a peripherally-located virtual instrument landing system (ILS) head-up display (HUD) affects landing precision. Our experiment compared three spatial formats of a peripherally-located virtual ILS HUD: a) a *large-format display* located within rectangular regions defined relative to the center of the HUD, with the lateral flight indicator subtending +/-5 to 62.5° by +/-0 to 16.875° (HxV) and the vertical flight command indicator subtending +/-0 to 45° by +/-6.875 to 16.875° (HxV); b) *near-peripheral displays* comprised of roughly the inner half of the large format display; and c) *far-peripheral displays*, comprised of the remaining outer half. We found that restricting display locations and extents to either the near or far periphery provided landing precision statistically equivalent to the large-format displays, which suggests that HUD clutter could be reduced by moving virtual ILS displays into the far periphery without negatively impacting landing precision.

Non-alphanumeric displays incorporating more naturalistic and less symbolic representations of actual flight parameters—known as *virtual displays*—can enhance aircraft pilots' spatial orientation, increase overall performance, and reduce workload (Hettinger, Brickman, Roe, Nelson, & Haas, 1996). Bulkley, Dyre, Lew, & Caufield (2009) showed experimentally that an instrument landing system (ILS) using a peripherally-located head-up display (HUD) with virtual symbology composed of moving arrows has the potential to provide superior landing precision as compared to the military standard ILS symbology (MIL-STD-1787B) presented in its normal central location within the HUD. One unresolved question is whether the peripheral location and spatial extent of the ILS HUD developed by Bulkley et al. (2009) modulates participants' ability to use the display to minimize vertical and lateral errors during simulated instrument landing approaches. This question is particularly important for addressing issues of display clutter (Kaber, Alexander, Stelzer, Kim, Kaufmann, & Hsiang, 2008). Moving the virtual ILS symbology to smaller spatial extents within the far peripheral regions of a HUD makes it less likely to visually obscure other important flight information within the HUD or the visibility of environmental obstacles viewed directly through the windscreen or canopy. Our purpose here was to examine the effects of the location and spatial extent of the virtual ILS on the landing performance.

Even under the best circumstances, pilot fatigue and workload are highest during landing (Hart & Hauser, 1987). Final approach and landing phases of flight account for a disproportionately high number of accidents, particularly at night, despite instruments such as the ILS that are specifically designed to provide accurate information regarding deviations from the optimal approach path (Boeing, 2002; Khatwa, Collins, & Helreich, 1998; Ashford, 1998). To land safely, pilots must infer their position and movement by integrating the optical flow and horizon seen through the windscreen with symbolic representations of altitude, pitch, heading, and airspeed read from instruments, while simultaneously avoiding other aircraft and ground obstacles and verbally communicating with air traffic control. These activities place a high demand on focal visual and central processing resources (Wickens, 1991), which may be limited by fatigue. Indeed, Roscoe (1980) estimated that 90% of aircraft control is performed using the central visual field: to accurately read most aviation displays, pilots must directly fixate their gaze upon the display and allocate enough mental resources to interpret the alpha-numeric information presented.

However, a virtual display could present landing information to pilots without taxing overburdened central visual field and attentional resources by taking advantage of automatized orienting and motion coding processes that are particularly robust in peripheral vision. In the context of simulated flight, Cox (2000) showed that a virtual speed error indicator composed of moving fields of arrows projected to the visual periphery provided better flight path and airspeed control than a head-up display (HUD) speed indicator defined by MIL-STD-1787B, while simultaneously lowering subjective workload. In a follow-up study that measured eye and head movements, Schaudt, Caufield, and Dyre (2002) further showed that participants spent the greatest amount of time looking directly at the traditional MIL-STD HUD airspeed indicator during the simulation and that they rarely, if ever,

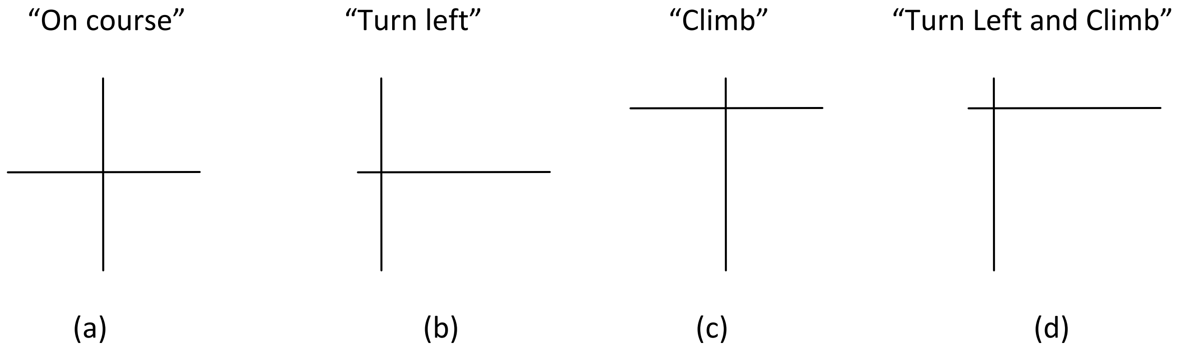


Figure 1. Format of MIL-STD 1387B HUD ILS command bars. The four panels show display configurations representing: (a) zero lateral and vertical error, on course, no control correction needed, (b) rightward lateral error, the aircraft needs to turn left to reestablish the optimal approach path, (c) downward vertical error, the aircraft needs to climb to reestablish the optimal approach path, and (d) both downward and rightward errors, the aircraft needs to turn left and climb to reestablish the optimal approach path.

looked directly at the peripheral virtual speed error indicator. These studies clearly demonstrated that the peripheral visual field can process the optical flow created by moving display elements and extract meaningful speed information with less attentional demand and central visual field resources than traditional symbolic displays.

More recently, Bulkley et al. (2009) found that a similar peripherally-located virtual display could be adapted to replace the flight command bars of a traditional ILS display, which are typically presented in the center of the HUD (as defined by MIL-STD 1787B). The ILS command bars instruct pilots to the direction the aircraft needs to be moved to attain proper lateral and vertical alignment with the optimal approach path during landing. The display contains a horizontally-oriented bar that moves up and down to indicate vertical angular errors from the ideal glide slope, and a vertically-oriented bar that moves left-to-right to indicate lateral angular errors from the ideal approach path (see Figure 1). When the bars form a symmetrical cross "+" the aircraft is optimally aligned with the ideal approach path. Movements of the bars from the "+" configuration indicate that control inputs are needed to regain the optimal path, and serve as command displays for how the pilot must maneuver the aircraft. In essence, pilots "chase the bars." When the vertical bar is left of the center, the pilot banks left to correct a rightward lateral error. When the horizontal bar is above the center, the pilot climbs to correct a downward glide slope error, and so on. Bulkey et al. (2009) adapted the peripherally-located virtual HUD format developed for representing airspeed by Cox (2000; see also Schaudt et al., 2002) to an ILS display by simply replacing the centrally-located flight command bars with fields of moving arrows located in the visual periphery (See Figure 2). When the arrows appear in the left visual field, the pilot responds by banking left to correct a rightward lateral error. When the arrows appear in the upper visual field, the pilot climbs to correct a downward glide slope error, and so on. The magnitude of lateral and vertical errors is redundantly coded by the speed of the moving arrows and their size, which both increase as errors increase up to a maximum limit. Importantly, size changes occur in discrete steps, creating transient events that can alert a pilot to increasing errors *pre-attentively* (Egeth & Yantis, 1997), even when the displays are presented in the visual periphery. Also, to help reduce display clutter, the arrows disappear completely when errors fall below a minimum threshold.

Indeed, the primary design goals of the virtual ILS were to reduce central visual field load, attentional demand, and display clutter within the HUD while still providing enhanced precision in maintaining glide path. Bulkley et al. (2009) demonstrated that a large-format virtual ILS can afford greater glide path precision than a traditional display, and other experiments underway in our laboratory are assessing whether central visual field load and attentional demand are reduced by such displays in a manner similar to that found for the virtual speed display examined by Cox (2000) and Schaudt et al. (2002). The focus of the research presented here was to evaluate different locations and spatial configurations of the virtual ILS display to examine whether display clutter can be reduced without negatively impacting landing performance. We report one experiment that examines three potential display configurations (see Figure 2): a) a *large-format display* identical to that tested by Bulkley et al. (2009), b) *near-peripheral displays* comprised of roughly the inner half of the large format display, and c) *far-peripheral displays*, comprised of the remaining outer half. Table 1 lists the specific spatial extents for each display configuration. These displays qualitatively varied central visual field load and display clutter as follows. Central visual field load and clutter was highest for the near-peripheral display because all display elements were presented within the

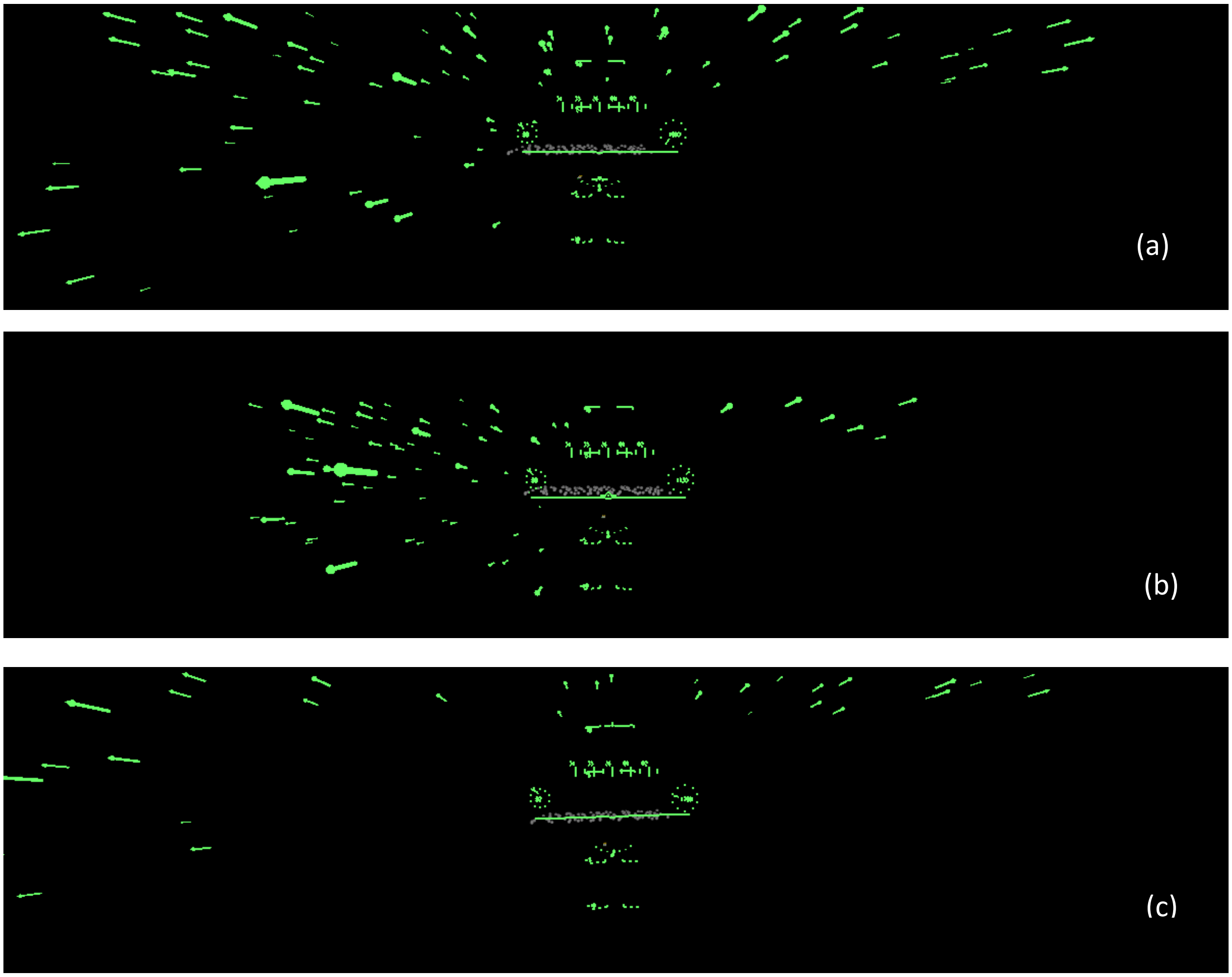


Figure 2. Screen shots of three spatial configurations of the virtual ILS HUD: (a) the large-format display equivalent to that developed by Bulkley et al. (2009), (b) the near-peripheral display, and (c) the far-peripheral display. All three panels represent both downward and rightward errors, the aircraft needs to turn left and climb to reestablish the optimal approach path.

relatively small areas surrounding the traditional HUD symbology. The large-format display represented intermediate central visual field load and clutter because the display elements were spread out into the periphery. The far-peripheral display represented the least central visual field load and display clutter because the virtual symbology was limited to areas of the visual field furthest from the traditional HUD symbology.

Method

Participants

The experiment tested 12 volunteer participants from the undergraduate, graduate and faculty population of the University of Idaho. Undergraduate participants were compensated with course credit. Participants reported normal or corrected-to-normal acuity and no previous aviation piloting experience. Bulkley et al. (2009) showed that non-pilots using simplified flight controls fly similar flight paths to those flown by pilots for simulated visual approaches under "blackhole" conditions (Gibb, Schvaneveldt, & Gray, 2008), which suggests that pilots and non-pilots are able to perceive their dynamic spatial orientation relative to the runway in a similar manner. Hence, we believe that our sample of non-pilots validly represents the same visual processes that pilots use to control their aircraft during landing.

Stimuli and Apparatus

Visual displays simulating flight were created by a set of four personal computers running ViEWER v2.23 (Dyre, Grimes, & Lew, 2009). One computer served as the simulation host, which coordinated the activity of three graphics computers over a local area network. Each graphics computer rendered to one of three display projectors, which front-projected images with a spatial resolution of 1024 x 768 pixels (H x V) at a refresh rate of 60 Hz onto three large screens arranged as three sides of an octagon with the design viewpoint at the center of the partial-octagon, 1.8 m from the center of each screen. Together, the three screens subtended 135 x 33.75 degrees of visual angle (H x V).

The simulated environment consisted of a large island extending 288,000 m in width (x) and 72,000 m in depth (z). A 30m wide runway started at the center of this island [x, y, z] = [0, 0, 0] and extended 950 m in depth [0, 0, -950]. Green, white, and red lights were spaced every 50m along the sides of the runway in a standard runway lighting configuration. The runway was texture-mapped with a pavement texture and the surrounding ground surface was texture-mapped with a high spatial frequency seamless texture resembling dirt and grass. At the far end of the runway the ground began to slant upward at 5 degrees for 18,000m before leveling off at a maximum elevation of 1600 m. To simulate the lights of a small city lying beyond the runway, white dots were randomly dispersed on this slanted area covering an area of 4000 x 8000 meters (width x depth) starting at 650m past the end of the runway and centered on the runway. Simulated ambient lighting was equivalent to an overcast, moonless night such that only the runway marker lights and city lights were clearly visible. Mild haze modeled as exponential fog was added to the displays to provide a sense of distance through aerial perspective. This haze did not affect the visibility of the runway or city lights.

The simulated view from the vehicle was open, no windscreen frame or other artifacts were used to create a simulated cockpit. The vehicle started 9889m short of the runway at an altitude of 720m, aligned laterally with the center of the runway and moving forward on a level trajectory at 45 ms^{-1}. The target glide-slope angle of 7.3 degrees was defined via two invisible waypoints. The first was located at the vehicle starting point [0, 720, 9889], and the second was centered on the runway, 2m from its near end [0, 0, -2].

Simulated lateral and vertical wind disturbances were produced by translating the vehicle along the lateral (x) and vertical (y) axes independently. Each disturbance was defined by a sum-of-sines with five prime frequencies. Amplitudes were chosen such that the maximum acceleration did not exceed 1 G (9.8 ms^{-2}). The disturbance frequencies for the x- and y-axes were 0.055, 0.085, 0.145, 0.185, 0.215, and 0.035, 0.065, 0.115, 0.155, 0.205 Hz, respectively; amplitudes were 10.47, 2.34, 1.37, 1.08, 0.93 and 25.85, 7.49, 2.39, 1.32, 0.75 m, respectively.

Since participants were non-pilots, simplified flight controls were used to control the lateral and vertical positions of the simulated aircraft. Left-right movements of a CF-F-16 joystick with the right hand controlled lateral velocity with a transfer function defined by first-order with gain = 25 ms^{-1} at maximum stick deflection and exponential lag constant = 1.0 s^{-1}. For simulation of banking while moving laterally, left-right stick movements also produced a zero-order roll with a gain = 15° at maximum deflection and exponential lag constant = 0.5 s^{-1}. Movements of a CF Pro Throttle configured as a first-order controller with gain = 25 ms^{-1} and exponential lag constant = 1.0 s^{-1} with the left hand controlled vertical velocity. Forward movement of the throttle caused the vehicle to go up (simulating more thrust), backward movements caused the vehicle to go down (less thrust).

Similar to Bulkley et al (2009), monochromatic green HUDs were superimposed over the terrain and environmental objects on the display and. included the following indicators from MIL-STD-1787B: aircraft pitch reference symbol, climb/dive marker, climb/dive ladder, airspeed indicator, target airspeed indicator, altitude indicator, heading indicator, and bank indicator. The ILS was implemented within the HUD as a peripherally-located virtual display of fields of moving arrows, which replaced the MIL-STD flight command bars. Fields of moving arrows randomly arranged within a volume of space appeared in the upper, lower, left or right peripheral areas of the HUD to provide control commands to overcome lateral and vertical deviations from the optimal flight path (see Figure 2). Similar to the MIL-STD ILS, the virtual ILS used a command format that informed pilots which direction to move to correct their course—in effect participants needed to "chase the arrows" to maneuver toward the optimal glide path. The arrows coded lateral and vertical deviation error magnitudes redundantly using both speed of movement and size. Zero course error resulted in zero speed and size--nothing was displayed. Small course errors resulted in small arrows moving slowly. As course errors increased, the speed of the arrows would increase proportionately, and the size of the arrows would increase in step-wise increments to produce sudden size-changes that naturally captured attention. The location and spatial extent of the fields of moving arrows varied across three conditions as listed in Table 1.

Table 1
Spatial Extents of Virtual ILS HUDs

	Horizontal Extent	Vertical Extent
Large-format		
Lateral Command	+/- 5.0 to 67.5	+/- 0.000 to 16.875
Vertical Command	+/- 0.0 to 45.0	+/- 6.875 to 16.875
Near-Peripheral		
Lateral Command	+/- 5.0 to 36.25	+/- 0.000 to 11.875
Vertical Command	+/- 0.0 to 36.25	+/- 6.875 to 11.875
Far-Peripheral		
Lateral Command	+/- 36.25 to 67.5	+/- 0.000 to 16.875
Vertical Command	+/- 0.0 to 45.0	+/- 11.875 to 16.875

Note. All units in degrees of visual angle from the HUD center

Table 2
Altitude Errors by Condition

Display	Mean Error (m)
Large-format	-1.62 (0.85)
Near-Peripheral	-0.39 (0.85)
Far-Peripheral	-0.65 (0.85)

Display	SD Error (m)
Large-format	13.40 (0.76)
Near-Peripheral	13.00 (0.76)
Far-Peripheral	12.29 (0.76)

Display RM	S Error (m)
Large-format	15.06 (0.84)
Near-Peripheral	14.57 (0.84)
Far-Peripheral	14.03 (0.84)

Note. Numbers in parentheses represent within-subject standard errors of the mean for the comparison of display configurations.

Experimental Design and Procedure

A 3 x5 within subject factorial design tested the effects of HUD format (large-format, near-peripheral, and far-peripheral) and block (1-5). The order of HUD formats was randomized within each block. Each non-crashing trial lasted approximately 3 minutes and 40 seconds and participants were able to complete all 15 trials, plus instructions and debriefing within a single 90 minute session.

The experimental session proceeded as follows. After obtaining informed consent, participants were instructed to land the plane along a linear glide slope connecting their starting position to the end of the runway. They received 2 training trials to learn how to land the simulated aircraft. First, to familiarize participants with the display of the ground and runway and to show them how the controls moved the vehicle laterally and vertically the experimenter demonstrated a visual approach, with environmental lighting turned on to reveal the terrain and runway and ILS indicators turned off. Participants then completed one practice visual approach using the same visual conditions, during which the experimenter provided additional instruction only if extreme deviations from the optimal approach path where observed. Following these training trials, participants were instructed that the experimental trials would simulate night landing using virtual indicators that would appear to help direct them to the optimal approach path and that these indicators would be made up of moving arrows in the display. In addition, they were informed that crashes were possible and to not be alarmed if they were to crash. If participants impacted the ground at any time the trial ended. After the experimental trials, participants were debriefed and informed of the purpose of the experiment.

Results

Altitude errors were defined as the difference between actual altitude and the target altitude defined by the linear glide slope at a particular point along the approach. Lateral errors were defined as the difference in position of the simulated aircraft relative to the runway centerline. Mean errors in altitude and lateral position computed over the entire trial duration represented constant error, or accuracy of control. Standard Deviations (SD) of altitude and lateral position errors computed over the entire trial duration represented variable error, or precision of control. Root-mean-squared (RMS) errors represented overall error.

All types of lateral and altitude errors were analyzed using 3x5 within-subjects analysis of variance (ANOVA) with display configuration (large-format, near-periphery, and far-periphery) and block (1-5) as the two variables. For altitude errors, no reliable differences were found between display configurations or blocks for mean, SD, or RMS errors ($p > .05$, see Table 2). These results show that all three display configurations afforded statistically equivalent control over glide slope—there was no reliable performance cost to moving the virtual display into the far periphery, indeed the non-significant trend was toward better performance with the more peripherally-located display. A similar lack of effect for display configuration was found for lateral errors. Overall, error magnitudes were quite similar to those found by Bulkley et al. (2009) for the virtual ILS display.

Discussion

While considerable caution is needed whenever interpreting a null result, we believe that the null effect found in the present study when considered together with the results of Bulkley et al. (2009) can be taken as evidence that the three virtual display configurations afforded equivalent perception and control of glidepath. The experimental measures used in this study were identical in all respects to those used by Bulkley and his colleagues who found a statistically reliable advantage for the same large-format virtual ILS examined here as compared to the MIL-STD 1387B ILS in its traditional central field location. This demonstrates that the error measures are reliable enough to detect differences in performance when they exist and that it is unlikely that our null result was due to unreliable measures of error. This point is further underscored by the fact that the present study tested a larger sample of participants than Bulkley et al. (2009) and thus had potentially greater statistical power for detecting reliable differences between the display conditions.

These results have important implications for the design of HUDs. Clearly, the peripheral visual field—even the far periphery—is a potentially important visual resource that is underused with current HUDs and may be a particularly valuable resource for processing flight parameters related to spatial orientation, such as ILS control of landing approaches. Peripheral virtual displays of ILS commands appear to provide pilots with information in a more natural, pre-attentive manner that has the potential to lessen attentional demand, central visual field load, and display clutter, while still affording performance that is equal to or better than traditional ILS displays. Further research is needed to assess these claims, particularly with pilots as participants to see if this potential can be realized.

References

Ashford, R. (1998). *Approach and Landing Fatal Accidents 1980–1996*, Safety Analysis Report No. SA-98-002, Civil Aviation Authority, London.

Boeing Commercial Airplane Group. (2002). Statistical summary of commercial jet airplane accidents, worldwide operations, 1959-2002. Retrieved December 8,2002, from http://www.boeing.com/news/techissues/pdf/statsum.pdf

Bulkley, N., Dyre, B. P., Lew, R., & Caufield, K. (2009). A Peripherally-located Virtual Instrument Landing display affords more precise control of approach path during simulated landings than traditional instrument landing displays. *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society,* 31-35. San Antonio, TX: HFES

Cox, E. M. (2000). The effect of a virtual indicator on speed control. (Unpublished Master's Thesis, University of Idaho).

Dyre, B. P., Grimes, J. G., & Lew, R. (2009). ViEWER 2.23: A Virtual Environment Workbench for Education and Research. *http://www.webpages.uidaho.edu/~bdyre/viewer.htm.*

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48,* 269-297.

Gibb, R., Schvaneveldt, R., & Gray, R. (2008). Visual misperception in aviation: glide path performance in a black hole environment. *Human Factors, 50*(4), 699-711.

Hart, S. G., & Hauser, J. R. (1987). Inflight application of three pilot workload measurement techniques. *Aviation, Space, & Environmental Medicine, 58*(5),402-410.

Hettinger, L. J., Brickman, B. J., Roe, M. M., Nelson, W. T., & Haas, M. W. (1996). Effects of virtually-augmented fighter cockpit displays on pilot performance, workload, and situation awareness. *Proceedings of the Human Factors Society 40th Meeting*, 30-33. Philadelphia, PA: HFES.

Kaber, D. B., Alexander, A. L., Stelzer, E. M., Kim, S., Kaufmann, K., & Hsiang, S. (2008). Defining and psychological modeling of perceived clutter in advanced cockpit displays. *Aviation Space and Environmental Medicine, 79,* 1007-1018.

Khatwa R., Collins R., & Helmreich, R.L. (1999). Data acquisition and analysis working group final report. *Flight Safety Digest,*1-77.

Roscoe, S. (1980). *Aviation Psychology.* Ames, IA: Iowa St. University Press.

Schaudt, W.A., Caufield, K.J., and Dyre, B.P. (2002). Effects of a virtual speed indicator on guidance accuracy and eye movement control during simulated flight. *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society,* 318-322. Baltimore, MD: HFES

Wickens, C. D. (1991). Processing resources and attention. In D. Damos (Ed.), *Multiple task performance.* London: Taylor & Francis.

VIBROTACTILE 'ON THIGH' ALERTING SYSTEM IN THE COCKPIT?

Shira Yosef , Yael Salzer & Tal Oron-Gilad
Human Factors Laboratory
Dept. of Industrial Engineering and Management
Ben-Gurion University of the Negev, Israel

The objective of this research was to examine the utility of a novel placement for a vibrotactile display in the cockpit. This objective was pursued in stages through a number of research phases, here we report on the final phase concerning the benefit of the vibrotactile display in a visually loaded environment. Results support placing a directional alerting vibrotactile display on the thigh of a seated operator.

Since the visual and to a lesser extent, the auditory modalities have been exhausted in the cockpit, the tactile modality has become a relevant candidate to counteract information overload. It is assumed that the tactile modality does not entirely compete with the visual and auditory modalities and generally requires little to no cognitive effort to analyze spatial directionality (Brill et al., 2004; Eriksson et al., 2006; Jennings et al., 2004). As such, tactile displays may introduce solutions to impending limitations in visual perception and processing in the cockpit (Eriksson et al. 2006; van Erp et al. 2002; 2006). Specifically, Salzer, Oron-Gilad and Ronen (2010) proposed the thigh as a potential platform for orienting in the vertical plane (see Figure 1), demonstrating the ability to localize locations on the vertical plane stimulated by vibrotactors mounted on the thigh of a seated operator.

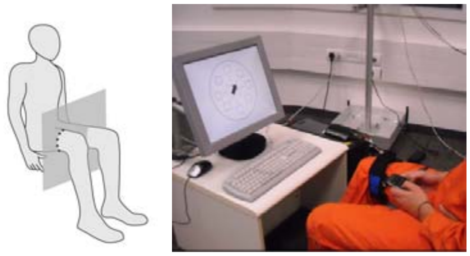


Figure 1. Orientation in vertical plane created by vibrotactors mounted on the thigh (illustration and in-practice).

In today’s cockpit's collision-alert system the visual and auditory cues orient the pilot which way to take to avoid collision (i.e., move away from the hazard). Thus, when considering how to implement tactile cue two contradicting display approaches emerge; to have the tactile cue compatible with the direction of the visual and auditory alerting cues, or to simulate the source of hazard (i.e., the direction of the object to avoid). Salzer, Oron-Gilad, Ronen, and Parmet, (in press) evaluated their on-thigh vibrotactile display when added to existing audio and visual alerting cues utilizing two tactile display modes: a) compatible; where the location of the tactile cue was compatible with the visual direction of response, (i.e., aiming away from the source of hazard). b) inverse; the location of the tactile cue was directed toward the direction to be avoided (i.e. generating a tendency to move away from the source of hazard). Their results revealed that directing the way to escape from hazardous situations, compatible with the visual alerting cues, was preferred. Nevertheless, advantages for adding tactile alerting cues over the visual alerting cues alone were not found, probably because these alerts were examined in a quiet environment with no additional cues or noise.

The current experiment aimed to examine the utility of the vibrotactile cues in a loaded environment. It introduced a more demanding environment where both visual load and working memory load were added over collision alerts alone. Overall, it was expected that the general benefit of the compatible mode versus the inverse mode will remain. Yet, it was expected that once the visual modality was more loaded, advantages for the tactile alert display will emerge.

Method

Participants

Five female and 5 male undergraduate students (age 24 to 29, mean 25 years) participated for course credit.

Apparatus

Experimental system. Tactile stimuli were displayed via a prototype developed by IAI-LAHAV (patent 11/968,405 pending), consisting of a tactor controller Eval2.0 (by Engineering Acoustics Inc. (EAI)) regulating eight EAI-C2 vibrotactors. The vibrotactors were stitched to an elastic fiber strip, 6 cm apart from each other, worn on the right thigh over a pilot suit. A designated program in E-prime2.0 running on a PC computer with Windows XP activated the experimental procedure and displayed the visual alerts on a 19" LCD screen. An additional PC computer was used for playing the simulated flight path on a dome projection screen allowing a field of view of 60 degrees. The 19" screen was placed on a table in front of the sitting participant, behind it was the dome projection screen. Response was collected by a standard keyboard.

Visual directional alert stimulus A compass rose was displayed on the 19" screen. The visual cue was a black arrow in the center of the compass rose pointing towards one of the eight directions (up, down, left, right and four diagonals) as shown in Figure 1.

Tactile directional alert stimulus Each of the vibrotactors represented one of the eight directions. The vibrotactile stimulus was a continuous 800ms pulse at 250Hz.

Flight movies and flying objects. Five movies of a flight path accompanied by the sound of a helicopter were allotted randomly to each block. Occasionally, throughout the flight simulation, red and yellow objects appeared flying toward the viewer (i.e. the flying aircraft) (see Figure 2). The participant was required to count the number of red objects observed. At unexpected intervals within the block, a question interrupted the visual cue on the 19" screen, asking the participant to enter the number of red objects counted up to that moment. Once a response was provided, the participant was instructed to restart counting.

SWAT (Subjective Workload Assessment Technique). The SWAT (Reid and Nygren, 1988) questionnaire was submitted upon completion of each experimental block.

Design and Procedure

The design consisted of two within-participant variables; visual stimulus (2; present or not), and tactile display mode (2; compatible or inverse). Each condition was displayed in a separate experimental block. An additional fifth block consisted of visual stimuli only (i.e. without vibrotactile stimuli). The order of the five blocks was counterbalanced among participants. Participants arrived at the lab, one at a time, changed to standard pilot suit, filed an informed consent and were debriefed. The participant was asked to perform two simultaneous tasks; the first task was to count and remember until asked the number of red objects in the video. The requirement to report the number of counted objects appeared three times throughout an experimental block, at random occasions. The second task was to respond accordingly to the spatial stimuli. The numeric keypad was used for collecting participants' directional response; each key 1-9 (except 5) corresponded to a directional visual and/or tactile stimuli according to their relative location. In the compatible tactile display mode, the direction of stimulus was corresponding with the location of the response key (e.g. left most tactor corresponds with key no. 4). In the inverse tactile display mode, the direction of stimulus was opposite to the location of the response key (e.g. left most vibrotactor corresponds with key no. 6). The direction of the visual stimulus always corresponded with direction of response. The participant was familiarized with the tasks by completing a practice blocks of 24 trials prior to each experimental block, of 32 trials. Ten practice trials preceded the 32 experimental trials for the visual-only display block. The participant was instructed to respond as fast and accurate as possible. Feedback was provided only in the practice trials. Upon completion of each block the participant completed a SWAT questionnaire.

Figure 2. The projected environment of the flight path. The red object is marked for emphasis.

Results

Performance on the Spatial Alerts

Response Time. A two way repeated measures ANOVA with visual stimulus (2; present or not), and tactile display mode (2; compatible or inverse) was analyzed over data from the compatible and inverse display blocks for response times (RT) was conducted. Mean estimates are specified in Table 1. Extreme values of RT, shorter than 150ms and longer than 2500ms, were excluded from analysis. A significant main effect for visual (F(1, 9)=17.231, p=.00248) was found. At the presence of the visual stimulus, RT was lower (visual present M=978 ms SD=322, visual not present M=1367 ms, SD=158). There was no significant difference between the compatible and the inverse (F<1). To evaluate the affect of the addition of vibrotactile stimulus, a repeated measure ANOVA compared the visual display block (V) with the tactile compatible display with visual condition (CTV), which had the shorter RT of the two tactile and visual combined display conditions. There was no significant difference between V and CTV.

Accuracy. A logistic regression model within the framework of generalized linear mixed model (GLMM) over visual stimulus (2; present or not), tactile display mode (2; compatible or inverse) for correct recognition was conducted. The full model included the two-way interaction and main effects. The main effects tactile display mode and visual stimuli presence were significant (Wald Chi-Square$_1$= 5.1, p<.024, and Wald Chi-Square= 53.7, p<.001, respectively). The interaction was not significant. Correct recognition rate (CRR) for compatible tactile display mode was higher than inverse tactile display mode (CRR=.93, SE=.02, CRR=.87, SE=.03) and the presence of visual stimulus improved CRR (CRR=.97, SE=.01, CRR=.71, SE=.05). Mean estimates are specified in Table 1. A Logistic regression on tactile display mode (3; compatible, inverse, none) for success rate in the unchanged presence of the accompanying visual signals did not reach significance, indicating that recognition was equally good with and without tactile stimulus when the visual stimulus was present.

Table 1

Response Time (RT) and Accuracy (ACC) by Modality and Tactile configuration.

		Tactile Blocks				**Visual Block**
N=10		**CT**	**CTV**	**IT**	**ITV**	**V**
RT(ms)	Mean	1367	872	1368	1085	864
	SD	158	61	122	134	197
ACC	CRR	.98	.76	.96	.66	.98
	SE	.008	.027	.020	.080	.009

*CT=compatible tactile, CTV=compatible tactile+visual, IT=inverse tactile, ITV=Inverse tactile +visual, CRR=correct recognition rate

Visual Loading Task

Participants were asked to count the number of red items that appeared in the environment and to specify this number when asked within the experimental interface. At the end, the total number of items registered by each participant was calculated. Zero was given if a participant failed to reach the appropriate number of items and 1 if the participant specified the correct number of items. The number of participants (out of 10) who correctly identified the total number of items in each experimental block was 8, 7, 10, 7 and 5 respectively for the CT, CTV, IT, ITV and V blocks. Thus, performance was worse in the visual-only condition where half (5 out of 10) of the participants failed to identify the correct number of objects. Performance was perfect in the tactile-only condition (IT) when no visual alerts were present, implying perhaps on the toll of visual load to task performance when visual alerts were present.

Workload (SWAT)

A GEE regression analysis was conducted with participant as the random effect and experimental block (5) as the main effect. The main effect for experimental block was only significant for the SWAT time dimension (Wald Chi-square (4) = 15.17, p<.004). CTV and ITV conditions were perceived as more temporally demanding (mean ratings were 1.5, 1.7, 1.4, 1.7, 1.5, respectively for the CT, CTV, IT, ITV and V blocks).

Discussion and Conclusions

The current experiment revealed benefits for tactile signaling in a loaded visual environment. With regard to the directional alerts, response time and CRR were both facilitated by the presence of visual cues. Comparing to the low-demanding environment where only the directional alerts existed (Salzer et al., in press), the loaded environment generated longer response times, as expected. Nevertheless, accuracy ranges remained unchanged.

The advantages of adding tactile cues became apparent in the loading task where the worse performance was found in the visual only condition, indicating that participants had difficulties in attending to two visually displayed tasks simultaneously. The CTV condition generated the best performance on the alerts, but was not better than the CT and IT conditions in the loading task, indicating perhaps that the presence of visual alerting cues may have disrupted performance on the loading task, or vice versa, that the presence of tactile cues enabled higher performance in the loading task. The benefits of the tactile stimulus with regard to the two tasks were most notable when compared with the visual only mode (V). The SWAT temporal demand scores confirmed that while the presence of visual and tactile cues combined was beneficial in the loaded environment in terms of overall mission performance, the combined CTV and ITV conditions generated higher perceived temporal workload. To conclude, the presence of tactile stimulus helped to maintain balance between the two tasks. Namely, the tactile stimulus contributed to improve situation awareness in the visually loaded environment.

Acknowledgement

This work was supported by the IAI Israeli Aircraft Industry, Uri Paz-Meidan, Technical Monitor (upaz-meidan@iai.co.il). The views expressed in this work are those of the authors and do not necessarily reflect official IAI policy.

References

Brill, C., & Terrence, P. (2005). Discrete tactile versus auditory target cueing. In Gilson, R. D. ;Redden, E. S. ;Elliot, L. R., *Remote Tactile Displays for Future Soldiers* (ARL-SR-0152), Aberdeen Proving Ground, MD.: Army Research Laboratory.

Eriksson, L., van Erp, J., Carlander, O., Levin, B., van Veen, H., & Veltman, H. (2006). Vibrotactile and Visual Threat Cueing with High G Threat Intercept in Dynamic Flight Simulator. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting 2006.*

Jennings, S., Craig, G., Cheung, B., Rupert, A., & Schultz, K. (2004). Flight-Test of a Tactile Situational Awareness System in a Land-based Deck Landing Task. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting.*

Reid, G.B., & Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 185–218). Amsterdam:Elsevier.

Salzer, Y., Oron-Gilad, T.& , Ronen, A., and Parmet, Y. (in press). VibroTactile "On-Thigh" Alerting System in The Cockpit, Human Factors.

Salzer Y., Oron-Gilad, T. & Ronen, A. (2010). Vibrotactor-belt on the Thigh – Directions in the Vertical Plane, A.M.L. Kappers et al. (Eds.): EuroHaptics 2010, Part II, LNCS 6192, pp. 359–364, 2010. Springer-Verlag Berlin Heidelberg.

van Veen, H. A., & van Erp, J. B. (2001). Tactile Information Presentation in Cockpit. In S. Brewster, R. Murray-Smith, *Haptic HCI 2000, LNCS 2058,* 174-181. Berlin Heidelberg: Springer-Verlag.

van Erp, J. B., Veltman, J. A., van Veen, H. A., & Oving, A. B. (2002). Tactile Torso Display as Countermeasure to Reduce Night Vision Goggles Induced Drift. *RTO HFM Symposium on "Spatial Disorientation in Military Vehicles: Causes, Consequences and Cures", RTO-MP-086,* Spain:

van Erp, J. B., Groen, E. L., Bos, J. E., & van Veen, H. A. (2006). A Tactile Cockpit Instrument Supports the Control of Self-Motion during Spatial Disorientation. *Human Factors, 42*(2), 219-228.

EFFECTS OF COMMUTING ON CREWMEMBER FATIGUE: A COMPREHENSIVE STUDY IN SUPPORT OF RISK MANAGEMENT

Lori J. Brown
Western Michigan University, College of Aviation
Kalamazoo, USA
Geoff Whitehurst
Western Michigan University, College of Aviation
Kalamazoo, USA

The need to reduce accidents and incidents caused by human fatigue in the aviation industry remains on the National Transportation Safety Boards' (NTSB), most wanted list. At many airlines, crewmembers are forced to work to the point of exhaustion because of: poorly scheduled duty time; lengthened duty days; minimum scheduled rest requirements; working the backside of the clock, multiple short-haul legs; and long commutes to work. Although, commuting, in the context of aviation, has yet to be defined, the U.S. Census Bureau defines an 'extreme commute' as a travel 90 minutes or more, each way to work. Americans who endure a daily "extreme commute" of 90 minutes or more each way to work, is a rapidly increasing number which is now in excess of 3.4 million (Alexander, 2009). A recent pilot study by Western Michigan University, the NTSB report following the Colgan Air crash (NTSB, 2010; pp. 47-48) and information from Airtran Airways in a workers' compensation case (WC, 07-00328, 2008) suggest that this could include a significant number of commuters from the aviation industry.

In a NTSB safety study of US major airline accidents involving flight crews from 1978 to 1990, one finding directly addressed the concern about how time since awake may contribute to fatigue. The study stated; "Half the captains for whom data were available had been awake for more than 12 hours prior to their accidents. Half the first officers had been awake for more than 11 hours. Crews comprising captains and first officers whose time since awake was above the median for their crew position made more errors overall and significantly more procedural and tactical decision errors" (NTSB, 1994).

Unfortunately, in some instances aircraft accidents with fatigue causation factors can prove to be fatal for all onboard, as in the crash of Colgan Air flight 3407, which crashed in Buffalo, New York, on February 12, 2009. According to the NTSB report, the probable cause of the accident was "the captain's inappropriate response" to a low speed condition (NTSB, 2010, p.155). The report cited several contributing factors; however commuting was not cited as a contributing factor. A pilot, who commutes, has to travel from the city or country in which they reside before checking in for duty at their base domicile.

According to the NTSB report (2010), "the pilots' performance was likely impaired because of fatigue" (NTSB, 2010, p. 108). This has raised a concern about the potential contribution to fatigue from time spent commuting to a domicile, which has been a safety concern- since both of the pilots flying the Colgan Air flight 3407, were 'commuting' pilots.

For most commuters in America, economic necessity can push them to the extremes of commuting such lengthy distances (Howlett, 2005). Typically, Americans commute via automobile- despite the rising fuel costs. People make these lengthy commutes for many reasons. A few may want a rural lifestyle, some are accommodating a spouse who works closer to home, housing prices, quality of schools, or economic reasons (Pisarski, 2006).

Several studies regarding automobile commutes have shown that long-distance commuters (90 minutes to 3 hours) suffer from psychosomatic disorders at a much higher rate. "Commuters who drive have it especially hard--bad weather, traffic jams and accidents all cause stress" (Schaefer, 2005). "The psychosomatic condition of these people was terrible," says Steffen Haefner, who led the study. The proportion who complained of symptoms such as pain, dizziness, exhaustion, and severe sleep deprivation was twice as high as in a control group of non-commuters. According to the study, (Schaefer, 2005), the mental ills of long distance automotive commuters, include sleep disturbances, fatigue and concentration problems.

Although commuting in the aviation context has not been clearly defined, the FAA recently published a notice of proposed rulemaking (NPRM) which incorporates the suggestion that a "local area" be defined as an area within a two-hour travel period- regardless of mode of transportation. Early studies, such as, "The Journey to Work" (Liepmann, 1944), offer an alternative view.

In the work "The Journey to Work", Liepmann (1944) notes: "It would be naive to assume that the magnitude of the commute to work can be measured by geographical distance alone- without any qualifications being made as to the mode of transportation"

. Crewmembers which transport via an automobile have to pay for fuel and depreciation of the car, whereas the crewmembers traveling on a jumpseat have no such expense. Therefore, we can include the measure of monetary cost. In addition, time spent commuting and time away from a family- is a hidden cost. There is no exact measure available for these costs, but they are important. This may cost the employee personal cost due to loss of personal and family time. Complicating the analysis is the fact that there may be hidden benefits as well as, costs in the commute. There is also comfort and convenience. For example, an aircraft may be inconvenient, no matter how fast it travels- as it may leave at the wrong time, or incur a delay.

A sufficiently attractive wage may compensate for a long commute, which the crewmember sacrifices time away from family and personal days off. This wage differential need not necessarily be a monetary one; it may be in terms of pleasant working conditions, retirement benefits, and the numerous factors, which make up 'job satisfaction (Liepman, 1944)." The current industry trend has been plagued with salary, benefit, and retirement reductions, which can make it difficult to recruit and retain the best and the brightest new hires.

As noted by the Airline Pilots Association, International: "The economics of the industry and the transient locations of crew bases and locations at airports in costly major cities necessitate the ability for pilots, particularly those with less seniority, to commute from less expensive communities" (ALPA, 2010). Airline mergers, base closures, and reductions can also necessitate the need to commute and often require an unpaid move for the crewmember to retain employment. Some airlines have commuting policies currently in place to assist the crewmembers when traveling to work at their domicile. A review of these policies is beyond of the scope of this paper; however, this is a crucial element surrounding commuting practices.

Regardless of whether the crewmembers' commute to work is personal choice or company imposed, it can be difficult to obtain 8 hours of restorative sleep in a given 24-hour period, with a 14 hour duty period-with or without a lengthy commute.

Although the table below allows minimal time for family and personal obligations, the 30-minute commute would <u>not</u> be realistic for most crewmembers. Most large airports have employee parking lots with shuttle service to the terminal building. Crewmembers would likely need at least 30 minutes to park their car, take the parking lot shuttle, and walk to the check-in area. This does not leave more than a few minutes to drive to work.

Table 1

Time calculated for routine activities following a 14-hr FDP getting less than 8 hr. of sleep.

Time	Activity
30 min	Wake-up, groom and dress
1.0 hr	Get Children off to school, take care of pets, etc.
30 min	Make and eat breakfast
30 min	Commute to work – drop kids at school/daycare on the way
14.0	Duty Period
30 min	Commute home – pick-up kids on the way home
1.0 hr	Family activities or household responsibilities
1.0 hr	Dinner
1.0 hr	Clean kitchen, check homework assignments, put kids to bed
1.0 hr	Unwind, read, pay bills, etc
30 min	Prepare for bed; brush teeth, hair, wash face, shower, etc.
21 30	Total hr of activities
2 30 Number of hr left for sleep in a 24 hour period	

Note. Table is adapted from the Flight Attendant Fatigue, Part V: A Comparative Study of International Flight Attendant Fatigue Regulations and Collective Bargaining Agreements (FAA, 2009)

If we remove, all of the personal time from the table above, we can dramatically reduce activities to allow for 8 hours of sleep (FAA, 2009). This is without eating dinner, and assuming that the crewmember could fall asleep immediately- with no interaction with his or her family. A layover at a hotel would only change the scenario slightly, as the travel time used to ride to the crewmembers home residence would be replaced with a hotel shuttle and check-in period.

Commuting responsibly and arriving fit for duty is the responsibility of the crewmember. Many crewmembers sleep in hotels and commuter apartments the night before checking in for duty at their domicile, allowing them to arrive well rested. This expense is generally the responsibility of the crewmember, which can be very difficult for a crewmember earning a low salary.

We must also consider the duty time may be extended to 16 hours under the current regulations. The FAA is proposing to amend its existing flight, duty, and rest regulations applicable to certificate holders and their flightcrew members. On August 1, 2010, the President signed the Airline Safety and Federal Aviation Administration Extension Act of 2010, P.L. 111-216 (the Act). In section 212 of the Act, Congress directed the FAA to issue regulations no later than August 1, 2011 to "specify limitations on the hours of flight and duty time allowed for pilots to address problems relating to pilot fatigue". The Act directed the FAA to consider several factors that could affect pilot alertness including time of day, number of takeoffs and landings, crossing multiple time zones, and the effects of commuting (P.L. 111-216, 2010).

Preliminary data from a pilot study conducted by Western Michigan University, funded by a research development award, (Brown, 2011), entitled "Effects of Commuting on Pilot Fatigue: A Comprehensive study in Support of Risk Management" contributes findings from the crewmembers perspective. In the study (Brown, 2011), preliminary findings, show that 42% of the professional pilot participants indicated that they have changed pilot domiciles 3-4 times in their career, and 17% have changed domiciles (unpaid) over six times throughout their career. Additionally, over 55% of the participants reported commuting to their current domicile, with 67% of the commuting crewmembers reporting travel over 4 hours and crossing multiple times zones, to get to their assigned domicile.

There is very little scientific data suggesting how many pilots and flight attendants commute on a whole.

In the WMU pilot study (2011), all of the participants indicated they have experienced unforeseen delays during their commute to work, which increased their travel time. Delays caused by weather, diversions, mechanical problems, flight delays, flight cancelations, and difficulty obtaining a jumpseat of standby seat were reported. All of the participants indicated that a mandatory move to a new domicile would cause hardships such as:

- Economic pressures, family stability, possible safety hazards
- Loss of relationships and decreased quality of life
- Added travel expenses not covered by airline, difficulty selling home
- Increased time away from family, increased risk of not being able to commute to work, spouse may lose their job
- Higher expenses, family turmoil, possible safety issues, general reduction in quality of life
- Move children out of schools

None of the participants in the WMU preliminary study indicated that the time spent commuting to their base domicile, would affect their ability to perform their duties safely.

The factors, which participants indicated, most affect their ability to perform their duties safely (Brown, 2011) include: inadequate crew rest; length of duty day; operations during the ‘backside of the clock’; trip pairing check-in times; construction of schedules; quality of rest prior to and during the trip; cumulative sleep debt; operations in various time zones; and lack of time for sustenance and hydration.

Congress has directed the FAA to contract the National Academy of Sciences (NAS) to conduct a study of the effects of pilot commuting on fatigue (NAS report 13097, 2011). The NAS study will review available information on: the prevalence of pilots commuting; characteristics of commuting by pilots; and the impact of commuting on pilot fatigue, sleep, and circadian rhythms; commuting policies of commercial air carriers (including passenger and all cargo air carriers), including pilot check-in requirements and sick leave and fatigue policies.

The NAS study will also:

- Define “commuting” in the context of pilot alertness and fatigue;
- Discuss the relationship between the available science on alertness, fatigue, sleep and circadian rhythms, cognitive and physiological performance, and safety;
- Discuss the policy, economic, and regulatory issues that affect pilot commuting;
- Discuss the commuting policies of commercial air carriers and to the extent possible, identify practices that are supported by the available research; and
- Outline potential next steps, including to the extent possible, recommendations for regulatory or administrative actions, or further research, by the FAA.

This has raised a salient point, which has caused industry and media uproar, which is the suggestion of regulating or administrative actions surrounding commuting. This controversial discussion has raised the questions of “how can the FAA regulate a crewmembers personal time?” The Federal Aviation regulations (FARs’) currently regulate drug and alcohol use prior to duty. Although, commuting, is generally a conducted during ‘the crewmembers time off, by their individual choice, and there is no compensation for this time. It is important to note that commuting to a pilot domicile, even while traveling on a different airline, has been considered “work related” in a recent workers compensation case. One example of this was the Airtran Airways case, in which a pilot commuting to work on another carrier was killed during the commute.

As stated in the Workers Compensation, case (2009-SC-000429-WC and No. 2008-CA-001223-WC, 07-00328):

> Clarence Fortney was a pilot employed by Airtran Airways, Inc. and a commuting pilot riding as a passenger on Comair Flight 5191, when he was killed when the plane crashed on takeoff in Lexington, Kentucky on August 27, 2006. Fortney resided in Lexington to be near his family. Airtran employed about 1450 pilots who resided throughout the United States in August 2006 and was required to know and follow the income tax laws of numerous states and localities' because 70% of the pilots resided outside the state of Georgia. Airtran incurred additional expense due to participating in a nationwide Transportation Security Administration database that was updated every 24 hours and due to verifying the identity of pilots seeking to fly free or at a reduced fare on Airtran flights.
>
> Fortney indicated when applying for employment with Airtran that he would be willing to relocate and that there were no restrictions on where he would locate, but Airtran never dictated where he or other pilots must reside. Consistent with industry practice, Airtran provided employees and their families with free or reduced-fare travel on Airtran flights and participated in reciprocal conveyance agreements with other airlines, which also provided free or reduced-fare travel on aircraft operated by those airlines. Nothing required Airtran pilots to fly when commuting to and from work, but those who lived outside Georgia generally used the free or reduced fare arrangements in order to be able to afford to commute. Pilots performed no work while commuting by air; were not paid until they checked in at the Atlanta hub for an assigned flight; and were not reimbursed for commuting expenses. Klaus Goersch, Vice-President of Flight Operations for Airtran, testified that any Airtran employee could choose where to live.
>
> Airtran did not operate in Kentucky in August 2006, but had a reciprocal arrangement with Comair, which permitted pilots to travel free or at a reduced fare in a cockpit jumpseat on a "'Space Available' basis".
>
> Fortney was commuting to Atlanta under Airtran's arrangement with Comair when he was killed. Evidence that the arrangement made it possible to do so financially while working for Airtran, which enabled Fortney to live where he chose, compelled legal conclusions that it was an inducement to Fortney to accept the employment, and that it benefited Airtran by accomplishing its purpose. The report concluded that his death was work-related because he was making such a trip when it occurred.

Summary

Although further study is required, the WMU pilot study (2011) suggests that long range commuting is more the “norm” than the exception for today’s aviation industry, and that the current 14-16 hour flight duty periods, make it difficult for crewmembers to achieve 8 hours of restorative sleep in a 24 hour period-regardless of how one may travel to work. In addition, the time spent commuting to work, regardless of the mode of transportation, comes at a price. It is important to understand that many crewmembers commute responsibly, and have so for decades. Careful thought needs to be put into schedules, fatigue risk management plans, the use of models, and commuting policies- so that we can give our next generation of pilots and flight attendants the tools necessary to operate safely, and enable business productivity.

Acknowledgements

The Western Michigan University, pilot study is funded by Western Michigan Universities’ Research Development Award, and approved under HSIRB 09-11-20 protocol. A special thanks to Captain John Gadzinski, Captain Steve Sevier, and all of the professionals who participated in the industry feedback group. Full study results will be published when data collection is completed.

References

Avers, K., King, J., Nesthes, T. , Thomas, S. , & Banks, J. (2009). Flight Attendant Fatigue, Part V: A Comparative Study of International Flight Attendant Fatigue Regulations and Collective Bargaining Agreements. (Final Report DOT/FAA/AM-09/22 DOT). Washington, DC: Federal Aviation Administration Office of Aerospace Medicine, (2009).

Alexander, C. H. (2009).The American Community Survey Issues and Initial Test Results. United States Census Bureau. Retrieved from http://factfinder.census.gov/home.

Howlett D, Overberg P., (2005). Extreme Commuting Illustrates Tradeoffs and Opportunity Cost. USA Today, 2005.

Liepmann, Kate K., *Journey to Work*, Oxford University Press, New York: 1944, p. 35.

National Research Council (2011). Issues in Commuting and Pilot Fatigue: Interim Report. Committee on Commuting and Pilot Fatigue, Board on Human-Systems Integration, Division of Behavioural and Social Sciences and Education. Washington, DC: The National Academies Press. Retrieved from **http://www.nap.edu/catalog/13097.html**.

National Transportation Safety Board. (1994). *A Review of Flightcrew-Involved, Major Accidents of U.S. Air Carriers, 1978 Through 1990*. Safety Study NTSB/SS-94/01. Retrieved from http://libraryonline.erau.edu/online-full-text/nstb/safety-studies/5594-01.pdf.

National Transportation Safety Board. (2010). *Loss of Control on Approach, Colgan Air, Inc. Operating as Continental Connection Flight 3407, Bombardier DHC-8-400, N200WQ, Clarence Center, New York, February 12, 2009.* Accident Report NTSB/AAR-10/01, PB2010-910401. Washington, DC: National Transportation Safety Board.

Pisarski, A.E. (2006). Commuting in America III: The Third National Report on Commuting Patterns and Trends. Transit Cooperative Research Program (TCRP) Report 110/NCHRP Report 550, published by the Transportation Research Board, Washington.

Public Law 111-216 (2010.) Airline Safety and Federal Aviation Administration Extension Act of 2010. Retrieved from http://www.gpo.gov/fdsys/pkg/PLAW-111publ216/pdf/PLAW-111publ216.pdf.

Schaefer, A., (2005). Workers are traveling ever longer to attain the job or home life they want, but the daily stress may outweigh the gains, Scientific American Mind (October 2005).

United States Department of Transportation. (2010). Federal Aviation Administration 14 CFR, Parts 117 and 121 Flightcrew Member Duty and Rest Requirements: Proposed Rule Federal Register/ Vol.75. No. 177/ Tuesday, September 14, 2010/Proposed Rules. (NPRM).

PERCEPTIONS OF GENDER-RELATED PILOT BEHAVIOR

Rosemarie Reynolds
Alexandru Milut
Joshua Hirschheimer
Bridget Cox
Bojan Milenkovic
Wangyang Xu
Embry-Riddle Aeronautical University
Daytona Beach, Florida

An online version of the 34-question Aviation Gender Attitude Questionnaire (AGAQ) was administered to 113 Federal Aviation Administration-certificated pilots. A statistically significant difference ($p < 0.05$) in perceptions of female pilots was found between male and female pilots, with male pilots viewing female pilots' flying proficiency, flight confidence and standards more negatively than did female pilots. These perceptions were not moderated by age, experience, or opportunities to fly with female pilots. Our findings replicated previous findings in South Africa, Australia, Norway, and South America. The paper concludes with a discussion of the implications for female pilots, and training implications for both genders.

Women have flown aircraft since the 18th century, when Jeanne Labrosse first soloed in a balloon. Five years after the Wright brothers' first flight at Kill Devil Hills, North Carolina, Therese Peltier soloed an aircraft in the Military Square in Turin, Italy. Two years later, Baroness Raymonde de Laroche became the first licensed female pilot. But these female aviation pioneers were the exception, rather than the rule; the aviation industry has historically discouraged women from becoming pilots. Changes in societal attitudes towards women as a whole, combined with the passing of anti-discrimination and equal opportunity legislation, have removed many of the legal barriers that women faced in pursuing flying, but not the hostility and suspicion from their male peers that can lead to isolation, sexism, and harassment (Davey & Davidson, 2000).

The purpose of this study was to examine perceptions of gender-related pilot behavior in the United States (US) pilot population. Such an assessment is important for several reasons, with safety of flight being the most important one; if the crewmembers of an aircraft distrust each others' abilities and judgment, they will not be able to function adequately as a team. Additionally, at a time when declining standards in professionalism and the quality of the pilot pool are at the forefront of the debate on pilot certification and hiring standards, the aviation community cannot afford to shun future members due to preconceptions and prejudices. Finally, an assessment of pilot attitudes can be helpful in either addressing the problem of negative perceptions if they exist, or in dispelling the myth if they do not.

Previous studies have examined gender differences in both pilot behavior and perceptions of pilot behavior. Fischer and Orasanu (1999) found gender differences in the way that female and male pilots communicated with their air crews. Baker, Lamb, Grabowski, Rebok, and Li (2001) examined accidents in general aviation, and found that male pilots were more likely to crash due to inattention or flawed decision-making, while female pilots were more likely to crash because they mishandle the aircraft. It should be noted, however, that McFadden (1996) found no difference in the accident rates of male and female airline pilots. Turney (1995) examined perceptions of gender-based differences in air crews, and found that male pilots were perceived as more task-oriented and confident, while female pilots were seen as better communicators and negotiators. Turney and Bishop (2004) $^($found gender-based differences in concepts such as command, leadership, effective communication, and decision-making in air crews. Vermeulen (2009) examined how flight instructors and commercial pilots in the Republic of South Africa differed in their perceptions of female pilots, finding that the two groups differed significantly in their assessment of female pilots' flying proficiency but not in their safety orientation.

Our purpose was to examine gender-based perceptions of pilot behavior, rather than actual differences in behavior, using the 34-question Aviation Gender Attitude Questionnaire (AGAQ). The AGAQ, designed by Wilson (2004), gauges attitudes, stereotypes and prejudices towards female aviators. Although the original questionnaire contained 72 questions, the questionnaire has since been refined to 34 questions covering four dimensions: 1) flying proficiency, 2) safety orientation, 3) flight confidence, and 4) flight standards.

The AGAQ has been shown to have acceptable factorial validity, internal consistency and unidimensionality, and appears to be a valid and culturally unbiased tool that could be used to ascertain gender-related perceptions in aviation across different Western, English speaking cultures (Kristovics, Mitchell, Schaap & Vermeulen, 2009; Vermeulen & Mitchell, 2007).

Hypotheses

We proposed that pilot perceptions of gender-related behavior would vary as a function of a number of variables, including: pilots' gender, age, flight experience, and opportunity to fly with the opposite gender. Specifically, we hypothesized that:

H_1: Perceptions of female pilots' behavior will vary depending on gender (same sex vs. opposite sex).
H_2: Male pilots' perceptions of female pilots' behavior will vary with the male pilots' age.
H_3: Male pilots' perceptions of female pilots' behavior will vary with the male pilots' experience as measured by total flight hours.
H_4: Male pilots' perceptions of female pilots' behavior will vary depending on the amount of opportunities male pilots have had to fly with female pilots.

Method

Our method was a survey, hosted online using the Survey Monkey software and its web hosting option. The Survey Monkey software automatically collected and catalogued the data, which was then be interpreted using SPSS statistical analysis software. Traffic to the website was generated by contacting various pilot and aviation organizations via e-mail and telephone calls, through a viral campaign on the Facebook social networking site, and through notices on various online pilot and aviation-related forums.

Sample

For the purposes of this study, we defined the pilot population as "anyone who holds a valid Federal Aviation Administration (FAA) pilot certificate." The total population consists of more than 612,000 pilots. We obtained a total of 113 complete responses to our survey. Of our respondents, 89 were male (79%) and 24 (21%) were female. The age of our sample ranged from 18 to 64, with an average age of 27. The majority of our sample (61%) had a commercial pilot or higher rating. The average flight experience in years was eight, and the average number of flight hours was 2,723.

Instrument

We used the 34-question Aviation Gender Attitude Questionnaire (AGAQ), which covers four dimensions: 1) flying proficiency, 2) safety orientation, 3) flight confidence, and 4) flight standards. The AGAQ has been shown to have acceptable factorial validity, and internal consistency (Vermeulen & Mitchell, 2007). The AGAQ uses a 5-point Likert scale, with lower scores indicating more positive attitudes.

Analysis

Where appropriate, the answers to the AGAQ were reverse-coded, and the individual question scores were then aggregated according to the four factors identified by Vermeulen and Mitchell (2007). Reliability analysis was carried out by calculating Cronbach's alpha (α =.94).

Hypothesis One

Out first hypothesis stated that perceptions of female pilots' behavior would vary by gender. T-tests were performed to test the hypothesis. The mean scores and results of the t-tests are shown in Table 1. There was no difference in perceptions of safety orientation, however; for the remaining three dimensions, there were significant differences in perceptions based on gender.

Table 1.

Perceptions of Female Pilots by Gender

Mean			*t*	*p*
	Males[1]	Females[2]		
Flying proficiency	43.37 2	6.96	6.5	<.01
Safety Orientation	19.97 1	8.46	1.14 ns	
Flight Confidence	22.63 1	6.75	4.52 <.0	1
Erosion of Standards	11.63 8.	21	4.58 <.0	1

Note. [1] n= 89 [2] n = 24

Hypothesis Two

Our second hypothesis stated that male pilots' perceptions of female pilots' behavior would vary with the male pilots' age. We divided our sample of male pilots into two groups; those between 18 and 30, and those between 31 and 64. T-tests were performed to test the hypothesis that mean scores would vary by group. The scores and results of the t-tests are shown in Table 2. There were no significant differences perceptions based on age.

Table 2.

Male Pilot Perceptions of Female Pilots by Age

Mean			*t*	*p*
	18-30[1]	31-64[2]		
Flying proficiency	44.56 3	9.52	1.74	ns
Safety Orientation	19.85 2	0.33	.34 n	s
Flight Confidence	23.07 2	1.19	1.32 ns	
Erosion of Standards	11.68 1	1.48	.23 n	s

Note. [1] n= 68, [2] n = 21

Hypothesis Three

Our third hypothesis focused on the effects of experience, as measured by fight hours, on male pilots' perceptions of female pilots' behavior. We divided our sample of male pilots into two groups; those with 350 hours or less, and those with 351 hours and more. T-tests were performed to test the hypothesis that mean scores would vary by group. The scores and results of the t-tests are shown in Table 3. There were no significant differences perceptions based on experience.

Hypothesis Four

Our fourth hypothesis stated that perceptions of female pilots' would vary depending on the amount of opportunities male pilots had to fly with the opposite gender. Male pilots were grouped into those who had fewer (or no) opportunities to fly with the opposite gender (50% of the time and below) and those who had more opportunities to fly the opposite gender (50% and above). T-tests were performed to test the hypotheses.

The mean scores and results of the t-tests are shown in Table 4. There were no significant differences in perceptions based on the opportunities male pilots had to fly with the opposite gender. Although not statistically significant, opportunities to fly with the opposite gender actually resulted in less favorable perceptions of female pilots.

Table 3.

Male Pilot Perceptions of Female Pilots by Number of Flight Hours

Mean			*t*	*p*
	350 hours or less[1]	351 hours or more[2]		
Flying proficiency	43.85 4	2.59	.49	ns
Safety Orientation	19.36 2	0.94	1.3 ns	
Flight Confidence	23.25 2	1.62	1.3 ns	
Erosion of Standards	11.58 1	1.71	.16 n	s

Note. [1] n = 55, [2] n = 34

Table 4.

Male Pilot Perceptions of Female Pilots by Opportunities to Fly with the Opposite Gender

Mean			*t*	*p*
	Few opportunities[1]	More opportunities[2]		
Flying proficiency	42.80 4	6.20	1.03	ns
Safety Orientation	19.86 2	0.47	.38 n	s
Flight Confidence	22.53 2	3.13	.37 n	s
Erosion of Standards	11.46 1	2.47	1.03 ns	

Note. [1] n= 74, [2] n = 15

Discussion

Our findings corroborate those found in previous research, indicating that there are indeed differences in perceptions of female pilots based on gender, with male pilots viewing female pilots in a more negative light than do female pilots (Kristovics, Martinussen, Mitchell, Vermeulen, & Wilson, 2009; Vermeulen, 2009; Wilsom, 2004). None of our other hypotheses were confirmed: neither age, experience, nor opportunities to fly with female pilots had an impact on the negative perceptions regarding female pilots. Even after 100 years of female participation in aviation, a certain level of distrust exists regarding female pilots. Although these findings are somewhat disheartening, they do point out areas in which further work needs to be done. Changes in societal attitudes towards women as a whole, combined with the passing of anti-discrimination and equal opportunity legislation, have not removed the hostility and suspicion that female pilots face from their male peers.This finding is important for several reasons, with safety of flight being the most important one; if the crewmembers of an aircraft distrust each others' abilities and judgment, they will not be able to function adequately as a team.

The main limitation of our study was the small and somewhat skewed nature of our sample, which was heavily weighted toward younger and less experienced pilots. Because of this, there are doubts as to how accurately our sample represents the pilot population of the United States as a whole.

Future research should focus integrating the results of this study with those of previous studies using the AGAQ in order to obtain a more complete view of pilot perceptions of gender-related pilot behavior. A complete and accurate picture of how various subsets of the pilot population view gender-related pilot behavior will allow for the roots of such attitudes to be identified and addressed.

References

Baker, S.P., Lamb, M.W., Grabowski, J.G., Rebok, G., & Li, G. (2001). Characteristics of general aviation crashes involving mature male and female pilots. *Aviation, Space, and Environmental Medicine*, *72*, 447-452.

Davey, C.L., & Davidson, M.J. (2000). The right of passage? The experiences of female pilots in commercial aviation. *Feminism and Psychology*, *10*, 195-225.

Fischer, U., & Orasanu, J. (1999). Confronting the boss indirectly: Study of cockpit crews finds copilots use 'hints' to correct captains. *Research News & Publications Office. Georgia Institute of Technology.* Retrieved 25 April, 2002, from http://gtresearchnews.gatech.edu/newsrelease/CREW.html

.Kristovics, A., Martinussen, M., Mitchell, J., Vermeulen, L., & Wilson, J. (2009). How pink is the sky? A cross national study of the gendered occupation of pilot. *Employment Relations Record,* l 5, 43-60.

Kristovics, A., Mitchell, J., Schaap, P., & Vermeulen, L. (2009). Exploring the equivalence of the aviation gender attitude questionnaire (AGAQ) for South African and Australian pilots: a cross-cultural comparison. *The International Journal of Aviation Psychology,* 19, 367-390.

McFadden, K.L. (1996). Comparing pilot-error accident rate of male and female airline pilots. *Omega, International Journal of Management Science*, *24*, 443-450.

Turney, M.A. (1995). Women's learning and leadership styles: Impact on CRM. In N. McDonald, N. Johnston & R. Fuller (Eds.), *Applications of psychology to the aviation system*. Proceedings of the 21st Conference of the European Association for Aviation Psychology (EAAP). Vol. 1 (pp. 262-268). Aldershot: Ashgate.

Turney, M.A., & Bishop, J.C. (2004). Women's learning and leadership styles: Implications for air crews. In M.A. Turney (Ed.), *Tapping diverse talents in aviation: Culture, gender, and diversity*, (pp. 61-75). Aldershot: Ashgate.

Vermeulen L.P. (2009) Flight instructors' perceptions of pilot behaviour related to gender. *SA Journal of Industrial Psychology,* .35.

Vermeulen, L., & Mitchell, J. (2007). The development and validation of a measure to assess perceptions regarding gender-related pilot behavior. *The International Journal of Aviation Psychology*, 17, 197-205.

Wilson, J. (2004). *Gender-based issues in aviation, attitudes towards female pilots: A cross-cultural analysis*. Unpublished doctoral thesis, University of Pretoria, Pretoria, South Africa.

Preliminary Program Assessment for Attracting Physically Disabled Individuals to Aviation

Erin E. Bowen
Department of Technology Leadership and Innovation, Purdue University
West Lafayette, IN, USA

Bernie Wulle
Department of Aviation Technology, Purdue University
West Lafayette, IN, USA

Michael E. Jednachowski
Department of Aviation Technology, Purdue University
West Lafayette, IN, USA

Developing effective aircraft piloting skills takes years of training with qualified experts to meet and exceed flight certification requirements. Not everyone who desires to can reach those standards, and the challenge is even greater for those with physical disabilities who desire to learn to fly. While considered uncommon, there are opportunities that seek to modify aircraft and train physically impaired students to be pilots. However, little research has been done on whether there also needs to be modification in the approach of flight and ground instructors when training impaired students. The present study presents preliminary findings of a qualitative study designed to assess the need to adapt flight training for those with physical handicaps. The authors review plans for an expanded follow-up study and discuss additional issues in opening aviation to those with physical or other limitations.

The history of aviation is filled with glamorous icons, colorful details of battles in the sky and the geniuses who made it all possible. What has been of less focus is the monumental force behind each of these stories; that is, the people behind the scenes who make it all happen. These positions are filled not with enchanting personalities or larger than life characters but rather experts at a position who got the job done well. Historically, aviation has consistently predominately of individuals who were male and physically fit. Even currently, applicants for aviation jobs, in particular commercial pilot positions, must meet very specific criteria. Those who do not fit those specifications may experience discouragement upon finding that they do not meet the requirements for specific aviation positions and leave the industry altogether. Those with a handicap or disability might not even consider a position within the aviation community at all, functioning under the perception that the physical requirements for the industry will exceed their capabilities.

Disabled Flight Training Support

Organizations exist to promote the contribution of those with disabilities to society by filling positions that are vital to everyday business. A paper written by Marianne Roche suggests that, since the time of U.S. President Lyndon Johnson, there has existed within Western society the assumption that every American citizen has a right to participate and contribute to that society; but that only very recently that developmentally disabled citizens have been “deemed qualified to be investors” (Roche, 1980). The physical requirements to fly an airplane has more than likely put certain individuals under the illusion that there is no spot for them in aviation. This assumption is not only detrimental to the individual, who excludes themselves from potential economic/employment opportunities in a global, high-technology industry, but also to the industry and its agencies themselves, who suffer from an absence of otherwise capable and qualified employees who have the potential to contribute greatly to industry success.

In response to a perceived need to increase opportunity and foster independent and unique skills among the physically disabled, the Able Flight program was developed and announced in 2006 with the intention to "offer people with disabilities a unique way to challenge themselves through flight training, and by doing so, to gain greater self-confidence and self-reliance" (Able Flight, 2011). An increase in technology that is accessible, less costly and more readily adaptable to the changing requirements of various physical limitations have allowed the Able Flight organization to provide training flight training for those with a variety of physical handicaps. Through the work of groups such as Able Flight, the love of aviation has been met with tangible success for those with physical disabilites.

Recently, Able Flight entered into a collaborative agreement with a collegiate aviation program at a large Midwestern university. The purpose of this collaboration was to provide a unique opportunity for physically disabled individuals to participate in sport pilot flight instruction within the structured framework and oversight of a large-scale collegiate aviation program. Trainees accomplished the goal of achieving a sport pilot certificate by passing the oral and flight test required for every sport pilot applicant.

Aviation Industry Potential Applications

It is important to note in this collaboration that there were no amendments to the flight certification process and no special consideration given to trainees because of their disabilities. Each passed federally-mandated performance requirements that have been generated on their own merit. Aviation training has been standardized with few changes for decades and has existed in a "one size fits all" environment. Further research is intended to be two-fold; first, to determine if there is a need to change the way flight training is given to students with a handicap; and second, to promote opportunities in the aviation industry for physically disabled individuals, using a flight training program as an entry point. Many qualified and intelligent individuals may not posses the money or ability to become a pilot, but aviation cannot advance without actively recruiting capable individuals who have historically not been represented in the industry.

With regard to the first research area, the ability to make accommodations for those with disabilities has increased in recent years because of technological advancement. The aircraft used for Able Flight pilot training in the collegiate setting was modified in such a way that the students were able to safely maneuver the aircraft using only hand controls. While training must be done to federally mandated certification standards, aviation is somewhat unique among industries because of the pre-existence of a highly personalized training environment. Pilot training, specifically in-aircraft training, occurs on a one-to-one basis. Students with a wide variety of physical or educational/learning style needs can be placed with a flight instructor or teacher properly equipped with the knowledge and technology to meet the situation. Constantly adjusting teaching methods and materials can benefit all students if that variety of learning is taken into consideration (Agogino & Hsi, 1995). However, there has not yet been a cohesive review of how physical disabilities may alter the teaching and learning strategies for a high-technology skillset such as piloting ability. The collaboration with Able Flight in a collegiate setting provides a unique opportunity to undertake methodical assessment of particular learning style trends or specific issues which may be indirectly or unintentionally limiting the success of a physically disabled individual in aviation.

Second, the aviation industry is a complex and multi-faceted system that extends far beyond serving as a pilot. Unfortunately, informal discussions with a number of aviation industry and government regulatory agency leaders suggest that few physically disabled individuals are participating in those branches of the industry in which their handicap would be no limitation whatsoever, or overcome with reasonable accommodation. The Able Flight/collegiate aviation joint flight training program offers a well-publicized entry point, a highly visible tool to raise awareness among industry leaders and among the physically disabled of the opportunities available in the aviation industry. Research establishing training requirements, accommodations, human factors issues, and other considerations that provide evidence-based knowledge for industry and regulatory leaders may go a long way to encourage more active recruitment of physically disabled persons to this industry. The potential for job growth and gains for both groups are substantial. The present study presents early steps being taken to establish such industry connections in support of opening the aviation industry to a significantly under-represented group of individuals.

Methodology

Given the small number of participants in the first year of the training program, the study was designed using a case-based methodology in order to capture a depth of information regarding the perceptions, experiences, and perspectives of program participants. *Participants* in this context refers both to the two physically disabled trainees as well as to the two Federal Aviation Association (FAA)-certified flight instructors assigned to work with them and oversee their flight training.

Trainees each completed the Learning Styles Inventory (Felder & Solomon, 1996) in order to assess the potential for interactivity between learning preferences and established flight training protocols. Given the small sample size, statistical analyses could not be conducted; however, review of LSI results and comparison to extant literature suggested no potential significant issues impacting the flight training process. In order to develop a system to recruit a population that has previously had a small role within aviation is first to determine if there exists any differences in learning for which to compensate. A learning styles inventory was created to determine how students perceive and receive information by David Kolb in 1976. Through multiple studies, it has been determined that students learn best when information is presented in a way that best correlates to their learning style (Dunn, 1986, 2000). Studies on aviation students and learning styles have shown that these students usually fit into one of two categories. A longitudinal study by Kanske, Brewster, and Fanjoy (2003) showed that by senior year, 72% of students fit into either the assimilating or converging learning style. Some general conclusions have emerged in how learning styles relate to pilot training and one says that when teaching accommodates various preferences, more students are successful (O’Connor, 1997).

Participant Reflection Analysis

Both trainees and flight instructors were asked to write journal-style reflections on their experiences, challenges, and perceptions of the flight training program. Participants completed reflections after the first day of the training program, at approximately halfway through training, and upon completion of the program. The list of reflection questions can be found in Appendix A.

Of particular note are the contrasts between the instructor and trainee perspective throughout the course of the program. Prior to program beginning, both instructors and (though to a lesser extent) trainees expressed concern with the fixed amount of time available to complete the sport pilot training; the same concern arose for both groups at the end of training. This concern lingered following training even though both trainees successfully earned sport pilot certification according to federally mandated standards. There appeared to be a perception that, although certification standards were met, complete comfort with skill acquisition for both instructors and pilots may not have been met in the established time limits.

Questions regarding the appropriate instructional strategy for the Able Flight program arose in reflections and in anecdotal discussions with study leaders. Instructors discussed at the beginning of the program the feeling that they should be more “laid back” in approach because students were training for fun rather than for professional development; yet by the end of the program reported they would be less laid back if instructing for this program in future. Trainees in turn reported at program completion a desire to have been pushed harder by instructors. The authors anticipate that future iterations of the Able Flight training will address this issue, particularly as the program expands to recruit physically disabled individuals for aviation careers.

One issue notable by its absence was the lack of reflective discussion on any limitations imposed by the physical handicaps possessed by the flight trainees. When questioned regarding training challenges, concerns, or physical skill challenges, trainees focused on technical issues (learning takeoffs/landings), the time limitations given the short training cycle (shared by instructors), the challenges faced in the cockpit regarding multitasking and regulatory awareness, and material mastery. In the training setting, flight students appeared to regard themselves as trainee pilots first and foremost, and participants in a program for the physically disabled second. Instructors may have also been unprepared for this attitude, as they expressed a level of surprise with the trainees' desire to socialize and celebrate training milestones like any typical college-aged student.

While the cases described here represent a very small sample of a preliminary introduction of the Able Flight program into the collegiate setting, the authors anticipate continuing and expanding the research program with the continued expansion of the training program over a multi-year period. Data will continue to be gathered and shared with interested research collaborators.

Conclusion

Preliminary assessment of the Able Flight training program in a collegiate aviation setting indicates strong potential for this program to redefine the opportunities for physically disabled students in the aviation industry. The collegiate setting provided a grounding and structure that trainees reported finding beneficial and supportive in the face of the training challenge. Currently, graduates of this pilot program are successfully continuing to participate in aviation activities, including demonstrating their flight skills by visiting U.S. military veterans' hospitals in their aircraft and showcasing the potential opportunities still open in aviation in the face of physical disability. The research team anticipates continuing and expanding programmatic research on the training, human factors, and implementation issues related to increasing the presence of physically disabled individuals in aviation through expansion of the Able Flight training program and integration with other existing aspects of the collegiate aviation training environment.

References

Able Flight: About Us (2011). Retrieved February 26, 2011 from http://ableflight.org/about-us.html

Agogino, Alice M., and Sherry Hsi. (1995). Learning style based innovations to improve retention of female engineering students in the Synthesis Coalition. In *ASEE/IEEE Frontiers in Education '95: Proceedings*. Purdue University.

Dunn, R. (2000). Learning styles: Theory, research, and practice. *National Forum of Applied Educational Research Journal*, 13, (1), 3-22.

Dunn, R. (1986). Learning style: State of the science. *Theory into Practice,* 24(1), 10-19.

Felder, R. M. and Barbara A. Soloman, B. A. (1996). *Index of Learning Styles*, http://www.ncsu.edu/felder-public/ILSpage.html, accessed May 5, 2010.

Kanske, C. A., Brewster, L. T., & Fanjoy, R. O. (2003). A longitudinal study of the learning styles of college aviation students. *International Journal of Applied Aviation Studies,* 3(1), 79-89.

Kolb, A., & Kolb, D. A. (2001). Learning styles. In Forest, J. & Kinser, K. (Eds.), *Encyclopedia of Higher Education in the United States*. Santa Barbara, CA: ABC-CLIO Publishers.

O'Connor, T. (1997). *Using Learning Styles to Adapt Technology for Higher Education*. Available: http://www.indstate.edu/ctl/styles/learning.html.

Roche, Marianne. (1980). Marc Gold's training in industrial work sites as a placement approach. In K. Mallik, & E. Shaver (Eds.), *Unmasking abilities hidden by developmental conditions (pp. 3-5)*. Washington, D.C.: George Washington University Job Development Lab.

APPENDIX A
Reflection Questions

Participant Reflections

Directions: For each question, please write at least 2-3 paragraphs that explore your perceptions and opinions as they regard the Able Flight training program. You are welcome to write more if you choose. These questions help the class instructors improve their training and help the research investigators find ways to attract a wider range of students to aviation.

Reflection Assignment #1 (end of first day's class):

1. *What are you most excited about learning in the Able Flight training program this summer?*
2. *What do you anticipate will be the greatest challenge of this training program, and why?*
3. *What do you hope to gain as a result of participating in flight training this summer? What do you think will be different for you after this training?*

Reflection Assignment #2 (halfway point of program):

1. *What has been the greatest challenge to this point in the program, and why was that so?*
2. *How have the instructors helped you in your learning so far?*
3. *What could the instructors do better in this first half of the course, and why?*
4. *Looking forward to the next half of the program, what are you most excited* **or** *anxious about learning, and why?*

Reflection Assignment #3 (end of course):

1. *What aspect of this training program was most enjoyable to you, and why did you enjoy it?*
2. *Looking back over the whole flight training program, what was the greatest challenge you faced, why was it a challenge, and how did you overcome it?*
3. *How did the flight instructors help you in your learning?*
4. *What would* **you** *have the instructors do differently if they were to teach this same program again?*
5. *How has this experience impacted your physical skills? Has it presented new challenges or played to strengths you already had, and why?*

Instructor Reflections

Directions: For each question, please write at least 2-3 paragraphs that explore your perceptions and opinions as they regard the Able Flight training program. You are welcome to write more if you choose. These questions help the research investigators find ways to attract a wider range of students to aviation.

Reflection Assignment #1 (end of first day's class):

1. *What are some of your concerns as an instructor in the Able Flight training program?*
2. *What do you anticipate will be the greatest challenge of this training program, and why?*
3. *How do you anticipate changing your "traditional" instructor approach for the Able Flight students, and why do you anticipate those changes?*

Reflection Assignment #2 (halfway point of program):

1. *What has been the greatest challenge to this point in the program, and why was that so?*
2. *What has been the most unexpected or unanticipated event so far in the program, and why?*
3. *How have you had to alter your training approach for the students, why did you make those changes, and how did they help or hinder training?*
4. *Looking forward to the next half of the program, what are you most excited* **or** *anxious about teaching, and why?*

Reflection Assignment #3 (end of course):

1. *What aspect of this training program was most enjoyable to you, and why did you enjoy it?*
2. *Looking back over the whole flight training program, what was the greatest challenge you faced, why was it a challenge, and how did you overcome it?*
3. *What would* **you** *as an instructor do differently if you were to teach this same program again?*
4. *How has this experience impacted your instructor skills/knowledge? Has it presented new challenges or played to strengths you already had, and why?*

PILOT CERTIFICATION: CONVERTING A HISTORICAL FRAMEWORK TO DIGITAL DATA UTILIZATION

Matthew J. Baranowski
Purdue University
West Lafayette, Indiana USA

Brian G. Dillman
Purdue University
West Lafayette, Indiana USA

Pilot certification relies on a single subjective flight exam to determine competency and safety. Evidence-based evaluation of an airman's historical flight performance through flight recorder data may offer increased validity and reliability in training and in certification. This study examined commercial pilot students (*N*=13) training in Cirrus SR-20 aircraft in a collegiate flight training program. Each student's performance of the steep turns maneuver before, during, and after a flight exam using digital flight recorder data was correlated with each other to examine the validity of single evaluation flight exams. The results indicate that each student's average historical performance before a flight exam is a more accurate indicator of future performance than the flight exam itself. This study investigates the feasibility of utilizing digital flight recorder data to objectively analyze students' flight performance.

The current process for pilot certification by the Federal Aviation Administration (FAA) is individually subjective, despite continuous efforts for standardization (FAA, 2010). There are hundreds of Designated Pilot Examiners (DPEs) across the country that are charged with applying the Practical Test Standards (PTS) developed by the FAA towards pilot certification for individual applicants. Each of these examiners has the responsibility of utilizing the PTS to strictly deliver a practical exam. It is difficult for examiners to be impartial in this process. Furthermore, stable and qualified pilots occasionally have flights outside tolerances allowed by the PTS while mediocre pilots may have flights in which they meet standards only for the duration of the exam. A DPE only has the opportunity to observe one flight for which they can base their determination of the satisfactory or unsatisfactory outcome for pilot certification. Digital parameters provided by a data collection system combined with the qualitative inputs from the instructor will provide flight instructors with more information in not only the training process but for determining whether a particular student is statistically prepared for a practical exam. Most importantly data collection will provide the DPE with a trend of information over multiple flights where a more rounded picture can be analyzed with a statistical probability and reliability of future performance. This paper will provide an initial analysis as to the feasibility of utilizing data collection systems as a mechanism to radically change the process during which pilots are trained and certified.

This study will use digital flight recorder data to compare commercial pilot students' historical performance of the steep turn maneuver during training flights to their performance during a phase check course completion exam and to post exam performance. The primary research question that this paper addresses is: Is there a correlation between students' historical performances of a maneuver and their performances during flight tests? A secondary research question is: Is there a correlation between human grading and evidence-based grading of a maneuver on students' flight tests?

Literature Review

A pilot flight test is designed to ensure that an applicant meets the standards for knowledge and skills listed in the Practical Test Standards (PTS) for the certificate or rating sought (FAA, 2010; FAA, 2002) which are based upon Part 61 of Title 14 of the Code of Federal Regulations. The practical flight exam, combined with a written test and flight instructor recommendation, is the final demonstration of ability in pilot certification. The role of the DPE is to subjectively evaluate a pilot's competency through a scenario-based practical test, with emphasis on Single Pilot Resource Management (SRM) skills. The scenario should be designed to include as many operational tasks of the PTS as possible while accurately replicating the implications of a realistic flight. This allows the examiner an opportunity to evaluate the applicant under realistic flight circumstances, and determine if he/she is capable of safe airmanship in the future. PTS tasks that cannot be incorporated into the scenario, such as steep turns shall be accomplished during a break or at the end of the scenario.

The demonstration of the aeronautical skill portion of a flight exam represents the completion of tasks, such as steep turns to evaluate an applicant's technical flying ability (FAA, 2010). The objectives and tolerances for the technical maneuvers are listed in the PTS (FAA, 2002). For example the steep turns maneuver shall consist of a coordinated 360° steep turn with at least 50° of bank, followed by a steep turn in the opposite direction at the recommended airspeed, not to exceed maneuvering speed (Va). The applicant shall divide attention between airplane control and orientation, and exhibit knowledge of the elements related to steep turns. The tolerances are to maintain the entry altitude, ±100 feet, airspeed, ±10 knots, bank, ±5°, and to roll out on the entry heading, ±10°. Unsatisfactory performance is to be determined by the examiner if the applicant consistently exceeds the tolerances or fails to take prompt corrective action when tolerances are exceeded. The DPE must subjectively evaluate whether or not an applicant consistently performs within the PTS tolerances, and takes prompt corrective action if deviations from the tolerances occur.

Subjective evaluations will always remain a part of pilot evaluation (Meister, 1999). It is assumed that all performance measurement will have some level of subjectivity. All methods of measurement and evaluation are subjective at some point in the development of any grading metric. A useful performance measurement metric should be valid, reliable, and quantitative while minimizing intrusiveness. Operationally collected data is preferred to maintain these qualities.

Evidence-based evaluation techniques have been utilized in many industries to improve the reliability and validity of performance measurement (Carpenter, Kane, Carter, Lucas, Wilbur, & Graffeo, 2010). The medical industry has already begun to use evidence-based evaluations in human skills performance testing (Kirby, Numnum, Kilgore, & Straughn, 2008). Gynecology residents at a teaching hospital undergo simulator training in laparoscopic surgery. Laparoscopic surgery is a minimally invasive surgery utilizing a camera inserted through a small incision to monitor the surgical instruments. The simulator training involves a series of six tasks to develop the residents' dexterity. The simulator grades each resident on the ability to perform the tasks within specified tolerances.

Early aircraft flight data monitoring and evaluation began in the aviation industry with British Airways and TAP Air Portugal in the 1960's adapted by the Flight Safety Foundation and the FAA in the 1990's (Lacagnina, 2007). The FAA implemented the Flight Operational Quality Assurance (FOQA) program in 2004. FOQA is designed to be a voluntary, non-punitive safety program that allows commercial airlines to monitor and share de-identified recorded flight data (FAA, 2004). Monitored data trends are evaluated to determine operational risk issues that may lead to potential safety hazards. The key to the program is that corrective actions to unsafe operations are applied and maintained. Proactive action to potential hazards can have significant impacts to aircraft accident reduction.

Flight data monitoring is not limited to commercial aviation. The Flight Safety Foundation began a three year demonstration project called Corporate Flight Operational Quality Assurance (C-FOQA) involving 27 business jets operated by two companies (Lacagnina, 2007). The two operators reported interest and benefit to the data analysis.

FOQA programs target operational trends of the entire flight department within an organization. Another direction would be to evaluate individual pilots, and determine exactly that individual's strengths and weaknesses. The Advanced Qualification Program (AQP) represents a data-driven system to evaluate individual performance (FAA, 2006). AQP is a voluntary program that offers an alternative method for the certification of aviation personnel. AQP training programs replace programmed flight hours with proficiency-based training and assessment. There is an emphasis on scenario based training and evaluation. The program utilizes evidence-based quality control and allows for flight data to be used in performance assessment.

AQP's alternate means for evaluation allows for evidence-based pilot certification. With the increased availability of General Aviation Flight Data Monitoring (GA-FDM), improved reliability and validity of flight tests are available to Part 91 general aviation flight organizations (Lau, 2007; ATA, 1998). An individual's collective flight record can be a more accurate indicator of future performance than a single flight test. Flight data is limited to measuring technical performance and would be difficult replace an examiner's judgment of human factors, SRM, ADM, and overall airman competency. Future application of this research may allow an examiner to use a student's entire performance history, to supplement or replace portions of a flight exam.

Methodology

This study examined commercial pilot students (n=13), training in Cirrus SR-20 aircraft, in a collegiate flight training program. An analysis was performed of each student's performance of the steep turns maneuver before, during, and after a course completion phase check. An automatic computer grading of digital flight recorder data provided an objective measurement of performance. The flight recorder data was collected from the Garmin G-1000 Perspective avionics system installed in the Cirrus aircraft.

Each student's pre-phase check performance will be the average of their ten most recent repetitions of the maneuver before their phase check. The post-phase check performance will be the average of each student's first five repetitions of the maneuver after their phase check. Four Pearson-R correlation tests were performed to compare the students' entire historical performance, their performance before the phase check, their performance during the phase check and their performance after the phase check. A fifth Pearson-R correlation test was performed to compare the objective grading metric to the subjective grades issued by the phase check examiners.

Each maneuver was objectively graded as the average normalized deviations in altitude, bank angle, and airspeed per second. The grading metric was based on the Commercial PTS tolerances for the steep turns maneuver. A score was calculated for each second of the maneuver by summing one point per foot of altitude deviation, ten points per knot of airspeed deviation, and twenty points per degree of bank deviation. These weights were applied to normalize the score such that a full deviation to the PTS tolerances of any one parameter would result in a score of 100 points. A 100 point penalty was assessed for any instance the pilot failed to roll out on the entry altitude within the tolerances. This one time penalty is assessed because a failure to roll out on entry heading is a pass/fail tolerance, indicating a failure of $1/4^{th}$ of the maneuver and penalized as $1/4^{th}$ the score. Level flight in this grading metric is considered any bank less than 10°. The bank parameter is only measured when the aircraft heading is more than 25° (half the bank angle for roll transitions) from the entry heading, excluding the transition period from level flight to the established steep turn. A score of zero would indicate zero parameter deviations throughout the maneuver and therefore considered perfectly performed. An average normalized score of 400 would indicate a maneuver performed entirely at the tolerances established by the PTS standards. Any score in excess of the 400 would indicate a maneuver performed largely outside of all of the tolerances of the maneuver. The subjective grade issued by the phase check examiner was based on the examiners in-flight observations of altitude, airspeed, bank, and heading control. A perfect performance of the maneuver would receive a score of 20. Any deficiencies in performance during the maneuver would result in a reduction of the score from 20.

Results

The scores for each of the participants are included in Table 1. There was no correlation between the phase check performances and each student's historical data, including pre-phase check and post-phase check maneuvers ($r = -0.07$). There was no correlation between the pre-phase check performances and the phase check performances ($r = 0.08$). There was no correlation between the post-phase check performances and the phase check performances ($r = -.01$). There was a large correlation between the pre-phase check performances and the post-phase check performances ($r = 0.73$). There was also a large correlation between the objective phase check scores from the data analysis and the subjective phase check examiner scores ($r = -0.61$).

During the phase check, 69% of the students performed the maneuver better than their historical performance average, including pre-phase check and post-phase check repetitions. 85% of the students performed better than their pre-phase check average scores and 62% of the students performed better than their post-phase check scores. 62% of the students showed improvement in the post-phase check scores when compared to their pre-phase check scores.

Table 1

Average Steep Turns Maneuver Scores.

Subject	Pre-Phase Check Phase	Check	Post-Phase Check	Combined Pre/Post	Subjective Grade
1	219 157	186		208 16	
2	192 91		153	179 18	
3	198 176	147		181 18	
4	195 139	164		185 19	
5	210 89		194	205 19	
6	187 102	188		188 19	
7	268 263	146		228 16	
8	184 179	155		174 18	
9	200 218	205		202 15	
10	203 74		201	203 19	
11	164 239	194		176 17	
12	182 124	260		195 16	
13	355 141	458		372 16	

Discussion

The large correlation between pre-phase check performance and post-phase check performance indicates that the historical flight data provides an accurate indication of future performance on the steep turn maneuver. The results indicate that the phase check performance did not reflect either the historical or the future performances of the steep turn maneuver. Since good performance is indicated by a low score on the digital data evaluation and a high score on the subjective examiner evaluation, the large negative correlation indicates the validity of the digital data evaluation for measuring steep turn performances. The students' performances indicated that overall improvement had occurred from the pre-phase check performances to the post-phase check performances. The phase checks themselves however yielded better performances than either the pre-phase check and the post-phase check averages. It is important to mention that the amount of pressure to perform that the pilot is experiencing is different during some of these assessment periods. During pre-phase check and post-phase check the amount of pressure to perform by the pilot would be similar, but during the phase check the amount of pressure to perform is higher due to the fact that the pilot is being graded by a perceived superior. 40% of the final grade in the flight course is based solely on the performance of the pilot on this one flight. Even during the times that a flight instructor is observing the pilot during the pre-phase and post-phase checks there is a lower level of pressure because of the small impact a negative performance would yield. Performance of the steep turn maneuver critical for a good final grade so it is possible that the participants were paying a higher level of attention during the performance of each maneuver and therefore yielding a better performed maneuver. This could account for the discrepancy on the correlations during the various flight periods. The amount of pressure to perform on a phase check and during a practical exam for pilot certification would be similar, but this study did not compare the performance on the phase check versus a practical exam.

This study did not attempt to define quantitative standards for passing or failing a maneuver. Further research may use a larger population for an extended period of time to establish qualitative grading standards. This study also did not take into consideration students' attempts to correct for performance deviations. The Practical Test Standards do allow for an applicant on a checkride to exceed a limitation if prompt corrective action is taken to return the aircraft to within standards. The definition of "prompt corrective action" is determined by the particular Designated Pilot Examiner giving the practical exam. Further research should solidify an acceptable measure of "prompt corrective action" and then apply a measured time it takes for pilots to react and apply corrections to their deviations to be utilized in the evaluative process.

Summary and Conclusions

The question needs to be asked, "What is the end of course completion phase check (or pilot certification flight test) supposed to assess?" The results from this study indicate that the performance on the phase check for the steep turn maneuver is different than it is before or after the phase check. In all but three cases the performance during phase check was better than during the pre-phase and post-phase sessions. With the low number of pilots that were evaluated it will take more information before it can be surmised that a pilot will perform better during phase checks than on normal flights, but that is the initial finding.

Lastly, pilot certification is limited to a single subjective evaluation of technical performance and personal judgment by a Designated Pilot Examiner. The practical flight exam is designed to be a scenario-based evaluation of airman competency and a one-time indicator of a pilot's performance and safety. It is difficult for a single subjective evaluation of an individual's performance to be a valid and reliable indicator of actual level of competency. Evidence-based evaluations of recorded flight data offers an alternative method to enhance the adequacy of pilot training and evaluation. Future application of this research may allow instructors to statistically analyze whether or not a student is competent to perform technical maneuvers such as steep turns before a flight exam leading to the possibility of an AQP based training program being implemented in collegiate flight training. Furthermore, Designated Pilot Examiners may be able to use a student's entire performance history to supplement or replace portions of a flight exam. Overall this study provides evidence supporting a shift towards evidence-based pilot training and evaluation.

References

ATA. (1998). Advanced qualification program: Data management guide. Retrieved from http://www.faa.gov/training_testing/training/aqp/library/media/dmg.pdf

Carpenter, C. R., Kane, B. G., Carter, M., Lucas, R., Wilbur, L. G., & Graffeo, C. S. (2010). Incorporating evidence-based medicine into resident education: A CORD survey of faculty and resident expectations. *Academic Emergency Medicine, 17*(10), S54-S61

FAA. (2010). Order 8900.2 CHG 1: General aviation airman designee handbook. Retrieved from http://fsims.faa.gov/wdocs/orders/8900_2.htm

FAA. (2006). Advisory Circular 120-54A: Advanced qualification program. Retrieved from http://rgl.faa.gov/Regulatory_and_Guidance_Library/rgAdvisoryCircular.nsf/list/AC%20120-54A/$FILE/AC%20120-54a.pdf

FAA. (2004). Advisory Circular 120-82: Flight operational quality insurance. Retrieved from http://rgl.faa.gov/Regulatory_and_Guidance_Library/rgAdvisoryCircular.nsf/8ce3f88c034ae31a85256981007848e7/40c02fc39c1577b686256e8a005afb0a/$FILE/AC120-82.pdf

FAA. (2002). FAA-S-8081-12B: *Commercial pilot: Practical test standards for airplane*. Washington, DC: Flight Standards Service.

Kirby, T. O., Numnum, T. M., Kilgore, L. C., & Straughn, J. M. (2008). A prospective evaluation of a simulator-based laparoscopic training program for gynecology residents . *J Am Coll Surg, 206*, 343-348

Lacagnina, M. (2007, August). C-FOQA takes root: Foundation introduces a powerful corporate aviation safety tool. *Aerosafetyworld, 11-15.*

Lau, S. K. (2007). General aviation flight data monitoring. Retrieved from http://www.ga-fdm.com/pdf/GA%20FDM%20White%20Paper_v3%282%29.pdf

Meister, D. (1999). Measurement in aviation systems. In J. A. Wise, V. D. Hopkin, & D. J. Garland (Eds.), *Handbook of aviation human factors* (pp.33-49). Boca Raton, FL: Taylor and Francis Group, LLC.

CAN VIDEO GAMES BE VALID PREDICTORS OF SUCCESS FOR THE PRIVATE PILOT PHASE CHECK?

Laura E. Bourguignon, Donald A. Petrin, Brian G. Dillman, John P. Young
Purdue University
West Lafayette, Indiana

A study was conducted at a Midwestern university to determine whether or not video game performance can be an accurate predictor of success on the private pilot phase check. In this study, there were thirty-three potential participants enrolled in AT 145, the private pilot flight training course. Twenty-three (70%) students agreed to participate. These volunteer participants were asked to complete a video game (Super Mario Galaxy). The final score and time to finish a pre-selected level were recorded. These scores were correlated with the aggregated scores that the participants earned during their private pilot phase checks. Although a survey of literature suggested there might be a correlation between video game performance and pilot performance, the results of this study showed no significant correlation.

Collegiate flight training requires certain skills and abilities for students to be successful within the normal timeframe and expense of training. There are two different sets of competencies: motor skills and cognitive skills (Rao, 2005). A motor skill is the ability to make purposeful movements to perform a task. A skilled movement is the result of a combination of force, velocity, accuracy and purposefulness (Kent, 2006). On the other hand, cognitive skills are the mental processes used to think, analyze, and learn. Cognitive skills can be classified into different categories (Cognitive Definition, 2010):

- Processing speed refers to the efficiency the brain processes data.
- Auditory processing refers to skill of processing sounds.
- Visual processing describes the skill of receiving and manipulating visual data.
- Memory can be categorized into short and long term. Both are important parts of the learning process.
- Logic and reasoning are two skills required to compare data. These are particularly important for problem solving and planning.
- Attention skills refer to the ability to stay focused on a task while discarding irrelevant information.

Cognitive skills are important in a number of tasks, including flying, and therefore are vital to the future of a private pilot. However, the path to becoming a private pilot can be costly and time-consuming, especially for the students who might encounter learning plateaus during the training. Although all difficulties cannot be attributed to the lack of cognitive skills, cognitive deficiencies can be a factor. Many collegiate admissions offices currently use high school transcripts and Grade Point Average (GPA) as the main tools for selection of incoming students into flight education programs. Since there may be wide variability between high schools in determining what level of learning outcomes deserve an “A” or “B”. etc., the use of grades solely as a success predictor can be inconsistent. Clearly, it would be extremely useful to have additional metrics for predicting success. These additional selection tools might improve both the quality of admitted aviation professional flight students and increase their success rate.

Background

One aspect of cognitive skill crossovers was investigated by Hambrick, Oswald, Darowski, Rench, and Brou (2009). The authors investigated the determinants of success in a task designed to simulate multi-tasking, common to many occupations. The test subjects were given tests of working memory capacity and processing speed. The researchers were interested in their previous experience with video games. The results of this study uncovered a strong relationship between video game experience and multitasking, showing that multitasking is a trainable skill. The authors also found that individual differences in strategy accounted for a large proportion of the variance in multitasking.

Other aspects of cognitive skills involve visual processing. Green and Bavalier (2003) studied the underlying causes of the enhanced visual processing capabilities of action video game players. They measured the spatial resolution of a subject's visual field using the smallest distance a distractor could appear from a target and still be identified by the subject. The difficulty to differentiate between objects close together is known as crowding. The results suggested that video game players had an increased tolerance to crowding. The research also found that non-video-game players could enhance their visual resolution with training. This study provided a causative relationship between video-game play and augmented spatial resolution. As mentioned in Green and Bavelier (2003), it is possible that by selecting avid game players, the researchers chose a population that would have superior visual skills even without the difference in training.

More recently, hand-eye coordination has been a topic of interest for research focused on the benefits of video games. According to Green and Bavalier (2004), video games improve not only hand-eye coordination, hasten reaction time, and benefit peripheral vision, but also affect visual learning. The ability to infer the three-dimensional structure of an object from a two-dimensional representation is also critical for certain professions such as architects, engineers, pilots, mechanics and many more. These findings demonstrate that video game players can process a rapid stream of visual information with increased efficiency as compared to non-players.

Two years later, Green and Bavalier (2006) also conducted two studies to investigate the relationship between action video gamers and the ability to track multiple objects. Both groups, video game players and non-video game players, had a similar near-perfect performance until the number of objects became too great, at which point both groups' performances dropped drastically. However that critical number for the video game players was higher than for the non-video game players. Green and Bavalier found that the difference was due to an enhancement in the process of counting. A secondary study was set up to confirm that video games help in the processing of multiple objects at one time. Video game players were also able to allocate their attention to more items than non-video game players. The conclusion of the studies was that video game playing enhances the number of objects that can be apprehended due to changes in visual short-term memory skills.

Many studies have identified the cognitive skills acquired by video games, but few have investigated the relationship between video games and flying abilities. A study was accomplished at the school of the Israeli Air Force to test the transfer of skills from a complex computer game to flight (Gopher, Weil, & Bareket, 1994). Gopher et al. compared training strategies of two groups: one focusing on the specific skills required in performing the game, the other to help cope with the high attention load of the flight task. The two groups received 10 hours of training, while a third group received no training or game experience. The two trained groups scored higher in the following test flights than the untrained group. With large group of participants this study illustrated the benefits of video gaming to improve flight training. However, the results might not be applicable to U.S. civilian or military flight training.

Another way to consider the use of video games is the study of Mautone, Spiker, and Karp (2008). This study focused on improving training via the use of video game technology. A specific game was developed to train pilots on the use of the flight management system (FMS). The test subjects were students from Arizona State University's flight training program. Two groups were randomly set up: half of the subjects were given game-based training (GBT) FMS instruction, while the other half received standard computer-based training (CBT). Access to the training material was automatically recorded in a database. A test was then given to both groups to test the transferability of the training to an actual FMS. The results showed that the GBT group scored higher on the test and retained the knowledge better than the CBT group. An analysis of the training database also indicated that the GBT group accessed the training material more than the CBT group.

Video gaming has many positive outcomes, such as improvements in visual processing, short-term memory skills, quickness of reflexes, eye-hand coordination, counting process and target identification, and attention-management skills. Some of these previously discussed studies also showed better performance in flight training and check ride ratings for video game players than non-video game players. While some of these advantages might not be useful for some careers, clearly they would be of value to pilots in general aviation, the airlines, or military. All of these studies imply that there is a measurable correlation between video game performance and skills needed for flight. This paper aims to determine the possible impact of video game skills on flight training candidates, as a possible predictor for private pilot training success.

Research Question & Hypothesis

Recently, researchers have studied the relationship between video game performance and cognitive ability. Many agree that gaming can boost a player's cognitive and attention skills (Gagnon, 1985; Green & Bavalier, 2003, 2004, 2006; Triplett, 2008). However, very few studies have looked for a relationship between gaming success and success in early flight training. These earlier investigative efforts, often outdated and sometimes inapplicable to modern video games, served as a catalyst for this research effort. As a result, the research question became: Can video games serve as a valid predictor of success for private pilot flight progress checks?

After developing the research question, it was necessary to choose an appropriate video game for the experiment. In order to find a video game that might mimic the cognitive skills needed for flying, the types of video games were investigated. Video games are usually classified into action, adventure, role playing, arcade, strategy, simulation, driving, and puzzle. From previous research, action video games stood out as the type of game with the most transferable skills to pilots (Achtman, Green, and Bavelier, 2008). In order to choose an appropriate video game for this study, the video game ratings created by the Entertainment Software Rating Board were consulted. Some action video games are rated Mature or Adults Only. However, to avoid any harm to the subjects, this research used a video game rated E for everyone. According to Nintendo, Super Mario Galaxy is a game rated for everyone with only mild cartoon violence (Super Mario Galaxy Wii, 2010). The game meets the standard of an action video game without the negative aspects of a violent game.

After choosing a video game, the completed research question of this paper became:
Is Super Mario Galaxy a valid predictor of success for the private pilot phase check in a university flight program? The null hypothesis of this research is that Mario Galaxy is not a valid predictor of success for the private pilot phase check. The alternative hypothesis is that Mario Galaxy is a valid predictor of success for the private pilot phase check conducted at Purdue University.

Methodology

Participants

Participants for this study included 23 Professional Flight students (2 women and 21 men) enrolled in AT 145 (Private Pilot Flight) at Purdue University during the fall semester 2010. A total of 33 students were enrolled while 23 volunteered. Of the 23 volunteers, three did not complete both parts of the experiment and were not included in the statistical analysis.

Measures

The video game Mario Galaxy was administered using a television set, game console, and a controller. During the video game, players must move the main character "Mario" through a 3D spatial world while defeating foes, collecting coins that will award extra lives, and collecting jewel shards to be used throughout game play. For the purposes of this study, the items collected (coins and shards) and time to complete a specific level of the game were recorded. At the appropriate time, the aggregate score for each participant on the private pilot phase check was also recorded. This flight progress check normally occurred toward the end of the semester and was evaluated by a full-time faculty instructor employed by the university.

Procedures

This experiment was advertised via electronic mail to all the students enrolled in the AT 145 class in the fall 2010 semester. All of the students willing to participate were included in this study. The experiment was organized in two different parts over the entire fall semester.

The first part consisted of completing the video game exercise near the beginning of their fall flight training. The participants were required to finish only one level of the video game. However, prior to data collection, they were allowed one practice session to become familiar with the basic operation of the game. Data was then recorded during the following level.

The second part of the study was completed upon completion of the private pilot phase check. When a student met all the requirements to take the private pilot phase check, the instructor recommended him/her for the exam. Purdue University employed eight full-time flight instructors who were responsible for grading the phase checks. The participating students were divided equally between each phase check instructor. The phase check is graded according to a standardized list of maneuvers and an established grade scale. Once the phase check was completed, the grade was recorded for each participant. If a student failed the phase check, they were required to complete additional training. Afterwards, they could try to pass the phase check during a second try. However, the student only received one grade for the phase check, regardless of the number of tries. The data recorded was statistically analyzed to determine whether or not there was a correlation between a high score on the video game and the phase check score.

Results

The results of this study were analyzed using the tool Statistical Analysis Software (SAS). The data was assumed to be a normal distribution. A QQ plot confirmed this assumption. Another graph confirmed the equal variance assumption. The video game variables (time, coins, and shards) were not found to be correlated to the score of the phase check:

- Time: p-value of 0.1248
- Coin: p-value of 0.4964
- Shard: p-value of 0.1157

The p-values suggested that time to complete the video game level and shards collected might be more correlated to the score received on the phase check than the coin variable. However, a significance level of $\alpha = 0.05$ was used to access the p-value. All the p-values were much larger than the significance level of 0.05. Therefore, the data does not provide sufficient evidence to conclude that the ability to play Super Mario Galaxy on the Wii -video game system is correlated to the score received during the phase check.

Discussion & Recommendations

At the completion of this study, a better understanding of video games as they relate to the cognitive skills needed for success on the private pilot phase check was acquired. While not statistically significant the p-value of time indicates an ability to correlate the ability to play Mario Galaxy Wii to the success of acquiring flight skills by slightly less than 88%. While this is not high enough to generalize and base future decisions on these facts, it does provide more information than is currently available. The cost of flight training is significantly high and while an inability to play Super Mario Galaxy would not preclude an individual from pursuing his/her hope of flight, it would give a person enough information to re-evaluate their decision to proceed. That being said, there was insufficient evidence to reject the null hypothesis. Had the null hypothesis been rejected, the video game could have been incorporated into the selection process. Additionally, if the video game was not used as a selection tool, it could have been used to identify student cognitive shortcomings in those individuals who might struggle to complete training. This would suggest matching them with more experienced flight instructors or offering them additional training. However, the results were not conclusive enough to make such sweeping recommendations.

However, some improvements could be made for future study. The low number of participants (23) made it difficult to obtain significant results. A larger number of participants is recommended for future research. In addition, the low number of female participants (2) prevented the researcher from examining the differences in scores between males and females. Although it might be a difficult task in a largely male-dominated environment, efforts should be made to increase the number of female participants. The research could be expanded to a larger sample size as well as other flight schools' students in the United States to determine if the results are generalizable to a larger population. Finally, a future study could also consider the correlation between video game performance to other flight courses, such as commercial, instrument, and multi-engine training.

Super Mario Galaxy was chosen for this study. This specific game includes multiple variables to assess performance (time, items collected, etc). This study suggested that shard and time were the only variables that might be correlated, however not significantly. This might indicate that all variables of a video game might not transfer to flight training performance. The choice of other video games should also be considered.

Although a survey of literature suggested there might be a correlation between video game performance and pilot performance, the results of this study showed no significant correlation. Further research, taking the authors' suggestions into consideration, could lead to different results and enhance the body of knowledge in pilot performance prediction.

References

Achtman, R.L, Green, C.S., and Bavelier, D. (2008). *Restorative neurology and neuroscience, 26*(4/5), 435-446.

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

Green, C., & Bavelier, D. (2003). Action-video-game experience alters the spatial resolution of vision. *Psychological Science, 18*, 88-94.

Green, C., & Bavelier, D. (2004). The cognitive neuroscience of video games. In L. Humphreys & P. Messaries (Eds.), *Digital media: Transformations in human communication* (211-224). New York, NY: Peter Lang Publications.

Green, C., & Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, *101*(1), 217-245.

Hambrick, D., Oswald, F., Darowski, E., Rench, T., & Brou, R. (2009). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology.* doi: 10.1002/acp.1624

Kent, M. (2006). *Oxford Dictionary of Sports Science and Medicine*. Oxford, United Kingdom: Oxford University Press.

Cognitive Definition. (2010). Retrieved August 3, 2010, from http://www.learningrx.com/cognitive-definition-faq.htm

Mautone, T., Spiker, A., & Karp, R. (2008). Using serious game technology to improve aircrew training. *The Interservice/Industry Training, Simulation & Education Conference, 200.*

Rao, K. (2005). Cognition and motor skills. In a Henderson, A., & Pehoski, C. (Eds.), Hand function in the child: foundations for remediation (pp. 101-116). Philadelphia, PA: Mosby.

Super Mario Galaxy Wii. (2007). Retrieved July 31, 2010, from http://www.nintendo.com/sites/supermariogalaxy/

Triplett, J. (2008). *The effects of commercial video game playing: A comparison of skills and abilities for the Predator UAV.* Wright-Patterson Air Force Base, OH: Air Force Institute of Technology.

AIRSPACE DECONFLICTION FOR UAS OPERATIONS

Lisa Fern
San Jose State University Research Foundation
Moffett Field, CA
Susan R. Flaherty & Robert J. Shively
U.S. Army Aeroflightdynamics Directorate, AMRDEC[1]
Moffett Field, CA
Terry S. Turpin
Turpin Technologies
Foster City, CA

Increased use of unmanned aerial systems (UAS) in combat zones has put mounting pressure on airspace operations. Interviews were conducted with military helicopter pilots, air traffic controllers, and UAS operators to better understand the current concept of operations for managing potential conflict between manned and unmanned aircraft during combat operations. Interviews with the UAS operators revealed limited situation awareness of the low altitude airspace picture. To address these issues, a graphical airspace display with basic conflict detection and alerting logic was developed. In simulation, this display was compared with a baseline textual display, derived from current operations as identified during the interviews. Operators controlled a single UAS in densely populated airspace and deconflicted with other aircraft while conducting a surveillance mission. Operators were evaluated on their ability to maintain separation assurance standards and deconflict with other aircraft. The graphical interface was found to improve operators' ability to maintain separation and perceived lower workload.

Due to ambiguities in the current low altitude airspace picture, airspace coordination procedures have evolved beyond published U.S. Army doctrine and field manuals. With the cooperation of the U.S. Army Unmanned Aerial Systems Training Center (TRADOC), Ft. Huachuca, AZ and Aviation Technical Test Center (ATTC), Ft. Rucker, AL, six UAS operators, five pilots, and six air traffic controllers with warfighter experience were interviewed on the planning and execution of a combat mission throughout all aviation mission phases. Structured interviews were recorded and summarized. Subjective ratings were collected from pilots and UAS operators to identify areas of perceived high workload during intense mission phases, when air-to-air conflict was most likely.

Interview findings were not intended to supersede any official concept of operations (CONOPs), but instead provide an understanding of the innovations and risks that the U.S. Army aviator faces when operating in the low level altitude airspace. Primarily, manned and unmanned aircraft are successfully deconflicted at HIDACZ airfields through horizontal separation, which channels vehicles down predictable flight paths. Predictability exposes low flying Army aviators to enemy targeting. As a matter of procedure, airfield throughput can be impacted when aviators are forcibly halted from arrivals and departures for UAS launches and recoveries.

Beyond the airfield, blanket altitudes, killboxes, and keypads are sufficient in separating manned and unmanned aircraft above the coordination altitude. mIRC chat is used effectively to communicate airspace clearances between controlling agencies and UAS operators. Discrete transponder squawk codes allow for identification, control, and deconfliction of aircraft by approach controllers. However, in aircraft without transponders (e.g., Raven UAV) or aircraft intentionally not transmitting position information below the coordination altitude, the risk of air-to-air collision increases. Below the coordination altitude, deconfliction is see-and-avoid using a Common Traffic Advisory Frequency (CTAF). Typically, small UAS operations and rotary wing aviators shared a common mission area and therefore were at risk for mid-air collision. Pilots are often unaware of small UAS daily operations because they are not listed in the Air Tasking Order (ATO) or Air Control Orders (ACOs).

---

[1] DISCLAIMER: Reference herein to any specific commercial, private or public products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the United States Government. The viewing of the presentation by the Government shall not be used as a basis of advertising.

Recommendations from the aviators, UAS operators, and ATS personnel were collected and outlined to mitigate these low altitude airspace risks. The recommended technical solutions for aircraft deconfliction were: 1) an active traffic alert and collision avoidance system (TCAS) that emits position information and receives position information from transponder-equipped aircraft; 2) low altitude approach control radar; 3) look-down radar from an airborne platform; 4) fully integrated, near-real time Blue Force Tracker in all aircraft to include allied forces and civilian support aircraft; 5) UAS operator airspace displays to include manned and unmanned aircraft position with conflicting flight path trajectory and active ROZs affecting the planned mission; 6) an encrypted mode of communication for UAS and pilot communities to speak directly to deconflict flight paths.

As a direct result of these interviews, a graphical airspace display with basic conflict detection and alerting logic was developed to increase the amount and availability of airspace information to the pilot. It was hypothesized that the graphics display would decrease the total number and duration of conflict events (as detected by the detection and alerting logic) with other aircraft. Similarly, it was expected that since conflicts should be detected sooner, reaction time to initiate a flight path change to deconflict would also be faster in the graphics display condition. Lastly, the authors hypothesized that operators would be able to deconflict from other aircraft faster in the graphics display condition since they would be able to actually see the location of the conflicting aircraft when making flight path changes (rather than having to approximate it's actual location from the chat messages). In addition to the two display conditions, an audio condition was introduced in order to examine the possibility of interaction effects when a simple auditory tone was presented to the operator with a caution warning alert. It was expected that the auditory alert would likely improve performance, particularly in reaction time, in both display conditions, however it was not known if the performance benefits would be equivalent across the text and graphics-based displays.

Method

Participants

Twelve pilots were recruited to participate in this study. All 12 participants were male, with an age range of 20 to 40 years of age (*M* = 29.08). Participants were required to hold, at minimum, an active Private Pilot License. Total flight hours ranged from 540 to 6038 hours (*M* = 2754.42). None of the pilots had any military flight or UAS control experience. Eligibility was limited to participants who were right-handed and had normal or corrected-to-normal vision.

Multiple UAS Simulator (MUSIM)

Ground Control Station Hardware. This simulation was generated with a quad-core CPU using an NVidia GeForce GTX 280 video card, and 2GB RAM. The monitor used was a 30" Apple Cinema Display, with a display resolution of 2560X1600 and 24-bit color.

Software. This experiment was run on the openSuse 11.1 Linux operating system. MUSIM has the following software dependencies: 1) OpenSceneGraph for graphics and 2) FLTK for graphical user interface.

Terrain Database. A visual database was created using Creator Terrain Studio 2.0.2 and Creator 2.5.1. Terrain imagery was obtained from U.S. Geological Survey satellite photography. The simulation utilized 30-meter elevation data with 45-meter texture data in the lower resolution areas and 0.7-meter texture data in the high resolution areas. Three designated areas within the database were utilized for this experiment. These areas can be characterized as dense, urban terrain, or medium density, industrial terrain.

UAS Flight Model. Four generic flight models were used to emulate the ownship and traffic in this simulation: 1) a Shadow 200 with an average cruise speed of 80-90 kts, and a speed range of 70-130 kts was used to simulate the ownship as well as tactical UAS traffic; 2) a Predator B model with an average cruise speed of 75 kts and max speed of 220 kts was used to simulate a general aviation fixed wing aircraft; 3) a C-130 model (modified to simulate a Boeing 747) with an average cruise speed of 200 kts was used to simulate commercial aircraft; and 4) a generic helicopter model with the same mass as an OH-6A and variable cruise speed capabilities was used to simulate VTOL aircraft.

Operator Interface. This simulation utilized a 1:1 operator/vehicle ratio user interface, consisting of one sensor view, a 2D top-down map (or tactical situation) display with gridded overlay and manual waypoint editing GUI, an Air Vehicle (AV) control panel, and a chat room window (see figure 1). An optical mouse was used for navigation of operator control panels in the operator interface.

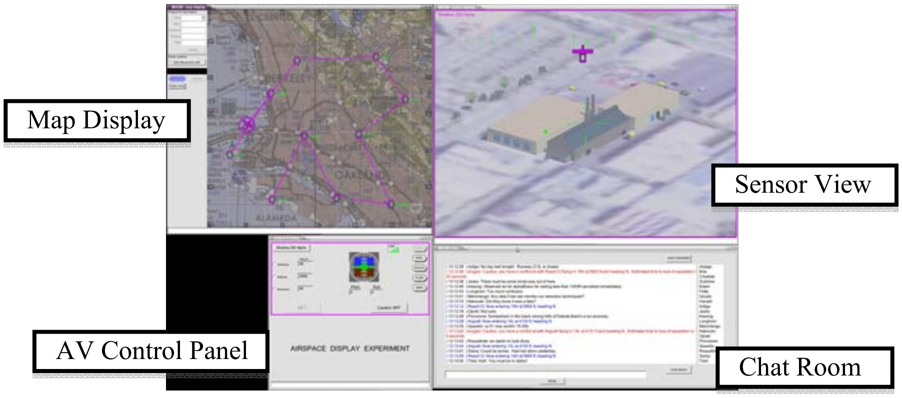


Figure 1. The MUSIM interface.

Payload and Cursor Controllers. The simulated sensor payload on each UAS was electro-optical only with three degrees of freedom (DOF) sensing capability, including 360deg pan capability and +45/-135 pitch limits. Zoom (y-axis) capabilities supported a progressive change in FOV from 2 to 16deg (x – 8x). Sensor slew rate was set at 60deg/second. The gimbaled sensor was operator-controlled via a 6-DOF 3Dconnexion SpaceExplorer input device. The SpaceExplorer utilizes 6-DOF sensing technology (x-, y-, z-axes, pitch, heading, and roll) in a pressure-sensitive device that requires right/left and up/down twisting motions to control the stare point of the payload.

Experimental Design

The present study utilized a within-subjects, repeated measures design to examine task performance and workload measures while conducting a reconnaissance mission and deconflicting with traffic in high density airspace. Two different airspace displays were compared across two audio alert conditions.

Airspace Displays. Two different airspace displays were developed for experimental comparison: textual and graphical. The textual was an Internet Relay Chat (mIRC) display that was designed based on current military UAS airspace operations as identified in the operational interviews. The graphical display was an integrated airspace display developed by overlaying graphical traffic information on the tactical map display.

In the text condition, operators were required to monitor the airspace mIRC chat room window located below the payload window (see Figure 1). All aircraft within the mission area of the UAS reported their location, via the chat room, as they transited grid locations. If the caution warning algorithm detected either a caution or warning level conflict, a message was sent to the operator from the air traffic controller, and repeated at 30 second intervals for as long as the conflict lasted.

In the graphical display condition, traffic and conflict information was presented to the operator on the tactical map display (see Figure 1). Non-conflicting aircraft were represented by white icons with rollover data tag information displaying aircraft identification (i.e., call sign), altitude and speed. Conflicting aircraft with 2 minute trajectory lines were displayed with color-coding according to alert level (e.g. yellow for caution, red for warning).

Audio alert. Two audio conditions were used for comparison: and off. In the audio alert condition, a simple tone was presented to the operator with the initial detection of all caution and warning level conflicts, regardless of display type

Caution Warning Algorithm. For this experiment, a predictive conflict detection caution/warning algorithm was developed based W.E. Kelly's work (1999), where conflict is defined as a "predicted violation of a separation assurance standard." The conflict algorithm uses instantaneous state vectors to calculate the closest point of approach and time remaining until separation standards are violated. If the closest point of approach is less than a stated minimum and time remaining to loss of separation is within a specified look-ahead window, then a conflict is declared.

Based on the Shadow Air Crew Training Manual separation standards, a cylindrical protection zone measuring 1000ft above and below with a 1000ft radius was defined around the ownship UAV. If another aircraft was within two minutes of entering the protection zone, based on instantaneous state vectors, a caution, predictive-level conflict was declared. If an aircraft breached the protection zone of the ownship, separation standards were lost and a warning, loss of separation-level conflict was declared. The algorithm was intended only to give conflict alerts, and not suggest conflict resolutions. Figure 2 illustrates the protection zone around the ownship UAS.

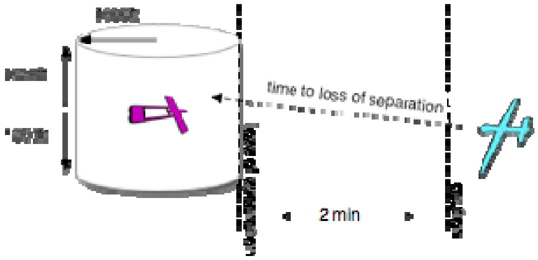


Figure 2. Shadow UAS separation assurance standards used to develop the caution warning algorithm.

Missions. Operators were tasked with flying a reconnaissance mission in an urban terrain scenario with 10 Named Areas of Interest (NAIs) and dense air traffic. The reconnaissance mission required the operator to set a flight path around the urban area in order to acquire imagery on each NAI. The ownship (Shadow 200) maintained a mission altitude of 6000ft and encountered aircraft of various types. One 10-minute practice session and eight 20-minute experimental missions were flown. Various target configurations at the NAIs and timing of conflict events were established. Fifteen conflict events with different aircraft types were programmed for each experimental mission. Operators' primary mission objective was to plan and fly a flight path around the 10 NAIs while maintaining safe separation from other aircraft. Deconfliction with other air traffic was managed by changing the ownship's waypoint headings. Operators were instructed that they were solely responsible for deconflicting with other aircraft (i.e. the aircraft would not change their own trajectories to move out of the way of the UAS). Their secondary mission objective was to collect intelligence, by identifying the number of targets at each of the NAIs.

Procedure

All participants were required to fill out an informed consent for minimal risk form and demographic survey intended to elicit information regarding participants flying experience and computing/gaming experience. Participant were then given a pilot briefing describing the MUSIM environment and the mission requirements. After the training material was presented, pilots were given the opportunity to fly MUSIM with no traffic in order to familiarize themselves with the simulation. Once participants indicated feeling comfortable with the simulation, they started the experimental sessions.

The experimental sessions were blocked and counter-balanced by display type so that participants received both the audio and non-audio conditions for one display type before being exposed to the second display type; the audio condition was also blocked within the display type. At the beginning of a testing block, the participant was trained for that particular display block (either mIRC or graphics), followed by a 10-minute practice mission. The practice mission was then followed by three 20-minute missions each for the audio and non-audio conditions for a total of six experimental missions per display and 12 for the entire session. Following each auditory alert block, the participants were asked to fill out a NASA-TLX workload rating form (Hart & Staveland, 1988).

Data Collection

Primary task performance. The primary task for this experiment was maintaining separation and managing deconfliction with other aircraft. The average number and duration (in seconds) of both caution and warning level conflicts were recorded. Reaction time and time to deconflict were also collected. Reaction time was measured from the time that the conflict alert was presented to the operator until he initiated a flight path change in

MUSIM. Time to deconflict was defined as the time from this initial change input until the alerting logic stopped, indicating that there was no longer a conflict.

Secondary task performance. The secondary task for this experiment was to identify the number of targets at each of the 10 NAIs in a mission. The percentage (%) of NAIs visited and the accuracy (%) of target identification were collected for this task.

Workload. A NASA-TLX was administered to the participants after each experimental block. Participants rated six dimensions of workload (mental, physical, temporal, performance, effort, and frustration) on a five-point scale. A composite score of was also calculated.

Results

The data was analyzed using a 2 (display: text, graphics) X 2 (audio alert: on, off) repeated measures analysis of variance (ANOVA). Post hoc analyses utilized Bonferonni pairwise comparisons. There were no significant effects of display or audio the secondary task measures, and significant interactions between display type and audio condition were found across any measures.

Primary Task

Number of conflicts. There was not a main effect of display type on the number of predictive conflicts. The average number of predictive conflicts encountered in the mIRC display condition did not differ significantly from the number encountered in the graphics display condition. However there was a main effect of display type on the number of loss of separation conflicts, $F(1, 11) = 30.415$, $p < .001$. There were significantly more loss of separation events in the mIRC condition ($M = 2.3$; $SD = 2.1$) than in the graphics condition ($M = .6$; $SD = .9$) (see Figure 3). There was no main effect of audio alert on the number of predictive or loss of separation conflicts.

Conflict Duration. There was no significant difference in the duration of predictive conflicts between the mIRC and graphics display conditions. However, as shown in Figure 4, the duration of loss of separation conflicts was significantly longer in the mIRC condition ($M = 18.19$; $SD = 11.09$) than in the graphics condition ($M = 11.31$; $SD = 5.27$), $F(1, 11) = 43.504$, $p < .001$. There was not a main effect of audio on duration of predictive or loss of separation conflicts.

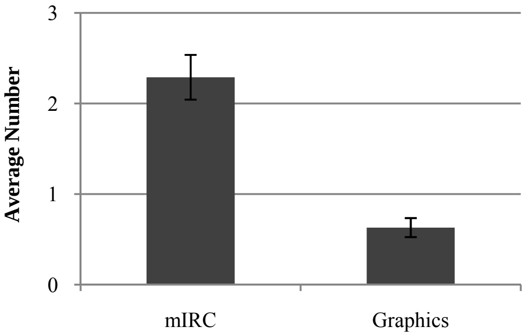


Figure 3. Number of loss of separation conflicts by display type.

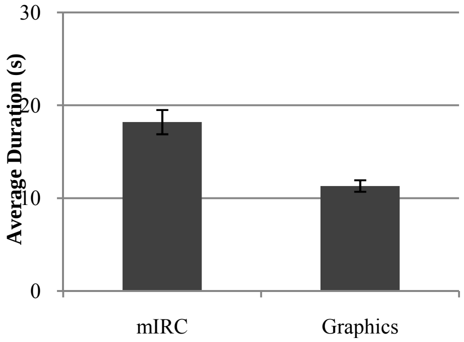


Figure 4. Average duration of loss of separation conflicts by display type.

Reaction time. There was a main effect of audio on reaction time to deconflict, $F(1, 11) = 12.213$, $p < .01$. Operators were significantly faster to react to conflict alerts when audio was present ($M = 17.53$; $SD = 8.19$) than when it was not ($M = 22.40$; $SD = 10.75$). There was not a significant main effect of display time on reaction time to deconflict from a conflict.

Time to deconflict. There was no main effect of display on the amount of time operators took to deconflict from other aircraft. Although operators appeared slightly slower with mIRC ($M = 20.13$; $SD = 15.77$) compared to graphics ($M = 15.26$; $SD = 6.74$), this difference was not significant. There was no significant main effect of audio on the amount of time it took operators to deconflict.

Workload

On the six dimensions of workload, there were significant main effects of display type on temporal, *F*(1, 11) = 4.569, *p* < .05, effort, *F*(1, 11) = 10.160, *p* < .01, and frustration ratings, *F*(1, 11) = 12.294, *p* < .01 (see Figure 4). Mean workload ratings on these dimensions were consistently higher in the mIRC condition (3.6, 4.1, and 3.1, respectively) than in the graphics conditions (2.9, 3.2, and 1.9). In addition, the overall composite score for workload was higher for mIRC (*M* = 3.4; *SD* = .7) than for graphics (*M* = 2.6; *SD* = .6), *F*(1, 11) = 11.097, *p* < .01. There were no significant main effects of audio on any of the six dimension or the composite score.

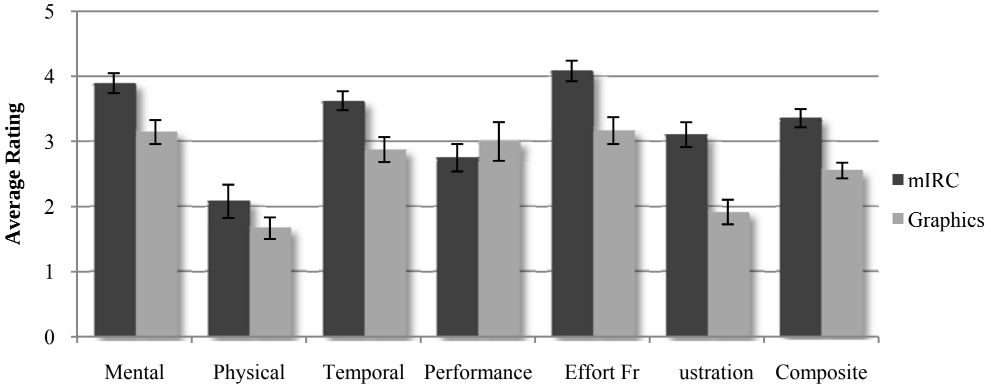


Figure 4. NASA-TLX ratings on the six workload dimensions and the overall composite score.

Discussion

The results of this experiment demonstrate the utility of a graphical airspace display to improve UAS operators' ability to maintain separation assurance standards and deconflict with other aircraft, while at the same time decreasing operator workload. This finding was evidenced by a reduced number of loss of separation events as well as a reduction in the duration of loss of separation conflicts. In addition, there were significant reductions in operator effort, time pressure and frustration when using the graphical display compared to the text display. Although there were no significant effects of display on reaction time to deconflict, time to deconflict or any of the secondary task measures, closer examination of the secondary task data reveals very high performance levels in both display groups. On average, operators visited 98% of the NAIs and reached 86% target identification accuracy across all missions. This data suggests that the secondary task was not difficult enough, and the lack of significant differences in performance measures may be a result of ceiling effects. Further evidence of this possibility is seen in the relatively low workload results; although there were significant differences in workload ratings, in general, workload scores were not very high.

Despite these challenges with the experimental design, the primary task and workload results give strong evidence to the need for a graphical user interface that presents airspace information to UAS operators, and serves as a basis for improving UAS operator airspace awareness both for military theater and civil operations. Follow on efforts will be needed to examine the effects of more challenging tasks, identify critical components of airspace SA, as well as explore different layouts and configurations that support optimal performance. Although mission profiles and requirements will differ, the results of this work will also inform new research on the design of user interfaces for civil airspace operations.

References

Ewing, L. (2009, June). National airspace system, here we come. *Unmanned Systems*, *27*(6), 22-25.

Hart, S. G., & Staveland, L.E. (1988). Development of a NASA-TLX (task load index): results of empirical and theoretical research. In P.S. Hancock & N. Meshkati (Eds.), *Human Mental Workload.* Amsterdam: North-Holland.

Kelly, W. E. (1999). Conflict detection and alerting for separation assurance systems. *Proceedings of the 18th Digital Avionics Systems Conference*, St. Louis, MO.

DECISION SUPPORT TOOLS AND LAYOUTS FOR MOMU (MULTIPLE OPERATOR MULTIPLE UAV) ENVIRONMENTS

Tal Oron-Gilad[1], Talya Porat[1], Jacob Silbiger [2] and Michal Rottem-Hovev[3]
[1] Human Factors Laboratory, Dept. of Industrial Engineering and Management, Ben-Gurion University of the Negev, Israel [2]Synergy Integration Ltd. and [3] Israel Air Force

The objective of this research is to design and develop tools, layouts, techniques and procedures to aid operators in handling Multiple Operator Multiple UAV (unmanned aerial vehicle) environments. In this paper we describe a study conducted on proficient operators, examining a new tool - 'Maintain Video Quality'. This tool aims to facilitate handoffs and UAV switching among operators. Experimental methodology and preliminary findings are discussed.

Multiple operators controlling multiple UAVs (MOMU) is an operational setup that has been shown to be beneficial for covering areas of interest, particularly in reconnaissance missions, and highly relevant to homeland security and surveillance operations. With the increase in UAVs' self-control tasks requiring less human execution, current UAV systems transit to one operator supervising a team of semi-autonomous UAVs, as opposed to the converse (Brzezinski et al., 2007). However, this mode of operation often increases the cognitive burdens of its operators. Besides the challenge of preventing high operator workload and low situation awareness, caused by the need to attend to multiple sources of information at once, this mode also requires switching of information sources, i.e., tasks, missions, video feeds or camera manipulations, and coordination among operators. Switching is a time-critical and cognitively demanding task. Cognitive costs of switching may be loss of orientation and situation awareness, increase in workload, and decrease in efficient verbal team communication. Consequently, switching between sources can disrupt operator performance (Draper et al., 2008). As the autonomy of the video feed source increases and interfaces improve, switch costs gradually become the bottleneck which limits the number of source feeds that a single operator can manage or be aware of (Hancock et al., 2007).

The aim of this entire research project, which is a US-Israel collaboration, is to identify information, and develop tools and layouts which may facilitate quick and efficient task switching and coordination in MOMU environments, in order to decrease switch-costs and improve mission performance. Previous studies (Porat et al., 2011) have shown that there might be a tradeoff between the screen layout and the number of zoom-related operations that the operator has to perform. As such, the Israeli research team aimed to develop tools that will facilitate a more optimal way for window size changes without affecting the zoom level. Specifically, in the study reported here, experienced UAV operators examined a new tool – the 'Maintain Video Quality' tool. This tool further explored the relationship between zoom operations, window size, layout of multiple video windows, and mission components (e.g., coverage area). In the next section we describe shortly the pre-study examining three different layouts, followed by the methodology and findings of our recent study examining the 'Maintain Video Quality' tool.

Layout Manipulation Preliminary Study

This pre-study dealt with manipulation of window size (Porat et al., 2011). When the operator controls or needs to be aware of multiple video feeds, it is possible that these feeds should be presented to him/her in a way that conveys their importance/relevance to the mission at hand. As such, three display layout configurations were examined: fixed, adaptive (automation-controlled) and 'user control' (see Figure 1). In the adaptive and user controlled layouts, the video feed window which is most in use (e.g., time on window, mouse clicks) enlarges on account of the other windows.

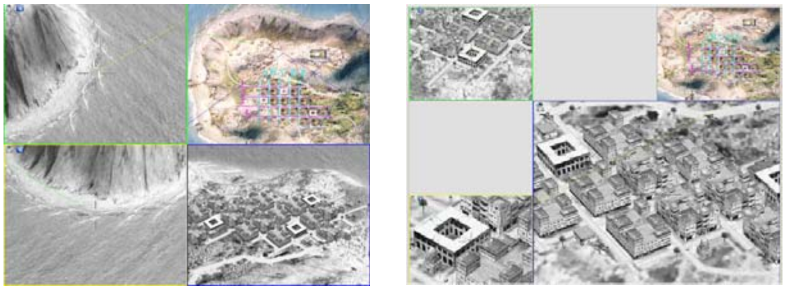


Figure 1. Fixed layout contained four same sized windows (Left). Adaptive/user control layout, one window enlarges on account of the other two (Right).

Preliminary findings revealed an interesting interaction between the zoom value (the total number of zoom operations) and the layout (F(2, 4)=11.74, p=.021). In the fixed layout, the values of the zoom were significantly higher than in the adaptive and user control layouts. This could imply that having a larger video feed window for the main task could reduce the need for zoom manipulations. Thus, there might be an interesting interaction between the video feed window size, necessity for zoom manipulations and desired target size. Since one of the goals is to reduce operators' workload, reducing the zoom manipulations needed by the operator may reduce the amount of workload he experiences while performing the mission. According to the above, the following "Maintain Video Quality" tool was developed to further examine the relation and interaction between the video feed window size, necessity for zoom manipulations and desired target size.

'Maintain Video Quality' Tool and Study

The Maintain Video Quality tool is a dynamic layout feature which manipulates the relationship between window size, zoom and field of view. It enables operators to define a minimum desired video quality, which is defined as window size (pixels) divided by footprint size (meters) (see Figure 2). The system will preserve this quality, as long as it can, by increasing the available window size or/and changing the zoom. The tool contains two sliders: Zoom value (on the left for display only) and Video quality value (to its right), which serves as an interactive slider. The operator defines the minimum video quality she/he is willing to absorb by clicking on the desired value (a yellow mark will be displayed). This feature is important in surveillance tasks, when the target needs to be seen continuously at a certain level of detail and therefore within a certain size.

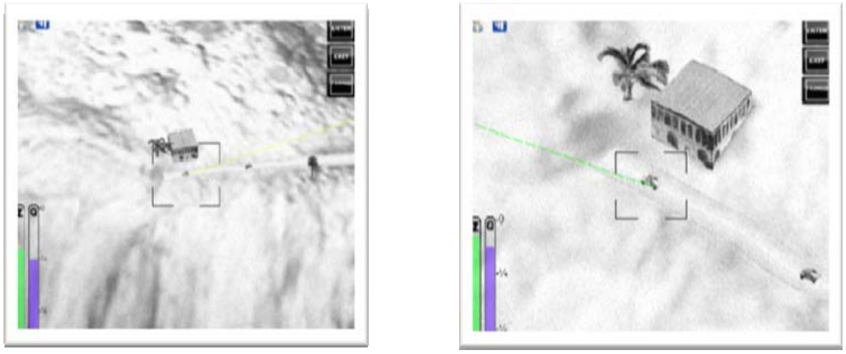


Figure 2. Maintain Video Quality tool. Video quality = 1/8 (Left). Video quality = 1/4 (Right).

Figure 3 shows several use-examples of the 'maintain video quality' tool. In the first example (Figure 3; top-left), the operator defines a minimum video quality constraint in the upper left window. While performing the task the user zooms-out in this window, an action which is threatening to break the defined constraint. The system automatically increases the window, to its maximum possible size (top-right). The second example (Figure 3; mid-left) shows that if a video is defined with a quality constraint for example on the bottom-right window and the size of the window decreases, which threatens to break the constraint, the system will zoom-in in order to maintain the constraint up to the possible maximum zoom-in (mid-right). The third example (Figure 3; bottom-left) shows the advantage of the tool when switching payloads. The work assumption is that the operator would like to view the target from a different point of view, but maintain the defined video quality. Therefore, when switching payloads, the system will make the necessary adjustments (zoom and window size) to the switched payload video, to maintain the desired video quality (bottom-right).

Methodology

Participants

Four highly experienced male UAV operators with similar skills and experience. Their age ranged from 25 to 28.

Operational Mission

To force switching of attention among payloads, the mission took place in two separate geographical locations: airport and city. The operator guarded a house and an airport utilizing 3 UAVs. The operator had to report upon the occurrence of 6 types of events, 5 events occurred in the city (vehicle exit, vehicle enter, vehicle framed, caravan exit, caravan framed) and one event in the airport (plane take-off). Events differed in their importance and caravan events were valued as most important. The aim of the mission was to identify maximum real events, ignore distracters and avoid false events.

Figure 3. Example 1 (top): user defines a quality constraint on the upper left video (yellow line) (left). Window grows to maintain quality constraint (right). Example 2 (mid): bottom-right window has a quality constraint (left). Bottom-right window size decreases, system automatically zooms-in to maintain constraint (right). Example 3 (bottom): The operator desires to switch between the top-left payload (green) to the bottom-right payload (blue) (left). The video quality of the blue payload (zoom and window size) changes to maintain the quality definitions (right).

Procedure

Figure 4 describes the study procedure for each operator in each layout. Operators commenced with the fixed layout and then proceeded to the dynamic layout. The total duration of the study for each participant was 3 hours (1 hour training and 2 hours experiment). Each scenario contained 18 events (11 real and 7 false) and lasted 15 minutes.

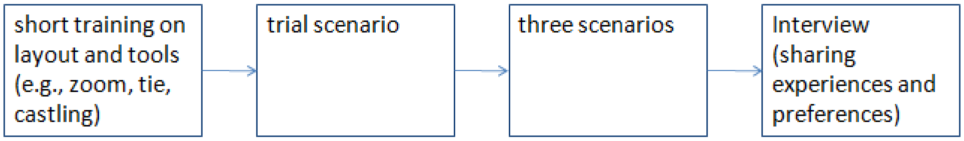


Figure 4. Study procedure for each operator in each layout.

Performance measures

Objective performance included success rate and detection time. Operational metrics indicating on the quality of payload utilization, zoom-value, double-clicks (manual following of a target), and locks of the payload on target were also collected. These measures evaluate the task-switching and payload manipulation efficiency. In addition, subjective assessment of workload was assessed using SWAT (Reid et al., 1988).

Results

Participants thought that the task was difficult and would normally not be performed by a single operator. This was also reflected in the relatively low Success rate - 62%. Preliminary results revealed no significant differences in success rate between the fixed (.63) and the dynamic (.61) layouts, in neither scenario nor per events analysis (see Figure 5). However, for both layouts, a learning curve was found and performance on the third scenario was always better than on the first and second ones. SWAT results also did not differ among layouts. Operators felt that the dynamic layout was not beneficial over the fixed layout. Operators were instructed to use the 'maintain video quality tool' as much as possible but were not forced to maintain it throughout the experiment. Overall 85% of the time the quality maintenance tool was on (hence, 15% of the time operators turned it off), and 50 % of the time, operators specified a different value than the default minimum value.

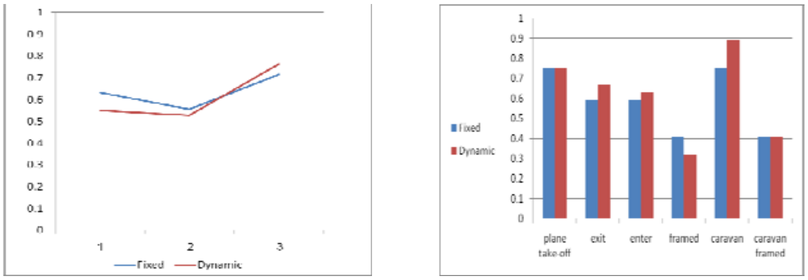


Figure 5. Scenario by Layout interaction for success rate (Left). Event by Layout interaction for success rate (Right).

Conclusions

The results indicate the need to further examine the utility of the 'maintain video quality tool', as this tool did not reveal superiority for the dynamic layouts in the examined scenarios. Future studies should examine tool features more specifically, while controlling layout and environment. To illustrate, in this study, the use of the 'maintain video quality tool' is inseparable from the dynamic layout and, perhaps, removing the dynamic change of window size, which operators felt was not beneficial in the current context, would have influenced the results and allowed the operators to utilize it more efficiently. In addition, the learning curve from the first to the third scenario could imply that operators need more time to learn and get used to such complex tools before improvements in their performance occur (as they also indicated in their subjective feedback).

Structured interviews with experienced operators strengthen the necessity and importance of layouts and tools in reducing operators' workload and improving mission performance. These studies mentioned here are only two of the many studies performed within the framework of this research project. The results of the study, although preliminary, revealed interesting concerns regarding fixed versus adaptive window size and the interaction between zoom and window size. Proposed future studies should, for example, examine automatic changes of zoom while switching or castling payloads (Porat et al., 2010); fitting the zoom to the task; examine the interaction of target size by zoom and window size and enable the operators to define manually the optimal window size for performing the task. Hopefully, results of the current and future studies will encourage researchers in the MOMU community to further develop decision support tools and layouts aimed to reduce operators' workload, increase situation awareness and improve mission performance, specifically for facilitating video-feeds switching tasks.

Acknowledgement

This work is part of the US/Israel MOA/FMF on Rotorcraft Aeromechanics and Man/Machine Integration under Task 24, Rapid Immersion tools/technique for Coordination and Hand-offs (RICH).

References

Brzezinski, A. S., Seybold, A.L., and Cummings, M. L. (2007). Decision Support Visualizations for Schedule Management of Multiple Unmanned Aerial Vehicles. Presented at the AIAA InfoTech@Aerospace, Rohnert Park, CA, May, 2007.

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. Retrieved from http://conferences.telecom-bretagne.eu/.

Hancock P.A., Mouloua, M., Gilson, R., Szalma, J., and Oron-Gilad, T. (2007). Is the UAV control ratio the right question? Ergonomics in Design, 15 (1), 7, pp. 30-31.

Porat, T., Oron-Gilad, T., Silbiger, J., and Rottem-Hovev, M. (2010). 'Castling Rays' a Decision Support Tool for UAV-Switching Tasks. CHI 2010 Conference Proceedings, Atlanta, April 10-15, 2010.

Porat, T., Oron-Gilad, T., Silbiger, J., and Rottem-Hovev, M. (2011). Switch and Deliver: display layouts for MOMV (Multiple Operators Multiple Video feeds) environments. IEEE CogSIMA 2011 Conference Proceedings, Miami, February, 2011.

Reid, G. B. and Nygren, T. E. The subjective workload assessment technique: a scaling procedure for measuring mental workload. In P. A. Hancock and N. Meshkati (eds.), Human Mental Workload (1988), 185 - 218.

IDENTIFYING TRAINING GAPS IN RQ-7B SHADOW: A U.S. ARMY UNMANNED AIRCRAFT SYSTEM

John E. Stewart

U.S. Army Research Institute, Ft. Rucker, AL, USA

Martin L. Bink

U.S. Army Research Institute, Ft. Benning, GA, USA

The mission of the RQ-7B has changed radically since 2003, when unmanned aircraft system (UAS) ownership shifted from Military Intelligence (MI) to Army Aviation. Instead of passive observation, RQ-7B operators must now acquire active scout-reconnaissance skills. Initial training takes place at Ft. Huachuca, AZ, the Army's main MI installation. Operators then report to their unit, usually a Brigade Combat Team (BCT). This research focused on (a) type training received at the schoolhouse (MI/scout-reconnaissance), (b) training received at the BCT, and (c) opportunities for training. It was found that schoolhouse training still was primarily MI, (e.g., image analysis and vehicle identification). Interviews with BCT staff officers identified 10 critical scout-reconnaissance skills not trained in the schoolhouse. These required additional training, mostly on the job in the unit, because opportunities for scout-reconnaissance training at home station were limited. The research concluded that more scout-reconnaissance training should take place at the schoolhouse.

The RQ-7B configuration of the Shadow is a 380 lb aircraft, with a 14 ft wingspan. It is powered by a 38 hp Wankel rotary engine, can operate at altitudes up to 14,000 ft., has a range of over 200 mi, and an endurance exceeding 5 hr. As part of the shift of its mission to aviation, RQ-7Bs are being retrofitted with laser target designators to supplement the electro-optical and thermal vision devices that comprise the mission payload (U.S Army Roadmap for UAS, 2010-2035, p. 76). Shadow currently is a Brigade level asset and located within the Special Troops Support Battalion, Military Intelligence (MI) Company of both light and heavy Brigade Combat Teams (BCT).

The original mission of the RQ-7B Shadow UAS was intelligence, surveillance, and reconnaissance, (ISR), in which the air vehicle operator (AO) and mission payload operator (PO) proceed to predetermined coordinates, observe and report on activities within this area, and await instructions to proceed to another location or return to base. Since 2003, the mission has changed considerably to one of scout-reconnaissance more similar to missions flown by manned Army scout helicopters, such as the OH-58D. The scout-reconnaissance mission differs from ISR in that the AO and PO play a more active role in its execution, for example, they may be directed to actively search on either side of a roadway to determine if a convoy's route is free of potential threats. If threats are detected, they are reported to the Tactical Operations Center (TOC), which relays the information to a ground commander or the crew of an OH-58D armed helicopter. For manned-unmanned operations, the RQ-7B operators could report the target coordinates directly to the crew of the OH-58D. In short, the RQ-7B is no longer a data collection platform, but a pilotless scout aircraft.

The current research effort was initiated at the request of the Training and Doctrine Command (TRADOC) Capabilities Manager-UAS (TCM-UAS). That is, it was unclear at the outset to what degree RQ-7B aircrews were trained to execute scout and reconnaissance missions. With the emerging scout-reconnaissance role of RQ-7B, TCM-UAS was convinced that UAS operators must now be trained to think and act like scouts, whose role is active observation, interpretation, and reporting of potentially hostile activities in contrast to collecting ISR data for later imagery analysis. Given the apparent gap between the skills required to execute scout-reconnaissance missions and RQ-7B training, the current research effort was conducted to determine what aspects of the scout-reconnaissance mission could be performed by RQ-7B aircrews, and to identify how and where RQ-7B operators could be trained on the identified scout-reconnaissance skills. Little is known about the efficacy of scout-reconnaissance training for UAS aircrews. The bulk of the research literature has addressed training in the context of aircrew errors and their contribution to UAS mishaps (e.g., Thompson, Tvaryanas, & Constable, 2005), rather than tactical skills training.

A preliminary assessment of RQ-7B aircrew collective training, at the Joint Readiness Training Center (JRTC) at Ft. Polk, LA (i.e., Stewart Barker, & Bink, 2010) provided an inkling that collective training needs existed for the RQ-7B. The aggregate perspective of senior training personnel was that there were few opportunities for RQ-7B teams to train for scout-reconnaissance missions, with most missions at JRTC tending to be ISR. That research also underscored the need for a closer and more detailed look at both schoolhouse training and the actual employment of the RQ-7B in the BCT, while in theatre, and at home station.

Method

Identification of Scout-reconnaissance Skills for Cavalry Scout and OH-58D Pilot

The first step was to identify possible scout-reconnaissance skills for RQ-7B operators. To do so, programs of instruction (POI) for two Army career specialties (i.e., Cavalry ground scout and OH-58D Common Core and aircraft qualification Track Course) primarily engaged in the scout-reconnaissance role were analyzed. The skill sets in these three POIs were vetted by three subject-matter experts (SME), retired Army aviators familiar with scout and reconnaissance operations, for relevance to tactical employment of the RQ-7B .

Determining Relevance of Scout-reconnaissance Skills for RQ-7B Operators

The second step was to define the training-critical scout-reconnaissance skills for RQ-7B operators from the skills identified in the first step. A list of 25 critical scout-reconnaissance skills derived from analyses of the POIs were determined by SMEs to be relevant to RQ-7B missions. These were incorporated into an interview protocol in which they were to be rated for criticality by respondents, from command positions within a BCT, and in leadership positions within an RQ-7B platoon.

Determining the Most Critical Scout-Reconnaissance Skills for RQ-7B Mission

Assessment of these two dimensions of criticality (i.e., skill importance and operator preparedness) came from structured interviews with nine operational unit primary staff officers

(Lieutenant Colonel though Captain) at three BCTs located at their home stations, and six principal members of two RQ-7B platoons from two of these BCTs (Warrant Officer 2 /Sergeant First Class/Specialist 4). BCT officers' and platoon members' input was used to identify and rank order the most critical scout-reconnaissance skills for RQ-7B aircrews. In each structured interview, two researchers presented each set of 25 skills to respondents and recorded their verbal responses. In general, there were few disagreements in the records of each researcher. In those instances of disagreement, the first author resolved the discrepancy after reviewing the audio record. In particular, responses to the interview question for, "What additional training do (aircrews) receive at the unit to support this skill?" were used to indicate training gaps, and responses to the interview question for, "What additional training is required (beyond current) to prepare (aircrews) for this skill?" were used to indicate how the critical scout-reconnaissance skills might be trained. "Training critical" skills were defined as those skills deemed by respondents to be of high importance and for which UAS aircrews were perceived as being poorly prepared to perform (requiring additional training) when reporting to the unit (BCT).

Determining How Scout-Reconnaissance Skills are Trained in the Schoolhouse

The final step was to determine (a) if the identified critical scout-reconnaissance skills, ascertained from the scout POIs and structured interviews, were trained in the schoolhouse (b) if they were, then how these critical skills were trained, and (c) if not, what scout-reconnaissance skills actually were trained. To accomplish this, a qualitative analysis of schoolhouse training POIs (UAS Common Core and [RQ-7] Shadow Operator) was conducted. The training program was also examined to determine the training environment (i.e., classroom, simulator or aircraft, or field exercises).

Results and Discussion

Relevance of Cavalry Scout and OH-58D Pilot Skills to RQ-7B

The Ground Cavalry Scout and IERW OH-58D(R) Track Courses include not only target detection and recognition, but also practical hands-on training and application in reconnaissance techniques for area and route reconnaissance missions. These missions resemble those that RQ-7B aircrews would execute. Skills trained for these missions were further analyzed for application to RQ-7B. Importantly, one of the 25 mission-relevant skills was not included in any of the reviewed POIs but was mentioned as important by TCM-UAS and SMEs, as well as most of those BCT officers interviewed. This non-POI skill was Tactical Operations Center (TOC) operations.

Ten Most Training-Critical Scout-Reconnaissance Skills for RQ-7B

Of the total 25 scout-reconnaissance skills previously identified, 10 were distinguished as most critical in terms of being ranked high in importance and low in preparedness across all respondents. These critical scout-reconnaissance skills appear in Table 1 below.

Table 1

Scout-Reconnaissance Skills Rated as Most Critical by Respondents

Skill Areas
Tactical Operations Center (TOC) operations.
SPOT and SALUTE reports (size, activity, location, uniform, time and equipment).
Actions on contact.
Target handover (visual/non laser).
Fundamentals of security.
Fundamentals of reconnaissance
Aerial observation
Downed aircraft recovery operations
UAS integration into the BCT
Laser target handoff to the ground.

Respondents were also asked what additional scout-reconnaissance training supporting these critical skills RQ-7B operators received when reporting to the unit. A glance at Table 2 shows that the bulk of the responses concerned training at the Combat Training Center (e.g., JRTC), formal training drills in the unit, and on the job training at the unit level, most while deployed in theatre. Two mentioned schoolhouse, even though the question was about unit training. When asked what additional training is required beyond that currently available, the modal responses, as evidenced in Table 3, are formal unit training and training in the schoolhouse. It appears that respondents believed that the schoolhouse should incorporate a greater share of the burden for scout-reconnaissance training. In brief, it seems that Table 2 reflects the *current* status of RQ-7B training, whereas Table 3 implies the *preferred* status of this training.

Table 2

Response Frequencies for Additional Types of Training Currently Received on Ten Critical Skills

Type of Training			
On the Job	Formal Unit	Combat Training Center	Schoolhouse
19	22	27	2

Table 3

Response Frequencies for Additional Types of Training Required Beyond Current Training on Ten Critical Skills

Type of Training			
On the Job	Formal Unit	Combat Training Center	Schoolhouse
4	10	2	10

How Critical Scout-Reconnaissance Skills are Trained in the Schoolhouse

The UAS Common Core POI included instruction in a classroom environment on the principles of Reconnaissance, Surveillance and Target Acquisition (RSTA) which comprised lessons in map reading, operational terms and graphics, UAS employment in military operations, introduction to tactical imagery intelligence, intelligence preparation of the battlefield, and associated exams. These tasks cannot be defined as scout-reconnaissance, in the context of Army Aviation RSTA operations, in which aircrews must play an active role in developing the situation once a target has been identified.

Common Core also included instruction, in a classroom environment on skills and knowledge needed to interpret UAS electro-optical and infrared video to provide detailed and rapid feedback on the status of enemy targets, significant activities, and areas of interest. The module also contains lessons on imagery identification techniques, positive battle damage assessment, and positive identification of a variety of Soviet era military equipment.

The Shadow UAS Operator Course is focused on RQ-7B-specific training, which includes overviews of unmanned aircraft operator/payload operator systems, simulator and air vehicle flight, and a field capstone exercise. Scout-reconnaissance training is not documented in the POI for this course. According to the Common Core and Shadow Operator POIs, none of the five primary scout-reconnaissance skills (i.e., TOC operations, SPOT and SALUTE reports, Actions on Contact, Target Handover, and Fundamentals of Security) were consistently trained in any systematic way in the schoolhouse, and it was difficult to discern exactly which of these skills were taught in the simulator.

Conclusions and Recommendations

Training at the UAS Training Battalion (UASTB) at Fort Huachuca, AZ still maintains a heavy MI orientation even though it has been eight years since Proponency for UAS passed from MI to the Army Aviation Branch. As a consequence, some skills required by the scout-reconnaissance deployment of the RQ-7B were not included in the schoolhouse training of RQ-7B aircrews. Of the 25 scout-reconnaissance skills identified as relevant to RQ-7B aircrews, ten were determined to be critical to the training of aircrews. In the absence of schoolhouse training, it was the task of the unit to train aircrews on the identified scout-reconnaissance skills. The absence of schoolhouse training on critical scout-reconnaissance skills imposed the onus on the RQ-7B platoon because opportunities for unit-level training often did not present themselves prior to deployment to theatre or to live exercises, such as JRTC.

One statement by interview respondents (i.e., BCT staff officers) and supported by POI analyses summarized one issue with schoolhouse training as it now exists. The RQ-7B aircrews are trained as image analysts, not as scout-reconnaissance aircrews. This perception was evidenced by the extensive classroom training on map reading, tactical military intelligence, identification of Soviet era combat vehicles, and the analysis of electro-optical imagery. The lag in organizational cultures in which an MI climate still persists in the schoolhouse phase of RQ-7B training was documented by this research effort. A qualitative analysis of the schoolhouse program showed that little of the current training is optimized for the current scout-reconnaissance role of the Shadow UAS operators. There was widespread consensus among

BCT and RQ-7B platoon respondents that a portion of the scout-reconnaissance training burden that currently rests with the unit can be trained successfully in the schoolhouse.

Beyond the current research, additional issues need to be addressed, one of which is the increasing importance of manned-unmanned missions. In manned-unmanned missions, both manned and unmanned aircrews will have to coordinate and communicate for the scout-reconnaissance training topics covered in this report. This type of teaming will involve the acquisition of new skill sets by both UAS and manned aircraft communities. Aircrews will not only have to become proficient at scout-reconnaissance tasks, but each type of aircrew must understand their respective roles, capabilities, and limitations.

References

Stewart, J.E., Barker, W.C., & Bink, M.L. (2010). *Army RQ*-7B *training issues*. Paper presented at Meeting 64, Department of Defense Human Factors Engineering Technical Advisory Group Meeting, Tempe, AZ, May 3-6.

Thompson, W.T., Tvaryanas, A.P., & Constable, S.H. (2005). *U.S. Military Unmanned Aerial Vehicle mishaps: Assessment of the role of human factors using Human Factors Analysis and Classification System (HFACS).* (HSW-PE-BR-TR-2005-0001). Brooks City-Base, TX: Air Force Research Laboratory, 311th Performance Enhancement Directorate. (DTIC No. AD A435063).

U.S. Army Training and Doctrine Command (2010). *U.S. Army Roadmap for UAS 2010-20*35. U.S. Army UAS Center of Excellence, Ft. Rucker, AL.

TRUST IN UNMANNED AERIAL SYSTEMS: A SYNTHETIC, DISTRIBUTED TRUST MODEL

Kristin E. Oleson
Institute for Simulation & Training, University of Central Florida
Orlando, FL
P. A. Hancock
Department of Psychology, Institute for Simulation & Training, University of Central Florida
Orlando, FL
Deborah R. Billings
Institute for Simulation & Training, University of Central Florida
Orlando, FL
Christopher D. Schesser
Department of Psychology, University of Central Florida
Orlando, FL

An integrated unmanned aerial system (UAS) team is composed of multi-tiered partnerships between at least four distributed entities: the vehicle itself, the vehicle operator(s), the military command team, and the ground team. In this team structure, effective communication, information exchange, and situation awareness within and amongst each entity is critical. While miscommunications can be mitigated through current training strategies and vehicle design, trust plays an important role in these interactions. We have constructed a model of integrated team trust. An extension of this model, directed at distributed trust, is shown to be essential for military teams with integrated UAS. Consequently, we demonstrate the importance of calibrating trust within the team, and in particular the human-UAS partnership.

The utilization of unmanned aerial systems (UAS) is increasing due to their demonstrated benefits in extending the capabilities of unaided human operators in multiple operational realms. Here, we use the term UAS to refer to any unmanned aerial vehicle (UAV) or unmanned combat aerial vehicle (UCAV), including all control technology and monitoring systems. While unmanned vehicles carry no human onboard, these systems are controlled (or monitored) by one or more human operator(s). Therefore, human interaction is required to an extent in order to successfully use the UAS. Many different multimodal control interfaces and design guidelines for customization are available (see Oron-Gilad & Minkov, 2010; Chen, Haas, Pillalamarri, & Jacobson, 2006).

Unmanned Aerial Systems

UAS should be used when machines provided better sustained attention than a human; there is a lower political and human cost in high-risk environments; and there is an increase in the likelihood for mission success (Unmanned Aircraft Systems Roadmap 2005-2030, 2005). Consequently, these unmanned systems are currently being integrated and used for a range of missions, including reconnaissance and surveillance, tactical strike, force protection (e.g. detection of improvised explosive devices; crowd control), and electronic signals collection (Billings & Durlach, 2008; Cavett, Coker, Jiminez, & Yaacoubi, 2007; Unmanned Aircraft Systems Roadmap 2005-2030, 2005). Due to task diversity, unmanned vehicles, also referred to as assets, are available in a wide range of platforms (e.g., large, small, fixed wing, ducted fan) and capabilities (e.g., ability to carry different payloads, acquire targets, and flight). These systems range from small man-portable micro-aerial vehicles (MAV) to larger self-landing UCAV (for a detailed description of MAV see Billings & Durach, 2008; for Mini Air Vehicles, see Coffey & Montgomery, 2002; and for UAV and UCAV, see Unmanned Aircraft Systems Roadmap 2005-2030, 2005).

Military operations at the platoon level, company level, battalion level, and brigade level each have access to specific assets specific to the tasking required per level (for a detailed description of integration of UASs in the military hierarchy, see Barnes, 2003). The autonomy and capabilities of an asset are designed specifically for a given team and task type (see Table 1). The degree of human intervention necessary in the operation of the asset is primarily determined by its level of autonomy (Adams et al., 2007). As such, UAS operators may act in supervisory and/or teleoperator roles, and multiple operators are often required. For example, a highly automated UAV may require one (or several) operators to interpret information as the UAV flies through predetermined waypoints. Conversely, a teleoperated UAV may require one operator to manually control it, while another operator interprets the acquired sensory information (Billings & Durlach, 2008). The operators in both examples can vary in the extent of their geographical separation from the asset, from the ground team to the United States launch site.

Table 1.
Examples of unmanned aerial systems: task, flight information, operation, level of automation

Asset	Task	Flight Information	Operator	Level of Automation
 RQ-11 Raven	Intelligence gathering	1.5 hours 1,000 ft 4 lb	Controlled by ground team	Can fly along GPS waypoints or be controlled by ground operator
 RQ-2B Pioneer	Intelligence gathering	5 hours 12,000 ft 452 lb	Controlled from launching airfield or ship	Can fly preprogramed waypoints but is usually controlled by an operator out to a range of 100nm
 RQ-1 Predator	Intelligence gathering	29 hours 40,000 ft	Remote station or airfield	Directly flown by operating team
 RQ-4A Global Hawk	Intelligence gathering	32 hours 65,000 ft 32,250 lb	United States or point of launch (far from recon sight)	Follows a programmed set of waypoints that can be upgraded during operation
 MQ-1: Armed Predator	Tactical Strike	24+ hours 25,000 ft 2,250 lb	Control center	Pilot in the loop: Operators at control center would control target designation and firing of weapons.
 MQ-9B: Reaper	Tactical Strike	29 hours 40,000 ft 10,500 lb	United States or point of launch (far from recon sight)	Monitored by the ground crew and target selection and firing are controlled by the ground crew.
 X-45	Tactical Strike	7 hours 35,000 – 40,000 ft 36,500 lb	United States or point of launch (far from recon sight)	Autonomous takeoff/landing; designated waypoints through a combat zone. Ordinance likely to be GPS guided bombs, allowing for hands-off weapons delivery.

Note. Images obtained from Wikipedia.com. For more detailed information see *Unmanned Aircraft Systems Roadmap 2005-2030* (2005).

Integrated UAS Team

The UAS asset is currently designated as a tool with programmed tasking to extend the capabilities of the operator. The operator's extensive tasking can include deployment and retrieval of the asset, monitoring and control of the asset (e.g., navigation, perception and manipulation of remote environments), managing mission and status (e.g., communication, decision-making, monitoring human and robot tasks, coordinating social interaction tasks), and managing camera feeds (Adams et al., 2007; Cavett et al., 2007; Chen et al., 2006). However, the operator-asset partnership cannot be considered in isolation; the operator works in conjunction with command personnel (responsible for planning and allocating resources), aircraft pilots, other operators (Mouloua et al., 2001), and a ground team (responsible for carrying out actions in the combat zone). Therefore, a team with an integrated asset(s) incorporates a unique, multi-tiered partnership between at least four distributed entities: the asset (e.g., UAV, UCAV), the human operator(s), military command team, and the ground team. Chen and colleagues (2006) have described this partnership as an interdependence as seen through defining the mission and tasks, allocating tasks, two-way feedback between operator and robot, controller input, and analysis of information. Burke and colleagues (2004) discuss three possible relationships between the human team members and the asset: the human-robot ratio (i.e., number of humans assigned to an asset), the spatial relationship (i.e., distance between the human and asset, as well as point-of-view), and the authority relationship (i.e., defining the roles of the team). These relationships can impact the team's workload, communication, and situation awareness within a mission.

Communication

One of the major issues in the operation of an asset is the proper communication of information across the levels of the integrated UAS team. This communication requires some degree of audio, visual or tactile relay of information from an asset to an operator across an often large geographical distance. Due to the large amount of information that is relayed, it is important to ensure that the information is exchanged efficiently. Yet, different assets have different sensors and relay different types and amounts of information. Some issues related to the transfer of information directly from an asset include the complexity, organization and time lag in communication, as well as the flexibility, adaptability, and cognitive controllability of the bandwidth and frequency (Chen et al., 2006; Unmanned Aircraft Systems Roadmap 2005-2030, 2005; Burke, Murphy, Rogers, Lumelsky, & Scholtz, 2004). Here we are interested in the information flow and distributed communication between team members as it relates to type of task.

According to Adams and colleagues (2007) there are three main paradigms for types of tasks that utilize an UAS integrated team: Information Only, Ground-Led, and UAV-Led. While these task types are specific for search and rescue, they can be extended to other domains. Within the Information Only task, there is no direct partnership between the asset and the ground team. There is a direct flow of information from the asset to operator to commander to ground team. The operators utilize the asset to gather information. This if followed by a command team dispatching the ground team. In the Ground-Led task, the asset follows a ground led search and is used as a support when the ground team loses a trail. The asset can increase the effectual field of view and to provide increased information to the ground team without corrupting the trail. The UAV-Led task incorporates a direct partnership between the asset and the ground team. The ground team has a direct link to the data from the asset, and both teams provide a target location to the command team. The asset's path is preselected by waypoints to match the ground team, and can be updated based on collected target information. Communication is very different across these three types of tasks (see Figure 1 for communication links in terms of data flow and commands/requests). Communication in this teaming incorporates data flow as well as commands and requests. Data flow consists of the audio, visual or tactile information that is relayed through the team. Commands and requests require human interpretation of this data. Situation awareness allows the filtering of data at each level of the integrated team.

There are three levels of SA that must be considered at each level of the team: perception, comprehension or understanding, and projection into the future (Endsley, 1995). UAS extend an operator's SA by increasing the amount of information an operator has access too at both the local (operator station) and remote (asset) environment. In this way, a UAS operator can filter sensory information acquired from the asset, only relaying information vital to the success of the mission to the other members of the team. This process can be hampered by exceeding amounts of information and elevated workload. Therefore, future work is examining how to improving interfaces and automation to assist operators "absorb, assimilate, and track relevant information over time" (Riley, Strater, Chappell, Connors & Endsley, 2010) in order to increase SA at the asset level. Until these design, enhancement in communication between team members can lead to the more accurate interpretation of information and more effective situation awareness (Cavett et al., 2007).

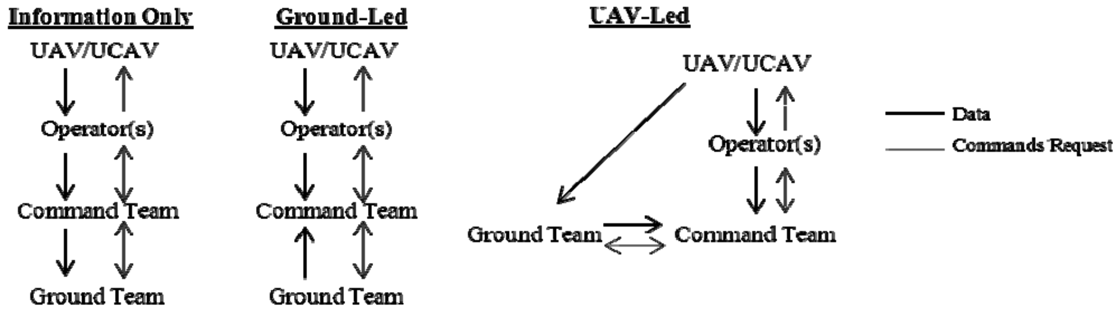


Figure 1. Communication of information among a multi-tiered team in three types of tasks.

Design and Training

System design and training guidelines continue to lead to the reduction of operator workload and enhanced performance in areas such as communication and situation awareness. A main goal of UAS teams is to reduce operator workload to such a degree that a reduction in the operator-asset ratio will lead to a single operator being able to control multiple assets. In attempting to meet these goals, the asset is moving away from a tool that is an extension of the operator to more of a human-asset partnership. Schulte and Meitinger (2010) refer to this teaming as co-operative control, in which interaction occurs through communication and a shared understanding of the situation. Therefore, critical attention is placed on the design of interfaces, displays, staffing of control stations (Billings & Durach, 2008; Mouloua, Gilson, Kring & Hancock, 2001). Further design of the autonomy and operator interfaces should be intuitive, facilitate control and data interpretation, promote the efficient use of time, have a high tolerance for workload, and span multiple application domains with minimal training requirements (Oron-Gilad & Minkov, 2010; Billings & Durlach, 2008; Adams et al, 2007; Burke et al, 2004). Training methods should be systematic, standardized, modular and flexible, with short training for simple tasks and main efforts on mission implementation (Oron-Gilad & Minkov 2010; Billings & Durach, 2008).

Trust

While the design of these unmanned systems, partnered with more systematic and standardized training protocols, are steadily enhancing the assets ability to complement and extend the capabilities of the human team members (e.g., situation awareness, decision making), the inappropriate calibration of trust within a human-robot team can lead to misuse or disuse in the field. A meta-analysis on human-robot trust (see Hancock et al, 2011) has provided support for inclusion of trust as a factor in human-robot teaming (d = +0.71), as well as for our three-factor model (see Figure 2) of trust in human-robot teams. From this meta-analysis, we found robot characteristics (d = +0.67) to be the presently, primary driving influence of trust in human-robot teams, followed by environmental characteristics (d = +0.47) with little influence from human characteristics. These findings suggest that calibration of human-robot trust could occur through specific integration of trust characteristics into the design and training protocols.

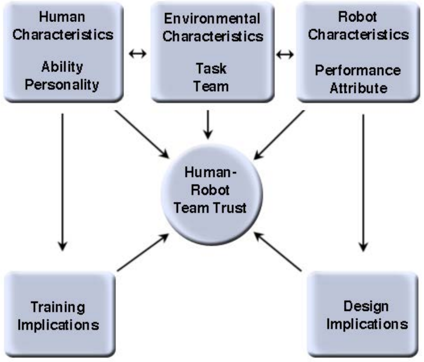


Figure 2. Three factor trust model: The direct and indirect factors that influence trust between human and robot team members.

Distributed Trust

These meta-analytic findings can be directly applied to the integrated UAS team at the operator-asset level. However an extension of this model specific to the multi-tiered partnerships and tasking is necessary. Integration of an asset affects (both directly and indirectly) the task completion and flow of information to the rest of the team. Therefore the application of system based trust (primary-level, secondary-level, tertiary-level, and so forth) is instituted. Primary-level trust can occur for entities that receive direct communication from another entity. Conversely, secondary-level trust may be present among team members across all lines of indirect communication. Tertiary-level trust is a more complex interaction between multiple team members. This distributed trust model only incorporates the notions of the primary-level and secondary-level trust within UAS teams for each type of task (Information Only, Ground-Led, and UAV-Led tasks; see Figure 3). A certain threshold of trust between team members needs to be reached so that effective interactions can occur. Calibration of trust helps to optimize this threshold between levels of under-trust (disuse) and over-trust (misuse). Total system trust can then be determined through the summation of the individual trust relationships (e.g., primary, secondary, tertiary level trust) between team members within a given task.

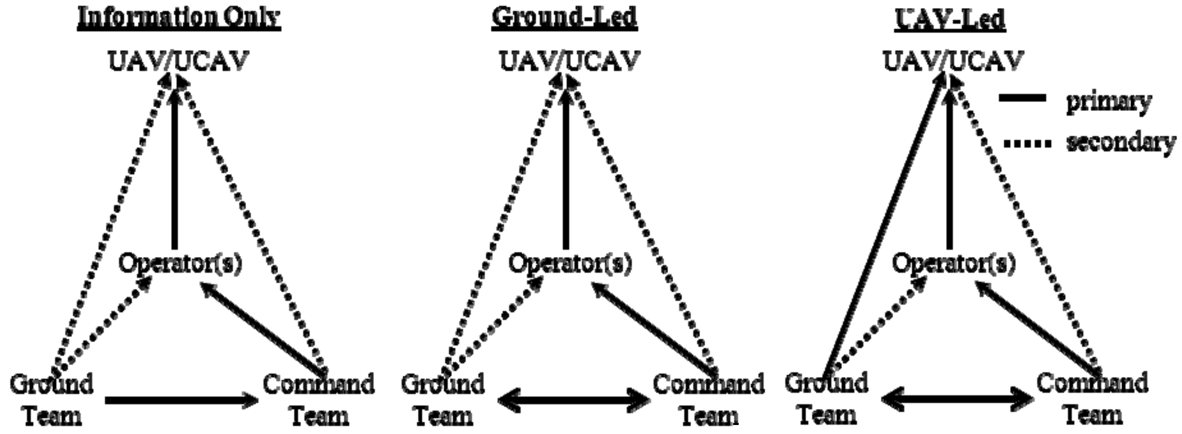


Figure 3. Primary and secondary trust links among a multi-tiered team in three types of tasks.

The calibration of trust between team members in a UAS teaming environment is important to consider in tandem with team communication. Trust calibration refers to match between an individual's perception of the robot and the robot's actual capabilities and performance (Hancock, Billings, & Oleson, 2011; Lee & See, 2004). Based on the above identified flow of communication and trust, we suggest that the calibration of trust at the operator-asset level will improve this direct interaction, as well as interaction, trust, and communication across the other team levels. The reasoning for calibration of trust at this level takes into account both the initial level of trust in an asset along with the high operator workload in this relationship. Therefore, ensuring appropriate trust at the operator-asset level can, in part, assist with issues in operator workload and situation awareness. Reduction in workload and an increase in situation awareness can, in turn, increase the efficiency of data information flow between the operator-asset and the ground team. In this way, trust in the operator-asset relationship trickles down and affects communication at these other levels.

Acknowledgements

The research reported in this document was performed in connection with Contract Number W911NF-10-2-0016 with the U.S. Army Research Laboratory. The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the U.S. Army Research Laboratory, or the U.S. Government unless so designated by other authorized documents. Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

Adams, J. A., Cooper, J. L., Goodrich, M. A., Humphrey, C., Quigley, M., Buss, B. G., & Morse, B. S. (2007). *Camera-equipped mini UAVs for wilderness search support: Task analysis and lessons from field trials* (BYUHCMI Technical Report 2007-1). Brigham Young University.

Barnes, D. W. (2003). *Considerations for the Maneuver Commander: Could the Emergence of Unmanned Aerial Vehicles Spell the Demise of the Army's RAH-66 Comanche in the Armed Reconnaissance Role?* (Master's thesis, U. S. Army Command and General Staff College). Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA416261.

Billings, D. R., & Durlach, P. J. (2008). *Effects of input device and latency on performance while training to pilot a simulated micro-unmanned aerial vehicle* (Technical Report 1234). Arlington, VA: Army Research Institute for the Behavioral and Social Sciences.

Burke, J.L., Murphy, R.R., Rogers, E., Lumelsky, V.J., & Scholtz, J. (2004). Final report for the DARPA/NSF interdisciplinary study on human robot interaction. *IEEE Systems, Man and Cybernetics Part C: Applications and Reviews, 34*(2), 103-112.

Cavett, D., Coker, M., Jiminez, R., Yaacoubi, B. (2007). Human-computer interface for control of unmanned aerial vehicles. *IEEE Systems and Information Engineering Design Symposium* (pp. 1-6). Charlottesville, VA.

Chen, J. Y. C., Haas, E. C., Pillalamarri, K., & Jacobson, C. N. (2006). *Human-robot interface: Issues in operator performance, interface design, and technologies* (ARL-TR-3834). Aberdeen Proving Ground, MD: Army Research Laboratory.

Coffey, T., & Montgomery, J. A. (2002). The emergence of mini UAVs for military applications. *Defense Horizons*, 22, 1-8.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Hancock, P. A., Billings, D. R., & Oleson, K. E. (in press). Can you trust your robot? *Ergonomics in Design.*

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50-80.

Mouloua, M., Gilson, R., Kring, J. & Hancock, P.A. (2001). Workload, situation awareness, and teaming issues for UAV/UCAV operations. *Proceedings of the 45th Human Factors and Ergonomics Society* (pp. 162-165). Minneapolis, MN.

Oron-Gilan, T., & Minkov, Y. (2010). Remotely operated vehicles (ROVs) from the bottom-up operational perspective. In M. Barnes, & F. Jentsch (Eds.), *Human-Robot Interactions in Future Military Operations* (pp. 211-228). Burlington, VT: Ashgate.

Riley, J. M., Strater, L. D., Chappell, S. L., Connors, E. S., & Endsley, M. R. (2010). Situation awareness in human-robot interaction: Challenges and user interface requirements. In M. Barnes, & F. Jentsch (Eds.), *Human-Robot Interactions in Future Military Operations* (pp. 171-192). Burlington, VT: Ashgate.

Schulte, A., & Meitinger, C. (2010). Introducing cognitive and co-operative automation into uninhabited aerial vehicle guidance work systems. In M. Barnes, & F. Jentsch (Eds.), *Human-Robot Interactions in Future Military Operations* (pp. 145-169). Burlington, VT: Ashgate.

Unmanned Aircraft Systems Roadmap 2005-2030 (2005). Washington, DC: Office of the Secretary of Defense. Available from http://handle.dtic.mil/100.2/ADA445081.

SOLUTION SPACE-BASED COMPLEXITY ANALYSIS FOR CONTEXT AWARE AUTOMATION

J. Comans, C. Borst, M.M. van Paassen and M. Mulder
Delft University of Technology, Faculty of Aerospace Engineering
Kluyverweg 1, 2629HS Delft, The Netherlands

Despite the huge advances in aviation automation and avionics over the last decades, the flight deck is sometimes still unaware of the 'bigger picture'. Current automation is not yet able to understand the intentions of the pilots or the intricacies of the environment it is operating in. When we want to increase the level of automation support, automation systems will need more context awareness. This paper will only focus on a small subset of the operational context, namely the complexity induced by surrounding traffic. To capture the constraints imposed by the surrounding traffic, we use a tool called the 'Solution Space'. The solution space is the subset of all speed and heading combinations for an aircraft that will not result in a conflict. In previous research, the solution space is visualized to provide a pilot or air traffic controller with an overview of the situation. Our goal is different in that we are trying to identify parameters in the solution space that correlate with the complexity of a traffic situation and how well they fit with the concept of traffic complexity as perceived by pilots. In this paper, the potential of the Solution Space as a complexity metric is evaluated in a preliminary offline analysis.

Modern avionics and automation systems are fairly oblivious to the world around them. Generally, only information relevant to their specific task is shared between systems. As an example, an autopilot instructed to climb to an altitude will mainly look at the altimeter data and initiate a climb; even though the Traffic Collision Avoidance System, TCAS, might show that there is another aircraft in the way. This example shows how automation operates in a fairly static way following basic predetermined rules.

If it were possible to treat an aircraft as a closed system (Vincente & Rasmussen, 1992), in which all input-output relations and disturbances were known, it would be straightforward to automate everything and remove the pilot from the cockpit. Reality is rather different. Aircraft are open systems, mainly due to unknown disturbances such as weather and due to complex ill-defined input-output relationships in an aviation system as a whole. Fortunately, humans are well suited to deal with this level of uncertainty. They are able to interpret different information sources and come up with creative solutions that cannot be provided by static automation.

When looking at the strengths and weaknesses of humans and automation, they are more or less each other's opposites. This indicates that they are well suited to work as a team to combine their strengths. Looking at current aircraft, humans and automation do not really work as a team (Billings, 1991). Automation behaves itself somewhat as a star player who, ones he gets the ball, does his tricks and scores without involving his team members. This might work in the majority of the situations, but when the situation does not work out as expected, it breaks down.

One of the recurring problems with modern automation is the lack of communication between automated agents and humans (Sarter & Woods, 2000). In the same way as automation operates in a predefined way, it also communicates in a predefined way in the form of annunciations and indications on the cockpit displays. When looking at human communication, we can see that the way in which we communicate a specific piece of information depends on the context. The way in which a passenger will alert a driver to a pedestrian who is about to cross the road will differ depending on, for example, whether there is an actual chance of running into this pedestrian or not.

In the cockpit, the way in which information should be presented to a pilot will also depend on the operational context at that time. During cruise for example, a pilot will generally have the time to interpret information and process it to determine meaning and importance. During a complex low visibility approach, the pilot will probably be too busy to let the information 'sink in' and will most likely benefit from a simple and unambiguous instruction provided by an automated agent. This would require automated agents to adapt to the operational context.

Being able to interpret the operational context will be key in creating automated agents that could become team players. It will be obvious that the information contained in the operational context can be rich. This paper will present a preliminary study on analyzing context information related to airborne traffic situations in the approach phase of business jet aircraft.

The approach chosen in this paper is to use the metrics derived from the Solution Space (Van Dam, Mulder, & van Paassen, 2008). The Solution Space is a representation of the constraints imposed by traffic on the velocity of an aircraft under control. Its visual representation can be used by pilots to resolve conflicts with other aircraft and to interpret a traffic situation. Next to the visual representation, the mathematical properties have been shown to correlate with the workload perceived by Air Traffic Controllers (ATCos) (Hermes, Mulder, van Paassen, Boering, & Huisman, 2009).

The aim of this paper is to show that the Solution Space reveals information about the traffic context from an aircrafts point of view. A number of off-line simulations of scenarios, with varying complexity, will be used to analyze which information is available from the Solution Space.

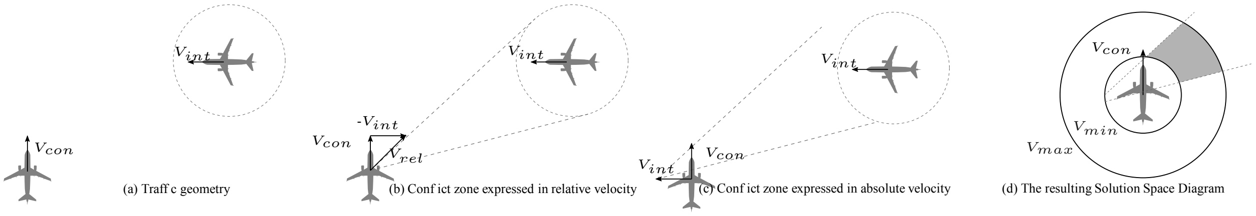
Complexity in Human-Machine Systems

In general terms, a complex system can be defined as a system that is so complicated or intricate as to be hard to understand or deal with. From a user centered human-machine system point of view, complexity could be viewed on two different levels: the operational complexity and the apparent complexity . The operational complexity is related to the inherent complexity of the system, while the apparent complexity is related to the complexity perceived by the user.

From a cognitive engineering point of view, the complexity introduced by the functional demands of the work domain is another source of complexity. This complexity will propagate through the human-machine interface (HMI) to the operator, unless the HMI is able to detect and reduce this increase in complexity. The next section will describe how we intend to extract information about complexity from a representation based on a work domain analysis.

Solution Space Diagram

Before we can start capturing context information, we must first understand the world. The basic definition of a traffic situation is shown in Figure 1 (a). The aircraft that we are interested in —the controlled aircraft— is flying with a velocity V_{con}. The intruding aircraft is flying with velocity V_{int}. The trajectory of this intruder might conflict with the trajectory of the controlled aircraft. The circle around the intruding aircraft shows the protected zone, a region in which no other aircraft can enter, typically with a radius of 5 Nm. Two aircraft are in conflict with each other when they are flying along a trajectory that will eventually result in the aircraft entering each others protected zones. Upon entering the protected zone, the conflict becomes a loss of separation. A TCAS system will interpret this situation by looking at the rate of closure and the distance to the intruding aircraft (Kayton & Fried, 1997). At close range this produces an accurate indication whether a collision is about to happen in a short time. In essence, TCAS reduces a three dimensional problem to a one dimensional problem. While TCAS is well suited as a collision avoidance tool, its algorithm is not suitable to discover context information. The approach we chose to look at context information follows from the work on Ecological Interface Design (EID) at our section (Van Dam et al., 2008). EID is a design methodology that focuses on the constraints imposed by the environment on an operator (Vincente & Rasmussen, 1992). EID starts by constructing a complete representation of the work domain. Based on this work domain analysis, a designer needs to find ways to communicate and visualize this representation.



(a) Traffic geometry (b) Conflict zone expressed in relative velocity (c) Conflict zone expressed in absolute velocity (d) The resulting Solution Space Diagram

Figure 1: Construction of the Solution Space Diagram

The actual work domain analysis of this example is beyond the scope of this paper, but Figure 1 shows the results of the visual mapping of the concepts and constraints identified. Figure 1 (b) shows how a conflict can be easily identified when we look at the relative velocity of the controlled aircraft with respect to the intruding aircraft. When the relative velocity, V_{rel}, points into the protected zone, there is a conflict. This enables us to define a Conflict

Zone (CZ) that indicates which relative velocities lead to a conflict. When looking at relative velocities, it is easy to identify a conflict; however, the relative velocity depends on the velocities of both the controlled and the intruding aircraft. Pilots and air traffic controllers work with absolute velocities since those are controlable. By shifting the CZ with V_{int} we can see the conflict represented in absolute velocity as shown in Figure 1 (c). With this representation, the magnitude and direction of V_{con} can be directly related to the CZ. Another advantage of using the absolute velocity representation is that this procedure can be repeated for multiple intruding aircraft. The last step is to create a Solution Space Diagram with the CZ. An example is shown in Figure 1 (d). By showing the minimum and maximum velocity that can be flown, the solution space diagram shows all possible combinations of speed and heading an aircraft can use. The CZ's of different intruding aircraft can be added to this diagram. In this way, a pilot or air traffic controller can immediately see what combinations of speed and heading could lead to a conflict, or how to get out of a conflict by changing their velocity vector until it is outside the CZ. The way in which this information is presented reveals information about the traffic geometry. Several parameters can be used to assess the situation the controlled airplane is in. Among them are the area occupied by the CZ's, the amount of free space, the distance to the closest conflict, the rate of change of the conflict area, etc. The goal of this paper will be to assess whether information extracted from the Solution Space can be used to determine context.

Offline Analysis

Pilot Interview

Two pilots were briefly interviewed in order to be able to create realistic scenarios. Both pilots are test pilots flying the Cessna 550 Citation operated by the Delft University of Technology (DUT) and the National Aerospace Laboratory (NLR). They have experience in operating at large international airports such as Schiphol Amsterdam Airport and regional airports such as Rotterdam Airport.

The main sources of traffic complexity involves approaches to smaller regional airports. This is because approaches at large airports are streamlined by Air Traffic Control and there are no significant differences in performance between the approaching aircraft.

When operating a business jet type aircraft, one of the main sources of complexity is the mix of aircraft with different performance envelopes. Maintaining proper separation can become difficult when an aircraft is unable to fly slower than 100 kts and is on an approach behind an aircraft flying at 60 kts.

During closely separated visual approaches, if a leading aircraft drifts away from its intended approach path, the complexity for the trailing aircraft's pilots will increase because next to their normal tasks, they have to evaluate whether the deviation poses a threat or not.

Assumptions

All scenarios considered will use only two aircraft. Both are simulated using a simple two dimensional kinematic model. The flown trajectory is calculated based on a list of way-points. The way-points define the tracks followed by the aircraft, and are connected with circular segments. The radius of these circle segments is based on the resulting radius for a rate one turn at the airspeed used for the trajectory (Ruijgrok, 1996).

The aircraft are assumed to fly at a constant speed, V_{app}. Wind and atmosphere effects are neglected. The position and orientation of the aircraft are determined by calculating the along-track-distance based on V_{app}.

For the calculated SSDs, a Protected Zone radius of $1500\ m$ is used. The SSDs are calculated with a lookahead time between $5\ s$ and $180\ s$, to limit the scope of the solution space to conflicts that are relevant in the near future.

Scenarios

Based on the the pilot interviews, three scenarios were defined. The aim of all three scenarios is to create situations where complexity varies over time.

The first scenario consist of two airplanes approaching parallel runways, separated by $800\ ft$ as shown on the left of Figure 2. The leading aircraft is on the right approach track flying at 70 kts. The trailing aircraft is a faster jet with an approach speed of 120 kts. The trailing aircraft starts at the beginning of the trajectory $280\ s$ after the leading aircraft. The trailing pilot will have to deal with the fact that she is caching up with the leading aircraft. In addition to her usual tasks during the approach, she will have to estimate her closure rate in order to keep a safe distance. When the distance is large, there is no urgency, but it is difficult to assess the closure rate. At closer distances, it will be easier for a pilot to estimate the closure rate, but the situation develops faster making it a time critical task.

In the second scenario, the trailing aircraft flies the same trajectory as in the previous scenario, a straight track towards the runway at $120\ kts$. The second aircraft is crossing the thresholds in front of the trailing aircraft as shown in Figure 3. Crossing in front of an approaching aircraft is not a standard procedure, but can occur due to miss communication or mistakes.

The final scenario assumes that the leading aircraft overshoots the interception of the approach track towards the right runway. This overshoot puts the leading aircraft temporarily in the path of the trailing aircraft. The lateral distance between the runways is slightly exaggerated in Figure 4 to make the figure clearer. In the actual scenario, the leading aircraft will fly into the trjacetory of the trailing aircraft.

Results

Figure 2 shows the results for the first scenario at three significant points along the trajectories. The corresponding SSDs are shown on the right of the trajectory plot. It is important to note that only the areas between the minimum and maximum velocity circles are part of the solution space. In this case however, the full conflict zones are drawn in order to better show the evolution through time.

The top row of Figure 2 shows the solution spaces for the faster trailing aircraft. Two effects are visible. The conflict zone shifts towards the center of the diagram, and the cone defined by the conflict zone widens. The shifting conflict zone shows that the trailing aircraft is catching up with the leading aircraft. The widening of the cone is related to the actions required to steer clear of the conflict. When the cone is small, only small heading corrections are required to get out of conflict.

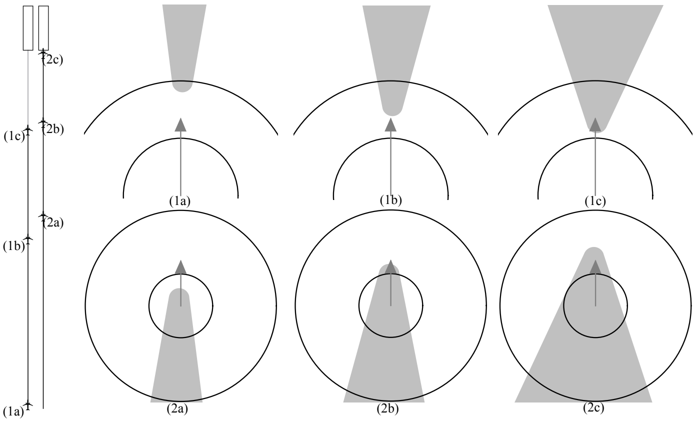


Figure 2: A parallel approach scenario with the resulting SSDs

The crossing target scenario is shown in Figure 3. In this scenario, the shifting and widening seen in the parallel scenario are also visible. From this figure, it is clear that the conflict zone does not shift towards the center of the diagram. It shifts towards the velocity of the observed aircraft, which is directed to the left in this case. The widening related to the decreasing distance between the two aircraft is also present, indicating that this distance is decreasing. This time, however, the solution space also rotates. This is caused by the fact that the relative direction towards the leading aircraft changes over time.

The combination of widening and rotation creates an interesting effect. In situation (1a), the conflict zone is narrow and tilted to the right. As the scenario develops into situation (1b), the solution space widens and rotates, but the distance between the velocity of the trailing aircraft and the solution space does not change significantly. When the leading aircraft has crossed (1c), the distance between the two aircraft has further decrease, as seen in the width of the solution space, but the distance between the velocity and the solution space has increased significantly showing that the conflict has been resolved.

The two previous scenarios described were fairly static in the sense that there were no heading changes. The overshoot scenario shown in Figure 4 is more dynamic. At the point of the overshoot, the heading of the leading

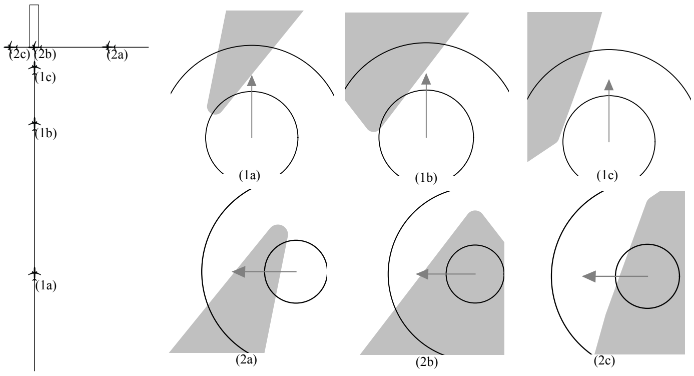


Figure 3: Two aircraft with crossing trajectories and their resulting SSDs

aircraft will change twice within approximately 30 s. Situation (1a) shows the situation before the intercept. There is no conflict, but the distance between the velocity is small, indicating that small disturbances can result in a conflict. After the overshoot, the leading aircraft will correct its heading to return to the intended approach track. (1b) shows the situation after the initial correction. At this point, the solution space has shifted from the left side of the velocity vector to the right side. It is important to note that in the transition from (1a) to (1b), the conflict zone will slide through the velocity vector, indicating there is a conflict for a short amount of time. Once the aircraft is back on its required approach track it will change its heading to the runway heading. This results in a full blown conflict as seen in (1c).

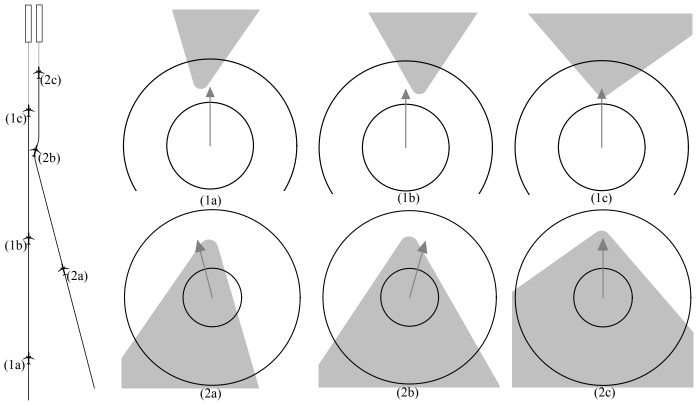


Figure 4: An overshoot during an approach capture with the resulting SSDs

Discussion

The parallel approach scenario shows two fundamental properties of the solution space. When the velocities of both aircraft are constant, the shift of the conflict zone has a direct relationship with the distance between the two aircraft under consideration. In the same way, the rate of change in distance can be related to the rate of shift of the conflict zone. Both properties relate to the direct goal of the pilots to maintain separation. They mainly provide a measure for the time it takes before two aircraft get too close to each other.

When the distance becomes smaller, the pilot have to take more drastic measures to avoid or get out of a conflict. At larger distances, the pilot has more options, hence the situation is less difficult. This property is captured by the width of the solution space. The rate of change in width could also be related to how fast a situation develops.

The previously described properties also show up in the crossing scenario. In this scenario, the pilot will need to evaluate whether the crossing aircraft has crossed his path once he gets to the intersection. The main difficulty lies in the fact that both aircraft are converging towards the intersection. From the pilots point of view, a safe way out is to turn right and pass behind the leading aircraft. This shows up in the solution space as an empty right hand side. Figure 3 shows a situation where the leading aircraft passes just in time. In this situation, it will not be clear for pilots that the aircraft will definitely have cleared the crossing by the time they get there. The closer they get to the crossing, the more important it becomes to monitor the leading aircraft, the complexer the situation gets. The solution space reflects this mainly in the change of the area. The distance between the conflict zone and the velocity shows that the aircraft is close to a conflict. Once the aircraft passes the crossing, it will be clear for the pilots that there is no conflict. Once again, this is reflected by the conflict zone moving away from the velocity vector.

The overshoot scenario presents a more difficult case. The previous remarks on area and proximity still hold, but the way in which the solution space moves is not predictable when looking only at the current state of both aircraft. This might be solved by noting that the behavior of the observed aircraft is goal driven. It will attempt to land on the runway. If this goal is known, we would be able to predict that the solution space will end up in the middle, creating a conflict. Any kind of information about the goals of the observed aircraft might be helpful, from a basic heading to a full list of way-points.

As an example, there are techniques to calculate solution spaces that take into account the future trajectory of the observed aircraft (d' Engelbronner, Mulder, van Paassen, de Stigter, & Huisman, 2010). These kind of methods assume that detailed information about the future trajectory of the observed aircraft is available. It is however doubtful that this kind of information will be available in the near future and therefore these methods were not used here.

Conclusions

This paper presented a preliminary study on the possibilities to use the solution space to quantify a subset of complexity. The goal is to use the geometric properties of the solution space to find metrics for complexity. An offline analysis was performed to investigate the effects of close proximity traffic on the solution space.

The preliminary results show that in scenarios with limited heading changes, the rate of change and proximity of a conflict zone can be related to the complexity of a scenario. In a scenario that includes significant heading changes, the solution space by itself is not sufficient. Goal related information will be needed to get an accurate representation of complexity.

References

Billings, C. (1991). *Aviation Automation – The Search for a Human-Centered Approach.* Lawrence Erlbaum Associates, Inc.

d' Engelbronner, J., Mulder, M., van Paassen, M., de Stigter, S., & Huisman, H. (2010, August 2-5). The Use of the Dynamic Solution Space to Assess Air Traffic Controller Workload. In *Proceedings of the AIAA Guidance, Navigation and Control Conference.* Toronto, Canada.

Hermes, P., Mulder, M., van Paassen, M. M., Boering, J. H. L., & Huisman, H. (2009, November). Solution-Space-Based Analysis of the Difficulty of Aircraft Merging Tasks. *Journal of Aircraft*, *46*(6), 1995-2015.

Kayton, M., & Fried, W. R. (1997). *Avionics Navigation Systems* (2nd ed.). Wiley Interscience.

Ruijgrok, G. J. J. (1996). *Elements of Airplane Performance.* Delft University Press.

Sarter, N., & Woods, D. (2000). Team Play with a Powerful and Independent Agent: A Full-Mission Simulation Study. *Human Factors*, *42*, 390–402.

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, and Cybernetics - Part A: Systems and Humans*, *38*(6), 1221–1233.

Vincente, K. J., & Rasmussen, J. (1992). Ecological Interface Design: Theoretical Foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, *22*(4), 589–606.

Integrating aircrew resources variability into the design of future cockpit

Barbara Mawhin
Airbus SAS Operations, Toulouse, France-LATI Université Paris Descartes, Paris France
Toulouse France
Philippe Cabon
LATI- Université Paris Descartes, Paris, France
Paris France
Florence Buratto
Airbus SAS Operations
Toulouse France

Cockpit design may have major consequences on future pilot's tasks. Functions needed to ensure flight safety are shared between pilots and aircraft systems. Today, cockpits are designed considering a theoretical minimum level of crew resources availability although it is widely acknowledged that availability of crew resources may vary from time to time because of aircrew internal state. Therefore, it is crucial to ensure that future cockpits are designed while taking into consideration this variability. This work aims at developing a methodology enabling designer to systematically integrate crew resources availability in the design process. In order to assess the impact of different sources of variability, the principles of the systemic model FRAM is used. The present work is analyzing the impact of crew resources availability on several use cases using FRAM principles. The data are collected by the means of focus group with operational and human factors experts. Some preliminary results are presented and discussed.

Aviation is considered as a high reliability system where safety is the result of complex interactions between human, machine and organization. Aviation evolves in a dynamic environment (reorganization of air traffic management in Europe, growth of traffic…), implying a constant development of new adapted design solutions to respect the required safety level . This implies a better human-machine cooperation and an evolution of pilot role and task from low level tasks to high level one. Historically, aircraft automation was developed based on the best compromise between performance and safety at an acceptable cost. System automation has grown on the assumption that human are better than machines on strategic tasks but not on repetitive one. But the growing complexity of aircraft systems makes them less transparent for users, leading to other problems of adaptation of pilots interactions with glass-cockpit (Amalberti, 1998). Furthermore, cockpits are designed considering a theoretical constant level of crew resources availability although it is widely acknowledged that availability of crew resources may vary from time to time because of endogenous factors such as fatigue, stress, sickness…. In this context, we propose to extend the traditional aircrew incapacitation definition to any situation that leads to a decrease of aircrew resources to a level lower than the required level of resources.

From a design perspective, the risk of incapacitation should be managed at 3 levels: prevention (avoid a decrease of aircrew resources), detection (detect the decrease of aircrew resources) and recovery (compensate for the decrease of resources).

The aim of this work is to develop a methodology that enable designers to integrate incapacitation in the design of future cockpits. The proposed methodology relies on two main stages, a risk analysis phase and a risk management phase. This paper focused on the risk analysis phase. After an overview of the theoretical background, some preliminary results are provided.

Theoretical Background

Models that account for human variability

Some models of cognitive psychology and physiology bring useful elements to understand human resources variability. Among cognitive models, the computational model of resources developed and revisited by Wickens (2008) is the most appropriate for this research as it defines resources according to the type of information process. Other useful models are relative to decision making (Klein, 2008), situational awareness (Endlsey, 2000) and problem solving (heuristics, strategies). Furthermore one of the critical resources that support the above cognitive function is alertness that is known to vary as a function of a wide range of factors. Recent progress in modeling has shown that alertness is regulated by 3 processes, i.e. sleep homeostasis, circadian factor and sleep inertia (Akerstedt, Folkard & Portin, 2004) Therefore from these scientific findings it becomes now possible to predict rather accurately what could be the alertness level of an operator during a given duty.

Theories on human machine interaction

In order to adapt aircraft systems to the variability of crew resources, several theories might be useful to bring elements in order to build aircraft artifacts adapted to human variability. In the automation design, Parasuraman, Sheridan & Wickens (2000) proposed a model to set the human performance regarding the types and levels of automation. Another theory proposed by Dinadis et Vicente (1999) suggested design principles based on the Rasmussen's model SRK – Skills Rules-Knowledge- (Vicente & Rasmussen, 1992) in order to adapt the type of information provided to the operator linked with the situation. The joint cognitive system theory brings a fruitful framework in the development of effective decision support to human activity (Hollnagel, 2003). The framework of adaptive automation is also considered as it refers to systems that can adjust their functioning or level of operation dynamically. Moreover they are linked with biocybernetics theory that suggests monitoring changes in workload to adapt the systems (Pope, 1995).

Evolution of safety models

In order to be able to integrate human variability in the design of future cockpit it is also necessary to understand the evolution of safety models in complex sociotechnical systems. These systems are characterized by multiple interactions (Perrow in Hollnagel, 2004). To address the complexity of these models, safety models have evolve from sequential and linear (e.g. the domino models from Heinrich, 1931) to systemic models (e.g. stochastic resonance model from Hollnagel, 2004). Table 2 shows safety models evolution and its consequences on design.

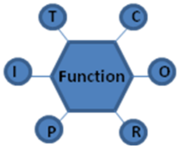
Table 1.*Evolution of safety models in complex sociotechnical systems (adapted from Hollnagel, 2004)*

Type of safety model	**Example**	**Main assumptions**	**Consequences on design**
Sequential models	Domino model, event trees	Identify specific causes and contain them	Automation is used as much as possible to improve performance and safety level
Epidemiological models	Latent conditions, Pathological systems	Identify latent conditions, multiple causes for an accident	Systemic approach and automation to make defenses and barriers stronger
Systemic models	Chaos model, stochastic resonance	Complex interactions and performance variability lead to the event	Monitor and control performance variability using adaptive automation and systemic barriers

Systemic models allow accounting for complexity of sociotechnical systems. They are composed of multiple sub-systems, each one comprising several functions. Functions are defined as the goals the system has to achieve to ensure its good functioning. Each function may have a variable performance in achieving its own goal. Variability is inherent to a system, specifically at the front line operators level, who are adapting their work to the environment in order to maintain the required level of efficiency. In complex socio-technical systems, activity is always under-prescribed facing the possible changing conditions in which task must be performed. Necessary adjustments are made by the operator at the best compromise between thoroughness and efficiency as described in the ETTO principle (Hollnagel, 2009). Internal state of the operator is an important criteria in this trade-off.

Cognitive and physiological limitations are the main reason why in a complex sociotechnical system, human is the most important source of variability facing a dynamic environment (more than technology or organization).

These principles have been recently developed in the framework of Resilience Engineering. In this context, accidents are seen as the result of the resonance between the variability of sociotechnical systems functions rather than the results of failure or human errors . In Resilience Engineering theory, variability of performance should be monitored and managed. A methodology is proposed by the Resilience Engineering theory in order to understand and model functional variability. Nowadays, the only method focused on performance variability analysis is FRAM - Functional Resonance Analysis Method - (Hollnagel, 2004) that is based on a functional approach. The first application step of this method is to define functional entities not based on system structures (even if in some cases a functional entity may be close to a structural unit). Each function is characterized by six parameters: input, output, time, control, preconditions and resources. (figure 1)



I : Input : That which the function uses or transforms to produce the outputs
O: Output: That which the function produces
T: Time: Time available and factors that affect time availability
C: Control: Control and protective systems that exist to supervise or restrict function
P: Preconditions: That must be fulfilled to perform a function
R: Resources: That which the function needs or consumes

Figure 1. One function visual representation in FRAM, applied to each function description.

The potential variability of each one is assessed using criteria and the identification of dependencies among them leads to the functional resonance assessment. Once potential variability is described, barriers should be defined by assessing the required performance level. The three axes for barrier design are prevention, detection and recovery. This study will focus specifically on recovery aspects using two solutions principles:

- decrease level of required human resources
- and/or reallocate available resources

Preliminary results

The current research is conducted according to the following steps :

- incapacitation categorization
- selection of a use case
- instantiation of FRAM model (i.e. the simulation of the impact of a given incapacitation on a selected use case).

Incapacitation categorization

Because of the very wide spectrum of incapacitation (from a light fatigue to a sudden death), incapacitations have been gathered according to their potential consequences in terms of safety rather than on the basis of their causes. From the various possible incapacitations a total of 6 classes (table 2) have been identified in the literature in order to assess the potential sources of human variability and so the risk classes.

Table 2. *Presentation of the Incapacitation classes.*

Incapacitation class	Definition	Example
C1	Sudden and total incapacitation of one pilot	Heart attack, AVC…
C2	Progressive and partial incapacitation but with impact on activity	Gastro intestinal disease, infectious disease, drugs…
C3	Subtle and partial incapacitation no visible for other crew member	Attention, fatigue…
C4	Progressive, in a first time partial then total incapacitation of crew	Hypoxia...
C5	Inappropriate behavior	Psychiatric decompensation…
C6	Rapid and total incapacitation of crew	Barotraumas, depressurization…

Use case description

As already mentioned, the research is based on the instantiation of the model on several use cases. A use case is defined as an operational situation that requires a replanification by the aircrew, a dynamic phase with a temporal pressure and a high workload. The first use case chosen is the Late Runway Change during the approach flight phase as:

- it is considered as a normal operation,
- a replanification process is required,
- it can generate events,
- it accounts for a supplementary cost for the airline (passengers are later at gate, higher volume of fuel used),
- the performance on task may be influenced by the crew state and external conditions (temporal pressure...).

Analysis of variability

The purpose of the present analysis is to propose solutions for barriers implementation in order to maintain balance between required resources and available resources.

The purpose of the approach phase after a late runway change is to fly the aircraft according to a new trajectory given by the ATCo (Air Traffic Controller) to the DH (Decision Height) with required performance.

The system under analysis is a late runway change operation as a whole. It is composed of three sub-systems, each one having an implication in the late runway change operation. The first one is the ATCo (as the use case is a Late Runway Change proposed by ATC). This means that the crew has to make the decision process to accept or not and need to update the action plan. The second and third sub-systems are the crew, PF (Pilot Flying) and PNF (Pilot Non Flying) and the aircraft.

Functional boundaries are defined in order to include the hardware artifacts of the cockpit, the procedures and humans. The operational boundaries of this study define the starting point of a late runway change when ATC asks the crew whether they accept it (after the Top of Descent and the first approach briefing) and ends at the DH where the missed approach procedure or landing is performed. Finally, the hardware boundaries are FCU (Flight Control Unit) and FMA (Flight Mode Annunciator), PFD (Primary Flight Display) and ND (Navigation Display). Once the system and its boundaries have been described, The FRAM method is used to perform an analysis of variability of each function comparing this variability in nominal conditions and making several instantiations by varying aircrew resources.

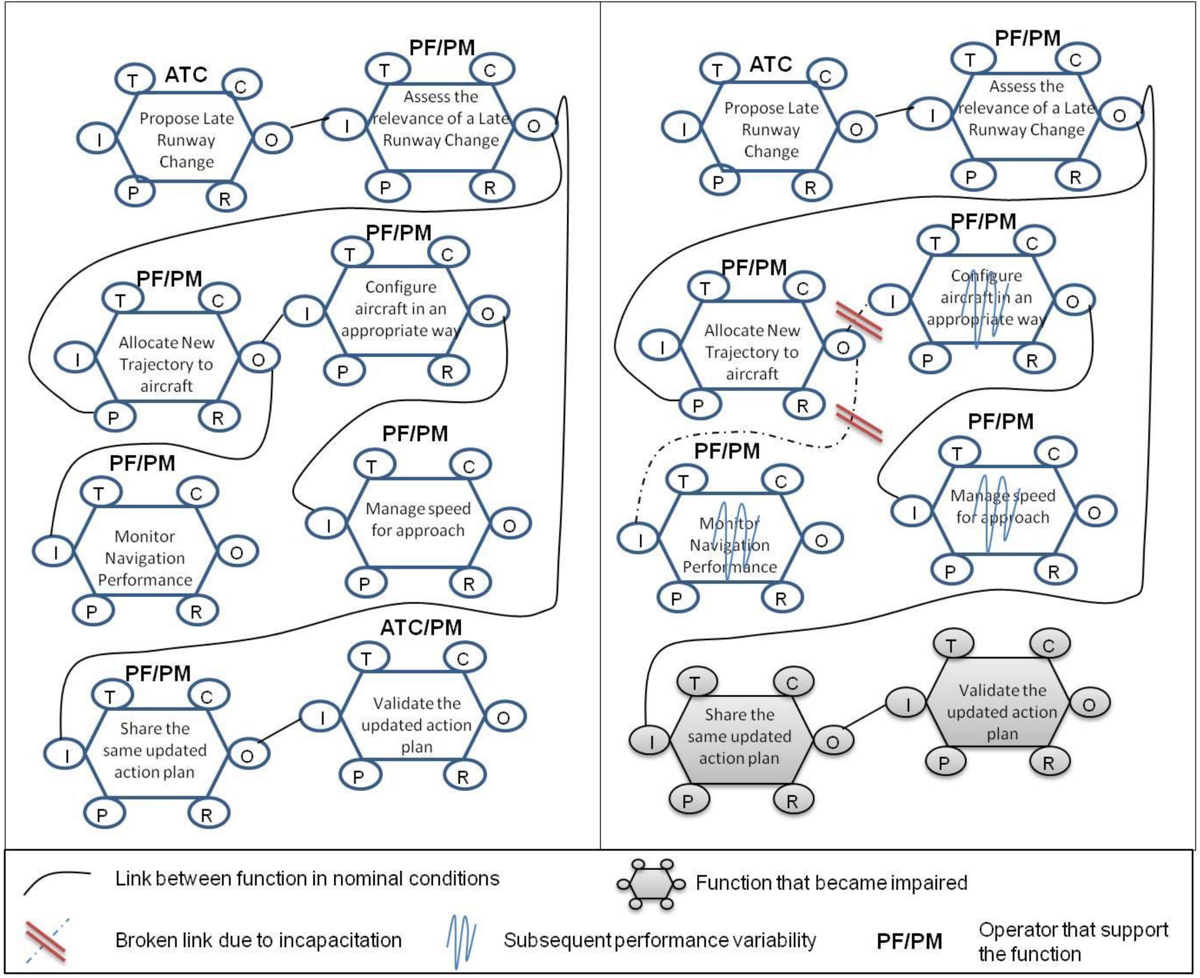


Figure 2.FRAM instantiation in nominal condition versus FRAM instantiation with incapacitation class C1.

The instantiation presented in figure 2 as an illustration shows the potential variability of several function if the performance of one of them (the function ‘Allocate new trajectory to aircraft) becomes unpredictable due to an incapacitation. In this illustration, the incapacitation concerns the PF and is a sudden and total one, corresponding to class C1 of incapacitation. In order to fit with the use case, the incapacitation arrives after the function ‘Crew accept the Late Runway Change’ has been performed. Comparing the nominal condition and the instantiation with incapacitation, FRAM allows to show a resonance process leading to performance variability on functions ‘configure aircraft in appropriate way', 'monitor navigation performance' and 'manage speed for approach. Moreover, the functions 'share the same updated action plan' and 'validate the updated action plan' became impaired due to the lack of resources provided by the pilot suffering of the incapacitation. As described in the theoretical background of FRAM, the next step aims at defining the type of barrier needed : physical, functional, symbolic or incorporeal.

Conclusion

The study presented in this article aims at defining methodological principles in order to take into account human resources variability in the design of future cockpits. In order to assess the performance variability, one of the methodological choice is to assess the benefits of the only method available to assess variability, FRAM. Even if the link between functions is an illustration, it seems interesting to use FRAM in order to assess potential variability of other functions if the performance of one is weakened. This preliminary results suggest that FRAM can be a useful element in the overall methodology as an enabler to focus on variability sources. In order to ease the risk management phase process, pilots strategies will be apprehended. Once the methodology for risk analysis phase validated, it will be applied on several use cases.

The second step of the research will be a risk management phase lying on focus groups with experts in order to propose barriers to improve the stability of the sociotechnical system with a particular focus on recovery. To improve the quality of inputs of the focus groups, tools provided by the creativity theories will be used. All this will lead to the identification of principles for design of future cockpits, hopefully leading to the development of a systemic method taking into account crew resources variability.

References

Akerstedt, T., Folkard, S., Portin, C. (2004). Predictions from the three-process model of alertness. *Aviation, Space and Environmental Medicine*, *75,* 75-83.

Amalberti, R. (1998). Automation in aviation : A human factors perspective. In D. Garland, J. Wise & D. Hopkin (Eds.), *Aviation human factors* (pp. 173-192, chapter 7). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dinadis, N., Vicente, K.J. (1999). Designing functional visualizations for aircraft systems status displays. *International Journal of Aviation Psychology*, *9*,241-269.

Hollnagel, E. (2009). The ETTO principle: Efficiency-thoroughness trade-off: Why things that go right sometimes go wrong. Farnham, UK: Ashgate.

Hollnagel, E. (2004). Barriers and accident prevention. Aldershot, UK: Ashgate.

Hollnagel, E. (2003). Cognition as control: A pragmatic approach to the modeling of joint cognitive systems *IEEE Transactions on Systems, Man, and Cybernetics*, Special Issue.

Hollnagel, E., Goteman, Ö. (2004). The functional resonance accident model. *In Proceedings of cognitive system engineering in Process Plant*, CSEPC 2004, 155-161.

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,* 286-297.

Pope, A.T. (1995). Biocybernetics system evaluates indices of operator engagement in automated task. *Biological psychology*, *40,* 187-195.

Klein, G. (2008). Naturalistic decision making. *Human Factors, 50,* 456-460 Doi: 10.1518/001872008X288385.

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

Wickens, C.D. (2008). Multiple resources and mental workload. *Human Factors*, 50, 449-455. Doi: 10.1518/001872008X288394.

APPLICATION OF COMMUNICATION GROUNDING FRAMEWORK TO ASSESS EFFECTIVENESS OF HUMAN-AUTOMATION INTERFACE DESIGN: A TCAS CASE STUDY

Doug Glussich, Jonathan M. Histon, Stacey D. Scott
Systems Design Engineering
University of Waterloo
Waterloo, ON Canada

The role of the human operator in automation augmented domains has shifted from primary decision-maker to collaborative partner, where the human often has to understand and manage state changes that result from the automation itself. Due to the challenges of these progressively complex states, there is increasing demand for automation systems that provide effective human-automation interfaces that keep the human more “in-the-loop”. Effective human-automation interaction in this situation is akin to effective human-human communication: an effective conversation occurs when people use commonly understood verbal and non-verbal mechanisms that lead to shared understanding, or *common ground*. In this paper, we demonstrate how the application of a communication grounding framework, typically used to describe the practices of effective human communication, can be used as an analytic tool to assess human-automation interfaces. This analytic tool can be used to highlight design flaws likely to result in breakdowns in human-automation interaction, and ultimately lead to human error.

Aviation has been a strong domain for the advancement of automation. In the century since first flight in North Carolina, aviation has helped advance technologies in the domains of metallurgy, electronics, navigation and engine design. From the early days of the diminutive Wright Flyer, to the Airbus A380, aviation has been largely a successful collaboration of man with technology. However, the introduction of more, and increasingly sophisticated, automation into aircraft systems has not always gone as smoothly, or safely as intended. As aircraft designs have advanced, and begun to incorporate more complex automation systems, the pilot’s role has shifted from one of manual controller of the aircraft systems to a higher-level role of managing the various automation systems now controlling the aircraft systems, and if necessary, intervene in flight operations to respond to abnormal situations.

Pilots of modern aircraft must work collaboratively with the onboard automation systems to maintain efficient performance and safe operations. Unfortunately, this human-automation collaboration is not always successful. Several accidents investigated by aviation safety bodies have been attributed to ineffective human-automation interaction (AAIASB 1987). In many cases, the aircrew was unable to understand the message the automation was communicating. They failed to recognize the overall situation and instead dealt with a secondary condition of attempting to ascertain what the automation was trying to convey; a condition known as operator “Out-Of-The-Loop” (OOTL) syndrome (Endsley 1995). An example of this situation was the case of the fatal crash of Helios Airways flight 522 in August 2005.

Helios Airways 522 was a Boeing 737-400 on a charter flight from the island of Crete to Prague, Czech Republic. During the initial climb, the aircraft failed to pressurize due to a pressurization switch that had been left in the incorrect position. Subsequently, a warning horn had sounded identifying the condition. The crew failed to recognize the horn as dealing with a pressurization problem and misidentified the condition. The pilots, unaware of the now hypoxic environment, lost consciousness. The aircraft later crashed outside of Athens at the cost of 121 lives. The crash was attributed, at least in part, to the aircrew’s failure to successfully identify and interpret the aircraft warning systems activation (AAIASB 1987).

A challenge for automation designers is to ensure that their systems can effectively communicate warnings and alerts to pilots so that appropriate actions can be taken. In many automation systems, however, it is assumed that simply conveying the warning or alert message is sufficient for effective communication to occur, and for the problem to be resolved. For example, the crew alerting system (CAS) mentioned in the above example is designed to alert the pilot of abnormal conditions based on various aircraft parameters; however, it does not have the capacity to monitor the reaction of the pilots to any warning it conveys. Anyone who has ever sent directives via an email or written memo knows that simply conveying a message to someone does not necessarily guarantee that the recipient will understand the message and take appropriate action – even highly trained pilots. Messages can often be interpreted in multiple ways, depending on the context, and, thus, misunderstandings can occur. Experiences in human-human communication show that successful communication typically requires much more interaction between parties, through an interactive back-and-forth process to establish that a message is understood (Clark and Brennan, 1991).

The inherent limitations of the one-way only communication model used in the CAS system has been recognized by automation designers, and more sophisticated alerting systems have begun to appear that incorporate improved communication processes that promote better human-automation collaboration. However, few analytic tools currently exist to assist designers in assessing how well their system designs promote effective human-automation interaction. Design limitations are often only discovered after expensive field testing has occurred or during accident investigations. Recognizing the parallels between human-human communication and human-automation interaction can open new possibilities for analytic tools, as one recognizes that decades of research has been dedicated to analyzing human communication processes and understanding how these processes can promote effective communication.

A particularly relevant communication analysis tool is Clark and Brennan's (1991) communication grounding framework, as it describes the process by which people reach a mutual understanding or, *common ground*, when communicating. Such "grounding" is also critical in the aviation context, as pilots must clearly understand the message conveyed by an automation system in order to implement an appropriate response action. Grounding mechanisms identified by Clark and Brennan's framework help communicating parties identify when misunderstandings have occurred, and help repair those misunderstandings. In the aviation context, it is important, for instance, that an automation system recognize when a pilot has misunderstood an alert and responded inappropriately, so that communication repairs can be made. In this paper, we show how the communication grounding framework can be used as an analysis tool to assess the effectiveness of the human-automation interaction processes enabled by the features of an automation system. We demonstrate this analysis method through a case study of a current aviation automation system, the Traffic Collision Avoidance System (TCAS).

To set the context for the case study, we first describe the TCAS automation system and then overview the communication grounding framework. Finally, we present the case study which analyzes the evolution of the TCAS design, and identifies how the design iterations introduced improved grounding mechanisms between the pilot and the automation system.

Traffic Collision Avoidance System (TCAS)

TCAS is an example of a device that warns pilots to the possible threat of collision with another aircraft. The system is a display situated in the flightdeck that is either a stand-alone device or one that is integrated with the navigation display. The display is comprised of an overview of the host aircraft as well as distance and relative altitudes of close proximity threat aircraft. If an aircraft enters a predefined range, the threat level of the aircraft is identified and registered on the display. In the application of TCAS to an impending collision scenario, time critical functions must be met with a response equal to the criticality of the scenario. This is accomplished by the use of aural and visual cues to increase the flightcrew's awareness of the possible threat situation. In TCAS, levels of communication between automation and humans are divided into two classifications; either an information level or an immediate action response. The first state, known as a traffic advisory, deals with the possibility of an aircraft becoming a threat due to its close approximation to the host aircraft. As the threat aircraft approaches the subject aircraft, their indication on the host aircraft display will turn color from white to amber and an aural "traffic, traffic" alert will sound. The purpose of this message is to inform the pilots of potential threat traffic in close proximity to the host aircraft. If the threat aircraft continues to approach the threat aircraft, the system will provide an immediate action command, for example, a "Climb, Climb Now" (known as a resolution advisory), in case the pilots must execute in a expeditious manner.

The TCAS system is an example of an automation system that actively communicates to the pilot (human operator) across a two-way channel. The system recognizes actions undertaken by pilots at the direction of the resolution advocated by the particular warning system.

These warning systems were designed through an examination of the weaknesses of pilot response to environmental threats. In the time preceding TCAS, aircraft threat recognition and avoidance was primarily done through aural communication between pilots and ground controllers transmitting threat location relative to the respondent aircraft, or between the pilots themselves via aural and non verbal gesturing (e.g., pointing to a certain location relative to their aircraft). After a long history of midair collisions, the United States Congress enacted legislation following the crash of Aeromexico flight 498 in August 1986 (NTSB, 2007) mandating that transport aircraft would require automated collision avoidance systems. In essence, the intent of the TCAS automation system is to augment both the communications between air traffic control and pilots and between the flightcrew themselves with another aid in collision detection and avoidance.

Initial TCAS technology provided for very few states. The original system (TCAS I) informed the pilots of a potential threat (the aural "traffic" call with accompanying visual display) and a "Clear of Conflict" aural presentation. The original TCAS system did not have the capacity to understand if the message was interpreted correctly by the pilots and had no provision to update or revise the message if the pilots failed to understand the original message. TCAS I provided the same type of one-way communication model as the original CAS technology, and failed to advance the need for confirmation of understanding of the intended message. The case study will discuss this design limitation in more detail and how the revised TCAS II system addressed this design issue.

Clark and Brennan's Communication Grounding Framework

Clark and Brennan's (1991) theory of communication grounding is one of the most fundamental and influential theories and frameworks to arise from the human communication literature. It has been widely used by communication and collaboration researchers to understand human communication behavior (e.g., Stahl, 2006), as well as an analytic tool to assess the ability of collaboration technologies to support successful communication (e.g., Vandergriff, 2006, Beers et al., 2007). Clark and Brennan's work has primarily been applied to the exchange of information between humans; however, researchers have previously utilized it to examine interaction between autonomous systems. Billard and Dautenhahn (1998) examined the use of common ground in the education of autonomous robots; specifically to examine information exchange between student and teacher robots.

The fundamental concept underlying the communication grounding framework is that effective communication relies on people's ability to reach a mutual understanding, or *common ground*, of the messages being conveyed to one another during a conversation. Reach this common ground, often referred to as "being on the same page", is accomplished through a process called grounding (Clark and Brennan, 1991). A core component of the grounding process is that conversation proceeds in a series of presentation and acceptance phases (Figure 1), during which the conversing parties attempt to achieve a mutual understanding of the message conveyed (or presented) before moving onto the next presentation and acceptance conversational segment.

Presentation phase:	**A** presents a message to **B**
Acceptance phase:	**B** accepts the message by giving *evidence* that he/she believes what **A** means by that message.

Figure 1. Conversational segment (Clark and Brennan, 1991)

The grounding process begins, and is a key aspect of, the acceptance phase. The message receiver, B in Figure 1, attempts to determine whether she understands the message that the presenter, A in Figure 1, has conveyed. If B believes she does not understand the message she will provide negative evidence of grounding by requesting clarification from A, or providing a similar response. If B believes she understands the message, but in fact does not, A will monitor B's response during the acceptance phase (and, in fact, throughout the rest of the conversation) for any evidence of a misunderstanding. If such evidence presents itself, then A will attempt to repair the miscommunication in order to complete the acceptance phase.

The communication grounding framework describes several forms of evidence people use to assess whether grounding has been achieved. The most common, and most relevant for our purposes include:

- **Lack of negative evidence**, i.e., evidence one was misheard or misunderstood
- **Relevant next turn**, e.g., receiving an appropriate answer to a question

In the human-automation communication context, an example of a relevant next turn would be for a pilot to implement an appropriate action in response to an automation alert, and thereby providing positive evidence to the automation system (if it is capable of monitoring the pilot's actions) that the alert was understood correctly. If instead, the pilot took no action when needed or an incorrect action this would provide negative evidence of grounding, which the automation system could then identify and respond to in order to clarify the original message (and complete the acceptance phase of grounding).

Measuring the degree of successful automation, from the communication grounding framework perspective, in human-automation interactions should answer the following questions: Is the message clearly conveyed? Has the message been understood and accepted by the receiver. If not, does the receiver have the ability to tell the original sender if the message is not understood? Does the original sender have the capacity to revise that original message in order to repair the misunderstanding?

Another fundamental concept of communication grounding is the *principle of least collaborative effort*: "In conversation, the participants try to minimize their collaborative effort – the work that both have to do from the initiation of each contribution to its mutual acceptance." (Clark and Brennan, 1991). This principle dictates that people will work *together* to minimize the overall effort expended during a conversation. On one hand this means that participants will attempt to minimize their own personal effort in transmitting and receiving information. On the other hand, it also establishes a social contract between parties to help each other efficiently reach a mutual understanding. In essence, this means that the "receiver" of the information will not make the "sender" do all the work to convey a clear and precise message. The receive will help the sender by identifying misunderstandings, and perhaps even suggesting alternative messages for the sender to consider to repair misunderstandings should they occur. This is a highly interactive process, but allows people to communicate very efficiently overall.

In the aviation context, brief alert messages are typically preferred, for example, the aural warning message "traffic", is typically used over a more elaborate, but unnecessary, "another aircraft in close proximity to you". Likewise, instead of requiring a pilot to acknowledge the "traffic" alert via input into the system before taking action, simply responding with the appropriate action will provide positive evidence of understanding and minimizes overall communication effort between the human and the automation.

The presentation and acceptance phases of a conversation have their own associated challenges unique to the medium in which they are taking place. For example, a face-to-face conversation requires less mental taxation than that of a text conversation from distant participants. These challenges, or constraints, are one of the dimensions that affect grounding. They are the basis by which parties increase overall performance through different environmental conditions. If augmentation of these dimensions can be achieved, the reduction of collaborative effort would be the overall outcome. These constraints include, but are not restricted to (Clark and Brennan, 1991):

- **Visibility**, an environment in which A and B are visible to each other.
- **Audibility**, an environment in which A and B can hear each other.
- **Reviewability**, an environment / media in which B can review A's messages.

Costs, on the other hand, provide alternative measures to address weaknesses in constraints. The penalty for these costs can retard the effective communication between communication participants. Communication costs include, but are not restricted to (Clark and Brennan, 1991):

- **Reception costs** are associated with receiving a message. Listening is generally easier (lower cost) than reading.
- **Understanding costs** are associated with understanding a message. The more complex the message or words used to convey the message, the higher the understanding cost.
- **Start-Up costs** are associated with initiating a communication. When co-present, start-up costs tend to be low, except if the environment is chaotic, and then getting the receiver's attention may be costly.

Automation designers must be aware of the potential constraints and costs associated with their system design and with the environment in which the system will be deployed. Each of these factors can influence the ability of the system to support effective grounding during human-automation interaction.

Case Study

The introduction of the initial TCAS I traffic warning system provided a significant advancement in collision avoidance, certainly above the historical "look out the window" approach. Analyzing the TCAS I design using the communication grounding framework, however, reveals limitations in its ability to support effective communication between the automation and the flightcrew. When a threat aircraft is recognized, the TCAS I system communicates this threat by presenting the relevant information to the flightcrew in both aural and visual formats; in this case an amber target is displayed on the TCAS display and an aural "Traffic, Traffic" message is provided. This is the "presentation phase" of the communication between the automation and the pilot. In human-human communication, this phase would then be followed by an "acceptance phase", during which the message receiver would demonstrate to the message conveyer that they have understood the message and the conveyer would monitor the receiver's behavior for evidence of understanding. The TCAS I system enables only a limited type of acceptance phase; it only monitors for positive grounding evidence, indicated by the pilots implementing the appropriate maneuver away from the threat aircraft. Once a positive outcome is detected, the TCAS I system would issue the aural message "Clear of Conflict". While this represents one type of "presentation" and "acceptance" phase, the grounding framework demonstrates that other grounding mechanisms are often necessary to facility successful communication after an initial message is presented.

From a grounding perspective, the main limitation of TCAS I is that it cannot detect miscommunications, and consequently, cannot attempt to repair those miscommunications. There is no provision in the system to monitor the flightcrew's actions for negative evidence of grounding (e.g., no action or an incorrect action taken). Thus, TCAS I has no means to repair a miscommunication, through, for example a revised cautionary command that might facilitate a relevant next turn by a flightcrew who has misinterpreted the significance of the initial presentation.

The revised TCAS II system addresses some of these issues by providing more sophisticated communication possibilities between the automation and the flightcrew after the initial presentation of the traffic advisory warning message ("Traffic, Traffic"). Once the initial warning is issued, TCAS II monitors the flightcrew's actions for evidence of grounding. In particular, it can recognize whether actions have been taken to maneuver the aircraft away from the threat, and whether these actions will be sufficient. If not, TCAS II can revise the initial message with the resolution advisory, or RA ("Climb, Climb Now"), in order to repair the miscommunication. The manner in which this message is conveyed also facilitates grounding. The brief aural command minimizes reception costs (listening to a warning is easier than reading a display) and understanding costs (the short climb command is easier and less time restrictive than detailing the action for the aircraft to increase altitude). Also, the 'climb' directive is also visualized on the TCAS display, providing reviewability, which further reduces understanding costs (visibility and audibility versus audibility alone). If the TCAS II system determines that the rate of climb or descent is insufficient after issuing an RA to the flightcrew, an aural "Adjust Vertical Speed, Adjust" will be broadcast. As with TCAS I, the communication loop is closed with the "Clear of Conflict" acceptance message.

Overall, the revised TCAS II design provides a much more interactive communication process between the automation system and the flightcrew, more closely aligned with human-human communication, primarily through expanded monitoring and communication repair capabilities in the automation design. Miscommunications are recognized and repaired to achieve a positive overall outcome.

In summary, TCAS II expands on the initial presentation / acceptance framework in TCAS I by introducing the ability of TCAS II to revise miscommunications. TCAS II provides for the automation to look for positive and negative evidence of grounding from the flightcrew that the initial presentation has been accepted and understood.

Challenges Looking Forward

Although the design revisions introduced in the TCAS II system provided expanded communication potential for human-automation interaction, its grounding capabilities could be improved. A recent traffic safety study by Eurocontrol (2010) identified key shortcomings in the TCAS II system. The most prominent failings identified in the study related to the system's inability of revise a resolution advisory in order to resolve a potential collision. From a grounding perspective, this indicates the system's inability to recognize a wider variety of negative evidence from the flightcrew's actions.

In the case of the traffic advisory, negative evidence occurs when no actions are taken in response to the initial presentation of the warning. In the TCAS I system, there was no ability to repair this miscommunication. TCAS II addressed this challenge with the RA ("Climb, Climb Now") command. The new challenge occurs while the aircraft is following an RA. When in RA mode, repairs must address not only the performance of the host aircraft (whether or not to climb or descent) but also the degree of execution. Although this is evident with the "Adjust Vertical Speed, Adjust", the repair message is not explicit enough as to whether the adjustment should be more or less aggressive, potentially introducing more misunderstanding. The performance profile is currently only included on the TCAS display. A message that more explicitly indicates these factors would facilitate the grounding process. The result would be a decrease in the understanding costs and a net positive performance.

The Eurocontrol study also identified the need for an RA reversal (e.g., from climb to descent) if the actions of the threat aircraft contravene its own TCAS system. This capability would require further design changes to the TCAS system to enable the presentation of a new, alternative message to the flightcrew, and further monitoring of the crew's understanding of this new message.

Conclusion

The effort to formulate messages in human-human conversations shared a close association with human-automation interaction. In this paper we examined how theories that were once identified as constituting the minutia of human-human conversations can be applied to human-automation communication. By examining the communication grounding framework, we identified this as an analytic tool to help designers assess the effectiveness of their automation system for promoting effective human-automation communication. The presented case study on the TCAS automation system highlights and explains crucial design improvements that contribute to improved human-automation communication in the current TCAS II system. The communication ground framework also helped identify design deficiencies that were addressed in proposed TCAS display improvements.

References

Air Accident Investigation & Aviation Safety Board (AAIASB) (2005). Aircraft Accident Report Helios Airways Flight HCY522 Boeing 737-31S at Grammatiko, Hellas on 14 august 2005. AAIASB Report No. 2006-11.

Beers, P.J., Boshuizen, H.P.A., Kirschner, P.A., Gijselaers, W.H. (2007). The analysis of negotiation of common ground in CSCL, *Learning and Instruction*, 17 (3), 427-435.

Billard, A., & Dautenhahn, K. (1997). Grounding communication in situated, social robots. In *Proceedings of TIMR97, Towards Intelligent Mobile Robots Conference*. Technical Report Series, Department of Computer Science, Manchester University, ISSN 13616161. Report number UMCS-97-9-1.

Clark, H. H. & Brennan, S. A. (1991). Grounding in communication. In L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.). *Perspectives on socially shared cognition*. Washington: APA Books, 127-149.

Endsley M., Bolte B., Jones D. (2003). *Designing for situation awareness: an approach to user-centered design*, CRC Press.

Eurocontrol (2010). Collision Risk due to TCAS Safety Issues: Investigations and analysis of TCAS II safety issues in the European Airspace. Report No. Eurocontrol Report TCAS-SAF/T2/07/D2a, v. 1.2.

Kaber, D. B., Tan, K-W and Riley, J. (2002). Improved usability of aviation automation through direct manipulation and graphical user interface design. International Journal of Aviation Psychology, 12(2), 153-180.

National Transportation Safety Board (1987). Collision of Aeronaves De Mexico, S.A. McDonnell Douglas DC-9-32, XA-JED and Piper PA-28-181, N4891F, Cerritos, California, August 31, 1986. NTSB/AAR -87-07.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 30(3), 286-297.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

Vandergriff, I. (2006). Negotiating Common Ground in Computer-Mediated versus Face-to-Face Discussions, *Language Learning & Technology*, 10 (1), 110-138.

CHARACTERIZING PILOTS' INTERACTIONS WITH THE AIRCRAFT COLLISION AVOIDANCE SYSTEM

Alexandra E. Coso, Elizabeth S. Fleming, Dr. Amy R. Pritchett
Georgia Institute of Technology
Atlanta, GA

Collision avoidance on large transport aircraft involves many components: Air Traffic Control (ATC), the pilot, and the Traffic alert and Collision Avoidance System (TCAS). This research explores pilots' interactions with ATC, the environment, and TCAS. Collision avoidance reports from NASA's Aviation Safety Reporting System (ASRS) were used to examine the encounter conditions surrounding collision avoidance incidents, including airspace, environment, and type of aircraft involved as well as pilot perceptions of the event. A coding scheme, developed in the early stages of this research, captured details regarding the traffic encounter, the role that ATC and TCAS played within the encounter, and the flight crew's response. This analysis spanned TCAS-related ASRS incident reports from 2008 to 2010. The results illustrate that the availability and presentation of traffic information impact pilot agreement (and disagreement) as well as their compliance (and noncompliance) with ATC and TCAS issue maneuvers.

The Traffic alert and Collision Avoidance System (TCAS) delivers a two-stage advisory and avoidance maneuver to the pilot when it predicts loss of aircraft separation. The first stage, "Traffic Advisory" (TA), advises the pilot to a situation, but does not command (or authorize) an avoidance maneuver. The second stage, "Resolution Advisory" (RA), delivers a vertical avoidance maneuver (or limits on vertical speed) as required to maintain separation. These stages are supported by the TCAS traffic situation display (TSD), which provides a horizontal spatial presentation of nearby traffic as an aid to visual acquisition.

When a pilot encounters a TCAS advisory, he or she has seconds to decide how to respond. The Federal Aviation Administration (FAA) Advisory Circular 120-55B mandates that the pilot in command should maneuver as a TCAS RA directs unless the maneuver would endanger safe flight operations. It is important to note that pilots generally receive these RAs in high-density air traffic environments, and thus the time period leading up to, and spanning, a TCAS RA may also include many other events. For instance, a pilot may visually acquire another aircraft, which may or may not be the advised traffic, or a pilot may receive a traffic call-out from the controller. In the same instance, the pilot may overhear other communications on the 'party-line' or incur non-collision avoidance related events and alerts. The pilot also may receive air traffic instructions, which may be perceived as creating or resolving the traffic situation. Still, the Federal Aviation Regulation (FAR) 91.3 explains that a pilot in command has the ultimate authority and responsibility for the safe flight of the aircraft. Therefore, to extensively examine pilot interactions with TCAS, it is critical to examine not only the interactions themselves, but also the context of the air traffic environment in which pilots may be influenced by multiple, sometimes-competing factors.

To explore the collision avoidance environment in context, reports from NASA's Aviation Safety Reporting System (ASRS) were examined. The ASRS is a database containing voluntarily submitted reports filed by any personnel (including pilots) to describe incidents relating to potential aviation safety concerns. This database provides a unique method of observation and analysis regarding the perceptions of pilots during events involving TCAS. Earlier studies, which examined pilot reported use of TCAS via ASRS reports, identified an unexpected degree of reported noncompliance to TCAS RAs, a wide range of roles that the pilots attributed to TCAS, and potential reliance on the TCAS traffic situation display beyond its intended role as an aid to visual acquisition (Mellone, 1993; Pritchett, 2001; Rantanen, 2003).

The purpose of this study is to provide a current review of ASRS reports with a more detailed analysis of the reported factors affecting a pilot's response to TCAS RAs and ATC instructions. This paper summarizes a broad analysis of cases where pilots report agreement (or disagreement) and compliance (or noncompliance) with TCAS and ATC instructions. While the narratives provide a window into pilots' perceptions of the events, it is important to note that the reports may reflect incomplete or inaccurate assessments of the events. Thus, the emphasis of this paper is on factors perceived by pilots and reported by pilots as influencing their responses.

Method

The ASRS database was accessed on July 20th, 2010, and relevant reports from January 2008 - April 2010 were selected using a pre-defined list of collision avoidance-related search terms: TCAS, ACAS, collision avoidance, traffic advisory, resolution advisory, avoid a collision, evasive action, and mid-air collision. An example is shown in Figure 1.

> "I was crewing as [Second-In-Command] of an aircraft cruising at [Flight Level] 190. There was a traffic alert on TCAS. The alert was an amber target at +3,300, twelve o'clock and descending rapidly. Both pilots' eyes [were] on the situation. We noticed the aircraft moved in a zigzag. [We] received an RA to descend on TCAS. Noticed the aircraft was at +300 feet. Reported to ATC immediately that we had an RA and were descending. At the same time of reporting to ATC, I looked out for traffic and spotted a B-52 at our two o'clock position and less than one mile horizontal. Looks like if we hadn't taken evasive action there could have been a collision. We descended until the TCAS advised us that we were clear of conflict. We descended to 18,300. Once clear of traffic we climbed back to assigned altitude of [Flight Level] 190. I listened to ATC tell the other aircraft he should have been at block altitude of 22,000 to 20,000 feet, thank goodness for TCAS."
>
> *(ACN: 879699, 2010)*

Figure 1. The narrative is an example of an ASRS report found in the July 20th, 2010 search, describing a pilot's interactions with TCAS.

A coding scheme was developed and tailored to the analysis of collision avoidance events involving TCAS from a preliminary analysis of reports from 2009 and 2010. The coding scheme is comprised of four dimensions. The first, *Encounter Conditions,* identifies the weather conditions, the airport, the types of aircraft involved and their respective flight paths. *Incident Description* and *Traffic Situation Awareness* record which type of advisory (TA, RA or both) that the pilot reports acting upon, whether the pilot was impacted by the possible visual acquisition of another aircraft, 'party-line' communications, and/or air traffic controller call-outs of traffic, as well as the response of the pilot to the advisory. Finally, *Perceptions of the Reporting Individual* captures descriptions framing important factors as positive or negative, perceived communication breakdowns, and recommendations related to TCAS or collision avoidance (as shown in Table 1).

Table 1.

Coding Example of ACN: 879699, 2010

Statement	**Incident Description**	**Traffic Situation Awareness**	**Perceptions**
There was a traffic alert on TCAS.	TA		
The alert was an amber target at +3,300, twelve o'clock and descending rapidly		Traffic Situation Display	
[We] received an RA to descend on TCAS	Descend RA		
Reported to ATC immediately that we had an RA and were descending.	Complied with RA		
At the same time of reporting to ATC, I looked out for traffic and spotted a B-52 at our two o'clock position and less than one mile horizontal.		After RA	
I listened to ATC tell the other aircraft he should have been at block altitude of 22,000 to 20,000 feet, thank goodness for TCAS.			Positive Perception of TCAS

Note: The dimensions of incident description, visual acquisition and perceptions were used to code the narrative itself, while encounter conditions were used to code information not included in the narrative.

Two coders independently coded the 278 ASRS reports with an inter-rater reliability goal of 80 percent. Using Cohen's kappa test, the resulting kappa value indicated 96% agreement, and coding disagreements between the raters were discussed until a consensus was reached. Subsequently, cases reporting noncompliance to instructions issued by TCAS or ATC were re-examined for common themes.

Results

As shown in Table 2 and Table 3, the reported incidents occurred in a variety of conditions and had a variety of outcomes in terms of reported compliance to TCAS RAs and ATC instructions. These categorizations were formed from the reports provided by the pilot, which in some cases were incomplete or may have been based on a biased perception of the situation.

Table 2.

Encounter conditions of the reported incidents

Year (n=278)			**Respondent Aircraft Type (n=278)**		
2008	109	39.20%	Commercial Jet	28	10.10%
2009	128	46.00%	Commercial Fixed Wing	35	12.60%
2010 (January – April)	41	14.80%	Commercial Jet Low Range	31	11.20%
Time of Day (n=262)			Commercial Jet Medium & Short Range	96	34.50%
Early Morning (12:01am – 6:00am)	20	7.60%			
Morning (6:01am – 12:00pm)	60	22.90%	Corporate Jet	51	18.40%
Mid-Day (12:01pm to 6:00pm)	131	50.00%	Military Aircraft	3	1.10%
Evening (6:01pm to 12:00am)	51	19.50%	Regional Jet	26	9.40%
Weather (n=226)			Small Personal Aircraft	8	2.90%
IMC	31	13.70%	**Other Aircraft Type (n=183)**		
Mixed	16	7.10%	Commercial Jet	44	24.10%
VMC	179	79.20%	Corporate Jet	6	3.30%
Phase of Flight (n=292, allowing for multiple flight phases per report)			Helicopter	6	3.30%
			Military Aircraft	9	4.90%
Climb	73	25.00%	Regional Jet	3	1.60%
Cruise	46	15.80%	Small Personal Aircraft	50	27.30%
Descent	44	15.10%	Visual Flight Rules Aircraft	18	9.80%
Approach	129	44.20%	Unknown Aircraft	47	25.70%

Table 3.

Reported compliance to TCAS RAs and ATC instructions

Reported Compliance to RA (n=248)		
Compliance	192	77.40%
Partial Compliance	25	10.10%
Noncompliance	19	7.70%
Unspecified	12	4.80%
Reported Compliance to ATC Instructions (n=78)		
Compliance	43	55.10%
Partial Compliance	15	19.20%
Noncompliance	20	25.60%

Of the reported RA encounters, 77% (n=192) of pilots reported complying with the TCAS instructions, while reported noncompliance to an RA occurred in less than 8% of the reports (n=19). Statistical analysis revealed no statistically significant relationships between the conditions described in Table 2 (year, time of day, weather, respondent aircraft, other aircraft, and phase of flight) and reported compliance to either TCAS RAs or ATC instructions, described in Table 3. Examining compliance in more detail, many pilots reported being already clear of the conflict when the RA was delivered, and thus did not comply with its instructions. For example, one pilot noted, "*I noticed that the TCAS depicted traffic was slightly behind us and to our left on my NAV display. The Captain immediately called something to the effect of, I've still got him, we're clear*" (ACN: 841821, 2009). Others reported using visual acquisition as justification for reported RA noncompliance: "*I elected, with the Captain's concurrence, to keep the descent so as to keep the MD80 in sight*" (ACN: 838285, 2009). In a few cases (15% of reported RA noncompliance, n=3) pilots viewed the TCAS instruction as directing them into traffic. "*Just then our TCAS gave an RA, 'Descend, crossing, descend.' The Captain said something to the effect of, 'I'm not doing that. He's descending, we'll descend right into him" and did not follow the TCAS RA*" (ACN: 854982, 2009). In 10% (n=25) of the analyzed reports, pilots conveyed partially complying with an RA. For these cases, pilots typically performed the vertical maneuver instructed by a TCAS RA, but added a horizontal component. These narratives suggest that the pilots believed their response was appropriate, and it followed standard procedure: "*As the Pilot Flying, the First Officer appropriately initiated a descending left-hand turn away from target per the aural and visual guidance from the TCAS*" (ACN: 802766, 2008).

Chi-square tests revealed that the relationship between reported compliance to TCAS RAs and any awareness of the location of other aircraft (i.e., from either the TCAS traffic situation display or visually out the window) is statistically significant, ($\chi^2(2, N = 233) = 10.990, p < 0.01$). Additionally, there exists a relationship between reported compliance and visual acquisition, without mention of the TCAS traffic situation display ($\chi^2(2, N = 233) = 7.291, p < 0.05$). As shown in Table 4, pilots reported 31 instances of maneuvering after receiving the precautionary TCAS TA and before receiving the RA. In these cases, pilots reported disconnecting the autopilot, performing a horizontal maneuver, or performing a vertical maneuver. For example, "*Pilot not flying reduced the scale of our TCAS display, and seeing traffic below, we reduced our descent rate to 300 FPM*" (ACN: 834304, 2009). During these maneuvers, pilots reported having awareness of the other aircraft 84% of the time. The relationship between a pilot's decision to maneuver on a TCAS TA and their reported awareness of the other aircraft, on the traffic situation display or visually is significant, ($\chi^2(1, N = 278) = 6.952, p < 0.01$). In the situation where a pilot reported maneuvering on a TCAS TA, it is likely he or she reported having awareness of the location of another aircraft.

Table 4.

Pilot reported response correlated to when a TCAS advisory was received

Reported Timing of Pilot Response to an Event (n=230)		
Pilot Maneuvered Before a TCAS Advisory	23	10.00%
Pilot Maneuvered After a TA and Before an RA	31	13.48%
Pilot Maneuvered After an RA	176	75.22%

In 78 of the narratives, pilots reported receiving collision avoidance instructions from ATC and in 26% of these reports (n = 20) pilots reported not complying with those instructions (shown in Table 3). Many pilots explained that they chose to follow a TCAS RA, which conflicted with air traffic instructions, and expressed a belief that the air traffic instructions would not resolve the traffic situation. For example, "*After we began the climb, ATC said to increase descent. Had we followed his instructions versus the TCAS RA, it would have ended in a midair collision*" (ACN: 852998, 2009). In some cases, pilots described relying on their awareness of other aircraft, based on the TCAS traffic situation display or visual acquisition. "*Traffic was depicted on TCAS, as we were converging traffic continued to head directly towards us and climbing up to our altitude. ATC issued a turning and climbing clearance to avoid conflicting traffic. I refused that clearance as I felt that would have caused a near midair or worse*" (ACN: 862593, 2009). Cases of "partial compliance" to air traffic instructions were also noted when pilots began to follow air traffic instructions but then received and complied with a TCAS RA(19%, n = 15). In several cases, the pilot chose to continue an ATC commanded turn while also following the RA vertical command. "*About 30 seconds went by before ATC told us to turn a heading of 270. I started the turn and the TCAS gave an RA to descend at a rate of 1500-2000 FPM. I turned off the autopilot, pulled the power levers to idle, and descended at a rate of 2000 FPM while continuing the turn*" (ACN: 849888, 2009).

As previously stated, collision avoidance cannot be examined by considering only the TCAS advisories and instructions or ATC traffic call outs and instructions. Throughout a collision avoidance situation, a pilot may receive and interpret the information presented (from their environment, by ATC advisories, or by TCAS advisories) and from that information, he or she may determine an avoidance maneuver is necessary. In the cases where a pilot chose to maneuver prior to receiving an RA or instructions from the controller, there is a high likelihood that the pilot had previously established awareness of another aircraft via the traffic situation display ($\chi2(2, N = 221) = 7.657$, $p < 0.03$ and ($\chi2(2, N = 256) = 10.403$, $p < 0.01$). For the instances when a pilot receives an RA, statistical analysis suggests, he or she will be more likely to comply with the TCAS instructions if he or she was first notified of the potential collision by TCAS, through either the RA itself or a TA ($\chi2(4, N = 229) = 14.059$, $p < 0.01$). Finally, if a pilot is directed to a traffic situation by either ATC or TCAS, they will most likely delay any response until prompted by an RA or ATC instructions ($\chi^2(4, N = 227) = 22.739$, $p < 0.01$ and $\chi^2(2, N = 267) = 9.266$, $p < 0.01$).

Pilots also frequently provided their assessments of the performance of TCAS and ATC (43%, n = 120), as shown in Table 5. In the case of TCAS, pilots focused their negative comments on the traffic situation display and their assumption of an error in the TCAS logic. For instance, one pilot explained that "*it was very hard to see [the other aircraft's] altitude as it was all cluttered together [on the traffic situation display]*" (ACN: 840426, 2009). Another pilot described his experience with TCAS, "*Descending into an airplane that is clearly descending? TCAS software clearly did not give appropriate guidance, nor did it self-correct when the initial guidance was so clearly wrong*" (ACN: 854982, 2009). Other pilots discussed feeling "overloaded" by the TCAS warnings. "*It was hard to hear instructions from ATC from the numerous RA callouts of the airplane and TA callouts which were shouting quite loud in our headsets – which made it difficult to understand the instructions given*" (ACN: 773537, 2008). Conversely, many narratives cited TCAS as the system that saved the day. "*The TCAS was what prevented this from being a potential midair*" (ACN: 802820, 2008).

Table 5.

Pilot perceptions of air traffic and TCAS performance

Perceptions of the Collision Avoidance System (N=278)			
Element	**Type of Perception**	**Sub Category**	**% of Sample**
ATC	**Positive (1.08% of all reports, n=3)**		
		Credit for save	100%
	Negative (35.25% of all reports, n=98)		
		Controller Assigned Collision Course	19.39%
		Lack of Traffic Call	16.33%
		Controller Error	12.24%
		Disinterest by the Controller	15.31%
		Lack of Situational Awareness	19.39%
		Other	17.35%
TCAS	**Positive (8.63% of reports, n=24)**		
		Credit for save	100%
	Negative (7.91% of all reports, n=22)		
		Unclear Information on TSD	13.64%
		TCAS Assigned Collision Course	50.00%
		Pilot was Overloaded	9.09%
		Other	27.27%

The overall perception of ATC as described in these ASRS reports was comparatively negative. Thirty-five percent of the analyzed reports included negative comments from the pilots regarding his or her interactions with the controller. Within these responses, pilots reported perceiving that the instructions provided by the air traffic controller, if complied with, would have resulted in a mid-air collision. These comments were common among the reports also describing noncompliance to air traffic instructions. Other pilots discussed the failure of the controller to provide traffic call-outs prior to the incident. "*I called the Tower after landing and told them it would have been helpful to get an advisory upon initial contact so we could have been more prepared. TCAS system was the only alert we had as Tower told us of traffic after the traffic had passed*" (ACN: 861931, 2009). Additionally, pilots noted instances where the air traffic controller appeared disinterested, unaware or not concerned about the traffic situation. "*It did not seem that the Tower Controller was very concerned about the event*" (ACN: 862312, 2009). Only three reports included positive comments, with one report stating "*THANK THE CONTROLLER and see if it could be counted as a 'save*'" (ACN: 858151, 2009).

Conclusions

The purpose of this study was to begin to explore the factors which affect pilots' agreement (or disagreement) and compliance (or noncompliance) with collision avoidance instructions. In an analysis of ASRS reports relating to collision avoidance and TCAS, pilots most often reported compliance with ATC and TCAS instructions. However, there were still many reports of noncompliance and partial compliance. In a large number of these cases, pilots perceived his or her actions as appropriate and aligned with standard procedure. Through further examination of these instances, the qualitative and quantitative findings indicate that pilots may perceive TCAS and ATC issued collision avoidance maneuvers as placing their flight into a near miss situation. Pilots also criticized ATC for not issuing traffic call-outs in a timely manner. In addition, the results suggest the information on the traffic situation display may be misleading. For instance, pilots' awareness of a second aircraft on the traffic situation display impacted their response to the potential collision. The information presented to a pilot about a collision avoidance situation is especially crucial in their decision making process. Both visual awareness and whether the pilot was directed to the potential incident by ATC, TCAS, or their environment were found to have a direct effect on a pilot's decision to maneuver.

In the complex environment which surrounds the collision avoidance system, it is necessary to understand the factors which affect a pilot's response to collision avoidance advisories and instructions. This study focuses on encounter conditions, traffic situation awareness, and other factors to begin to characterize patterns within pilots' interactions with TCAS and ATC. Future work in this area should consider a wider range of first-hand narratives, including those from the National Transportation Safety Board. Additionally, the results presented suggest further research is needed to determine different methods for presenting information in this dynamic and time-sensitive collision avoidance system.

Acknowledgements

This material is based upon work supported by a Cooperative Agreement (DTFAWA-10-C-00084) with the Federal Aviation Administration (FAA) Human Factors Research and Engineering Group, with Tom McCloy acting as technical manager.

References

Mellone, V. J. (1993). TCAS II, genie out of the bottle. *ASRS Directline,* (4). http://asrs.arc.nasa.gov/publications/directline/dl4_tcas.htm

Pritchett, A. R. (2001). Reviewing the Roles of Cockpit Alerting Systems. *Human Factors in Aerospace Safety, 1*(1), 5-38.

Rantanen, E. M., Wickens, C. D., Xu, X., & Thomas, L. C. (2004). Developing and validating human factors certification criteria for cockpit displays of traffic information avionics.

THE RELATIONSHIP OF AGE, EXPERIENCE AND COGNITIVE HEALTH TO PRIVATE PILOT SITUATION AWARENESS PERFORMANCE

Kathleen Van Benthem (kvbenthe@connect.carleton.ca)
Institute of Cognitive Science, Carleton University, 1125 Colonel By Drive
Ottawa, Ontario, K1S 0A3 Canada

Chris M. Herdman (chris_herdman@carleton.ca)
Visualization and Simulation Centre, Carleton University, 1125 Colonel By Drive
Ottawa, Ontario, K1S 0A3 Canada

Matthew Brown (mbrown5@connect.carleton.ca)
Visualization and Simulation Centre, Carleton University, 1125 Colonel By Drive
Ottawa, Ontario, K1S 0A3 Canada

Anne Barr (abarr@connect.carleton.ca)
Visualization and Simulation Centre, Carleton University, 1125 Colonel By Drive
Ottawa, Ontario, K1S 0A3 Canada

While cockpit navigational aids might assist general aviation pilots with awareness of terrain, aircraft and weather, the ability to maintain accurate and comprehensive cognitive representations of the aviation environment remains a critical cognitive aviation task. The present research reports on the relationship of age, pilot experience and cognitive health to situation awareness scores for 27 licensed private pilots in a simulated flight protocol. ANOVA and linear regression analyses revealed that *age*, recent *pilot-in-command hours* and the cognitive health indices of *working memory* performance and *reaction time* (auditory stimuli) were uniquely associated with three levels of pilots' situation awareness scores. Implications regarding private pilot situation awareness skill maintenance are discussed.

Despite assistance from cockpit navigational aids, the ability of the mind to form an accurate, comprehensive representation of the current and near future flight situation remains a critical cognitive task for all pilots. It is generally accepted that situation awareness builds upon three levels of cognitive processing (Endsley, 1988; 1995). The first level is the perception and processing of multi-modal stimuli into meaningful units of information. At the second level this information is assembled into a comprehensive schema of the aviation environment. The third level projects current information into a proposed future schema that anticipates a near future state of the environment.

Pilot situation awareness performance has been examined with respect to age, working memory processes and pilot experience (Endsley, 2000). Age has been proposed as a factor associated with all three levels of situation awareness performance due to the reductions in information processing speed, working memory capacity, inhibition and attentional processes that accompany normal aging (Bolstad & Hess, 2000). In a study by Coffey, Herdman, Brown and Wade (2007) older pilots were found to miss more critical events (nearby air traffic and instrument malfunctions) in a simulated flight environment than younger pilots. Similarly, Kennedy, Taylor, Reade and Yesavage (2010) reported that older pilots had a greater likelihood of landing in unsafe weather conditions and performed less well on flight control tasks than their younger cohorts.

Working memory, as described by Baddeley and Hitch (1974) is a model of short-term memory processes which encode, store and manipulate information from the environment in order to achieve a goal (e.g. memorize a phone number). Endsley (2000) suggests that working memory supports the main processes related to situation awareness. Morrow, Leirer, Altieri and Fitzsimmons (1994) found that recall of auditory material was most likely to be forgotten by older pilots and this phenomenon was not mediated by expertise. Morrow et al. suggest that declines in working memory were associated with poorer performance by the older participants, as it was the storage of the recent auditory information that appeared to significantly impact the performance of older participants. Corroboration for the findings involving working memory is reported by Taylor, O'Hara, Mumenthaler, Rosen and Yesavage (2005) who proposed that reduced ability to store and manipulate representations in working memory was associated with older age and that this effect resulted from declines in speed of processing and reduced ability to inhibit less relevant stimuli.

Some researchers have investigated the mediating effects of total flight hours and expertise on pilot situation awareness. Kennedy et al. (2010) studied older and younger novice and expert pilot groups and found age-related reductions in performance for most of the simulated flight tasks. Expertise did not reduce the effect of age, except in the case of banking performance in a holding pattern. Morrow et al. (1994) and Morrow, Menard and Stine-Morrow (1999) observed benefits of expertise for older pilots when reading back visually presented material, but no significant benefit was afforded by expertise for material with an auditory presentation.

While the literature indicates that age, experience and cognitive factors might be important factors in all levels of situation awareness, it is not clear how each factor uniquely contributes to performance in each level of situation awareness. In the present study we examined the relationship of age, experience and cognitive health to situation awareness scores of licensed private pilots using a simulated experimental flight protocol. We examined these factors individually at each level of situation awareness and found that age and recent pilot-in-command hours were associated with level 1; age, working memory, reaction time (auditory stimuli) and recent pilot-in-command hours were associated with level 2; and, working memory and reaction time (auditory stimuli) were associated with level 3 situation awareness performance.

Method

Participants

As part of a larger general aviation study (*N*=54) we examined the relationship of age, experience and cognitive health to three levels of situation awareness for 27 licensed private pilots (female pilots = 3).

Table 1.

Pilot Characteristics by Age Group

	Pilot Characteristics by Age Groups			
	Younger Pilot Group (*N*=16)		Older Pilot Group (*N*=11)	
	Range (Mean)	Standard Deviation	Range (Mean)	Standard Deviation
Age (years)	27-50 (40.1)	7.1	52-76 (59.4)	7.6
Length of Pilot Certification (years)	.5-25 (9.3)	7.2	.5-40 (22.3)	13.7
Pilot-in-Command in past 12 months (Hours)	10-60 (31.8)	16.2	0-32 (8.2)	11.9
Total Flight Time (Hours)	99-1000(322.6)	249.3	90-1309 (397.2)	370.4

Pilot ages ranged from 27 to 76 years and for purposes of analysis were categorized into younger (27 to 50 years) and older (52 to 76 years) pilot groups. Table 1 displays the mean age, years certified as a pilot, recent pilot-in-command hours and total flight hours logged. Participants in this analysis were certified to a maximum of private pilot and possessed a current private pilot license.

Procedure

Four main measures were administered in the order described here. The DCAT^{TM} and the auditory Perceptual Detection Test were our indices of cognitive health.

Pilot demographics and flight experience questionnaire. A pilot demographic and flight experience questionnaire was completed at the start of each session. Total years licensed and recent pilot-in-command hours (previous 12 months) were obtained from the pilots' flight logs. As expected, older pilots had flown more years than younger pilots (22.3 and 9.3 years, respectively), $F(1,25)=10.57$, $p=.003$. Older pilots, however, had significantly fewer recent pilot-in-command hours (8.2 and 31.8 hours, respectively), $F(1,25)=17.05$, $p<.001$.

Cognitive health measure 1: DCAT™. The DCAT™ is a computerized touch-screen system comprised of six individually scored sub-tests. Responses are made by touching the screen or depressing switches at desk top height. The DCAT™ produces *z*-scores (standard deviations from the age-group mean) that reflect both accuracy and timing of the responses (DriveABLE, 1997). Only subtest 6, the index of working memory, is reported on here. Subtest 6, *Identification of Driving Situations*, consists of selecting one of four response options to questions regarding judgment and situation awareness pertaining to brief (5 to 10 seconds) video clips of driving scenes. Information is presented in visual and auditory form during the video clip. The questions pertain to how the participant should respond in a situation, or what was the most dangerous element of the situation. While subtest 6 does not include scenarios found in general aviation, responses are supported by working memory and efficient integration of auditory and visual information. An analyses showed that older pilots ($M= -.16$) demonstrated significantly lower z-scores on the DCAT^{TM} Subtest 6 index of working memory, than did the younger pilots ($M= +.78$), $F(1,25)=5.20$, $p=.031$.

Flight simulator environment. Pilots flew a Cessna 172 medium-fidelity simulator. The simulator was an actual Cessna 172 cockpit with instruments and controlled linked with Microsoft® Flight Simulator ®. The system incorporates three large screens for approximately 120 degrees of horizontal visual field of view and 45 degrees of vertical field of view. All participants spent approximately half an hour in a warm-up phase of simulated flight in order to become accustomed to the flight simulator controls and to reduce learning effects. The experimental protocol required the pilot to fly three left-hand circuits in a low cognitive workload condition. For this condition the airfield was uncontrolled, the terrain was unremarkable and the pilot interacted with no other aircraft during the first circuit, one other aircraft during the second circuit, and two other aircraft during the third circuit. Pilots were required to provide details of their call sign, aircraft type and location at routine points during the circuit via radio communication. The simulated aircraft in the scenarios also provided this information through scripted radio calls.

Situation awareness tasks. In order to assess pilots' situation awareness a protocol, based on the Situation Awareness Global Assessment Technique by Endsley (1998; 2000) was utilized. SAGAT is structured around probes, which are interjected during the task and require participants to respond to questions pertaining to situation awareness. Probe questions were developed to reflect three levels of situation awareness. Pilots indicated (from memory) their current heading, airspeed and altitude, and the call sign and aircraft type of the other aircraft in the circuit (Level 1); where they believed they were in the circuit at the time of the simulator freeze and where all additional aircraft were located (Level 2); and where they predicted their aircraft would be in 2 minutes, and where the other circuit aircraft would be in three minutes (Level 3). The probe occurred at the beginning of the "base" leg of the circuit when there was only one other aircraft in the circuit. Raw scores for each level of situation awareness were converted to percent correct for comparison purposes.

Cognitive health measure 2: Auditory perceptual detection task. An auditory perceptual detection task was used to measure perceptual-motor responses to an external stimulus. The auditory perceptual detection task required thumb switch responses to randomly occurring beeps (range: 10 to 20 seconds) presented via a headset. Reaction time between stimulus onset and response was recorded for all correct hits.

Results

Age Differences and Situation Awareness

Figure 1 shows the mean situation awareness scores as a function of age. Separate one-way ANOVAs were used to compare younger vs. older pilots for each of the situation awareness scores. Level 1 situation awareness scores were marginally higher for younger (*M*=59) than for older pilots (*M*=48), *F*(1,23)= 3.3, *p*=.085. Level 2 situation awareness scores were significantly higher for the younger (*M*=98) than the older (*M*=82) pilots, *F*(1,23)=7.9, *p*=.01. There were no significant differences between level 3 situation awareness scores for younger (*M*=85) and older (*M*=79) pilots.

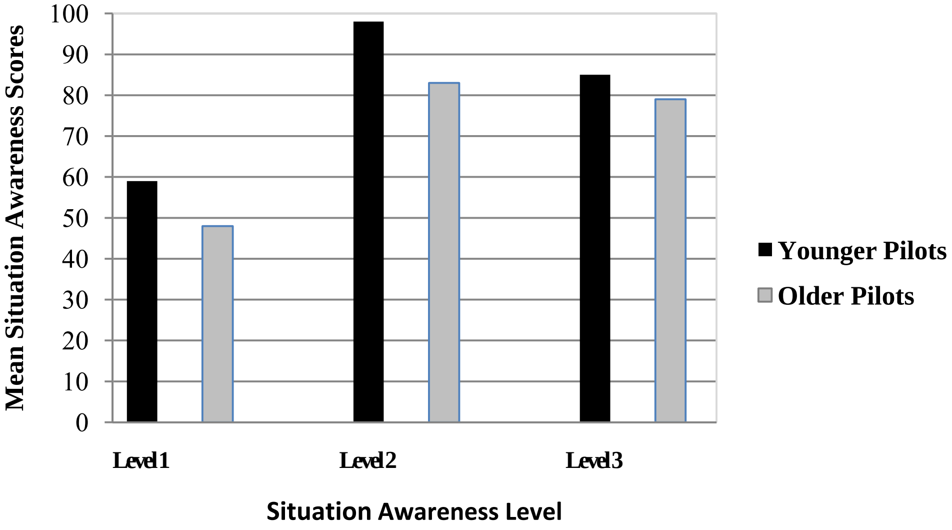


Figure 1. Mean Situation Awareness Performance by Age Groups and Situation Awareness Level

Correlation of Situation Awareness Performance with Pilot Age, Experience and Cognitive Health

As shown in Table 2, there was a moderate negative relationship of pilot age to level 1 situation awareness performance and a moderate positive relationship of recent pilot-in-command hours to level 1 situation awareness. For situation awareness levels 2 and 3, a moderate negative relationship was found between situation awareness performance and auditory reaction times and age. In addition, for both situation awareness levels 2 and 3, moderate positive correlations were found between working memory scores and recent pilot-in-command hours.

Table 2.

Relationship of Situation Awareness Performance with Pilot Age, Experience and Cognitive Health

Pearson Bivariate Correlation Analysis				
Situation Awareness Level	Perceptual Detection Task *Cognitive health: reaction time*	DCAT^{TM} 6 *Cognitive health: working memory*	Pilot Age	Recent Pilot-in-command Hours
Level1: *perception* -.49				.44
Level 2: *schema* -.50		.46	-.49	.40
Level 3 *future schema* -.41		.50	-.50	

Note All correlations are significant at the *p*< .05 level (two-tailed)

A partial correlation analysis of age and recent pilot-in-command hours with situation awareness levels 1 and 2 showed that when controlling for pilot-in-command hours the relationship between age and situation awareness remained marginally significantly (p<.1). However, when controlling for age, the pilot-in-command and situation awareness performance relationship was no longer significant (p>.1). This indicated that age might be a more reliable predictor of level 1 or 2 situation awareness performance than pilot-in-command hours.

Linear Regression Analysis

A model utilizing recent pilot-in-command hours and age accounted for 28% of the variance in level 1 situation awareness, F(2,22)=4.30, p=.027. A model predicting 54% of the variance in situation awareness level 2 scores was produced using cognitive health indices and age, F(3,20)=7.84, p=.001. Similarly, a model produced with our two cognitive health indices accounted for 35% of the variance in level 3 situation awareness scores.

Discussion

This study examined the relationship of pilot age, experience and cognitive health to three levels of situation awareness. Age and recent pilot-in-command hours uniquely predicted 28% of the variance in perception and storage of environmental stimuli as represented by level 1 situation awareness scores. Age and cognitive health measures predicted more than half the variance in level 2 scores, which reflected the pilots' comprehensive schema of the current situation. Approximately one third of the variance in scores pertaining to accurate representation of the future aviation situation was predicted by our cognitive health measures which incorporated working memory and reaction time to auditory stimuli.

The present findings are useful to stakeholders designing education programs with the goal of maintaining or increasing situation awareness skills for private pilots. Stakeholders developing private pilot education programs addressing situation awareness should consider the impact of older age, fewer recent flight hours by older pilots and reductions in cognitive health on the individual levels of situation awareness. Targeted strategies such as maintaining flight currency and engaging in cognitive activities to enhance working memory and reaction time might be useful in maintaining the situation awareness abilities of private pilots. These strategies appear important for maintenance of level 1 and 2 situation awareness, which is integral to forming accurate current mental representations of the environment and the basis for producing useful schema of future situations.

Acknowledgements

Funding for this research is provided in part by scholarships to K. Van Benthem from the Social Sciences and Humanities Research Council of Canada and the J. James Mackie Endowment for Graduate Scholarships, Ottawa, Canada. Research infrastructure was supported through the Canadian Foundation for Innovation and the Ontario Innovation Trust. We thank Andrew Staples of the Carleton University, ACE Laboratory for his simulation engineering expertise.

References

Baddeley, A. & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The Psychology of learning and motivation: Advances in research and theory, 8*, 47–89. New York: Academic Press.

Bolstad, C. & Hess, T. (2000). Situation awareness and aging. In M.R. Endsley & D.J. Garland (Eds.), *Situation awareness analysis and measurement*. Mahwah, NJ: LEA.

Coffey, E., Herdman, C.M. & Brown, M. (2007). Age-related changes in detecting unexpected air traffic and instrument malfunction. *Proceedings of the 14th International Symposium on Aviation Psychology, Columbus, OH.*

DriveABLE Testing Ltd. (1997). *Evaluations for at-risk experienced drivers*. DriveABLE Testing Ltd, Edmonton, Alberta.

Endsley, M.R. (1988). Situation awareness global assessment technique (SAGAT). *Proceedings of the National Aerospace and Electronics Conference (NAECON)*, 789–795. New York: IEEE.

Endsley, M.R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65–84.

Endsley, M.R. (2000). Theoretical underpinnings of situation awareness: A critical review. In M.R. Endsley & D.J. Garland (Eds.), *Situation awareness analysis and measurement*. Mahwah, NJ: LEA.

Morrow, D., Leirer, V. &Altieri, P. (1992). Aging, expertise and narrative processing. *Psychology of Aging, 7*(3), 376-388.

Morrow, D., Leirer, V., Altieri, P., & Fitzsimmons, C. (1994). When expertise reduces age differences in performance. *Psychology of Aging, 9*, 134-148.

Morrow, D., Menard, W. & Stine-Morrow, E. (1999). *The influence of aging and expertise on pilot communication.* Paper presented at the 10th International Symposium on Aviation Psychology, Columbus, OH.

Kennedy, Q., Taylor, J. L., Reade, G. &Yesavage, J. (2010). Age and expertise effects in aviation decision making and flight control in a flight simulator. *Aviation, Space, and Environmental Medicine, 81*, 489-497.

Taylor, J., O'Hara, R., Mumenthaler, M., Rosen, A. &Yesavage, J. (2005). Cognitive ability, expertise, and age differences in following age-traffic control instructions. *Psychology of Aging, 20*, 17–33.

Designated Pilot Examiner's Perception of Private Pilot Certification in General Aviation Advanced Technology Cockpits

Michael Friday
Oklahoma State University
Noble, Oklahoma
Frederick D Hansen Ph.D.
Oklahoma State University
Tulsa, Oklahoma

All aircraft require some degree of instrumentation. With the explosion of computer and flat panel display technology, the "glass cockpit" has entered aviation. In the past, pilots transitioned into the new technology after years of flight experience. These pilots already understood: the regulations, principals of flight, navigation, and the performance characteristics of their aircraft. The new glass equipped cockpits are now entering the environment of primary flight training where pilots are still learning the basics. Additionally, there is no structured or generally accepted methodology for training in advanced cockpits. The FAA has not established specific new guidelines for pilot-in-command for these aircraft nor much guidance to Flight Examiners (FEs) that must perform the actual certification of new pilots. The basic research question being explored is whether Designated Flight Examiners perceive whether the current FAA certification process is adequate for a private pilot to safely operate advanced technology in GA aircraft?

All aircraft require some degree of instrumentation in order to operate, but they differ in degree of complexity. The instrumentation can be categorized into: engine and aircraft performance, navigation, communication, and flight management. The complexity of the instrumentation is a function of the aircraft type and the flight environment. The necessity for "safety of flight" requires redundancy for many of these devices, which further increases cockpit complexity and density.

Until recently these individual equipment items were typically self-contained with their own displays and dedicated controls. Each was also individually connected to the central electrical buss. Other instruments, such as: altimeters, air speed and rate of climb indicator instruments utilized only vacuum and/or static pressure sources in order to operate. Despite being produced by different manufacturers, these devices had the same basic appearance, operated in a similar manner, and were certified under FAA Technical Standard Orders (TSOs).

The term "glass cockpit" refers to the flat panel displays common in modern laptop computers. The glass cockpit originated in military aircraft where it became a necessary in order to display the multitudes of information required to fly military missions. The finite real estate in military aircraft also made these compact/high density displays a necessity. With the advancement of microelectronics/microprocessors, glass cockpit displays have taken on even more functionality including flight management and mission planning. As the glass cockpit technology matured, these devices moved into the commercial sector. They first appeared in the Boeing 757 aircraft. Glass cockpits have become even more wide spread. In fact, these devices are now available: on new production general aviation aircraft and as retrofits kits for older general aviation aircraft.

Statement of the Problem

Recent advances in glass cockpit systems include enhanced situational awareness. The new features include: terrain, XM weather, traffic information service, airways, airport, and IFR approaches all of which can be displayed in a variety of formats and overlays. This variety and density of informational content, multiple display-formats, new symbology, and computational capability of Technology Advanced Aircraft (TAA) creates concerns about the effectiveness of current training methodology and certification process. "The common denominator in all these changes is the need to have an adaptable flight training system that will not only maintain but greatly improve the safety and utility of general aviation flight operations" (Wright, 2002).

The integrated glass systems have a vast amount of flight information, databases, and presets available. The databases required frequent updates. System software is also routinely updated to correct errors or add new features. Also these systems permit tailoring of their presentation at the user's discretion. A pilot using rental aircraft that had limited training on these systems could be placed in a challenging or confusing situation (AOPA 2007). This could lead to a potential safety hazard. The enhanced situation awareness within the cockpit further decreases the amount of time pilots to look outside the aircraft.

These high-tech aircraft are being placed into service at many flight-training facilities yet many training programs have not been adapted to reflect the required changes in learning strategy. In an Aviation Monthly 2004 Safety Report, it stated that pilots were "on their own" with respects to learning the new technology. The article points out that "the one size fits all approach" or the traditional method of training is no longer adequate. An APOA Safety Foundation report stated "training to use nontraditional avionics using traditional methods is not optimal" an goes on further to say "any training institution or CFI that attempts to do in-the-air training on advanced IFR GPS navigators, FMSs, or glass cockpit aircraft before having a through introduction and practice on ground via similar, ground powered aircraft, or at the very least with computer based instruction, is just not performing in the best interests of the client" (AOPA 2007).

The FAA has recognized the changing environment of advanced avionics and TAA and created the FITS program. While FITS is a step forward for training in advanced technology, it is not a mandatory requirement. Also, the FAA has begun updating certain publication to reflect the changing environment of TAA. Specifically, the Instrument Flying Handbook (FAA-H-8083-15A) has been revised to include the depiction and interpretation of flight information on glass systems. Discussions with multiple flight centers indicate no structured or generally accepted methodology for training TAA. According to one survey, reading printed media (manuals) are not found to be helpful with advanced avionics because they are not interactive (AOPA 2007).

Research Questions

To date, the FAA has not established specific new guidelines for pilot-in-command for these aircraft: no special endorsement or sign-off is required. Related to this matter, the FAA has provided little guidance to Flight Examiners (FEs) that must perform the actual certification of new pilots. Contact with several FEs in Oklahoma has showed this to be a concern.

An FAA Aviation News article addressed concerns of FAA's GA OPS inspectors or FE not having sufficient training in TAA aircraft to affective fly these aircraft and utilize the onboard systems. Specifically, the article pointed out that one manufacturer's glass system does not necessarily respond or display information the same as another's. Also, there is an inability to demonstrate certain system failures without experiencing a true failure.

The governing regulations concerning general aviation flight training are contained in the Federal Aviation Regulations (FAR) Part 61 for certification of pilots, flight instructors, and ground instructors and Part 141 for pilot schools (Wright, 2002). These regulations have not seen substantial changes since 1977 even though FAA officials have noted: "emerging changes in system safety philosophy and changes in NAS flight procedures and in flight technologies may call for a new approach to flight training" (Wright, 2002).

There is a need to know the extent to which Flight Examiners perceive a problem with the current private pilot certification process with respect to the operation of TAA that could have a negative impact on aircraft safety. The primary research question of this study is "Do Designated Flight Examiners perceive that the current FAA certification process is adequate for a private pilot to safely operate in the National Airspace System with the introduction of advanced technology in GA aircraft?"

Methodology

This research consisted of a survey instrument mailed to a random sample of the FAA DPE population of 1076 examiners. Every member of the DPE population had an equal chance of being selected (Fraenkel & Wallen, 2003, p270). The Stat Trek website was utilized to generate a table of 250 random numbers ranging from 1 to 1076.

A similar survey of all DPE's conducted by the FAA received a 64% response rate (Hackworth, King, Cruz, Thomas, Roberts, Bate, and Moore, 2007). It was anticipated that a similar response rate was achievable for this study because of the apparent high interest in this topic and the ability of this research to allow DPEs to voice their concerns and potential influence the FAA to take action.

The two-fold goal of the survey instrument was first to profile the general population of DPEs exposure to advanced technology. This profile identified what aircraft flown, what advanced technology, how they prepared themselves for exploiting the technology, number of practical test given, and their perceptions of the current requirements for pilot certification in advanced technology and its impact on safety. The second goal was to select DPEs for an in depth interview into what may be needed to improve the process for preparing pilot for the advancing technology.

Results

Surveys were mailed to one hundred randomly selected DPEs. Forty six valid responses were received. Based on the consistency of the results from the returned surveys, the decision was made that additional surveys would not need to be mailed. The results from the survey are discussed below.

The questionnaire profiled the DFEs' experience and qualification as FAA examiners. DPE experience range from 2 years to 61years with an average of 17.4 years. Within the past 12 months, DFEs average 70 practical tests each of which 8.8 were given in ADT/TAA. Worth noting, 17 of 46 DFEs or 37% reported giving no practical tests in ADT/TAA within the past 12 months.

The heart of the questionnaire dealt with DFEs' perceptions of the certification process and performing Practical Tests in Advanced Technology equipped aircraft. The questionnaire was sectioned to explore perceptions on: FAA guidance & regulations, safety impacts, knowledge & training requirements, performing practical test, and examiner training.

The DFEs were asked whether they were satisfied with the current FAA guidance and regulations for certifying new Airmen in Advanced Display equipped aircraft. Of the DFEs responding to the questionnaire, the majority or 64% are indeed satisfied.

The DFEs were asked whether the experience, practical test, and knowledge requirements were adequate. Again, the majority of DFEs responding agree that these requirements are adequate at 56%, 66%, and 67% respectively (*Figure 1*. It is worth noting that the adequacy of pilot experience requirements is approximately 10% below the practical test and knowledge requirements. Several DPEs commented in the remarks section of the questionnaire that the insurance companies often dictate the experience requirements for flying TAA/ADT equipped aircraft.

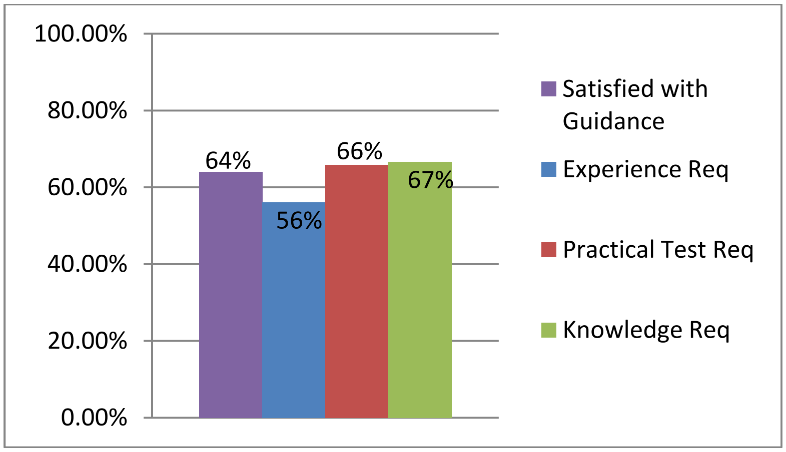


Figure 1. Adequacy of FAA guidance and regulations.

The DFEs were asked whether pilot licenses should specify either traditional or ADT equipped aircraft. The response was overwhelmingly "NO" at 86%. Asked if a Flight Instructor's logbook endorsement should be required for ADT equipped aircraft, 65% of DFEs responding agree to a one-time logbook endorsement but only 44% agree that the endorsement should be model specific (*Figure 2)*.

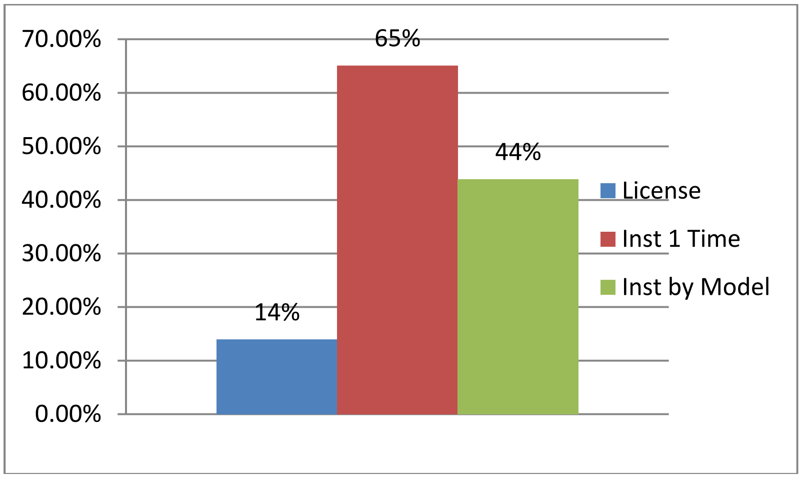


Figure 2. Technology endorsement recommendations.

With the ability of Advanced Display Technology to alert pilots of potential danger, examiners were asked whether they perceived an effect on risk-taking behavior with usage of this technology in various flight conditions. The DPEs believe that ADT effects is to fly in lower visibility (23 DPEs), fly in hazardous weather (21 DPEs), and fly lower altitude (12 DPEs), fly closer to terrain (11 DPEs), and closer to other aircraft (6 DPEs). 12 DPEs believe ADT had no effect on risk-taking behavior (*Figure 3)*.

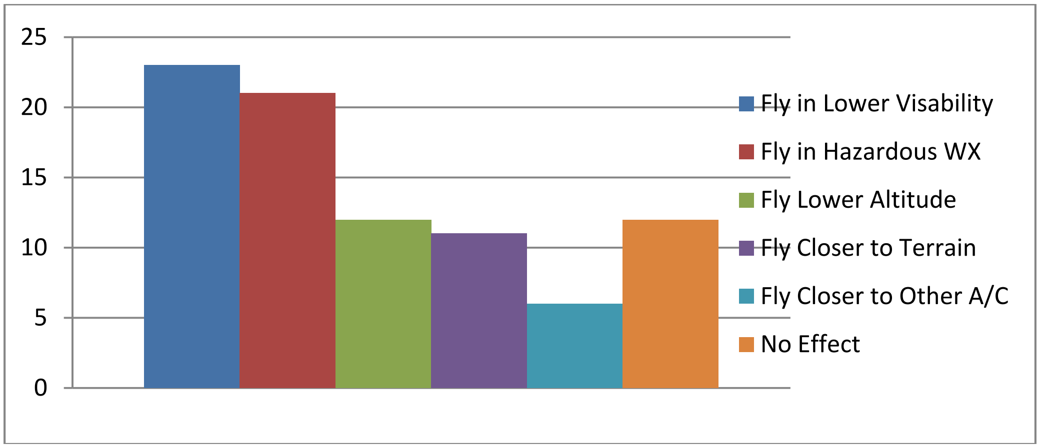


Figure 3. Flight examiner perceptions concerning safety of flight with ADT aircraft

Examiners were asked if they were required to demonstrate a specific feature or task associated with ADT during an FAA flight check and 48.8% agreed they had. When asked if they required demonstration a specific feature or task associated with ADT during a PT, 92.7% agreed they do require a specific demonstration. DPEs were asked if there were tasks or procedures that were difficult to perform/demonstrate in ADT equipped aircraft and 72.2% agreed there were task difficult to perform. One examiner commented that performing partial panel operations was difficult in ADT. Asked if a procedural/aircraft simulator would be more suitable for demonstrating certain features or task associated with advanced avionics, 71% agreed a simulator would be more appropriate.

Summary of the Findings

The DPEs are generally satisfied with the FAA's guidance and regulations pertaining to certifying new airmen specifically with new cockpit technology and ATA. The examiners overwhelming agree that a pilot should have a logbook endorsement for the technology flown. Most DPEs perceive that ADT has created additional risk taking on the part of pilots. It appears that a standardized ADT training curriculum is needed. DPEs need to ensure applicants meet the requirements of the Practical Test Standards especially with respect to conventional navigation skills. Examiners and Certified Flight Instructors need to take responsibility for their own training and be proficient in the technology flown.

Acknowledgements

The results of the initial survey confirm the need for the next stage of this research – qualitative interviews with flight examiners to further explore the perceptions of the flight examiners. When asked whether they would be willing to participate in a follow-up telephone interview 87 percent of those who completed the survey were willing to assist further and only five declined outright. The complete results of the survey along with the results of the interviews will be published in the dissertation of Michael Friday.

References

AOPA Air Safety Foundation (2007). Technologically advanced aircraft: Safety and training. Electronically retrieved from http://www.aopa.org

FAA Aviation News (May/June 2007). *FAA TAA Inspector training*. Electronically retrieved from http://www.faa.gov/news/aviation_news/

Fraenkel, Jack, R. & Norman E. Wallen (2003). *How to design and evaluate research in education* (5th ed.). Boston: McGraw-Hill.

Hackworth, C. A., King, S. J., Cruz, C., Thomas, S., Roberts, C., Bates, C., & Moore, R. (2007). *The private pilot practical test: Survey Results from designated pilot examiners and newly certificated private pilots.* Oklahoma City, OK: Civil Aerospace Medical Institute.

Krey, N. *Automation, airmanship, and the next generation of training*. Retrieved April 3, 2006, from http://s92270093.onlinehome.us/FlightDeck/pubs/saepaper/index.htm

FLIGHT INSTRUCTORS AND PILOT EXAMINERS PERCEPTIONS OF OLDER AND YOUNGER PILOT COMPETENCY AND SAFETY CONCERNS

Kathleen Van Benthem (kvbenthe@connect.carleton.ca)
Institute of Cognitive Science, Carleton University, 1125 Colonel By Drive
Ottawa, Ontario, K1S 0A3 Canada

Chris M. Herdman (chris_herdman@carleton.ca)
Visualization and Simulation Centre, Carleton University, 1125 Colonel By Drive
Ottawa, Ontario, K1S 0A3 Canada

Simon Garrett (simon@rfc.ca)
Chief Flight Instructor/Pilot Examiner, Rockcliffe Flying Club
Ottawa, Ontario

A national online survey was developed to examine differences in competency issues between younger and older pilot groups as perceived by flight instructors and pilot examiners. With their expertise in pilot competencies and a broad age-range of clientele, flight instructors and pilot examiners were considered a rich source of untapped data. Results indicated that respondents were more likely to identify *decision-making* and *problem solving* as a concern for younger pilots. *Error recognition* and *procedural knowledge* were more likely to be highlighted as concerns for older pilots. In addition, *physical ability to fly the aircraft* during training and currency review was three times more likely to be selected as a concern for older pilots. Patterns in perceived competency differences between age groups can inform targeted flight safety strategies.

This study examined how flight instructors' and pilot examiners' perceptions of competency and safety issues differed for younger (<25 years) and older (>59 years) general aviation pilots. Literature pertaining to general effects of age on cognition would specify cognitive slowing, disinhibition, and reduced working memory capacity (Bolstad & Hess, 2000), and provides insight into what factors might be associated with older pilot competency concerns. In general, poorer performance by older pilots is reported for tasks such as recall of air traffic control information (Morrow et al., 1992; 1994) and missed critical cockpit and nearby traffic events (Coffey, Brown & Herdman, 2007). Taylor, Kennedy, Noda and Yesavage (2007) conducted a three-year longitudinal simulator study of general aviation pilots and found that older pilots performed significantly lower at baseline and during the study period for following air-traffic controller commands, traffic avoidance, cockpit instrument scanning, and an approach and landing task. Kennedy, Taylor, Reade and Yesavage (2010) studied older and younger novice and expert pilot groups and found that older pilots had a greater likelihood of making unsafe landing decisions (due to weather); performed less well on flight control tasks than their younger cohorts; and, expertise did not reduce the effect of age, except in the case of banking performance in a holding pattern. Morrow et al. (1994) observed benefits of expertise for older pilots when reading back visually presented material, but no significant benefit was afforded by expertise for auditory material.

Federal Aviation Administration commissioned studies examining pilot age and accident rates were summarized by Broach (2004) and in cases where pilots over the age of 60 were included in the analysis there were increases in pilot accident rates associated with older age. Converging evidence is reported by Li, Baker, Qiang, Crabowski and McCarthy (2005) who found significant increased relative risk of aviation crashes for older, male, inexperienced pilots. Eyraud and Borowsky (1985) studied naval pilots and found that older pilots tended to display more procedural type errors than younger pilots: this pattern was considered a factor of over-confidence rather than cognitive aging effects (in this study older pilots were less than 60 years of age). In a study that included some pilots over the age of 60, Rebok, Qiang, Baker and Li (2005) found no discriminating pattern of crash-related errors between older and younger air taxi pilots.

The search for accident error pattern differences between younger and older pilot groups has not yielded consistent results to date. Taking a different approach, we looked for patterns of competency and safety concerns for older and younger pilot groups as perceived by flight instructors and pilot examiners. Flight instructors, in particular, might have the opportunity to develop a long-term professional relationship with their clientele and might fly with the same pilot over a number of years. This longitudinal perspective, along with the subject matter expertise gained from promoting and evaluating the competency of pilots of various ages, was considered a potentially rich source of data which might inform the issue of age-based patterns of pilot error.

Method

A national online survey of flight instructors and pilot examiners was launched in February 2011. The following results are based on 20 days of online data collection. The survey was disseminated to flight instructors and pilot examiners in the general aviation domain.

Participants

Survey dissemination methods targeted flight instructors and pilot examiners working in Canada. After disqualified responses were removed from the dataset, 61 complete responses were obtained. A *Human Resource Study of Commercial Pilots in Canada* (Air Transport Association of Canada, 2001) reported that there were approximately 500 general aviation flight instructors in Canada. This provided us with a conservative response rate of 12.2%. The median respondent was male (87%), between the age of 25 and 39 years (36.4%) and reported a mix of younger and older general aviation and commercial clientele (50%). Almost half the respondents had worked as a flight instructor or pilot examiner for ten years or more.

Procedure

Flight instructors and pilot examiners working in the general aviation domain were targeted using a national email campaign. The survey used a well-known internet-based platform and was accessible only via a hypertext link in the recruitment email. The survey recruitment email was successfully delivered to approximately 120 addresses (those not returned as undeliverable). To maximize recruitment of eligible respondents, the recruitment email encouraged qualified respondents to forward the survey invitation email to their flight instructor and pilot examiner colleagues.

Survey

The online survey consisted of 11 (for flight instructors) or 16 (for pilot examiners) questions. The questions and the response options were designed in collaboration with subject matter experts comprised of pilot examiners and flight instructors with more than 80 years flying experience and 10,000 hours of instruction time combined. Respondents were asked to reflect upon key times during flight training when they most seriously weighed the safety and competency concerns of their clients and to consider what concerns were frequently associated with younger and older pilot groups. Respondents were asked to select as many of the suggested areas of safety and competency, which they associated most with younger and then with older client groups, as they deemed appropriate. Pilot examiners were asked four additional questions which pertained to reasons why clients might fail a portion of their flight test, or reasons why clients might not be recommended for licensing after a flight test. Response options pertained to procedural and general aviation knowledge, physical ability to operate an aircraft, cognitive factors such as situation awareness, problem-solving, error recognition and decision-making, as well as confidence, our psychosocial factor. Two questions pertaining to currency review also included the response options of total flight and recent flight hours. Respondents were also asked if they agreed with the statement that they rarely or never saw competency differences between younger and older pilots: either generally (flight instructors and pilot examiners) or when completing a flight test (pilot examiners only).

Results

Flight Instructors and Pilot Examiners Perceptions of Age Group Differences in Flight Training or Flight Currency Review Responses

Sixty-nine percent of respondents indicated that they observed differences between younger and older pilot groups when considering competency and safety issues related to the training or currency aspects of flight instruction. Sixty percent of respondents indicated that they observed differences between younger and older pilot groups when considering competency and safety issues related to the pilot examination aspects of flight instruction.

Competency and Safety Concerns during Flight Training and Currency Review

As shown in Figure 1, there was a clear difference in how frequently respondents chose decision-making and judgement as a concern for each pilot group: with 63.5% of respondents selecting this response for younger pilots and only 27.7 % of respondents selecting this option for older pilots. As will be shown, this pattern was consistent throughout the survey. Problem solving was also perceived as a greater issue for younger pilots (34.6%) than older pilots (14.9%). Many more respondents selected physical ability to operate the aircraft as a concern for older pilots (51.1%) than for younger pilots (9.6%). Error recognition was also cited more often for older pilots (38.3%) than younger (26.9%). During the flight training phase, procedural knowledge concerns (but not crystallized intelligence such as knowledge of the aircraft parameters or general aviation domain specific facts) was cited more often for older (46.8%) than younger pilots (28.8%).

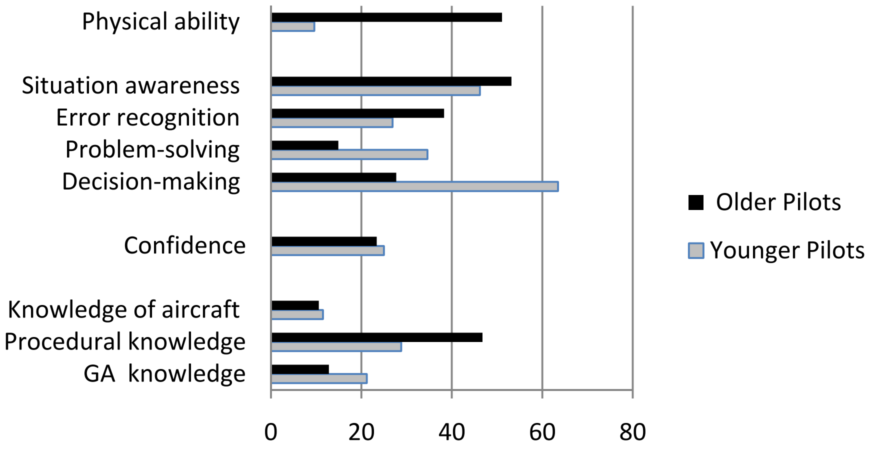


Figure 1. Competency and safety concerns during flight training for younger and older pilot groups.

Competency and Safety Concerns for Younger and Older Pilot Groups during Currency Reviews

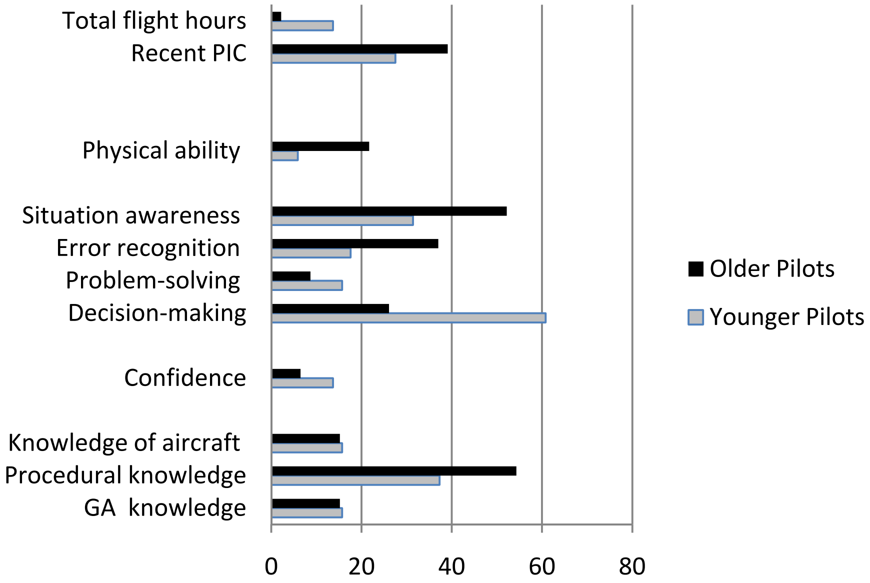


Figure 2**.** Competency and safety concerns for younger and older pilot groups during currency reviews

As shown in Figure 2, when considering currency reviews a similar pattern of flight training competency issues was found. Situation awareness and error recognition were greater concerns for older versus younger pilots: 52.2% and 31.4%, respectively for situation awareness; and 37.0% and 17.6% respectively for error recognition. As was also seen in Figure 1, respondents tended to identify problem solving and decision-making more often for the younger pilots than the older pilots. Decision-making and judgment was more than twice as likely to be a concern for younger (60.8%) than for older pilots (26.1%). As expected, total flight hours was not a concern for older pilots, however, recent flight time was more of a concern for older (39.1%) than younger (27.5%) pilots. As was found for flight training concerns, procedural knowledge was more often perceived as a concern for older (53.4%) rather than younger pilots (37.3%).

Reasons for Failing Part of or All of a Flight Test: Comparisons of Younger and Older Pilot Groups.

As shown in Figure 3, the age group differences for reasons a pilot must repeat a segment of the flight test, or reasons for not recommending licensing after the final flight test are more subtle than differences found during the previously described training or currency reviews (Figures 1 and 2).

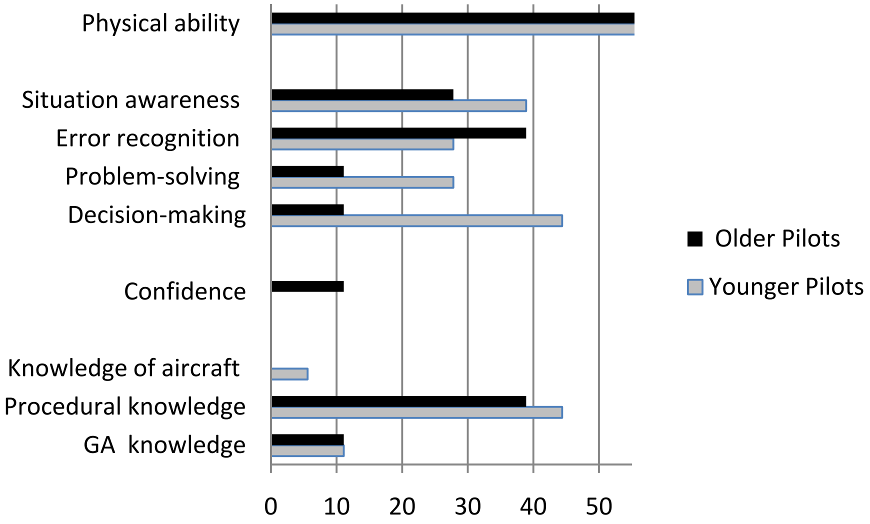


Figure 3. Reasons for failing part of a flight test: comparisons of younger and older pilot groups.

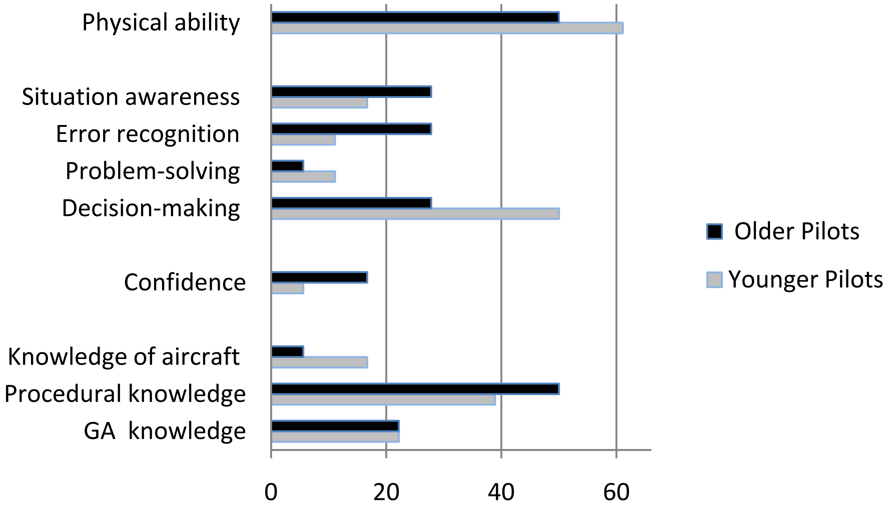


Figure 4. Reasons for failing a flight test entirely: comparisons of younger and older pilot groups.

Pilot examiners clearly highlighted decision-making and judgment as a reason for partial (44%) or full failure (50%) of a flight test for younger pilots. Situation awareness, error recognition confidence and procedural knowledge tended to be cited as reasons for failing a flight test moderately more frequently for older pilots.

Age Group Patterns: Significant Differences between Older and Younger Pilot Group Competency Concerns

Applying single factor analysis of variance we analyzed all response options across each competency question to determine if any of the response patterns showed statistically significant differences between the younger and older pilot groups. As shown in Figure 4, we found that error recognition (*M*=35.5% older, *M*=20.9% younger) and procedural knowledge (*M*=47.5% older, *M*=37.4% younger) were significantly more likely to be perceived as a concern for older pilots: $F(1,6)=9.49$, $p<.05$ for error recognition, and $F(1,6)=4.90$, $p=.068$ for procedural knowledge.

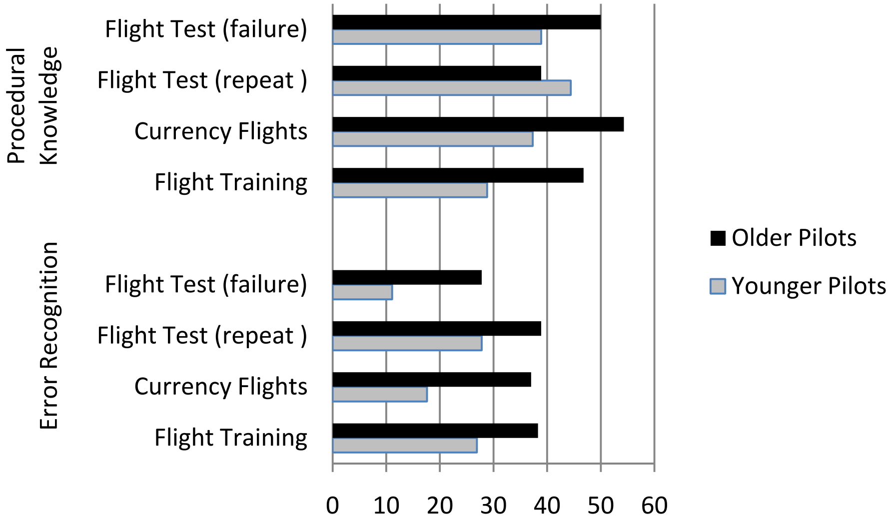


Figure 5. Procedural knowledge and error recognition selection patterns.

As shown in Figure 5, we found that respondents were very clear in their pattern of responses regarding decision-making and judgment: younger pilots were selected 54.7% of the time and older pilots 23.1%, $F(1,6)=27.11$, $p<.01$.

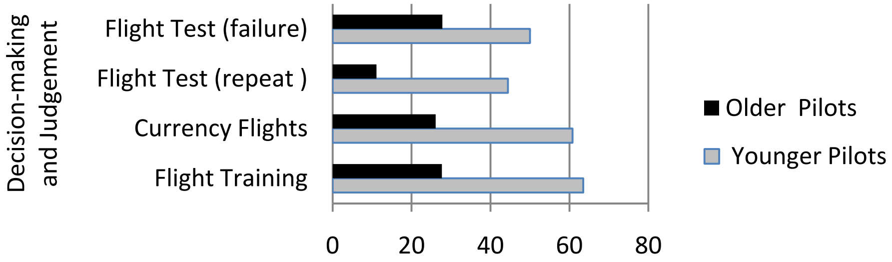


Figure 6. Pattern of decision-making and judgment selections for younger and older pilots.

Discussion

This study examined patterns in flight instructors' and pilot examiners' perceptions of competency and safety issues for younger and older general aviation pilots. Stakeholders developing strategies to improve safety for pilots across the adult life span can utilize these findings and target pilot groups with content tailored to their particular needs. The majority of flight instructors and pilot examiners indicated that they observed differences in competency and safety concerns for younger and older pilot groups. More specifically, the following patterns arose from the survey responses: for flight training, flight currency and pilot examination situations decision-making and judgement skills were significantly more likely to be identified as a concern for younger pilots; in contrast, respondents were more likely to identify procedural knowledge and error recognition as key concerns for older pilots. This differs from the results of Kennedy et al. (2010) who found more decision-making and judgement errors for older rather than younger pilots (this difference may arise from our very distinct pilot age groups: pilots under age 25 and pilots over age 60). Other dominant patterns identify physical ability to operate the aircraft and situation awareness skills as more of a concern for older pilot group during training and currency flights. These results are supported by Morrow et al. (1992; 1994), Coffey et al. (2007) and Taylor et al. (2007) who report age effects for errors in the procedural and situation awareness domains.

During pilot examination the majority of pilot examiners selected physical ability to operate the aircraft as a reason why a pilot might fail a portion of the flight examination for both younger and older pilots.

Confidence appeared as a concern more often for older pilots in pilot examination situations but not during training or currency review flights. We suggest that a strategic approach, incorporating age-related patterns of concern, to flight training and aviation skill maintenance will not only address safety from a pragmatic and efficient stance, but might also serve to reduce the anxiety and stress that can accompany flight currency and pilot examination processes, in particularly for older pilots.

Acknowledgements

Funding for this research is provided in part by scholarships to K. Van Benthem from the Social Sciences and Humanities Research Council of Canada and the J. James Mackie Endowment for Graduate Scholarships, Ottawa, Canada. We thank Jean René de Cotret and colleagues of the Rockcliffe Flying Club for their subject matter expertise and the Aviation Transport Association of Canada for their assistance.

References

Air Transport Association of Canada. (2001). *Human Resource Study of Commercial Pilots in Canada.* Retrieved February 1, 2011 from http://www.flying-colors.org/ATAC/2793-1.pdf.

Bolstad, C. & Hess, T. (2000). Situation awareness and aging. In M.R. Endsley & D.J. Garland (Eds.), *Situation awareness analysis and measurement*. Mahwah, NJ: LEA.

Broach, D. (2004). *Methodological issues in the study of airplane accident rates: Effects of accident and pilot inclusion criteria and analytic strategy*. (DOT/FAA/AM-04/8). Washington, DC: Federal Aviation Administration Office of Aerospace Medicine.

Coffey, E., Herdman, C.M. & Brown, M. (2007). Age-related changes in detecting unexpected air traffic and instrument malfunction. *Proceedings of the 14th International Symposium on Aviation Psychology, Columbus, OH.*

Eyraud, M. Y., & Borowsky, M. S. (1985). Age and pilot performance. *Aviation, Space, and Environmental Medicine, 56*, 553-558.

Kennedy, Q., Taylor, J. L., Reade, G., & Yesavage, J. (2010). Age and expertise effects in aviation decision making and flight control in a flight simulator. *Aviation, Space, and Environmental Medicine, 81*, 489-497.

Li, G., Baker, S., Qiang, Y., Crabowski, J. & McCarthy, M. (2005). Driving while intoxicated history as a risk marker for general aviation pilots. *Accident Analysis and Prevention, 37*, 179-184.

Morrow, D., Leirer, V. & Altieri, P. (1992). Aging, expertise and narrative processing. *Psychology and Aging, 7*(3), 376-388.

Morrow, D., Leirer, V., Altieri, P., & Fitzsimmons, C. (1994). When expertise reduces age differences in performance. *Psychology and Aging, 9*, 134-148.

Morrow, D., Menard, W., & Stine-Morrow, E. (1999). *The influence of aging and expertise on pilot communication.* Proceedings of the 10th International Symposium on Aviation Psychology, Columbus, OH.

Rebok, G., Qiang, Y., Baker, S., McCarthy, M., & Li, G. (2005). Age, flight experience, and violation risk in older pilots. *International Journal of Aviation Psychology, 15*, 363-374.

Taylor, J., Kennedy, Q., Noda, A. & Yesavage, J. (2007). Pilot age and expertise predict flight simulator performance: A 3-year longitudinal study. *Neurology, 68*, 648-654.

Taylor, J., O'Hara, R., Mumenthaler, M., Rosen A & Yesavage J. (2005). Cognitive ability, expertise, and age differences in following age-traffic control instructions. *Psychology and Aging, 20*, 17–33.

HUMAN FACTORS INVESTIGATION METHODOLOGY

Sarah Harris
Royal Air Force (RAF) Centre of Aviation Medicine
RAF Henlow, United Kingdom

There are many tools for assessing the Human Factors (HF) of accidents. Although these tools can aid the classification of factors in military accidents, they do not completely support the military *investigation* process. The RAF has therefore adapted the Human Factors Analysis and Classification System (HFACS) to develop a new Human Factors Investigation Methodology. This methodology combines some of the HFACS categories with the military operational process, providing a timeline to plot when accident factors may have emerged and a source framework to plot whether the factor emerged at the organisation, supervisory, task, equipment, environment or operator level. The benefit of the methodology is that it helps identify at which point the factor influenced one of the four key accident events; hazard entry, recovery, escape and survival, thus facilitating more effective recommendations. As it is generic, the methodology can be used for any investigation type, including preventative investigations.

This paper describes initial difficulties experienced with current investigation and classification tools and how a new methodology was developed to address these difficulties. Each part of the methodology is discussed in turn, along with recommendations for further development.

Requirement

The RAF initially used HFACS (Wiegmann and Shappell (2003)) to understand and classify accidents. Although this approach was effective at classifying accident factors and Reason's Swiss Cheese Model of Latent and Active Failures (Reason (1990)) was effective at understanding errors, neither approach was conducive for mapping to the standard military aviation process. Specifically, it did not reflect the timeline element of the military accident, which made it difficult to plot at which point each factor became a contributor, e.g. a pre-condition for an unsafe act can occur as a result of a previous unsafe act, making the standard template of supervision>preconditions>unsafe act difficult to use. Most importantly, the framework was not conducive for explaining to and integrating with the remaining Investigation Panel, who may have limited understanding of human factors and how HFACS may integrate with their part of the investigation.

Although other investigation tools are available, these are typically designed to provide inter-rater reliability and validity when offered across investigators to be applied to their accident scenario. Although these variables are important when non HF experts are attempting to understand their identified unsafe acts, these tools were found to be either too prescriptive and/or they focused mainly on the accident event itself. Again, they were not conducive for visual presentation to the remaining Investigating Panel as the models were visually different from their perception of the aviation process. Investigator buy-in and HF integration are key aspects to any investigation if the HF Investigation is to be effective.

To this end, as accidents and incidents occurred, a new methodology was developed. This included a timeline that visually represented the aviation process and allowed a more detailed understanding of the *route* and *timing* of risks and hazards.

Development

The Human Factors Investigation Methodology was developed to meet the investigation requirements. This methodology was developed throughout more than 30 military aircraft accidents, of which involved fixed-wing, rotary, multi-engine and unmanned aircraft and accident types from mid-air collisions to enemy action. This wide variety of accidents enabled the methodology to be refined as and when development issues arose.

As the methodology developed it was clear it would be based around some kind of timeline. It was also clear the factors from HFACS would also have to be represented as not only were these factors technically valid, they would also still be used in the post-accident phase of classification. This resulted in the identification of two axis: a timeline of events and a source for the type of factor. For this reason a *matrix* was developed to plot these two axis.

On the Y axis high level HFACS categories were plotted with some amendments:

1. A 'task' level was introduced to explicitly examine what it was the operator was trying to do at the time of the incident or accident. Specifically, this enabled explicit examination of the estimated margin for error for that task to assess its performance reliability.
2. The categories 'environment' and 'equipment' were separated out from technological environment and physical environment. This was easier to visually present to the remaining Investigation Panel.
3. Unsafe Acts or Acts were made into 'behaviours and actions' to include actions that were positive or benign, such as eating habits and secondary taskings.
4. The sub-categories within each level were broken out to include the most common areas within the military process. Not every sub-division within HFACS was represented at each relevant level as the methodology was designed to be an aid to detect routes, timings and relationships and not to be entirely prescriptive.

On the X axis the timeline was initially divided into 'Entry Conditions' to reflect those factors that occurred prior to the day of the event and the 'Accident Point', to reflect those factors that occurred on the day of the event. However, it was soon clear a distinction had to be made within the Accident Point to include a 'Readiness' stage, i.e. the stage the operator prepared themselves for what it was they were doing at the time of the accident. It then became apparent that the X axis timeline represented two 'Key Transition Points' (KTP): Entry Conditions > Accident Point and Readiness > In-Flight. These points were clear stages in the accident route when existing hazards and risks could have been mitigated. For this reason, estimation of the KTPs was included as a specific stage in the methodology.

Accounting for the X and Y axis, the matrix was termed 'Accident Route Matrix' (ARM).

During the refinement of the methodology additional factors were also realised. It became apparent there were certain questions that an HF investigator needed to answer to understand the sequence of events. By answering these questions it not only ensured the HF Investigator understood the accident but it also acted as a checklist for ensuring they had sufficient evidence. For this reason, the 10 important questions were termed 'Key Accident Characteristics' and also included as a specific stage in the methodology. Further, it became clear there were always four factors that occurred in the accident sequence: Hazard Entry, Recovery Response, Escape Response and Survival Response. For this reason, these factors were included in the ARM and 'Hazard Management' became another stage in the methodology. Finally, it was also realised that it was equally important to look at, from a HF perspective, how *detectable* the hazards were. For this reason, Detection became one of the final stages.

Results

The methodology consists of seven elements; Evidence Collection, ARM, Key Accident Characteristics, Hazard Management, KTPs, Detection and Advice (see Figure 1).

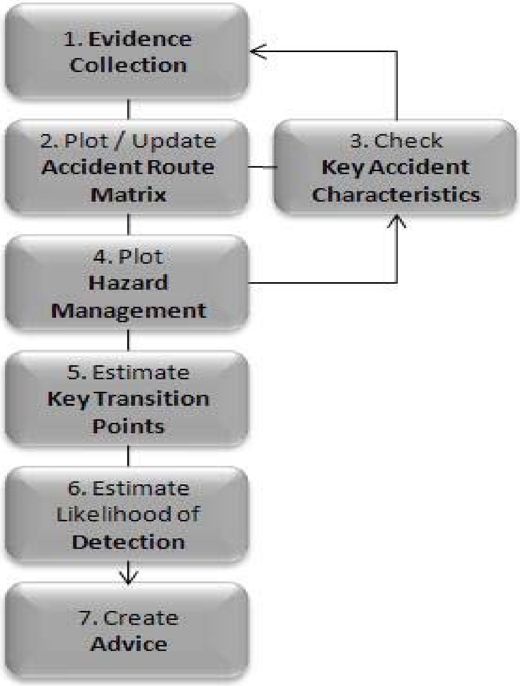


Figure 1.

All seven elements of the Human Factors Investigation Methodology

Evidence Collection

Subjective evidence. The initial stage of the methodology is to collect time-critical subjective evidence as soon as feasible post-crash. This usually involves one to one detailed interviews with the relevant operators using a semi-structured format. The types of questions usually follow the framework of the ARM (see Figure 2) but they can vary if required. Subjective data is collated for pilots within the aircraft (crew) and between aircraft (formation) to compare mission perspectives. Subjective data is also collated for other relevant parties such as Engineers, Supervisors and Air Traffic Controllers. Although not prescriptive, the subjective data can include performance data, which can be achieved by asking the operator to rate factors such as perceived arousal, demands, difficulty, frustration, pressure, awareness, understanding, communications and predicted success for each key phase of flight. For this methodology, individuals estimate these factors by pointing at a simple *Very Low > Very High* 5-point scale. The operator is also questioned on their point of focus, decisions and actions. Operators are then asked to estimate their views on Entry Condition and Readiness factors by pointing at a simple colour coded *Ideal > OK > Poor > Very Poor* 4-point scale. Although perceived estimates are not considered scientifically valid accounting for error in judgement, memory decay and distortion, the *change* in ratings, or the report of very negative ratings, has proved very useful for identifying areas for further assessment.

Objective Evidence. To validate subjective data, objective data from the crash site and Aircraft Data Recorder can be used to ascertain *actual* control inputs, crew communications, aircraft performance and hazard management.

Accident Route Matrix

The ARM forms the basis of the new HF methodology (see Figure 2). The ARM assists the HF investigation process by providing a framework to plot the subjective and objective data. Factors can be plotted by type (those at the Organisation, Supervision, Equipment, Environment or Operator level) and by time of effect (Entry Conditions or Accident Point). Data can be actually plotted using a blank ARM or the investigator can plot evidence in a report using the ARM as a framework. Once completed, the ARM can be read from top to bottom to understand how factors eventually influence Operator behaviours and operator Conditions, or it can be read left to right on each row to assess the route-cause of in-flight events, e.g. an unusual tasking in-flight may be explained by looking at typical tasking in the Entry Conditions and what tasks the operator is used to.

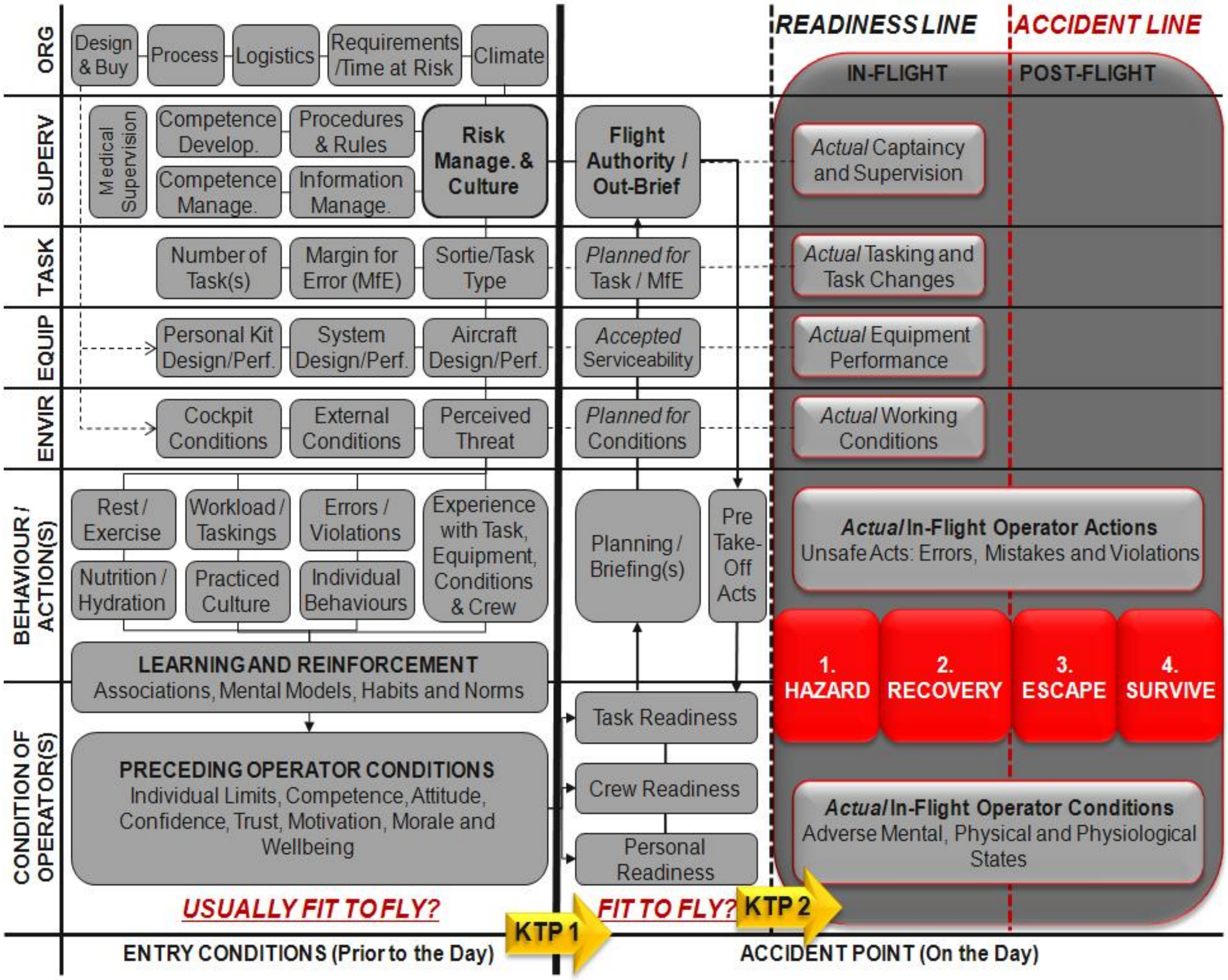


Figure 2.

Element 2 of the Human Factors Investigation Methodology: Accident Route Matrix

The ARM framework is based around a typical military aviation process. Within the Entry Conditions on the left-hand side, Organisation factors will influence the working environment in which the pilot will be working, Supervision and culture will influence the life the pilot will have on their Squadron and Equipment and Environment factors will influence their experience of flying.

In terms of the end state or effect of this, the Operator Behaviours section reflects what they have been doing, accounting for the conditions, tools and tasking, and how this has influenced their learning, associations and ultimately reinforcement of their behaviour. All factors eventually lead to Operator Conditions, i.e. the preceding condition the operator was in leading up to the accident day. Within the central Readiness column, the operator's Readiness will be influenced by the Entry Conditions + any over-night or on-the-day factors. At this point the operator may enter the planning and briefing stage and eventually, prior to take-off, their mission Readiness will be influenced by all Entry Conditions and Readiness factors. Post take-off the operator will enter the In-Flight stage and at some point, will pass through to the Post-Flight stage. This sequence can be applied to other operators, as it can be used to reflect generic on-shift Readiness and on-task performance.

Realistically in an accident sequence, subjective data is used to initially plot the centre and right-hand side of the ARM. As the investigation progresses, further subjective data may be collected and the ARM can be built up progressing to the left-hand side as appropriate. If available, further data from the aircraft, such as the Aircraft Data Recorder, Cockpit Voice Recorder and Head Up Display may be available to validate the ARM, turning any *suspect* or *reported* data into *actual* data.

Key Accident Characteristics

The investigator can now use the Key Accident Characteristics to ensure most significant data has been collected. Although this checklist is not exhaustive, it does act as a good prompt during the investigation. The Key Accident Characteristics include:

1. Was the operator suitably qualified?
2. Was the operator suitably current?
3. Had they recently practiced the task?
4. Had they been in the same situation before?
5. Was the task suitably planned?
6. Was the task achievable / reliably achievable?
7. Was the task justified (accounting for operational benefit v risk)?
8. Was the task correct (accounting for current procedures and rules)?
9. Was the task suitably authorised?
10. Was the behaviour intended and/or recognised?

Hazard Management

A Hazard Management assessment will help the HF Investigator identify how the crew managed the accident and how they progressed into the escape and survival stages. With both subjective and objective data, the investigator should be able to plot the points in Figure 3.

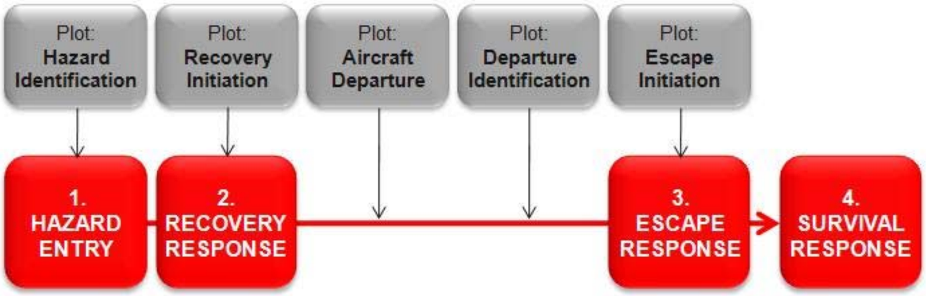


Figure 3.

Element 4 of the Human Factors Investigation Methodology: Hazard Management

Key Transition Points

Once the ARM has been completed and the HF Investigator is confident all available data has been assessed, the two KTPs can be qualitatively estimated (see Figure 2 for location of KTPs).

1. KTP 1 is based on the identified Entry Conditions and addresses the question - what was the level of risk carried for the entire Squadron, i.e. were all pilots at risk of this accident?
2. KTP 2 is based on the identified Readiness stage and addresses the question - what was the level of risk carried for the specific crew, i.e. were the crew at risk of this accident before they even took off?

These questions will help identify at which point the hazard/risk developed to a point it increased the probability of the accident being investigated.

Detection

Once the ARM and KTPs have been completed, the HF Investigator should use knowledge of information processing and perception to qualitatively estimate whether the hazards and risks could have been realistically detected by either the crew and/or the supervisory chain. Such information facilitates a better understanding of why the hazards and increased risks existed in the first place.

Advice

The final stage of the methodology is to provide advice based on the outputs of the investigation. Advice can be at any of the ARM levels and can state its probable improvement of one or more of the four key accident factors - Hazard Entry, Recovery Response, Escape Response and Survival Response.

Conclusion and Application of the Human Factors Investigation Methodology

The Human Factors Investigation Methodology has been developed to the stage it is usable for all accident sequences and aircraft types. It is also usable for all levels of operators, e.g. for a supervised individual and their supervisor. It can be used preceding use of HFACS and it can be used as a *preventative* methodology. The RAF currently use the methodology to conduct Operational Events Analysis, which includes application of the Human Factors Investigation Methodology but *before* an accident has happened.

However, for the methodology to be distributed as a tool, scientific validation would need to be conducted to assess factors such as inter-rater reliability. Further, should the methodology be required to produce quantitative results, the KTPs and Detection elements would need further work to ascertain if risk and error probabilities could be reliably used.

References

Reason, J. (1990). Human Error. New York Cambridge University press.

Wiegman, D. A., & Shappell, S. A. (2003). A human error approach to aviation accident anlaysis: The Human Factors Analysis and Classification System. Ashgate Publishing.

IDENTIFICATION OF AIR TRAFFIC CONTROL SECTORS WITH COMMON STRUCTURAL FEATURES

Annie Cho[1], Jonathan Histon[1], Emilio Alverne Falcão de Albuquerque Filho[2], R. John Hansman[2]

[1] Department of Systems Design Engineering, University of Waterloo

[2]Department of Aeronautics & Astronautics, Massachusetts Institute of Technology

In order to identify sectors supporting a minimum differences approach to generic airspace, traffic patterns in 360 high-altitude sectors were examined for common structural features. These structural features are used as the basis for two approaches to classifying current air traffic control sectors into groups which are expected to be similar to each other, and hence a basis for near-term deployment of generic airspace. The first classification approach is a holistic approach, based on emergent sector-wide traffic patterns in order to identify groups of sectors with shared structural features. The second, a decompositional classification approach, proposes using basic structural features (e.g. flows, merges, crosses) as building blocks, and classifies sectors based on combinations of those features. Initial classification results are presented for the holistic approach, and challenges and key steps are presented for the decompositional approach.

Introduction

In the United States the typical Certified Professional Controller (CPC) will maintain qualification on only a limited number of sectors (typically 5-7) within an area of specialization (Histon and Hansman, 2008). In order to move to and control sectors within a different area of specialization, significant and timely retraining activities are required. Consequently, staffing flexibility is limited and it is difficult for the Federal Aviation Authority (FAA) to respond to seasonal variations in staffing demands. Developing generic airspace, or airspace that is similar enough that no or minimal retraining is required to transfer a qualified controller between sectors is a possible means to address this challenge.

Analysis by MITRE has identified high altitude airspace as having the least number of airspace knowledge items (Levin, 2007). This, combined with the more homogenous mix of aircraft types and capabilities, makes high-altitude airspace attractive for an initial investigation of the potential of generic airspace. One way to deploy generic high-altitude airspace is to create sets of standardized or similar sectors. In order to create such set, though, it first must be determined what it means for a sector to be similar to others. The greater the standardization, the greater the flexibility; however, this comes at the cost of locally adapted sector-specific procedures and operations that provide locally tailored and more efficient operations. To balance these competing pressures, a minimal differences training approach to the development of generic airspace is being investigated. In this approach, classes of sectors are identified that could be made similar, but not necessarily identical, and controllers would then receive short targeted training on the relevant differences between the generic sectors in a particular class. This approach builds on our previous work which has identified the importance of supporting easily transferable mental models and abstractions in air traffic control (ATC) (Histon and Hansman, 2008).

To assess the potential of the minimal differences training approach, this paper presents a National Airspace System (NAS)-wide analysis of the similarity of existing high-altitude sectors. The analysis examines sectors from the perspective of structure, or “the physical and information elements that organize and arrange the air traffic control environment” (Histon and Hansman, 2008). First, key structural features that are thought to play a significant role in defining groups of similar sectors were identified. These structural features were then used as a basis for two classification methods for identifying groups of similar sectors. A holistic classification method, based on sector wide traffic patterns, is illustrated and preliminary classification of sectors is presented. A second method, based on explicitly decomposing sectors using the identified structural features is then presented.

Generic Airspace and Previous Work on Airspace Classifications

Interest in generic airspace has been driven by both experimental and operational considerations. Many experiments investigating new ATC operational concepts have been performed using generic sectors. Typically, a generic high-altitude sector is used to ensure that participants have similar backgrounds and previous experience with the airspace in the experiment, to avoid confounds associated with previous experience. These sectors are specifically designed to replicate the characteristics of a typical high-altitude sector, and to facilitate rapid learning for experimental participants (Guttman et al., 1995; Guttman and Stein, 1997). Generic sectors are also envisioned as part of new high-altitude airspace concepts, such as the Dynamic Airspace Super Sectors (Alipio et al., 2003).

However, classification and identification of similarities between existing airspace sectors has not been extensively studied. Christien (2003) used a Complexity Index (CI) as a basis for establishing common groups of sectors in European airspace. Much more research has focused on identifying complexity factors (Laudeman et al. 1998). Typical complexity factors include: aircraft density, the proportion of aircraft changing altitudes, sector size, and sector shape (comprehensive complexity factors lists can be found in reviews by Hilburn, 2004; Majumdar and Ochieng, 2001). Kopardekar and Magyarits (2003) found significant differences in the relative importance of complexity factors between en route facilities in the United States.

The breadth of potential complexity factors introduces the need to determine how the factors can be combined. Previous approaches have used weighted averages of identified factors to produce an overall complexity index (e.g. Kopardekar and Magyarits 2003, Laudeman et al. 1998). Other approaches have used algorithmic methods based on cluster identification techniques; groups of sectors (clusters) are identified by recursive splitting until the homogeneity of the resulting groups satisfy a pre-determined threshold (Christien, 2003).

These approaches, however, do not explicitly examine the potential of identifying common groups of sectors that would require reduced or minimal training for a controller to easily move amongst them. Structure has been shown to play an important role in controller cognitive complexity (Histon and Hansman, 2008) and is a useful perspective from which to identify similar sectors.

Approach

In order to identify key structural features and common groups of sectors, radar track data, collected through the Enhanced Traffic Management System (ETMS), were analyzed for two seven day periods (07/13/2009-07/19/2009 and 9/21/2009-9/27/2009). Radar tracks were plotted for flights that spent at least 10 minutes inside each high-altitude sector. As a first step, the radar track maps were reviewed for key structural patterns by manually going through the 360 high-altitude NAS-wide sector radar-track maps. After identifying several key structural patterns, these patterns were used as the basis for two approaches for classifying sectors into common groups. This classification exercise was first done by examining radar-track maps both on screen and using printed cards. Printing radar-track maps as mini-cards allowed easy physical maneuvring of the cards, which allowed quick grouping/de-grouping as different classification approaches were explored, similar to card-sorting techniques used in feature / requirement classification techniques used in design fields (Lafrenière et al., 2000).

Structural Features

To identify sector-specific elements and procedures that need to be similar and those that could be different in generic sectors, radar track maps depicting current sector operations were reviewed for key common structural features and recurring patterns. From this review, five key patterns were identified; Patterns 1 and 2 are consistent with previously reported structural patterns (e.g. Histon and Hansman, 2008) and are only briefly described.

Pattern 1 - Standard Flows. In most sectors, there are one or more distinct standard flows (Figure 1). Standard flows are the foundation for simplifying abstractions used by controllers to reduce cognitive complexity (Histon and Hansman, 2008). Hence, commonalities in the standard flows between sectors are thought to be important factor for identifying similar sectors.

Pattern 2 - Critical Points. Another key feature identified in multiple sectors was the presence of critical points, where flows cross, merges, and/or split (Figure 2). The relative location of the critical points, especially with respect to each other and sector boundaries, as well as the type (e.g. merge point vs. crossing point) can significantly impact cognitive complexity (Histon and Hansman, 2008; Hilburn, 2004). Similar to standard flows, critical points also support simplifying abstractions and are important considerations for identifying similar sectors.

Pattern 3 - Flow Trajectory Change Points. New structural features were identified in the review. Trajectory change points associated with flows (Figure 3) typically occur due to special conditions/restrictions such as keeping the flow within the lateral and/or vertical boundaries of the sector. The location of trajectory change points relative to other flows and the sector boundary is an important consideration for assessing sector similarity.

Pattern 4 - Vertical Handoffs. The radar track analysis also identified a key feature associated with aircraft being handed off and transitioning into or out of sectors vertically. In Figure 4, two flows can be seen terminating in the middle of the sector. The locations of the vertical handoffs, and their relationship with other flows in the sector (e.g. climbing or descending below a crossing flow) will likely affect how similar these characteristics need to be in order for two sectors to be considered similar.

Pattern 5 - Common Maneuvering Patterns. Two common maneuver patterns were also identified: the race-track holding pattern illustrated in Figure 5, and the path stretching pattern illustrated in Figure 6. Both of these features require free maneuvering airspace to be present in the sector. The location in the sector, and how it interacts with other elements such as military airspace, will likely affect how similar these features need to be in order for two sectors to be considered similar.

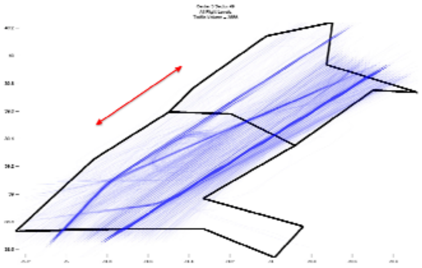


Figure 1. Standard flow

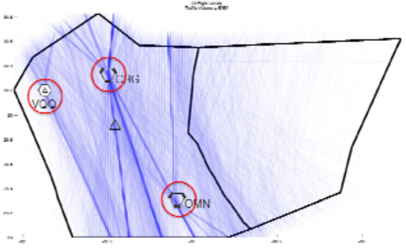


Figure 2. Critical points

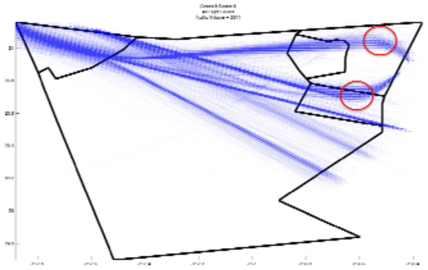


Figure 3. Flow trajectory change points

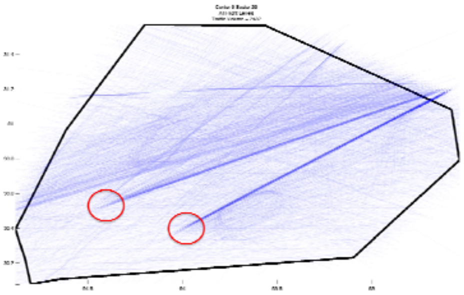


Figure 4. Vertical handoffs

Classification Approaches

The identified structural features provide a basis for identifying potential generic sectors. Sectors with similar structural features support similar simplifying abstractions, and have similar types of knowledge associated with them. These structural similarities should thus support the minimal differences approach to generic airspace. Two distinct approaches to identifying sets of sectors with common structural properties have been developed. The first, a holistic approach, is based on the overall structural appearance of a sector, without explicit accounting for individual structural features. The second, a decompositional approach, uses individual structural features as building blocks and explicitly accounts for combinations of structural features to classify sectors into common sets.

Holistic Classification Approach

The holistic approach identifies similar sectors based on the overall structural features. The same radar traffic maps used to identify structural features were used to categorize the 360 high-altitude sectors based on common patterns in the number and interactions between flows and perceived density of the flows. Each sector was classified into only one class. Table 1 shows the frequency of sectors in each of the 16 sector classes. Example radar traffic maps for two classes are shown in Figure 7. The value in the centre of each classification cell in Table 1 represents the percentage of sectors categorized into that class. The top 12 classes in the table represent different configurations of number and intensity of flows and non-standard flow traffic. Approximately 57% of the 360 NAS-wide high-altitude sectors were classified into these 12 classes.

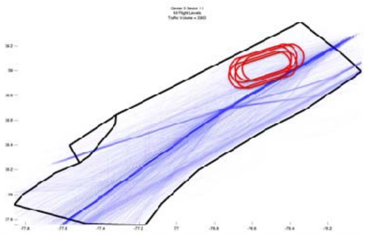


Figure 5. Holding pattern

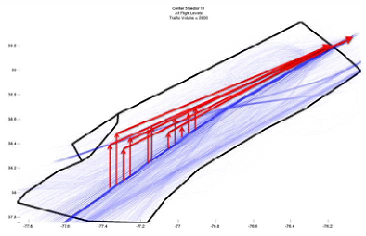
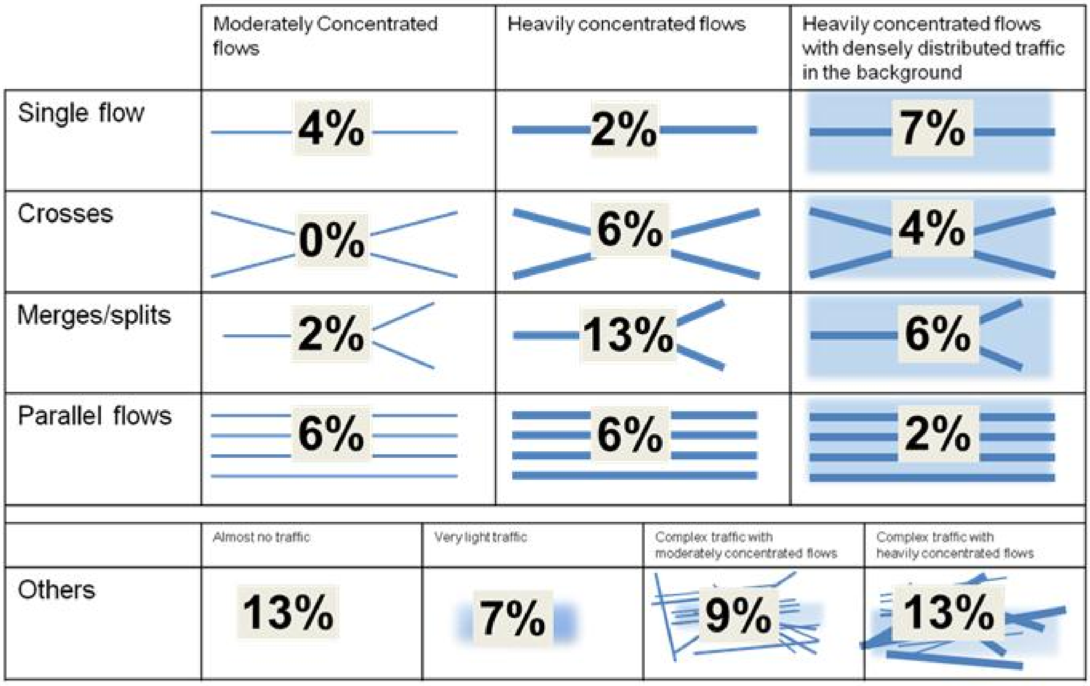


Figure 6. Path stretching pattern

Table 1. Visual and Canonical Guide for the Holistic Classification Approach



	Moderately Concentrated flows	Heavily concentrated flows	Heavily concentrated flows with densely distributed traffic in the background
Single flow	4%	2%	7%
Crosses	0%	6%	4%
Merges/splits	2%	13%	6%
Parallel flows	6%	6%	2%

	Almost no traffic	Very light traffic	Complex traffic with moderately concentrated flows	Complex traffic with heavily concentrated flows
Others	13%	7%	9%	13%

Not all sectors had structural features that cleanly mapped into these 12 classes. Four additional classes listed under "Others" in Table 1, were used for sectors with unique structural features. These sets are distinguished by the density of background (non-standard) traffic and the density of the primary flows. 13% of sectors had extremely low traffic counts and 7% of sectors had no dominant structural features with the traffic spread out evenly throughout a larger area of a sector rather than forming a concentrated flow as a route. 22% of sectors were composed of multiple dominant structural features (e.g. two crosses with a merge and a parallel flow). These sectors were classified as belonging to sets with *complex traffic with moderately or heavily concentrated flows.*

Several challenges were identified in using the holistic approach to classification. No attempt was made to account for altitude differences in aircraft trajectories. Including altitude distinctions would lead to additional classes being identified; features such as crosses would have different training implications if they are generated by traffic at varied and procedurally segregated altitudes. The representations used did not distinguish between directions of flight, making it difficult to definitively distinguish between merges and splits; other contextual cues can be used, but for the purpose of this preliminary analysis a single class was identified. Given the obvious differences between

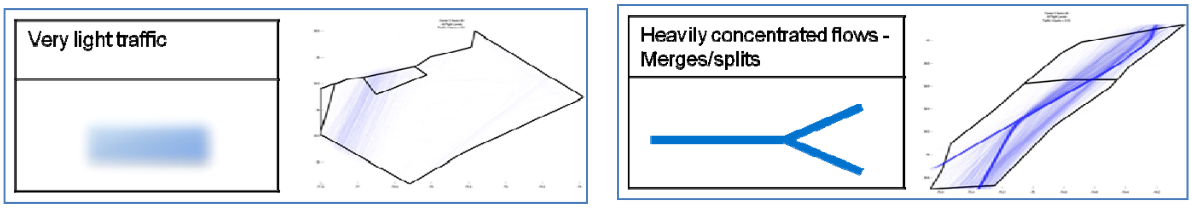


Figure 7. Example radar track maps: "Others – very light traffic" (left), "Merge / Split with heavily concentrated flows" (right).

merge and split operations, additional classes should be identified. Finally, the analysis was conducted using raw sector definitions; operations dictate that sectors are routinely combined during different parts of the day.

Nevertheless, Table 1 represents an initial break-out of the types of traffic patterns, and preliminary estimates of the relative frequency, that can be found across sectors in the NAS. The classes that are identified provide a basis for identifying groups of sectors that are expected to be similar enough to support a minimal differences approach to training in order to support controller qualification across the sectors in the class.

The Decompositional Classification Approach

Two shortcomings with the holistic approach motivated consideration of alternative approaches. First, the classes identified do not explicitly include the effects of key structural features such as the presence of standard maneuvering patterns. In addition, over 20% of sectors were classified as "complex traffic" sectors; however, there may be important opportunities for generic airspace sectors based on similarities between sectors within this class.

To address these challenges, a decompositional classification approach is being developed. In this approach, sectors are decomposed into elemental structural features. Similar sectors are identified based on the patterns of combinations of the elemental features. Examples of elemental features are shown in Figure 8: a crossing flow, a merge, parallel flows, a flow turn point, and a standard holding pattern. In Figure 9 three elemental features, a crossing flow, merge/split, and a holding pattern are identified in an example sector.

Figure 8. Examples of elemental structural features for the decompositional classification approach

In order to identify classes of similar sectors, combinations of these elemental features are added together using a notional sector algebra (Figure 9). This research is in a preliminary stage and several techniques are currently being investigated. The simplest is based on using a weighted combination of features, similar to the complexity based classification described by Christien (2003). Weights for elemental features in Figure 8 can be estimated based on their relative importance (e.g. cross assigned weight of "1" unit, a hold a weight of "2" units etc…) and then a Structure Score determined from the weighted sum of elements in each sector. Classes of sectors can then be determined by grouping sectors with similar Structure Scores. This has the advantage of simplicity and consistency with previous methods, but also loses much of the information gained by explicitly decomposing into individual elements. More sophisticated techniques, based on multi-dimensional clustering techniques and other formulations of multi-class classification algorithms are also being investigated.

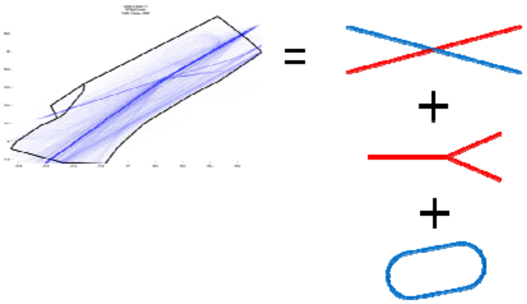


Figure 9. Elemental features in a sector (decompositional classification approach).

Whichever technique is used to aggregate the individual elements, there are several key challenges. The relevant elemental features to be used in the decomposition must be identified. Ideally these should form a mutually exclusive set, however, there will be some overlap as some structural features (standard flows) are integral parts of other structural features (crossing points). The relative importance of a structural feature for similarity may also be dependent on the spatial relationship with other structural features; in addition to number of features, relative distances, intensity, and frequency of use may need to be included as part of the decomposition.

Summary

Radar track data for 360 high-altitude sectors were used to identify five key structural features; similarities in structural features provide a basis for identifying classes of generic sectors. Similarities between sectors in the same class would support a "minimal differences training" approach to the deployment of generic airspace. Two distinct methods of using structural features to classify sectors were presented. The holistic approach, based on assessing the overall structural appearance of a sector, was used to identify 16 classes of high-altitude sectors. The second, decompositional, approach was proposed as the basis for comparative analyses of structural features of the sectors. The identification of classes of sectors with similar structure provides a basis for assessing the potential of near-term deployment of generic airspace. Having identified classes of sectors, future work will be further refining the classes, and using human-in-the-loop experiments to verify the relevance of the identified differences.

Acknowledgements

The authors gratefully acknowledge the FAA for funding this research through FAA Grant 06-G-006.

References

Alipio, J., Castro, P., Kaing, H., Shahd, N., Sheizai, O., Donohue G.L., and Grundmann K.. (2003) *Dynamic Airspace Super Sectors(DASS) as High-Density Highways in the Sky For A New US Air Traffic Management System*. in AIAA/IEEE Digital Avionics Systems Conference, 12-16 October 2003

Christien, R. (2003). *Air Traffic Complexity Indicators & ATC Sectors Classification*, 5th USA/Europe Air Traffic Management R&D Seminar, Budapest, Hungary. 23-27 June 2003.

Guttman, J.A., Stein, E.S., Gromelski, S. (1995). *The Influence of Generic Airspace on Air Traffic Controller Performance*, Report # DOT/FAA/CT-TN95/38, William J. Hughes Technical Center: Atlantic City International Airport.

Guttman, J. A., Stein, E. S. (1997).*En route generic airspace evaluation* (DOT/FAA/CT-TN97/7). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Hilburn, B.G. (2004) *Cognitive Complexity in Air Traffic Control: A Literature Review*, EEC Note No.04/04: Eurocontrol.

Histon, J.M. and Hansman, R.J. (2008).*Mitigating Complexity in Air Traffic Control: The Role of Structure-based Abstractions*. Massachusetts Institute of Technology: International Center for Air Transportation.

Kopardekar, P., Magyarits, S. (2003) *Measurement and Prediction of Dynamic Density*, 5th USA/Europe Air Traffic Management R&D Seminar (Budapest: Hungary).

Lafrenière, D., Dayton, T., and Muller, M.J. (2000). Variations on a theme: Card-based methods in participatory analysis and design. Tutorial at the CHI'00: ACM Conference on Human Factors in Computing Systems, April 1-6, 2000, The Hague, The Netherlands

Laudeman, I., Shelden, S.G., Branstrom, R., Brasil, C.L. (1998) *Dynamic Density: An Air Traffic Management Metric*, Report # NASA-TM-1998-112226, Ames Research Center, Moffett Field, California: NASA.

Levin, K. (2007) *Universal High Altitude Airspace*, Dynamic Resectorization Workshop, San Jose, CA

Majumdar, A., Ochieng, W.Y. (2001) *The Factors Affecting Air Traffic Controller Workload, A Multivariate Analysis Based Upon Simulation Modeling of Controller Workload*, 81st Annual Meeting of the Transportation Research Board. Washington, DC: Transportation Research Board.

DATA MINING TECHNIQUES TO DERIVE HUMAN AND SYSTEM PERFORMANCE MEASURES OF AIR TRAFFIC CONTROL FROM OPERATIONAL DATA

Esa Rantanen and Ernest Fokoue
Rochester Institute of Technology
Rochester, NY

The capabilities of the envisioned performance management services of the NextGen may offer means for testing of measures of human performance derived from operational data and allow for longitudinal studies on the effects of operational NextGen on human operators. Time-based metrics offer an attractive solution to measurement challenges of the NextGen, given the dynamic nature of the system, its dependence on severe time constraints, and the role of time in critical aspects of human performance, mainly workload and situation awareness. Due to potentially very high number of variables and complexity of the underlying data structure that render standard statistical techniques inadequate, novel techniques that perform nonlinear regression and pattern recognition along with feature selection and variable selection are potential candidates for such analyses. Promising statistical techniques to handle the many problems associated with these kinds of data include Generalized Linear and Generalized Linear Mixed Models, and Support Vector Machines.

Introduction

The current modernization efforts of the U.S. National Airspace System (NAS)—collectively known as NextGen—represent an unprecedented influx of new technologies and procedures to an immensely complex system that is an integral part of the nation's infrastructure and economy. As with all changes of this magnitude on systems as complex as the NAS, the final impact of modernization of the air traffic control (ATC) and –management (ATM) remains unknown, in particular on the task environments, working methods, strategies, workload, and performance of the human operators within the system.

Although there are many very good sources for human factors data and specifications for acquisition and implementation of new technologies in aviation systems (e.g., Ahlstrom & Longo, 2003; Cardosi & Murphy, 1995), the fact that the technologies and the procedures required for their use are indeed new will curtail the validity of existing standards. Validation efforts through controlled experiments are not without problems, either. It is difficult and often impossible to mimic operational task environments in simulations, demonstrations seldom and systematic experimentation requires large experimental designs, which in turn are expensive and time-consuming to run. Furthermore, ensuring sufficiently large numbers of participants from the target populations (e.g., airline pilots and air traffic controllers) is very difficult. An alternative to experimental validation of new technologies and procedures is to approach the problem through analyses of operational data in longitudinal studies.

Fortunately the new technologies associated with NextGen also allow for routine gathering of data that could be stored for a myriad of analyses. Radar data (the lat. and long. coordinates plus altitudes of aircraft) can be recorded at very high frequencies (e.g., every 10 s), providing a 4-dimensional picture of all air traffic for accurate reconstruction and analysis using software tools such a SATORI (Rodgers & Duke, 1993, and POWER (Manning, Mills, Fox, & Pfleiderer, 2001). Today, radar data are increasingly used to evaluate NAS operations through joint Federal Aviation Administration (FAA) and National Aeronautics and Space Administration (NASA) program called Performance Data Analysis and Reporting System (PDARS). As laudable and crucial to the NextGen overhaul of the U.S. air transportation infrastructure as these efforts are, however, the current measures predominantly focus on the system performance, with little regard on the human component within it.

About the Measurement of the Air Traffic Controller

Over 30 years after V. D. Hopkin's paper titled "The measurement of the air traffic controller" (Hopkin, 1980) appeared in the journal *Human Factors* the intricacies of measurement in ATC/M have, if possible, only been exacerbated by the increase of automation throughout the domain. Consequently, the complexity of the task environments in which controllers work has also increased, making scientific investigation regarding the impact of new technologies more and more difficult due to the escalating number of variables and their interactions in the operational environments as well as the shift from overt performance (i.e. manual control) to predominantly covert behavior (i.e. supervisory control) of the operators. It may also be argued that the traditional measurement techniques in ATC, that is, subjective, over-the-shoulder (OTS) evaluation of controller performance by other experienced controllers, have become inadequate in the face of the present challenges.

Hopkin (1980) concluded his paper by contemplating "prospects for progress" (p. 555) and laid out two different approaches to help in development of better measures of the air traffic controller. One approach listed several new measures to be developed, including time required to perform handovers and for a new controller to accept the position, time scale of decisions (under high workload controllers supposedly move from strategic planning to tactical decisions and opportunistic control), frequency and extent of use of automated aids, measures of boredom, and measures of trustworthiness of information presented to controllers. Another approach advocated by Hopkin was better application of basic psychological concepts to the measurement problems in ATC.

A Taxonomy of ATC/M Measures

Rantanen (2004; Rantanen & Nunes, 2003) took an emphatically systematic and comprehensive approach to the measurement problem in ATC. The approach served a dual purpose, (1) to perform periodic and thorough reviews of past and current research efforts and to organize the findings in a manner that facilitates the use of existing knowledge for a basis of future evolvement of ATC measurement, or, to avoid 'reinventing the wheel', and (2) to proceed cautiously on an issue as complex as ATC measurement and consider carefully all the constraints, assumptions, and threats to validity that may emerge. A result of this work was a taxonomy and a database for ATC measures.

There are several possible taxonomies for ATC measures. One is the dichotomy of measurement of system performance and the measurement of an individual controller or a team of controllers (Buckley et al., 1983; Hopkin, 1980). System measures are defined in system terms (i.e., capacity, throughput, delays, and channel occupancy times), and although they are greatly influenced by human performance, they are usually insufficient for the measurement of the performance of an individual controller. Possible measures of individual controller include identified task performance, human activity, errors, omissions, physiological and biochemical indices, and subjective assessment (Hopkin, 1995). Task performance measures compare the controller's output to that which is required in the task and encompass broad measures of errors and omissions. Human activity measures passively record what occurs in the task, such as radio transmissions, equipment usage, and communication and coordination with other sectors in terms of times, frequencies, and sequences of the activities.

Direct and Indirect Measures

The main division of measures in the ATC measures taxonomy (Rantanen, 2004) was between direct and indirect measures. Direct measures were defined as those that can be explicitly measured. Examples of such measures include a direct observation of a controller's action, measurement of a response latency, or count of aircraft in a sector at a given time. Indirect measures are those that cannot be measured directly but must be inferred from directly measurable variables. For example, certain actions of a controller may be indicative of his or her performance, response latency can be used to make inferences on some covert cognitive processes, and a number of aircraft in a sector can be used to signify sector complexity.

Identification of direct measures in the literature was a relatively simple task. Their classification was straightforward as well. In the resulting taxonomy, the separately reported direct measures were grouped under 65 distinct classes, down to a sixth level in some cases. Altogether 37 indirect measure classes were identified in the literature. The latter, however, rested on very narrow theoretical foundation for making the associations between direct and indirect measures. This is not to say that the measures reported in the literature are not valid; rather, insufficient information was provided to fully assess their validity. Only a minority of all articles reviewed explicitly justified making inferences about indirect variables based on direct measures by citing past research where such associations have been established. Those articles that did typically cited the same sources, making the body of supporting research literature for ATC measurement remarkably small. Finally, few articles took advantage of the opportunity to add to the foundational body of research in addition to reporting results on their particular topics and thus help validating measures used towards their ends (cf., Vicente & Torenvliet, 2000).

Primary and Secondary Measures

Yet another aspect of classification of measures that warrants discussion is the differentiation between what may be termed primary and secondary measures. Primary measures are those that are measured directly, for example, count of aircraft in a sector, or number of heading changes per aircraft. Secondary measures are those derived from primary measures, for example an average number of traffic in a sector, its variance, or range. In the case of the average, the criteria are implicit (the time duration or interval during which the aircraft were counted and the number of samples) but nevertheless have an impact on the eventual measure. It is clear that descriptive statistics are crucial for reducing and making sense of the data; yet, it is very important to acknowledge how such techniques might obscure some aspects of the data and skew others. These concerns are seldom effectively dealt with in the literature (Vicente & Torenvliet, 2000). This particular facet of measurement will be of especially great consequence when multiple measures representing various aspects of an indirect variable of interest are combined in some sort of an index, such as workload (Hart & Staveland, 1988) or dynamic density (Laudeman et al., 1998).

Operational ATC Data

Data from operational ATC has been available for detailed analysis for decades. The SATORI software was developed in the early 1990s to recreate air traffic situations for incident investigation purposes (Rodgers & Duke, 1993). The FAA also collected System Activity Recordings (SAR) that stored all flight and radar information in Air Route Traffic Control Centers (ARTCCs). These data were processed by two other software applications, the National Track Analysis Program (NTAP; FAA, 1991) and the Data Analysis and Reduction Tool (DART; FAA, 1993), which produced a number of text-based output files for further analysis.

Performance and Objective Workload Evaluation Research (POWER)

The NTAP and DART output files from the SAR recordings served as input data for yet another software tool, the Performance and Objective Workload Evaluation Research (POWER; Mills, Manning, & Pfleiderer, 1999; Manning et al., 2000, 2001; Manning, Mills, Fox, Pfleiderer, & Mogilka, 2002).). The POWER program derived over 40 separate measures that described a variety of aspects of ATC, including such measures as traffic count, control duration, and variability in aircraft headings, altitudes, and speeds, as well as latencies of handoff initiation and acceptance. A number of different controller activities were also recorded.

The POWER program was never—as far a we know—used beyond a couple validation studies. One such study used 12 traffic samples from four sectors in Kansas City ARTCC (Manning et al., 2002). In another study, Rantanen, Naseri and Neogi (2007) derived several additional measures of controller performance from operational data obtained from Indianapolis (ZID) ARTCC. For this research, the POWER program was augmented by the Medium Term Conflict Detection (MTCD) algorithm (MTCD Library, n. d.), which allowed for derivation of several critical metrics, including counts of aircraft pairs at the same altitude and time to loss of separation and duration of conflict. Another algorithm (Laudeman et al. 1998; Sridhar, Sheth, & Grabbe, 1998) implemented in the POWER program calculated a dynamic density value for every minute. For demonstration purposes, operational data from three different sectors of ZID ARTCC at two different (busy and slow) times were analyzed and different metrics derived from the data. Both the different sectors and the time periods were clearly distinguished by the measures, attesting to their sensitivity and validity.

Performance Data Analysis and Reporting System (PDARS)

PDARS is based on several FAA initiatives to measure the performance of the NAS and NASA Distributed Air/Ground (DAG) Traffic Management (TM), or DAG-TM, project (Prevot & al., 2003). The system automatically and continually collects radar track and flight plan data from both terminal and enroute ATC computer systems. From these data, variables such as traffic counts, travel times, travel distances, traffic flows, and in-trail separations can be derived. In addition, several events can also be identified and recorded, for example, takeoffs and landings, sector and facility boundary crossings, tops of climb and descent, fix crossings, reroutes, handoffs, and holding patterns. Several reports are also generated automatically (Den Braven & Schade, 2003; Nehl & Schade, 2007). Given the sophisticated data collection and analysis capabilities or PDARS, the system could plausibly be augmented by a number of additional metrics that focus on human variables.

Human Performance Measurement

An example of derivation of objective human performance metrics from ATC data is given in Rantanen (2009). The data were collected from the FAA's evaluation simulations of Future En route WorkStation (FEWS) under different air traffic load configurations at the FAA Technical Center at Atlantic City International Airport, NJ. The simulations were extremely realistic including a complete set of tasks for the participating controllers to perform. The simulations were conducted using the FAA Target Generation Facility (TGF), an emulator for the HOST computer system and the Display System Replacement (DSR), and the Center-Tracon Automation System. The simulation airspace was the Genera ARTCC with IFR in effect. The data were first processed by the FAA into event and trajectory files. The former contained a time line of all controller and pilot actions and events during the run to millisecond accuracy. A separate data processing program was developed to derive temporal task performance variables, the opening and closing of a window of opportunity to perform a given task, and when the task was initiated and completed. From these variables it was possible to determine air traffic controller's task prioritization schemes (first come, first served) as well as the effect traffic density had on their performance (Rantanen, 2009).

These examples demonstrate the feasibility of extracting useful metrics from complex, objective data. They fall short of systematic examination of the effects of different traffic characteristics (e.g., in different sectors at different times) or the effects of new tools of controllers had on their working methods or performance (e.g., FEWS), however, for they merely provide a snapshot of a particular hour or so. Furthermore, the data processing and algorithms employed were very time-consuming and thus not suitable for longitudinal research. New tools must therefore be developed to realize the goals advocated in this paper.

Challenges

Theoretical Obstacles

As has been discussed above, the primary challenge in deriving meaningful, reliable, and valid human measures from operational data lies in the matching of psychological research done in laboratories (controlled experiments) with the operational task environments where air traffic controllers work. This challenge is not insurmountable, however, as has been demonstrated before (Rantanen et al., 2007; Rantanen, 2009). The richness of data available from operational ATC through programs like PDARS plausibly allows for identification of common variables between the two domains, that is, independent and dependent variables in psychological research literature that are comparable to the task demands and performance indices identifiable in the operational world.

A prime example of such common variable is time. Time is also common to the human, the task, and the environment and thus offers a common unit of measurement of human performance in the context of the task. Time is central to several critical aspects of human performance. Time pressure is a key element of task load and one of the primary drivers of subsequent mental workload (Hendy, 1995; Hendy, Liao, & Milgram, 1997; Hancock & Chignell, 1988; Laudeman & Palmer, 1995; Loft, Sanderson, Neal, & Mooij, 2007). Time is also central to what is arguably the most critical construct of human performance in ATC, situation awareness (cf., Rantanen, 2009). Thus, accurate time stamps associated with myriad of variables collected from operational ATC provide a promising point of contact with vast amounts of existing psychological research.

Validation of human performance measures such as discussed here presents further challenges. Undoubtedly many metrics will measure approximately the same thing, or closely related things. Therefore, such metrics should be closely correlated when derived from the same data set. Metrics that do not agree with others supposedly measuring the same thing are suspect, warranting closer inspection of their validity. Similarly, comparison of the same metrics from data sets with known differences will allow for assessment of the sensitivity of the metrics (sensitive metrics are obviously preferred). Examination of corroborating evidence will thus be the primary means to validate the measures.

Statistical Challenges

Contexts such as ATC provide a very potent field of study for the application of data mining and machine learning techniques, primarily because of the potentially very high number of variables. A justification for the use of these techniques comes from the fact that it is not common to end up with situations where the complexity of the underlying structure of the data renders standard off the shelf statistical techniques obsolete and inadequate. Due to the potentially complex nature of the patterns underlying the observed data and considering the fact that in some cases the sample size is not large enough to adequately cover the input space of interest, nonstandard statistical analyses are most likely to be needed to better handle this kind of research. For example, it is very unlikely in most cases that a traditional multiple regression model would provide an adequate representation of the patterns of interest, even if one were to include all kinds of interactions between the variables.

Even under the assumption of a linear model, regularization techniques that enforce the shrinkage of some of the coefficients to zero might be needed because the resulting formulation of the statistical problem yields a severely under-determined system, and therefore an ill-posed statistical problem. It goes without saying that as p increases, it gets harder and harder, both computationally and statistically, to extract the information underlying the process at hand let alone fully understand it. The curse of dimensionality has been and still is one of the most important aspects to deal with in statistical analyses. Many techniques are known to work in low dimension, but fail miserably when p gets large. Therefore we need to more powerful tools to handle such situations.

The internal structure inherent in the input space quickly becomes a pitfall to any technique that does not take it into account. For example, a naive regression analysis that does address the correlation structure of the input space is bound to produce suboptimal (at best) results. Techniques such as principal component regression can be thought of here, but if one really desires to have a handle on the original variables—as one should for model identification—it is better to think of nonstandard techniques that concurrently achieve regression and variable selection.

Given the nontrivial, or at least nonstandard, nature of the structure potentially under consideration, it is very likely that expert's knowledge would be of great importance in zeroing in on the patterns underlying the data. The Bayesian provide tools and methods for incorporating such knowledge into modeling in order to achieve a better extraction of the true structure.

It would be very surprising if the patterns of workload management and situation awareness, for example, were the same (homogeneous) across the whole population under consideration. We expect large individual differences in controllers' techniques and tactics, differences due to different sector and traffic characteristics, and differences due to regulatory factors (e.g., SOPs and LOAs between sectors and facilities). It is reasonable to hypothesize that there might be clusters of patterns rather than a single pattern that captures the structure of the whole population. This is likely to impact even a regression analysis performed on the data. For instance, if one had to perform a regression analysis, one might have to consider modeling a mixture of regression models in order to account for the nonhomogeneity of the population being studied. Nonhomogeneity can also influence the correlation structure of the data, so that one might have to resort to such modern techniques of latent variable analysis as mixtures of factor analyzers and mixtures of principal component analyzers to extract meaningful new concepts and constructs from the inherently high dimensional data generated in operational ATC settings.

Modern data mining and machine learning that perform nonlinear regression and pattern recognition along with feature selection and variable selection are likely to be candidates for this kind analysis. Generalized Linear Models, Generalized Linear Mixed Models and modern techniques such Support Vector Machines should be attempted in this kind of studies as these techniques handle complex structures better than standard techniques.

Summary and Conclusion

In this paper we have argued for integration of results from an exhaustive literature review on the theoretical connections between constructs of interest (e.g., workload and situation awareness) and variables available from operational data. In other words, to examine human performance in operational contexts will require matching independent (or predictor) variables from research literature that have been shown to have valid relationship with dependent variables of interest in laboratory studies with those available from operational data. Such an effort could be built on the successes of existing systems of routine and largely automated data collection and analysis systems such as PDARS. Validation of the measures suggested here will require large amounts of data, also available from programs such as PDARS. All metrics derived from the data must be constantly evaluated for their coherence, consistency, and reliability, which will require research and development of new statistical methods.

References

Ahlstrom, V., & Longo, K. (2003). *Human factors design standards (HFDS).* (DOT/FAA/CT-03/05 HF-STD-001). Atlantic City, NJ: FAA Technical Center.

Buckley, E. P., DeBaryshe, B. D., Hitchner, N., & Kohn, P. (1983). *Methods and measurements in real-time air traffic control system simulation* (DOT/FAA/CT- 83/26). Atlantic City, NJ: FAA Technical Center.

Cardosi, K. M., & Murphy, E. D. (Eds.) (1995). *Human factors in the design and evaluation of air traffic control systems* (DOT/FAA/RD-95/3). Washington, DC: Office of Aviation Research.

Den Braden, W., & Schade, J. (2003). *Concept and operation of the performance data analysis and reporting system (PDARS)* (SAE Technical Paper No. 2003-01-2976). SAE International. DOI: 10.4271/2003-01-2976

Federal Aviation Administration (1991). *Multiple Virtual Storage (MVS); subprogram design document; National Track Analysis Program (NTAP)* (NASP-9247-PO2). Washington, DC: Author.

Federal Aviation Administration (1993). *Multiple Virtual Storage (MVS); user's manual; Data Analysis and Reduction Tool (DART)* (NASP-9114-H04). Washington, DC: Author.

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. 139-183). Amsterdam: North-Holland.

Hancock, P. A., & Chignell, M. H. (1988). Mental workload dynam- ics in adaptive interface design. *IEEE Transactions on Systems, Man, and Cybernetics, 18*(4), 647-658.

Hendy, K. C. (1995). Situation awareness and workload: Birds of a feather? *AGARD AMP Symposium on Situation Awareness: Limitations and Enhancements in the Aviation Environment* (21-1–21- 7). Brussels, April 24-28, 1995.

Hendy, K. C., Liao, J., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information processing load. *Human Factors, 39*(1), 30-47.

Hopkin, V. D. (1980). The measurement of the air traffic controller. *Human Factors, 22*(5), 547-560.

Hopkin, V. D. (1995). *Human factors in air traffic control.* Bristol, PA: Taylor & Francis.

Laudeman, I. V., Shelden, S. G., Branstrom, R., & Brasil, C. L. (1998). *Dynamic density: An air traffic management metric* (NASA-TM-1998-112226). Moffett Field, CA: NASA Ames Research Center.

Laudeman, I. V., & Palmer, E. A. (1995). Quantitative measurement of observed workload in the analysis of aircrew performance. *International Journal of Aviation Psychology, 5*(2), 187-197.

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). A review of mental workload in en route air traffic control. *Human Factors*. *49*(3), 376-399

Manning, C. A., Mills, S. H., Fox, C. M., & Pfleiderer, E. (2000). Investigating the validity of performance and objective workload evaluation research (POWER). *Paper presented in the 3rd USA/Europe Air Traffic Management R & D Seminar,* Neaples, Italy, June 13-16, 2000.

Manning, C. A., Mills, S. H., Fox, C. M., & Pfleiderer, E. (2001). *Investigating the validity of performance and objective workload evaluation research (POWER)* (DOT/FAA/AM-01/10). Washington, DC: Office of Aerospace Medicine.

Manning, C. A., Mills, S. H., Fox, C. M., Pfleiderer, E. M., & Mogilka, H. J. (2002). *Using air traffic control taskload measures and communication events to predict subjective workload* (DOT/FAA/AM-02/4). Washington, DC: FAA Office of Aerospace Medicine.

Manning, C., & Stein, E. (2005). Measuring air traffic controller performance in the 21st century. In B. Kirwan, M. Rodgers, & D. Schaefer (Eds.), *Human Factors Impacts in Air Traffic Management* (Chapter X). Ashgate.

Mills, S. H., Manning, C., & Pfleiderer, E. M. (1999). Computing en route baseline measures with POWER. *Poster presented at the 10th International Symposium on Aviation Psychology,* Columbus, OH, May 3-6, 1999.

MTCD Library (n.d.). Retrieved on June 19, 2005, from http://www.eurocontrol.int/asa/public/standard_page/ASA_MTCD_Library.html

Nehl, R., & Schade, J. (2007). Update: Concept and operation of the performance data analysis and reporting system (PDARS) (pp. 1-16). *IEEE Aerospace Conference*, 3-10 March 2007, Big Sky, MT.

Prevot, T., Shelden, S., Palmer, E., Johnson, W., Battiste. V., Smith, N., Callantine, T., Lee, P., & Mercer, J. (2003). Distributed air/ground traffic management simulation: results, progress, and plans. Paper presented at the *AIAA Modeling and Simulation Technologies Conference*, Austin, TX, August 2003 (AIAA-2003-5602-MST).

Rantanen, E. M. (2004). *Development and Validation of Objective Performance and Workload Measures in Air Traffic Control* (AHFD-04-19/FAA-04-7). Savoy, IL: Aviation Human Factors Division.

Rantanen, E. M. (2009). Measures of temporal awareness in air traffic control. *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 6–10). Santa Monica, CA: HFES.

Rantanen, E. M., & Nunes, A. (2003). Taxonomies of measures in air traffic control research. *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 977-980). April 14-17, 2003, Dayton, OH.

Rantanen, E. M., Naseri, A., & Neogi, N. (2007). Evaluation of airspace complexity and dynamic density metrics derived from operational data. Air Traffic Control Quarterly, 15(1), 65-88.

Rodgers, M. D., & Duke, D. A. (1993). SATORI: Situation assessment through recreation of incidents. *The Journal of Air traffic Control, 35*(4), 10-14.

Sridhar, B., Sheth, K. S., & Grabbe, S. (1998). Airspace complexity and its application in air traffic management. *Proceedings of the 2nd USA/Europe Air Traffic Management R&D Seminar*.

Vicente, K. J., & Torenvliet, G. L. (2000). The Earth is spherical ($p < 0.05$): Alternative methods of statistical inference. *Theoretical Issues in Ergonomics Science, 1*(3), 248-271.

A COGNITIVE MODEL OF THE CONTROL OF UNMANNED AERIAL VEHICLES

Dr. Chris Wickens, Tim Bagnall, Mala Gosakan, and Brett Walters
Alion Science and Technology
Boulder, CO

We describe a workload model of a single pilot unmanned aerial vehicle (UAV) control, which can provide the basis for extension to multiple UAV control and supervision. The model predicts multi-task capabilities based upon the multiple-resource model of human time sharing. Elements of the model are described related to task demand, resource conflict, and resource allocation (task priority). We then demonstrate its applicability to predicting pilot performance in the MQ-1 Predator, describing the "workload spikes" during a typical mission, and demonstrating ways in which high workload can be mitigated.

The control of multiple unmanned aerial vehicles (MUAVs) presents tremendous challenges first and foremost for the pilot, who must fly more than one aircraft at a time, in a manner that imposes extremely high workload. The extent to which such workload is excessive can be answered, in part, from pilot-in-the-loop simulations (Dixon, Wickens, & Chang, 2005; Wickens, Levinthal, & Rice, 2010). But also relevant are computational models (Cummings & Mitchell, 2007; Goodrich, 2010). However, the viability of such computational models of MUAV depends on valid models of the operator of a single UAV. We describe one such model below. In our description, we attempt to be both general, so that guidance can be provided for applications to other environments, as well as specific to the particular modeling project. This paper discusses the workload model that was used to study the MQ-1 Predator system and the findings of the study. The MQ-1 Predator UAV was designed to be controlled by two dedicated operators: a pilot and a sensor operator. Under this configuration, the pilot is responsible for aviation, navigation, strike, communication, and systems monitoring while the sensor operator is responsible for intelligence, surveillance, and reconnaissance (ISR), which includes target tracking, lasing (arming and firing of lasers for ordnance guidance and target illumination), systems monitoring, and communication. In addition to the complex tasks that need to be performed, what makes the work further challenging is that each of these operators has over five displays (e.g., a heads-up display (HUD), chat display, and a tracker) to obtain information necessary for the successful control and operation of the UAV. The work undertaken during the MQ-1 effort focused on developing a thorough workload model that predicted operators' overall workload associated with performing the mission tasks when looking at these displays.

Multiple Resource Model (MRM)

The fundamental model architecture that we harnessed was the multiple resource model of multi-task performance (Wickens, 2002, 2005, 2008), which is designed to predict the amount of interference imposed by two or more tasks performed concurrently. That is, it is a model specifically designed to address multi-task and task "overload" situations, very characteristic of the flight deck, UAV control station, or other high tempo task environments (including driving). The model architecture contains three fundamental elements:

A *task demand (scalar)* component that describes the workload, or demand for mental resources, of each task to be performed in the multi-task ensemble. As we will see, this total demand can be distributed across resources within each task, creating a vector. But at the simplest level, the interference between two component tasks is related to the sum of demands of the two, where each task demand is the sum across that task's resources.

A resource-specific *conflict component*, which penalizes a task-pair to the degree that they compete for common resources. As a simplifying example, two visual tasks will generally compete more than a visual and auditory task. Thus, the task demand and the resource conflict can combine, in a way described below, to predict the total task interference or dual task decrement.

A *resource-allocation component* that essentially allocates the decrement between the two tasks, in inverse proportion to task importance. Hence, a pilot will typically allow maintaining stability of the flight dynamics (aviate) to dominate addressing chat communications if the two are imposed at once. Multi-task researchers often speak of a primary and a secondary task. By definition, when these compete, the secondary task is that which is assumed to bear the most interference. Adhering to optimum resource allocation is related to good cockpit task management (Chou, Madhavan, & Funk, 1996; Funk, 1991).

Importantly, different aspects of this model have been validated in both driving (Horrey & Wickens, 2003) and to a lesser extent in aviation (Wickens et al., 2003; Sarno, 1995)

Application of the Model to UAVs

Our current application begins with parallel identification of demand and resource conflict components. This was accomplished by first referring to doctrine manuals to identify major meta-tasks in a UAV mission, such as handing-off, flying to target, surveillance of target, etc. (Eaton et al., 2006; Nagy, Muse, & Eaton, 2006). This was followed by extensive interaction with subject matter experts (SMEs), working with the crew station description to identify the workload associated with each meta-task and to break the meta-task down into subtasks such that each subtask could be associated uniquely with its own goal and workload characteristics. Because we were interested in the highest workload meta-tasks of the mission, we focused initially on the five subtasks that appeared to define most of the activity within the mission: aviating, navigating, chat communications (digitally with a display and keyboard), oral communications, and systems monitoring. We then used the workload values, assigned by the SME for each meta-task, to define the proportional workload of the tasks contained within, but these proportions were based on the resource-interface analysis described as follows.

Resource-interface analysis. As noted above, each task can be associated with demands on each of five resources within the current version of the multiple resource model: visual, auditory, cognitive, speech, and motor.[1] Any task must be associated with demands in one or more of these resources. Furthermore, each of these resources may be linked to one or more **interfaces** in the cockpit workstation. For example, a HUD is clearly associated with the visual channel, as is the chat room text display.

We also obtained data from the SMEs on the relative percentage of use that they assigned to each interface during the meta-task in question. Using reverse engineering, we were then able to associate the total demand of each task with a demand for each resource for the tasks, where proportion of time used was translated into a demand for the resource using the interface in question.

Computing task interference. Having completed this task description vector for all tasks performed concurrently during a meta-task, we examined each task pair that might be time-shared and computed its predicted interference by summing the total demand value for each task and summing the task interference across resources for the two tasks. For the two tasks shown in Table 1, these values, in red, are shown across the top (task A) and down the columns on the left (task B). Computing the resource conflict between the two tasks in question is described below.

Table 1. *Conflict Matrix*

		Task A				
		Visual	**Auditory**	**Cognitive**	**Speech**	**Motor**
	Task B	**1**	**0**	**1**	**0**	**2**
Visual	**2**	**0.8 or 1.0**	0.7	**0.6**	0.4	**0.6**
Auditory	**0**		0.9 0.	6 0.	8	0.4
Cognitive	**0**		0.	8	0.5	0.5
Speech	**0**			1.	0	0.8
Motor	**1**					**0.6 or 0.8 or 1.0**

Below and to the right of the red task A and B demand vectors, Table 1 represents a conflict matrix, the heart of any multiple resource approach. The numbers within each cell depict how much a resource used in the task across the columns (A) will compete with a task down the row (B). These values range from 0 (no conflict whatsoever) to 1.0 (maximum conflict). The specific assignment of values shown in the table is based on the multi-dimensions structure of multiple resource theory, and the reader is referred to Wickens (2002, 2005) for details. But, to cite two intuitive examples, the value of 0.4 between motor and auditory indicates that it is relatively easy to

---

[1] Other versions include tactile, spatial versus verbal cognition.

listen while manipulating with the hands (low conflict), whereas the high value (0.8) between auditory and speech indicates the great difficulty of listening while talking.

The table also highlights two additional complexities. Within visual/visual conflicts, the two values indicate the greater conflict (1.0) when two visual sources are separated (e.g., the view outside and on a head-down instrument panel) than when they are close (0.8) (e.g., view outside and a HUD). The latter is still high but not impossible. Within motor-motor, the lower value (0.6) results when two hands are used (e.g., joystick while keyboarding with the other hand) and the mid value results when a single hand is used in an integrated control (joystick manipulation with finger controls mounted as in a HOTAS). The highest value (1.0) is when a single hand must be used for controls in two different locations.

Using the conflict matrix in the above table, it is possible to assign the task vector of two concurrently performed tasks to the rows and columns. Task A represents the chat task, and Task B represents the task of aviating. From this representation, then, it is possible to compute the conflict score by summing the cell demands of all cells that are occupied by a non-zero entry in the column above and the row to the left (modulated as needed by the qualifiers for visual-visual and motor-motor interference). These cells are bold faced in the table. As is evident, this sum could range from 0.4 to 10.6 (all cells occupied).

Total predicted interference (sometimes referred to as predicted workload) is then computed by summing the weighted total demand and resource conflict component. This weighting is required because of the different maximum possible values of the two components: 10.6 for resource conflict and 30 for the sum of the maximum resource demand vectors across the two tasks. Regarding the latter, if, within each resource, the maximum is 3 (e.g., workload per resource could be 0, 1, 2, 3), then the maximum total demand is 30 (sum of maximum Task A demands is 3+3+3+3+3=15 and similarly the maximum demand vector sum for Task B is 15, for a total of 30) . Thus, to weight each component equally, the resource conflict score should be multiplied by 30/10.6 = 2.83.

The third component of the multiple resource conflict architecture is, of course, the task priority score. As the model adds the task demand and conflict to predict a total decrement, this decrement can be allocated proportionally to the row or column task in a manner inversely related to some judgment of task importance made either by the operator or imposed by a mission planner. Clearly, if the overall interference score is low, allocation will have little influence, as both tasks will be performed quite near their ceiling value in any case.

Regarding resource allocation, in the current application, we always designated one task (within each meta-task) as primary. All other tasks were secondary, and the model only computed interference between the primary task and one secondary task at a time. The proportion of task-pair interference assigned to the total interference score was based on the relative amount of time that each secondary task was being performed.

Task shedding and the “red line.” It became evident as we coded the tasks that there were times when a mission planner designated that certain sets of tasks should be performed concurrently, but they simply could not be, either because of their high demand values or because of conflict values =1.0 in occupied cells. Intuition also tells us that there are times when we must cease performance of one task altogether rather than perform it in a degraded manner, when demands become excessive. In the parlance of workload researchers, these are the circumstances in which workload has “exceeded the red line” and tasks are shed (Wickens, 2008). Such task shedding has two direct implications. First, it means that the workload predicted by summing tasks requested to be performed during a given period (we called this **prescriptive** workload) may be far less than that experienced by the operator, as s/he has shed tasks, in a closed loop fashion, to keep workload manageable. We called this **descriptive** workload. Second, it means that a comprehensive workload model must predict which task will be shed and how long it may remain unattended (or “unserviced” in the language of cueing theory) before it is resumed.

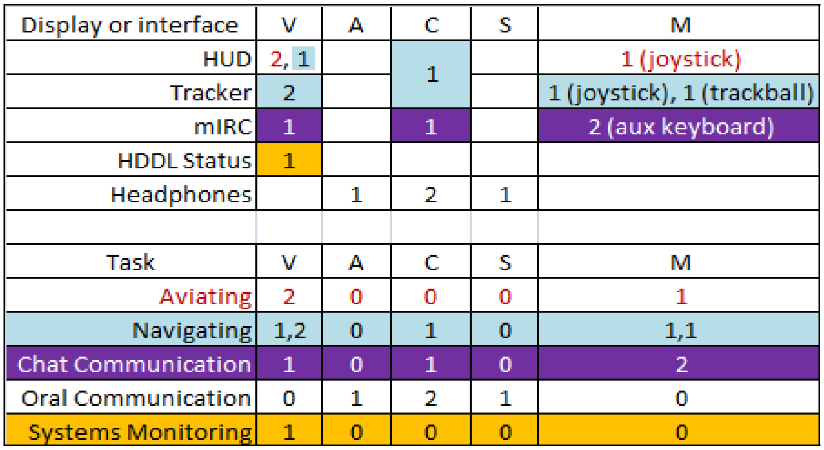
Human Performance Model (HPM)

For this effort, a human performance simulation model was developed to capture an entire mission of an MQ-1, from the pilot and sensor operator gaining control of the aircraft (“gaining handover”) back to relinquishing control of the aircraft (“losing handover”), with stages of en route to target, reconnaissance, strike, and en route (return to base) in between these two endpoints.

The Improved Performance Research Integration Tool (IMPRINT) was the software used to develop the human performance model. IMPRINT is a simulation and modeling tool that estimates manpower, personnel, and training (MPT) requirements and identifies constraints for new weapon systems early in the acquisition process. It is government-owned software and consists of a set of automated aids to assist analysts in conducting human performance analyses (Allender et al., 1995). Based on modeler's inputs, IMPRINT computes a time line of the workload experienced by an operator over a mission so that relative comparisons can be made within and between missions. It is important to note that for this effort, IMPRINT's built-in graphical user interface was not used to enter the workload values for each meta-task. Instead, code was written within each meta-task in the model to record workload.

Before the IMPRINT model was developed, the high-level functions and meta-tasks performed during the mission were identified using the results of our task analysis. Each meta-task was then decomposed into primary and secondary tasks. Next, we mapped the five mental resources that were needed to interact with each of the displays and/or interfaces to perform each task based on the meta-task description. Finally, the associated task demand values for each resource-interface pair were generated. An example of a meta-task within the Strike function was "The pilot discovers a target and obtains orders for a strike." This meta-task consisted of the following tasks: aviating, navigating, chat communication, oral communication, and system monitoring. The primary task was navigating, and the remaining tasks were considered secondary tasks.

Table 2. *Task Demand and Resource Allocation*



Display or interface	V	A	C	S	M
HUD	2, 1		1		1 (joystick)
Tracker	2				1 (joystick), 1 (trackball)
mIRC	1		1		2 (aux keyboard)
HDDL Status	1				
Headphones		1	2	1	
Task	V	A	C	S	M
Aviating	2	0	0	0	1
Navigating	1,2	0	1	0	1,1
Chat Communication	1	0	1	0	2
Oral Communication	0	1	2	1	0
Systems Monitoring	1	0	0	0	0

Table 2 shows the resource-interface and task demand assignments of this meta-task. The different tasks are color coded. Within each cell is the demand imposed by the task /resource combination in question. Total demand for each task (sum of resource demands) was based upon SME inputs, coupled with general workload ratings assigned in the documents provided (Eaton et al., 2006; Nagy, Muse, & Eaton2006). For each task, the analyst, working with the task description and SME consultant decided how to allocate the total task demands to the separate resource channels. As a simple example, the task of systems monitoring (yellow) had its entire (low) demand associated with the visual channel.

After all of the meta-tasks were identified, this information was used to develop the human performance model in IMPRINT. Workload analysis was conducted for 18 pilot meta-tasks (composed of 71 tasks) and 16 sensor operator meta-tasks (composed of 55 tasks; see Bagnall et al., 2010 for details).

Selected Results

Figure 1 shows a comparison between the prescriptive and descriptive workload for the pilot throughout the simulation run. The prescriptive workload for the pilot is the workload associated if all prescribed tasks for each meta-task were performed simultaneously. (Note that the scale of the y axis for the two graphs is different. This was done to see the "spike" in workload in the descriptive workload graph.). For the pilot, the highest workload occurred within the strike portion while performing the "Pilot maneuvers aircraft into proper position for strike" meta-task. This meta-task requires a substantial amount of cognitive and visual resources as well as fine motor control. The two displays that the pilot monitors during this meta-task are the HUD and the Tracker. The pilot's attention is divided between these two displays almost equally while performing the aviating and navigation tasks. It is also important to note that during the majority of this task (approximately 90% of the time) the pilot is either talking to, or listening to, oral communications over the headset. One suggestion to lower the pilot's workload during this meta-task is to limit the amount of conversations being monitored once the decision has been made to launch the weapon.

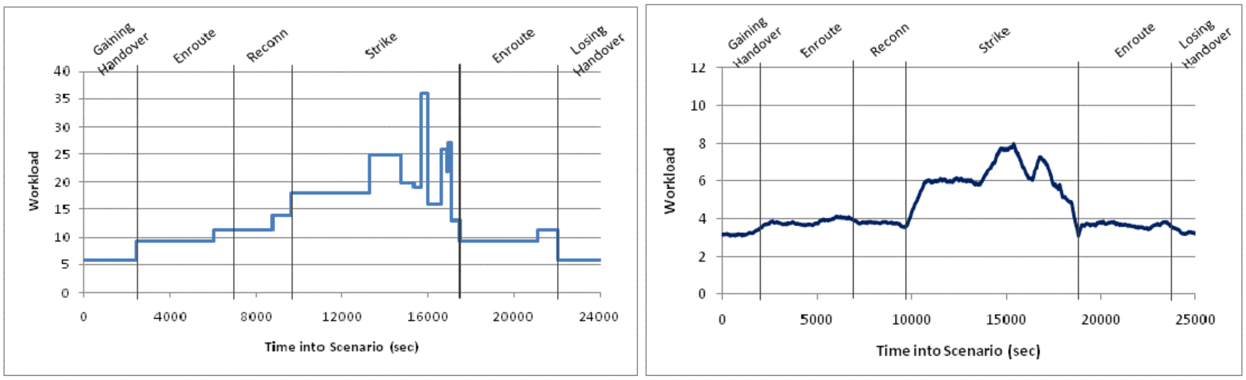


Figure 1. Prescriptive Workload (left) Versus Descriptive/Moving Average Workload (right) for the Pilot.

While the above discussion and findings apply only to one pilot meta-task ("Pilot maneuvers aircraft into proper position for strike"), it is to be noted that the HPM model has been set up such that an analyst can run different mission configurations (for example, nonstrike scenarios such as an emergency situation) and identify points (meta-tasks) of high workload for that particular mission profile and detect what tasks-mental resources-display combinations are contributing to those high levels. The analyst can then play what-if scenarios by either re-allocating or shedding some of these tasks to be addressed by automation and subsequently investigating the impact of that on operator workload. Similarly, the effects of experience on overall workload can be studied by either reducing or increasing (if a person with lesser experience is performing) the mental demands placed on a task.

In addition to the pilot workload, because the sensor operator is an integral part of the traditional configuration, the sensor operator's workload was also studied. For the sensor operator, the highest workload occurred while performing "SO guides the missile." This meta-task requires a substantial amount of cognitive and visual resources as well as fine motor control (i.e., the sensor operator uses a joystick to control the direction of the missile). During this task, the sensor operator is monitoring the location of the missile relative to the target on the HUD and controlling the missile's direction using the joystick. The sensor operator does not participate in any conversations (oral or chat) during this task. Therefore, this should serve as an indication to mission planners that adding any additional tasks such as a communication task would only further increase the workload levels experienced by the sensor operator for this meta-task.

As noted in the Introduction, with some additional assumptions regarding task switching and supervision, the core of this model can be embodied in a workload model of multiple UAV control, and an attempt at this has been made under the same effort. However, several assumptions were made for the MUAV configuration, and the results were not discussed in this paper. Further, it is important to note that the current application has not yet been validated against performance data in an actual Predator simulation.

Acknowledgments

The research reported in this document was performed in connection with contract MSIAC 09-0016/0035 for the Modeling and Simulation Information Analysis Center. The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the U.S. Government unless so designated by other authorized documents. Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon. The authors would like to acknowledge LtCol Roscoe Van Camp at Creech Air Force Base for his contributions to this effort.

References

Allender, L., Kelley, T. D., Salvi, L., Lockett, J., Headley, D. B., Promisel, D., Mitchell, D., Richer, C., & Feng, T. (1995). Verification, validation, and accreditation of a soldier-system modeling tool. In *Proceedings of the 39th Human Factors and Ergonomics Society Meeting*, San Diego, CA. Human Factors and Ergonomics Society, Santa Monica, CA.

Chou, C-C., Madhavan, D., & Funk, K. (1996). Studies of cockpit task management errors. *International Journal of Aviation Psychology, 6*, 307-320.

Cummings, M. L. & Mitchell, P. J. (2007). Operator scheduling strategies in supervisory control of multiple UAVs. *Aerospace Science & Technology, 11,* 339-348.

Dixon, S. R., Wickens, C. D., & Chang, D. (2005). Mission control of multiple unmanned aerial vehicles: a workload analysis. *Human Factors, 47,* 479-487.

Eaton, G., Geier, P., Kalita, S., Nagy, J., Palmer, B., Phillips, A., Probst, A., & Sanchez, C. (2006). *U.S. Air Force Unmanned Aircraft Systems Performance Analyses: Predator Pilot Front End Analysis (FEA) Report* (Unpublished Final Report). Wright-Patterson AFB, OH: SURVIAC.

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

Goodrich, M. (2010). On maximizing fan-out: towards controlling multiple unmanned vehicles. In M. Barnes & F. Jentsch (eds.), *Human-robot interactions in future military operations* (pp. 375-396). Farnham, U.K.: Ashgate.

Horrey, W. J. & Wickens, C. D. (2003). Multiple resource modeling of task interference in vehicle control, hazard awareness and in-vehicle task performance. *Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Park City, Utah*, 7-12.

Nagy, J., Muse, K., & Eaton, G. (2006). *U.S. Air Force Unmanned Aircraft Systems Performance Analyses: Predator Sensor Operator Front End Analysis (FEA) Report* (Unpublished Final Report). Wright-Patterson AFB, OH: SURVIAC.

Sarno, K. J. (1995). Role of multiple resources in predicting time-sharing efficiency: evaluation of three workload models in a multiple-task setting. *International Journal of Aviation Psychology*, *5*(1), 107-130.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theor. Issues in Ergon. Sci*., 3(2), 159-177.

Wickens, C. D. (2005). Multiple resource time sharing models. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas, H. Hendrick, (eds.), *Handbook of human factors and ergonomics methods* (pp. 401-407). Boca Raton, FL: CRC Press.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*, 449–455.

Wickens, C. D., Goh, J., Helleberg, J, Horrey, W. J., & Talleur, D. A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors, 45*(3), 360-380.

Wickens, C. D., Levinthal, B., & Rice, S. (2010). Imperfect reliability in unmanned air vehicle supervision and control. In M. Barnes & F. Jentsch (eds.), *Human-robot interactions in future military operations* (pp. 193-210). Farnham, U.K.: Ashgate.

COGNITIVE MODELING AS A TOOL FOR IMPROVING RUNWAY SAFETY

Michael J. Schoelles
Wayne D. Gray
Rensselaer Polytechnic Institute
Troy, New York

Runway incursions are low probability events resulting from complex combinations of cognitive and environmental factors, which can have deadly consequences. However, the development and evaluation of tools to reduce runway incursions are, ironically, hampered by the low incidence of such events. A possible path forward is the use of high-fidelity cognitive models to predict pilot performance under a wide variety of airport conditions and runway circumstances. We describe a fully embodied ACT-R 6.0 cognitive model, named SimPilot, of a pilot taxiing a simulated Boeing 737-800 aircraft. The goals of SimPilot are twofold. The first is automated testing of a new safety devices. The second goal is to show that modeling the multitasking inherent to taxiing in a cognitive plausible manner is an important step in predicting and preventing runway incursions.

New tools are continually being developed to increase aviation safety by reducing human error. Since it has been shown that small changes in the design of a system can lead to large changes in human performance, imagine the potential changes that could occur as the result of introducing a new tool. To ensure that the changes introduced are beneficial, extensive testing with subject matter experts (SMEs) is required, but in the aviation domain this is expensive. A potential solution to this problem is simulating SMEs with cognitive models. The type of models that can provide useful simulations are process models that can actually do the task and are based on a cognitive architecture.

In this paper we describe a cognitive model to test the Electronic Movable Map (EMM) tool that is intended to improve runway safety. Note the EMM used in this effort is not the real EMM developed for NASA but a functionally equivalent software version developed for testing this methodology. The cognitive model works in conjunction with the Aptima developed Performance Engine (PE), which collects and stores data from a data source and computes relevant performance data. To test the EMM, scenarios were developed to create runway incursions. The scenarios are to be followed by human pilots or model simulated pilots taxiing a Boeing 737 with and without the aid of an EMM.

Taxiing requires many cognitive and perceptual/motor interactions. For example, the pilot must steer the plane, monitor the taxiing speed, listen to directions from Air Traffic Control, watch for other aircraft, etc. Therefore a model of the pilot must be able to multitask. Recently, multitasking has received a lot of attention in the press and has become a research area for Cognitive Science and in particular, cognitive modeling. We choose ACT-R 6.0 as the cognitive architecture for this project in part because it supports multitasking. The version of ACT-R that we use incorporates an add-on that implements the Threaded Cognition theory (Salvucci & Taatgen, 2008, 2011) of multitasking. An additional reason for using ACT-R in this effort is the success by Byrne (Byrne & Kirlik, 2005) in modeling the taxiing task using ACT-R 5.0.

We use the X-Plane 9 Desktop Simulator™ that is available for MAC, Windows, and Linux operating systems. It supports an easy to use plug-in capability to connect the simulator to data collection systems and aircraft control systems. In the next section we describe the parts in more detail and show how all the parts fit together.

System Configuration and Operation

Figure 1 shows the configuration of the varied components in this test bed. There are five computers connected though a Network Switch.

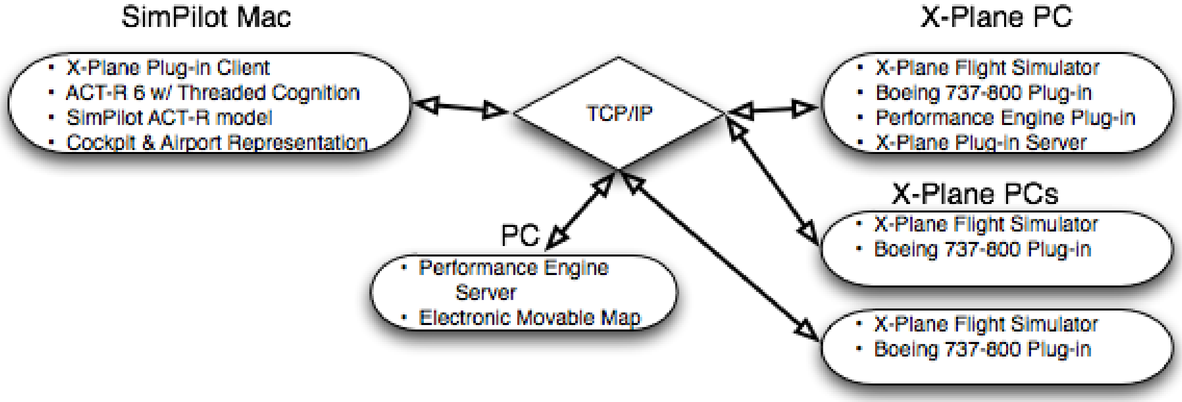


***Figure* 1. The SimPilot Mac, X-Plane PC, and Electronic Movable Map PC and two additional aircraft connected through a TCP/IP Switch.**

For the model to be able to simulate a pilot, it must be able to *see* what the pilot sees, *hear* what the pilot hears, and *manipulate* the aircraft controls. The Cockpit and Airport Representation modules enable these functionalities by providing a representation of the cockpit instruments and airport layout.

The airport representation module is constructed from data sent by the EMM. The model manipulates the aircraft controls by sending joystick commands, mouse movements and clicks to the Model Server running on the X-Plane PC. It receives data about the cockpit instruments from the Model Server. Figure 2 shows the cockpit.

***Figure* 2. The Boeing 737-800 cockpit.**

The SimPilot Mac also contains the scenario script, which it is to follow. The EMM and Performance Engine receive data about the Boeing 737 plane and other planes at the airport through the Performance Engine Server. The EMM sends aircraft positions and airport layout to the Airport Representation on the MAC. In this way the SimPilot achieves the functionality to see other airplanes, taxiways, and runways.

Model Description

The heart of our effort, SimPilot, is a work in progress. Most of the effort to date has gone into the development of the vitally necessary and excruciatingly detailed Cockpit and Airport Representations. However, our initial integration testing with the other components has been successful. In this section we will first describe the cognitive architecture on which the model is built. We will next describe the major change to the architecture required to perform the level of multitasking necessary for taxiing. The details of the model structure are then presented and the section ends with a discussion of the limitations and problems of our current approach.

ACT-R 6.0

ACT-R 6.0 (Anderson et al., 2004) is an embodied cognitive architecture that has perceptual and motor components along with cognitive processing, memory, and control components. The perceptual and motor components enable SimPilot to operate user interfaces by passing the interface software the same commands passed by the input devices used by humans. As the SimPilot is a cognitive, not an artificial intelligence model, its input commands mimic the speed and accuracy of human users.

In common with all ACT-R models, SimPilot consists of pattern matching and action rules. When a match is found, the action associated with that pattern is executed. ACT-R executes one rule at a time. ACT-R maintains simulated human time in that time for ACT-R processes and actions are set to the theoretical times for the corresponding human events. For example, to shift visual attention from one object to another takes 85 ms. Time to retrieve an item from memory varies as a function of the recency and frequency of that item's occurrence (Schooler & Anderson, 1997; Sims & Gray, 2004). When the model does a task ACT-R produces a trace that includes the action taken and a time stamp. The trace allows model performance to be compared with human performance.

ACT-R checks every 50ms (human time) all of its rules and executes one of the rules whose pattern is matched. If more than one rule can execute then ACT-R chooses the one it calculates would be the most useful at this time. As ACT-R makes this decision every 50 ms, this serial execution is not as constraining as it might seem and has been shown to be as accurate at simulating fine-grained human behavior as architectures that allow parallel firing of rules (Byrne & Anderson, 2001). If a rule could fire but did not because another one had a higher utility, chances are that in 50 ms it will be able to fire. The rules are intended to represent the fine-grained procedural steps that are executed to perform some task. As biological processes are inherently noisy in the signal processing sense of noise (Faisal, Selen, & Wolpert, 2008; Neri, 2010), ACT-R adds noise to the utility calculation to simulate the variability in time and performance that humans make. Besides a procedural memory ACT-R has a declarative memory that holds units or *chunks* of factual information. These chunks represent the portion of the simulated human's background knowledge necessary to perform the task that the model is attempting to perform. Like humans, errors can occur in the memory retrieval process due to random fluctuations (noise) in memory strength or activation (Sims & Gray, 2004). Either the wrong chunk is retrieved or the intended chunk is not "strong" enough to be remembered.

The perceptual components of ACT-R allow the model to see and hear. In common with the human brain, the visual component has *where* and *what* paths (Findlay & Gilchrist, 2003). The *where* path allows the model to detect features of an object such as color, size, and shape at a 2-D location in space. The *what* path moves visual attention to that location to encode the object with those features. ACT-R *hears* in much that same way that it sees in that sound events are detected and auditory attention is invoked to encode those sounds. By encoding objects and sounds in the environment the visual and auditory components add new declarative knowledge to the model. The motor component is the model's *hands* and *voice*. The manual component is capable of moving and clicking the mouse. Movement times are a based on Fitts' Law (Fitts, 1954). The vocal module is capable of speaking text and subvocalization (see, e.g., Huss & Byrne, 2003).

The *imaginal* component of ACT-R is intended to hold intermediate representations required in solving a problem or performing some task. New declarative chunks can be added by this component. The *temporal* component maintains an internal clock. The *goal* component in hold chunks that guide task execution. For the model presented in this paper the default goal component is replaced with a module that implements a form of Threaded Cognition (Salvucci & Taatgen, 2008) that implements the multitasking required for the taxiing task.

Threaded Cognition

Salvucci and Taatgen (Salvucci & Taatgen, 2008, 2011) propose Threaded Cognition as an integrated theory of concurrent multitasking. Multitasking is defined as doing 2 or more tasks at once. A thread is sequence of processing steps coordinated by a serial procedural resource and executed across perceptual and motor resources. The key claims of Threaded Cognition are that multiple active goals can exist. Associated with each goal is block of procedural processing. Processing conflicts can exist for procedural, declarative, perceptual, and motor resources. A thread will grab a resource if it needs the resource and the resource is available. It will release the resource when no longer needed. According to Salvucci and Taatgen, cognition favors the least recently processed thread. Declarative retrievals can be converted to hard coded rules over time thus reducing both declarative and procedural resource conflicts. The cognition requires no central and supervisory executive.

Most of the points in the paragraph above have been part of ACT-R since version ACT-R 5.0, thus the implementation of Threaded Cognition into ACT-R 6.0 requires only allowing two or more active goals and giving priority to the least recently processed goal. These changes are implemented in the version of ACT-R 6.0 used by our model by simply using a different goal component. Threaded Cognition is a relatively new theory and has not been tested in complex, dynamic environments. Also it has rarely been tried on more than two tasks.

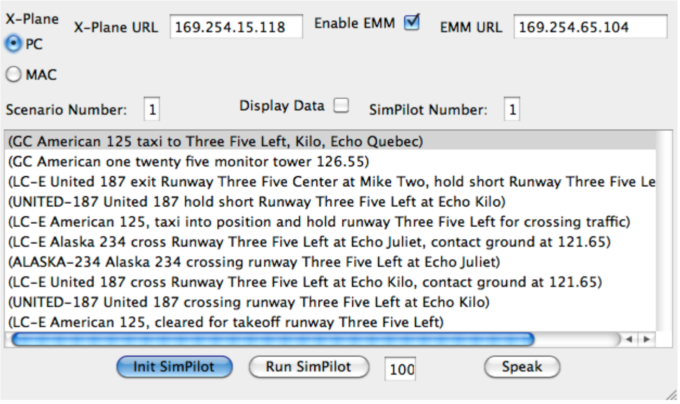
Model Processing

In this section we try to give a flavor of how the model operates and some of the other problems involved. The taxiing task is divided into subtasks. Each subtask is represented as a chunk, which has (a) an initial state, (b) the set of treads that can run simultaneously in the subtask, and (c) a final state. In general the final state of a subtask is the initial state of another subtask. The subtask chunks are put in the goal buffer and their associated productions initiate the threads specified in the chunks. In general the state of the other resources control the execution flow in the spirit of Taatgen's (2007) Minimal Control Principle.

The basic actions that the model must perform include setting switches, tuning radios and monitor indicators, working the throttle, monitoring aircraft speed, steering the aircraft, turning the aircraft, stopping the aircraft, listening to air traffic control, and watching for other aircraft.

All scenarios begin with preflight checks completed and the aircraft sitting on the tarmac waiting for instructions from Air Traffic Control (ATC). The execution of a scenario that provides ATC instructions requires experimenter action. The scenario text is shown in a window on the MAC screen, see Figure 3, and when a scenario event is to be executed it is clicked on. For example, to start the first scenario, the experiment would highlight the text "American 125 taxi to Three Five Left, Kilo, Echo Quebec" and the click the execute button. This causes a sound event within ACT-R. The model contains a rule that whose pattern simply matches any and all sound events and ignores any other environment or internal event. So when the event does occur a processing thread is initiated to attend and encode the sound. The sound is interpreted as the command to begin taxiing which sets a goal to execute the begin-taxiing-procedure that increases engine thrust, until the engines are at the proper N1 level and then releasing the parking brake. Increasing thrust and monitoring the N1 level is an example of the coordination required between manual actions and visual monitoring that is difficult to model. One reason is that it is difficult is the display of the N1 levels lags behind the manual action. In general, the model tends to overshoot and must make a series of corrective actions. Real pilots have a "feel" for doing this that which is beyond the current state-of-the-art in cognitive architectures.

To perform an action on an instrument control, such as releasing the parking brake and checking its state, the model must be able to see the control. Seeing is done through the cockpit representation, called the *visicon*. The visicon contains 77 entries, one for every light, switch, dial, lever, and display in Figure 2. For each entry, the visicon contains its location on the display, its size, its color, its value current value, and its type. These are the details that allow the vision module to function as if it sees these items. The visicon must be updated whenever cockpit display changes otherwise the model will not see the results of its action. Achieving the integration of these two software systems was a significant software engineering challenge as the X-Plane software only provides a third of the 77 item locations needed by the model. For the rest they must be manually calculated offline, which is not desirable since this is time-consuming and they may change with new versions of flight simulator. In addition, if the entire cockpit display can't fit on the screen then the locations in the visicon must be adjusted when the screen scrolls. This is problematic because the scrolling action is purely an interface issue and not relevant to the taxiing task.



X-Plane
● PC
○ MAC

X-Plane URL 169.254.15.118
Enable EMM ☑
EMM URL 169.254.65.104

Scenario Number: 1
Display Data ☐
SimPilot Number: 1

(GC American 125 taxi to Three Five Left, Kilo, Echo Quebec)
(GC American one twenty five monitor tower 126.55)
(LC-E United 187 exit Runway Three Five Center at Mike Two, hold short Runway Three Five Le
(UNITED-187 United 187 hold short Runway Three Five Left at Echo Kilo)
(LC-E American 125, taxi into position and hold runway Three Five Left for crossing traffic)
(LC-E Alaska 234 cross Runway Three Five Left at Echo Juliet, contact ground at 121.65)
(ALASKA-234 Alaska 234 crossing runway Three Five Left at Echo Juliet)
(LC-E United 187 cross Runway Three Five Left at Echo Kilo, contact ground at 121.65)
(UNITED-187 United 187 crossing runway Three Five Left at Echo Kilo)
(LC-E American 125, cleared for takeoff runway Three Five Left)

***Figure* 3. SimPilot and Scenario Control Screen**

The procedure that SimPilot uses to turns the aircraft is based on a task analysis generated a pilot. This analysis specifies speed changes and thrust changes for turning. An important data point in turning is when to start to turn. A pilot can see the intersection and through experience knows when to start turning. The model does have a representation of the airport that it acquires from the EMM, and it does have knowledge about where it is on the ground so the distance to the turn can be calculated, at this time it is not known whether this will be of sufficient fidelity to simulate real turns.

Human Performance Modeling in Aviation

In 2005, NASA chose five teams to develop human performance models (HPM) of pilots performing taxi operations with and without advanced displays (Foyle et al., 2005). Each team used a different modeling architecture. The Attention-Situation Awareness approach by Wickens and McCarley (Wickens et al., 2005) looked at attention allocation and situational awareness. Simulation data drove a model that predicted errors and the benefits of the T-NASA display. The ACT-R 5.0 model by (Byrne, Kirlik, & Fleetwood, 2008) is the closest to the model presented here. They were connected to an X-Plane Simulator but concentrated on decision-making strategies rather than multitasking because of the limitations of ACT-R at the time. Air-MIDAS by Corker (Foyle, et al., 2005) used working memory limits, interference processes and heuristics to predict errors. D-OMAR (Foyle, et al., 2005) by Deutsch and Pew is an event-based simulator with three different languages to develop perceptual, cognitive and motor processes, which they considered to be the building blocks of pilot expertise. They found that because errors are so infrequent, habit might intrude and lead to certain types of errors. IMPRINT by Lebiere and Archer (Foyle, et al., 2005) combined IMPRINT, which is a performance tool with ACT-R. IMPRINT provided the simulation environment and ACT-R acted as cognitive agent.

Conclusions

We have described a multitasking cognitive model based on the ACT-R architecture. The goals of the model are both applied and theoretical. On the applied side, we hope to advance the methodology of automated testing using simulated human experts rather than using actual people. This form of testing can be beneficial in testing design changes, particularly major changes that involve adding new tools and technology to an already proven system. On the theoretical side we are pushing the current theory in multitasking in order to indentify its weaknesses and what further changes to the architecture are required to make this a more robust theory.

Acknowledgements

This work was supported by NASA Ames Research Center through a Small Business Innovative Research Contract (Contract No. NNX09CB45C) awarded to Aptima, Inc. All opinions expressed in this article are those of the authors and do not necessarily reflect the official opinion or position of Rensselaer Polytechnic Institute, Aptima, Inc., or NASA. We would like to thank Dr. James Johnston and Dr. Michael Feary at NASA Ames Research Center for funding this research. We also would like to thank Gabe Ganberg, Mike Therrien, Joe Thoreson and Jamie Estock of Aptima, Inc.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. L. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036-1060.

Byrne, M. D., & Anderson, J. R. (2001). Serial modules in parallel: The psychological refractory period and perfect time-sharing. *Psychological Review, 108*(4), 847–869.

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

Byrne, M. D., Kirlik, A., & Fleetwood, M. D. (2008). 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, FL: CRC Press.

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nat Rev Neurosci, 9*(4), 292-303.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and series*. New York: Oxford University Press.

Fitts, P. M. (1954). The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. *Journal of Experimental Psychology, 47*(6), 381–391.

Foyle, D. C., Hooey, B. L., Byrne, M. D., Corker, K. M., Deutsch, S., Lebiere, C., et al. (2005, September 26-30, 2005). *Human Performance Models of Pilot Behavior.* Paper presented at the Human Factors and Ergonomics Society 49th Annual Meeting, Orlando, Florida.

Huss, D. G., & Byrne, M. D. (2003). An ACT-R/PM model of the articulatory loop. In F. Detje, D. Dörner & H. Schaub (Eds.), *Fifth International Conference on Cognitive Modeling* (pp. 135-140). Bamberg, GE: Universitas-Verlag Bamberg.

Neri, P. (2010). How inherently noisy is human sensory processing? *Psychonomic Bulletin & Review, 17*(6), 802-808.

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review, 115*(1), 101-130.

Salvucci, D. D., & Taatgen, N. A. (2011). *The multitasking mind*. New York: Oxford University Press.

Schooler, L. J., & Anderson, J. R. (1997). The role of process in the rational analysis of memory. *Cognitive Psychology, 32*(3), 219-250.

Sims, C. R., & Gray, W. D. (2004). Episodic versus semantic memory: An exploration of models of memory decay in the serial attention paradigm. In M. C. Lovett, C. D. Schunn, C. Lebiere & P. Munro (Eds.), *6th International Conference on Cognitive Modeling* (pp. 279-284). Pittsburgh, PA: Carnegie Mellon University/University of Pittsburgh.

Taatgen, N. A. (2007). The minimal control principle. In W. D. Gray (Ed.), *Integrated models of cognitive systems*. New York: Oxford University Press.

Wickens, C. D., McCarley, J. S., Alexander, A. L., Thomas, L. C., Ambinder, M., & Zheng, S. (2005). *Attention-Situation Awareness (A-SA) Model of Pilot Error*: University of Illinois at Urbana-Champaign, Institute of Aviation Human Factors Division.

A DISSONANCE THEORY EXPLANATION FOR VISUAL FLIGHT RULES (VFR) INTO INSTRUMENT METEOROLOGICAL CONDITIONS (IMC) ACCIDENTS

Steven Henderson
Transportation Safety Board of Canada
Gatineau, Quebec, Canada

In the U.S. in 2008, accidents resulting from VFR flight into IMC accounted for 2% of all general aviation (GA) accidents, but 8% of all fatal GA accidents. Furthermore, 88% of VFR into IMC accidents were fatal, compared to 17% of other aviation accidents. Dissonance theory is a model of attitude change associated with making difficult choices. Attitude change reduces cognitive dissonance arising from favourable aspects of a not-chosen alternative and unfavourable aspects of a chosen alternative, through spreading of alternatives. Under dissonance theory, pilots in marginal weather who repeatedly revisit their choice to either continue their flight or divert to an alternate destination progressively distort their perception of weather conditions, making them more likely to commit decision-making errors leading to VFR into IMC accidents. Many aspects of general aviation are consistent with factors that increase spreading of alternatives. Dissonance theory resolves inconsistent results from simulator-based studies of weather decision-making.

Visual flight rule (VFR) flight into instrument meteorological conditions (IMC) is a significant causal factor in general aviation (GA) accidents involving fixed-wing aircraft under 12,500 pounds maximum takeoff weight (MTW) in the United States, accounting for 2% of all GA accidents, but 8% of all fatal GA accidents. These accidents are disproportionately lethal. In 2008, 22 of the 25 VFR into IMC accidents were fatal, for a fatality rate of 88% compared to the 17% fatality rate of all other GA accidents.

Canadian VFR into IMC accident statistics are similar. Between 1995 and 2004, the Transportation Safety Board of Canada identified 80 aviation occurrences as VFR into IMC accidents. Although VFR into IMC accidents comprised only 2.5% of the 3,256 accidents involving Canadian-registered aircraft in that period, they comprised 12% of all fatal accidents, and took 96 lives (14% of all aviation fatalities). Furthermore, 55% of VFR-into-IMC accidents were fatal, compared to 10% of all other accidents involving private pilots.

Despite a substantial number of communications, tools and countermeasures training offered by the FAA, the NTSB, the AOPA and others to GA pilots regarding VFR into IMC accidents, these accidents continue almost unabated. Figure 1 (U.S. accident data from AOPA ASF database of NTSB data through 2008, flight hours from AOPA 2008 and 2009 NALL Reports) shows only a small but statistically significant decrease in the US VFR into IMC accident rate for GA aircraft under 12,500 lbs MTW ($R^2 = .439$, $p = .037$), and no significant decrease in the fatal VFR into IMC accident rate ($R^2 = .218$, $p = .17$).

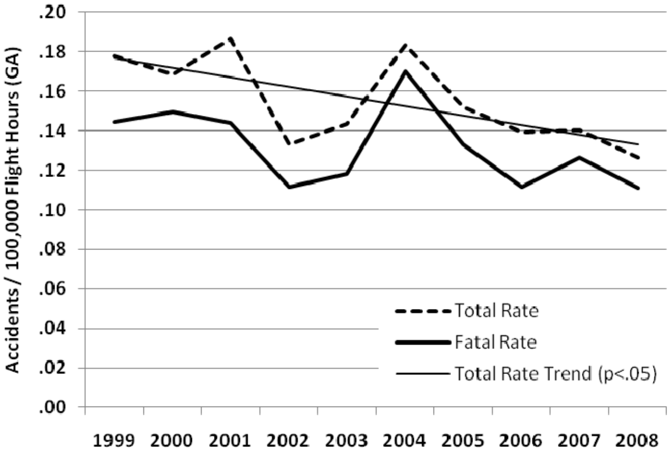


***Figure 1*. Total and Fatal VFR into IMC Accident Rates**

Simulator Research on Cognitive Models of Weather Decision-Making

Wiegmann, O'Hare, Goh and others have conducted a substantial body of pilot decision-making research using simulated cross-country flights (Goh and Wiegmann, 2001a, 2001b; O'Hare and Smitheram, 1995; O'Hare and Wiegmann 2003; Wiegmann, Goh, and O'Hare, 2001, 2002). They have suggested that a number of "failures at different stages of the decisional process" lead to VFR into IMC accidents. Several categories of human information processing factors and models that have been proposed and/or tested for VFR into IMC accidents are:

- **Situation assessment**. Inaccurate assessments of weather conditions. To counter this factor, Wiggins and O'Hare (2003) developed the "WeatherWise" computer-based training program for the FAA. The training program incorporates pictures and video clips to train pilots to identify critical weather cues during flight. The AOPA Air Safety Foundation and FAA sent the foundation's WeatherWise CD to all instrument-rated pilots in the United States (AOPA, 2008).
- **Risk perception**. Accurate assessment but without correct appreciation of risks, also often identified as pilot over-confidence (Goh and Wiegmann, 2001c).
- **Motivational factors**. "Get-home-itus", "sunk costs" or other personal or social errors, or corporate culture / operational / commercial pressures. A variant of the "sunk costs" model is the *prospect theory* (Kaneman and Tversky, 1979, 1984) explanation of decision framing offered by O'Hare and Smitheram (1995). Prospect theory offers a cognitive explanation for why decision-making is inherently riskier if framed in terms of losses than if framed in terms of gains. Indeed, O'Hare and Smitheram found that pilots in their decision scenario study made more conservative decisions if they framed their decision in terms of a gain from their current position than if they framed their decision as a loss (eg, "sunk costs" such as fuel used, or time and money spent) from their starting position (leading to risky decision-making). They recommended that new pilots should be trained to frame their decisions to continue or divert on the basis of gains or benefits from their current position rather than on the basis of losses or resources spent from the start of the flight.

However, none of these models can account for the startling inconsistency between the simulator study findings of Wiegmann, Goh, and O'Hare (2001, 2002) and O'Hare and Wiegmann (2003).

Pilots tested by Wiegman et al. (2001, 2002) flew in simulated scenarios under 5,000 foot ceilings and with 5 mile visibility until weather deteriorated over a distance of 15 miles to IMC either early or late in the flight. *Weather deterioration early in the flight led to more plan continuations than late weather deterioration,* which is contrary to the prospect theory or "sunk costs" prediction. Wiegmann et al. (2001) concluded that "VFR flight into IMC may be due in part to poor situation assessment and experience rather than to motivational factors and risk-taking behavior that increase with time and effort invested in the flight (p. 10)".

On the other hand, O'Hare and Wiegmann (2003) presented pilots with short and long scud-running simulator scenarios with more marginal weather, which also degraded more gradually. That is, pilots all flew (i.e., chose to continue) at or near their personal minima (1500 foot ceiling) for either 22 nm or 66 nm after weather deteriorated 20 nm into their flights. O'Hare and Wiegmann found that "*those who had covered the greater distance [were] much more likely to continue with the flight than those who had only come half as far* (p. 28, italics mine)". Their result is consistent with prospect theory and the "sunk costs" prediction.

None of the theories listed above can reconcile the contradictory findings of these two studies. However, dissonance theory, a model from social psychology, offers a compelling explanation for the contradiction.

A Different Model: Dissonance Theory

Dissonance theory is generally considered to be the most powerful theory to come out of social psychology within the last fifty years (Jones, 1985). This very general and powerful theory provides a compelling model for VFR into IMC decision-making, and may inform development of effective countermeasures for those accidents. Brehm (1956) and Festinger (1957) first proposed that a person's actions could generate psychological discomfort, or cognitive dissonance, which the decision-maker would then attempt to reduce. Cognitive dissonance of the type relevant to this model is generated by the *free-choice experimental paradigm*, and the results have been generalized to many real-world situations.

According to dissonance theory, making a free (and difficult) choice between alternatives generates dissonance, due to the negative aspects of the chosen alternative, and the positive aspects of the non-chosen alternative (ie., dissonant cognitions). Because those dissonant cognitions nearly overbalance the consonant cognitions, the decision-maker will tend to change the balance (and reduce dissonance) either by reducing dissonant cognitions (i.e., negative attributes of the chosen alternative and positive attributes of non-chosen alternatives), or by adding to or accentuating consonant cognitions (i.e., positive attributes of the chosen alternative and negative attributes of non-chosen alternatives), or both. Dissonance is thereby reduced by increasing the difference between the chosen and the non-chosen alternative by *spreading of alternatives* or *post-decision distortion*. This spreading is accomplished in many ways, such as by reducing the estimated probability of negative outcomes for the chosen alternative and increasing it for the non-chosen alternative, or by forgetting non-chosen alternatives.

> A favourite technique for reducing the postdecisional dissonance, according to the theory, is to change cognitions in such a manner as to increase the attractiveness of the chosen alternative relative to the unchosen alternative(s). (Knox and Inkster, 1968, p. 319)

Overconfidence is another dissonance reduction mechanism, although confidence of judgment is uncorrelated with decision accuracy (Blanton et al., 2001).

Brehm and Festinger's initial formulations allowed only for post-decision dissonance to be generated and reduced, but more recent investigations (see Brownstein (2003) for a comprehensive review) have unequivocally demonstrated that if one alternative is favored or preferred, pre-decision distortion may also occur, and may be up to two times more influential than post-decision spreading of alternatives (Russo, Medvec and Meloy, 1996).

Although much discussion and argument has been generated in the field of social psychology regarding the mechanisms underlying attitude change findings, those attitude change findings themselves, and in particular, those involving "spreading of alternatives" between chosen and not chosen alternatives are powerful and ubiquitous.

Dissonance Theory and VFR into IMC Accidents

Dissonance theory predicts that a pilot's successive difficult decisions to continue a flight into marginal weather conditions rather than diverting to another airport or returning to the airport of departure, may cause subsequent judgements of the chosen alternative (i.e., to continue the flight) to be more favourable or positive, and judgements of the other alternative (i.e., to divert to an alternate or make a precautionary landing) to be more negative. Therefore, decisions to continue a flight as weather worsens may well become less conservative than the initial decision to begin the flight, leading the pilot to believe that a decision to continue a flight into marginal (or less) weather conditions (i.e., nearly IMC) is reasonable and within his or her personal risk management limits. Furthermore, as the flight continuation decision is revisited repeatedly, both pre-decision distortion and post-decision distortion may affect decision-making simultaneously, and finally, many small distortions may well sum to deadly effect. Paradoxically, a pilot who revisits the continuation decision more often may generate more distortion and make riskier weather decisions.

Flight Simulation Support for Dissonance Theory

The earlier presented contrast between Wiegman, Goh, and O'Hare (2001, 2002) and O'Hare and Wiegmann (2003) offers the most compelling simulator study evidence for the influence of dissonance theory mechanisms on weather decision-making. Wiegmann et al. (2001, 2002) found that *late weather deterioration led to fewer plan continuations than early weather deterioration*, contrary to the predictions of the "sunk costs" theory or the "get-home-it-is" model, and concluded that

> VFR flight into IMC may be due in part to poor situation assessment and experience rather than to motivational factors and risk-taking behavior that increase with time and effort invested in the flight. (Wiegmann et al, 2001, p. 10)

In contrast, O'Hare and Wiegmann (2003) reported

> a significant difference between those who covered the longer and shorter distance before the critical weather change, with *those who had covered the greater distance being much more likely to continue with the flight than those who had only come half as far* (p. 28, italics mine).

The critical experimental difference between the two simulator studies, and the explanation of their opposite results, may be that pilots in the first study flew in the simulator with 5000' ceilings and 5 mile visibility until weather deteriorated to IMC early or late in their flights, while in the second study, all pilots (by choice) continued their flights when the ceiling dropped to 1500' (at or below their personal minimums in all cases) only 20 nm into the flight, and then dropped to IMC (800') either 22 nm or 66 nm later. That is, the pilots in the second study incurred spreading of alternatives for either 22 nm (early IMC) or 66 nm (late IMC), causing substantially more distortion for late IMC pilots than for early IMC pilots and making them much more likely to commit plan continuation errors. However, all pilots in the first study flew in conditions far above their own personal minimums until weather deteriorated to below IMC over 15 miles, giving them the same brief opportunity to be influenced by dissonance theory mechanisms regardless of their assigned experimental condition, so that weaker factors prevailed.

Another VFR into IMC flight simulation finding also supports the dissonance theory model. Goh and Wiegmann (2001b) found that pilots who diverted made more accurate *post-decision* assessments of visual conditions than pilots who continued simulated flights into deteriorating weather, and suggested that this finding demonstrated "errors early in the decision-making process in the form of inaccurate assessments of visibility, …

compounded by other factors such as their greater willingness to take risks, greater confidence in their flight skills, and a reduced sense of vulnerability to weather hazards and pilot error (p. 5)". However, the less accurate assessments made by the continuing pilots may instead have resulted from post-decision spreading of alternatives – that is, their initial assessments may have been equally accurate, but had become distorted by the time they were reported to the experimenters.

Mapping Dissonance Theory onto Weather-Related Decision-Making

The action-based model of cognitive dissonance (Harmon-Jones and Harmon-Jones, 2002) is an excellent fit for weather decision-making. "Spread of alternatives" is maximized by implementation of a decision (e.g., by taking off into marginally acceptable weather). Harmon-Jones et al. state that an essential function of the spreading of alternatives is to transform a decision into effective and unconflicted action, while noting that this may "be maladaptive and dysfunctional … when persons maintain and bolster a commitment to a decision that clearly harms themselves or others (p. 712)". This concern particularly applies to pilots' decisions regarding weather, in which the decision involves successive judgements about changeable conditions.

Numerous characteristics of dissonance theory map closely onto weather-related decision-making in aviation. Some aviation-related factors that may increase either the spreading of alternatives or pre-decision distortion or both, as shown in the literature, include:

- Difficulty of decision (Brehm, 1956; Harmon-Jones and Harmon-Jones, 2002) – weather close to pilot's personal minimums makes decisions most difficult (maximizing the probability of VFR into IMC);
- Immediacy of implementation (Harmon-Jones and Harmon-Jones, 2002) – pilots take off almost immediately after assessing weather conditions;
- Public commitment (Festinger, 1957) – pilots must communicate their decisions to ATC and passenger(s);
- Importance of decision (Brownstein, Read, and Simon, 2004) – An incorrect weather-related decision can lead to fatalities or serious injuries, and aircraft loss or damage;
- Action orientation increases spreading of alternatives (Harmon-Jones and Harmon-Jones, 2002) – Pilots are very action oriented:
  - a decision to take off leads to very rigorous action sequence(s);
  - continuing to implement a flight plan involves very busy and ongoing action requirements to continue the plan;
- Sequential presentation of information – increases pre-decision distortion or confirmation bias through seeking information favorable to preference (pre-decision distortion) or to decision (post-decision - cognitive dissonance "spreading of alternatives") (Jonas et al., 2001) – Pilot information scans and actions are sequentially organized (e.g., checklists, planning, instrument scans);
- Favored alternative or "tentative preference" will tend to increase pre-decision distortion (Brownstein, Read, and Simon, 2004; Russo, Meloy, and Medvec, 1998) – Clearly, a pilot's favored alternative will be to begin, continue, and complete a planned flight;
- Good mood increases predecisional bias in a free choice task (Meloy, 2000) – Most pilots love to fly, and anticipation enhances their mood.

Proposed Tests of Dissonance Theory Hypothesis

The cross-country decision-making simulation scenarios developed by O'Hare and Wiegmann (2003) could be adapted to present the critical experimental conditions of both Wiegman, Goh, and O'Hare (2001, 2002) and O'Hare and Wiegmann (2003), by crossing high ceiling versus scud-running with early versus late weather deterioration to test the dissonance hypothesis that only in the scud-running conditions will pilots have a greater tendency to continue the flight into late marginal (and worse) weather than into early marginal (and worse) weather.

Several other variables could also be manipulated to test other predictions of the dissonance theory hypothesis, and to test the effectiveness of potential countermeasures:

- Severity of weather at initial decision to depart (less marginal weather yields less dissonance, hence reduced subsequent spreading of alternatives and reduced likelihood of continuing into weather below minimums.)

- Gradual versus sudden onset of weather deterioration (fewer decisions regarding marginal conditions means less spreading of alternatives and reduced likelihood of continuing into weather below minimums.)
- Fewer versus more decisions (termination decisions offered) controlling for flight length (more decisions yield more opportunities to generate dissonance, thereby increasing subsequent spreading of alternatives and the likelihood of continuing into weather below minimums.)
- Pilot versus passenger perceptions (as passengers don't make the decisions to depart or to continue the flight, they should feel little or no dissonance, so should be more able to accurately judge weather conditions.)

High Risk Situations and Potential Countermeasures

Given that dissonance reduction mechanisms increase the likelihood of VFR into IMC accidents, informing pilots of relatively high-risk situations that increase the likelihood of pre-decision distortion and post-decision spreading of alternatives, and suggesting countermeasures developed in accordance with the large body of dissonance theory literature may reduce the incidence of VFR into IMC accidents and fatalities.

For example, more gradual onset of weather deterioration increases the number of decisions regarding marginal weather conditions, resulting in more spreading of alternatives and distortion, and increasing the likelihood of continuing a flight into IMC. (Indeed, a paradox of dissonance theory is that the more frequently a careful pilot scans weather conditions in flight, the greater their risk of continuing a flight into IMC, all other things being equal.) However, because passengers don't normally make the decisions to depart or to continue a flight, they should feel little or no dissonance, and may assess weather conditions more accurately. Therefore, pilots may make more accurate decisions if they solicit and consider the opinions of knowledgeable passengers.

Because perceptual distortion of weather assessment and decision-making is gradual and progressive, the strategy of separating a decision to divert from prior decisions to continue may offer an effective countermeasure for distortion. For example, if an in-flight pilot asks "Would I take off into these conditions?", or even, "Would I recommend that an average pilot and my child or other loved one take off in these conditions?" and responds "No", then the most reasonable choice is likely to divert to an alternate.

Another option for reducing the influences of prior and subsequent decisions may be for a pilot to carry pictures of weather conditions close to his or her personal weather minima, to provide an unchanging standard of comparison. Perhaps the WeatherWise training material (Wiggins and O'Hare, 2003) could be adapted to that end.

Dissonance theory literature contains substantial additional information regarding means for reducing or eliminating pre and post-decision distortion. For example, less biased weather decision-making may be facilitated by asking decision-makers to justify their information choice sources, by inducing accountability for the decision process (by auditing and evaluating that process) rather than accountability for decision outcomes, and in general, by focusing on information rather than on a prior decision or favored alternative (Jonas et al., 2001).

Dissonance may account for some other plan continuation errors as well. In 2008, fuel management accidents accounted for 73 non-commercial fixed-wing accidents, 9 of them fatal (AOPA 2009). Although accidents of this type have decreased by 50% over the last ten years, some still occur in part from "failure … to make timely decisions to divert for fuel in the face of changing circumstances. (ibid, p. 14)".

Conclusion

Research is needed to determine if many weather decision-making errors and accidents, and perhaps other plan continuation errors, result from an active, ubiquitous and powerful characteristic of human cognition that tends to bias even the most conscientious pilots toward distorted situation assessment and risky decision-making. If so, research is also needed to inform the development of effective tools and training to further reduce the number of VFR into IMC accidents, and perhaps the number of fuel management accidents as well.

References

AOPA Air Safety Foundation. *2008 Nall Report: Accident Trends and Factors for 2007*. Retrieved from http://www.aopa.org/asf/publications/nall.html

AOPA Air Safety Foundation. *2009 Nall Report: Accident Trends and Factors for 2008*. Retrieved from http://www.aopa.org/asf/publications/nall.html

AOPA Air Safety Foundation database of NTSB aviation accident information. Retrieved from http://www.aopa.org/asf/ntsb/search_ntsb.cfm

Blanton, H., Pelham, B.W., DeHart, T., and Carvallo, M. (2001). Overconfidence as dissonance reduction. *Journal of Experimental Social Psychology, 37,* 373-385.

Brehm, J. W. (1956). Post-decision changes in desirability of alternatives. *Journal of Abnormal Social Psychology, 52*, 384-389

Brownstein, A. L. (2003). Biased predecision processing. *Psychological Bulletin, 129*(4), 545-568.

Brownstein, A. L., Read, S. J., and Simon, D. (2004). Bias at the racetrack: Effects of individual expertise and task importance on predecision reevaluation of alternatives. *Personality & Social Psychology Bulletin, 30*(7), 891-904.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Goh, J. and Wiegmann, D. A. (2001a). Visual flight rules (VFR) into instrument meteorological conditions (IMC): A review of the accident data. *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus, OH: The Ohio State University.

Goh, J. and Wiegmann, D. A. (2001b). An investigation of the factors that contribute to pilots' decisions to continue visual flight rules flight into adverse weather. *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: Human Factors & Ergonomics Society. 2001.

Goh, J. and Wiegmann, D. A. (2001c). Visual flight rules flight into instrument meteorological conditions: An investigation of the possible causes. *The International Journal of Aviation Psychology, 11*(4), 359-379.

Harmon-Jones, E. and Harmon-Jones, C. (2002). Testing the action-based model of cognitive dissonance: The effect of action orientation on postdecisional attitudes. *Personality and Social Psychology Bulletin (PSPB), 28*(6), 711-723.

Jonas, E., Schulz-Hardt, S, Frey, D and Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology, 80* (4), 557-571.

Jones, E. E. (1985). Major developments in social psychology during the past five decades. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology: Vol. 1. Theory and methods* (3rd ed., 47-107). New York: Random House.

Knox, R. E. and Inkster, J. A. (1968). Postdecision dissonance at post time. *Journal of Personality and Social Psychology, 8*, 319-323.

Meloy, M. (2000). Mood-driven distortion of product information. *Journal of Consumer Research*, 27 (December), 345-359.

O'Hare, D. & Smitheram, T. (1995). "Pressing on" into deteriorating weather conditions: An application of behavioral decision theory to pilot decision making. *The International Journal of Aviation Psychology, 5*, 351-370.

O'Hare, D. and Wiegmann, D. A. (2003). *Continued VFR Flight into IMC: Situational Awareness or Risky Decision Making?* Final Report, April 17, 2003. FAA Grant: 2000-G-010.

Russo, J.E., Medvec, V.H., and Meloy, M.G. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes, 66*(1), 102-110.

Russo, J. E., Meloy, M. G. and Medvec, V. H. (1998). Predecisional distortion of product information. *Journal of Marketing Research,* 35(4), 438.

Wiegmann, D., Goh, J. and O'Hare, D. (2001). *Pilots' Decisions to Continue Visual Flight Rules (VFR) Flight into Adverse Weather: Effects of Distance Traveled and Flight Experience.* Technical Report ARL-01-11/FAA-01-3, September 2001, FAA Contract DTFA 00-G-010.

Wiegmann, D., Goh, J. and O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules flight into adverse weather. *Human Factors, 44*(2), 189-197.

THE ROLE OF EXPERTISE IN VFR FLIGHT DECISIONS WITH INCONSISENT WEATHER INFORMATION

Jordan Petry, Landon Thomas, Heewoong Park, Wai-Tat Fu
University of Illinois at Urbana-Champaign
Urbana, Illinois

To study the role of expertise in weather-related VFR flight decisions we conducted two experiments in which experts and novices were presented with inconsistent weather information. In the first experiment 51 pilots performed a series of pre-flight planning decision tasks. We found that experts were better at selecting reliable information than novices. In the second experiment, 24 pilots made a VFR flight using the XPlane flight simulator. Subjects were randomly assigned to one of four conditions with a different combination of good and bad flying (rendered in XPlane) and ATIS weather. We found that in poor weather conditions novices flew farther into bad weather than the experts before diverting. In the condition with bad ATIS and good flying weather, experts all made it to the destination while 67% of the novices diverted. Results have important implications to how dynamic decision-making skills can be included in pilot training.

Weather-related accidents are a significant problem in general aviation (GA). According to the AOPA Nall Report (2009), while weather-related accidents made up a small proportion of total GA accidents in 2008 (50 accidents, 4% of total), the fatality rate for these accidents was disproportionately high, at 70%. Previous studies (e.g., Wiegmann, Goh, and O'Hare, 2002) found that less experienced pilots were more likely to fly further into adverse weather than more experienced pilots, which indicates there are possibly opportunities for training to help reduce the prevalence of weather-related accidents. Because differences across weather products between experts and novices is not something that Wiegmann, Goh, and O'Hare address in their study, we wanted to investigate this particular aspect of weather decision-making. To that end, we designed two experiments to understand how expert and novice pilots access and integrate multiple sources of weather information to reach a go/no-go decision. In particular, we expect that decision errors will be more likely when there are inherent inconsistencies among weather reports. For example, when a local area weather report shows that the weather is good but a weather forecast report indicates that the flight will likely encounter instrument meteorological conditions (IMC), decisions by expert and novice pilots will differ. To determine this, we designed multiple sets of inconsistent weather reports to understand how expert and novice pilots resolved the inherent conflict when making a go/no-go decision. In particular, we are interested in how pilots utilize different sources of weather information, and how they prioritize each product (e.g. present weather v. forecast, human-produced v. machine-produced, etc), and if there are differences between how experts and novices use this information. Results will provide significant implication to pilot training.

Experiment 1

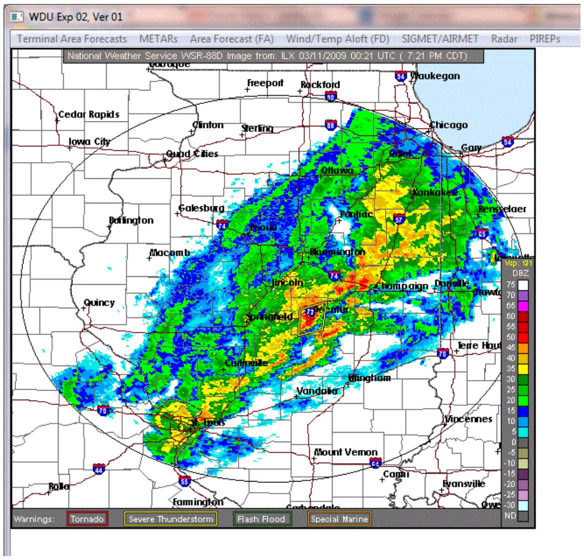
Our first experiment investigates differences in how expert and novice pilots retrieve and view pre-flight weather information. We developed a program to emulate computer-based weather sources such as Direct User Access Terminal (DUAT) and Aviation Digital Data Service (ADDS) which would display the weather to the subject and record information for later analysis. Given a set of seven text and graphical, present and forecast weather products, subjects were asked to make a go or no-go decision for a VFR (visual flight rules) cross-country flight in Central Illinois.

Materials

Weather Data Usage (WDU) is a program written to simulate a Direct User Access Terminal (DUAT) weather briefing and allow recording of data parameters such as mouse clicks, time on-screen, etc. WDU uses a menu-like interface to allow subjects to view their choice of seven weather products one at a time (See figure 1). We created seven kinds of weather reports (TAF, METAR, FA, FD, SIGMET/AIRMET, RADAR, and PIREPs). We divided these reports into multiple categories. First, we identified the set of reports that are judged to be most important based on a pilot study. We found that TAF and METAR seem to be most important, and we categorized these reports as "Major", and the other five reports as "Minor". We also divided the reports into whether they provide forecast information (FA, FD, TAF) and current weather information (the rest); and whether the current products were produced by humans (PIREP) or by machine (METAR, RADAR). We then created two sets (either good/bad or bad/good) of inconsistent weather reports for each of these 3 divisions of weather reports (i.e., Major/Minor, Present/Forecast, and Machine/Human), and two sets of weather reports where all 7 reports were consistent (all good and all bad). There were therefore a total of 16 sets of weather report. Each participant was given these 16 sets of weather reports in a random order.

Figure 1.

Weather Data Usage (WDU) Interface.



National Weather Service WSR-88D Image from: ILX 03/11/2009 00:21 UTC (7:21 PM CDT)

Note. Weather products are accessed via the menu on top.

Participants

Fifty-one subjects were recruited from the part 141 flight school at the University of Illinois Institute of Aviation. For the purposes of this experiment the researchers define expert and novice pilots as those who have more or less than 200 total flight hours, respectively. This consisted of twenty-eight experts and twenty-three novice pilots. All subjects hold at least a Private Pilot certificate. We obtained informed consent for each participant. Participants completed a questionnaire regarding demographics, pilot certification, total flight hours, and personal weather practices such as weather minimums and priority placed on certain weather products.

Task

Participants were asked to conduct sixteen simulated pre-flight weather briefing scenarios using the information provided to them through our "Weather Data Usage" (WDU) platform. Participants would have access to seven text and graphical weather products for each scenario and would make a "go" or "no-go" decision for a VFR (Visual flight rules) flight between Greater Kankakee, Illinois (KIKK) and Willard-Champaign, Illinois (KCMI) based on the information provided through WDU and using their knowledge of weather and of their personal and legal limits.

Procedure

Subjects were briefly shown the use of the menu system and other task requirements. Subjects were able to move freely between all weather products and stay on each for as long as they liked, until the point at which they made their go/no-go decision. Subjects were given task instructions and a headset microphone for recording verbal protocol, and were asked to "think aloud" to allow us to measure how they interpreted the information presented on the screen.

In each scenario certain products would present information contradictory to other products, which allowed us to see on which products subjects relied more heavily. At the bottom of the interface was an "end scenario" button which the subject was instructed to click once they had made their "go/no go" decision. At this point they selected their choice on the program as well as stated any reasons for their choice to the voice recording. Once they clicked on their "go/no go" choice they were invited by another button to continue to the next scenario (allowing them to take breaks between scenarios). Subjects would repeat this process through each scenario until completing all sixteen scenarios. The subjects were asked to make any final comments to the voice recording, then remunerated and dismissed.

Results

The mean estimated total flight hours for experts was 328.43 and the mean estimated total flight hours for novices was 110.33. When there was high consistency among products, twelve percent more experts than novices made a "go" decision in good weather. On average the experts made eight percent (three percent in bad weather) more "go" decisions than did novices. Table 1 below shows the total go and no-go decisions for each type of scenario, based on consistency among weather products.

Table 1.

Go/No-Go decision totals in Experiment 1.

		AG	AB	MG	MB	HG	HB	PG	PB
Expert	Go	26	16	38	17	43	5	28	20
	No Go	28	38	16	37	11	49	26	34
Novices	Go	15	13	19	10	31	4	19	13
	No Go	27	29	23	32	11	38	23	29

Note. In column headings, first letter A=all, M=inconsistent major/minor, H=inconsistent human/machine, and P=inconsistent present/forecast; second letter G=first set is Good, B=first set is Bad. Thus, AG=all good; MB=major set is bad, minor set is good, PG=present is good, forecast is bad, etc.

As shown in Table 1, the most pronounced difference between experts and novices was in scenarios with low consistency between "major" and "minor" weather products when the major weather reports indicated good weather, but the minor reports indicated bad weather (i.e., in the MG column). In fact, twenty-five percent more experts made a "go" decision in these scenarios compared to novices ($p<0.05$). Interestingly, the difference was not as large when the major set was bad and the minor set was good, in which case most expert and novice pilots decided not to go. Results indicated that expert pilots would more likely choose to ignore the minor reports, where novice pilots tended to equally weigh these reports (thus more likely making a "no-go" decision).

In addition, the difference between experts and novices was also large when the forecast weather reports were bad but present was good. Results indicated that experts would more likely decide to go, while novice pilots would more likely decide not to go in this set of inconsistent reports. However, this difference did not reach statistical significance ($p=0.2$). Results again indicated that experts tended to ignore forecast weather when they indicated bad weather, while novice pilots tended to equally weigh the reports.

Results of this experiment draw some interesting questions. Why are experts more likely to make a "go" decision when some weather products report the probability of entering IMC? Is this based on experience, and they truly will not enter IMC when their pre-flight planning says they might? Do they simply feel more comfortable with encountering IMC because they are instrument rated (all 28 experts are instrument rated, while only 7 novices are)? If these decisions to go when some reports show possible adverse weather, training should instill long-lasting focus on all weather products. If the experts experience shows that some weather products, when taken in context, are more reliable than others, than this can be applied to training as well. In any case further study should take place to investigate these outcomes application to flight training.

Experiment 2

Materials

Our simulator was made up of the following equipment: XPlane 8.60 simulator software, a control column manufactured by Precision flight controls (equipped with a control yoke, rudder pedal assembly, and a throttle input knob), and mouse capability for point-and-click access to various functions within the cockpit such as setting and identifying radio frequencies and navigational aids, as well as speakers for hearing ATIS information produced in the simulation. The simulator was set up to emulate a Piper Archer, which pilots in the flight program are familiar with. XPlane was modified for our experiment to allow the researchers to change weather conditions between "good" (calm winds, clear skies) and "bad" (strong winds, wind shear, overcast sky), and change output of "good" and "bad" ATIS weather reports for both KIKK and KCMI.

Participants

Twenty-four subjects were recruited from the part 141 flight school at the University Of Illinois Institute Of Aviation. This consisted of twelve experts and twelve novice pilots (defined previously). All subjects hold at least a Private Pilot certificate. After completing the informed consent and questionnaire mentioned above (Experiment 1, Participants), subjects were seated in our flight simulator.

Task

Participants were asked to make a simulated VFR (Visual flight rules) flight from Greater Kankakee, Illinois Airport (KIKK) to Willard-Champaign, Illinois Airport (KCMI). They were provided with the applicable VFR Sectional chart as well as use of navigational radios in the simulator. In flight they would also have access to supplied ATIS information for KIKK and KCMI.

Procedure

Subjects were given instructions to make a VFR flight from KIKK to KCMI. They were briefly shown the use of the simulator including how to change radio frequencies for ATIS briefings. Participants were told that the general radial from KIKK to KCMI was 190° if they wished to use VOR navigation. The subjects were asked to use verbal protocol methods for the duration of the flight. Participants were instructed that if, at any point during the flight, they felt that they would make a weather diversion, they should state those intentions explicitly on the verbal protocol and land the plane. Subjects were assigned to one of eight condition groups (three per group), described in Table 2.

Table 2

Experiment 2 subject groups.

	Group 1	G 2	G 3	G 4	G 5	G 6	G 7	G 8
Condition EG	$_{A}G_{W}$ EG	$_{A}B_{W}$ EB	$_{A}G_{W}$ EB	$_{A}B_{W}$ NG	$_{A}G_{W}$ NG	$_{A}B_{W}$ NB	$_{A}G_{W}$ NB	$_{A}B_{W}$

Note. E = expert, N = novice, G = good, B = bad, $_{A}$ = ATIS, $_{W}$ = weather trigger

Results

The mean estimated total flight hours for experts were 330.15 and the mean estimated total flight hours for novices was 125.91.The distance between Greater Kankakee, Illinois Airport (KIKK) and Willard-Champaign, Illinois (KCMI) is about 64 miles. Across all conditions, experts flew an average of 44.89 miles while novices flew an average of 39.95 miles. Table 3 shows the expert and novice average flight distance between conditions. The condition with the greatest performance difference between experts and novices was the condition in which subjects were presented with good weather, but an ATIS report with bad weather. All three experts in this condition completed the flight to KCMI (average distance 64.22 miles), while only one of three novices completed the flight (average distance 42.71 miles). Another condition which produced interesting results was that which had bad weather but a good ATIS report. Experts on average diverted 4.1 miles earlier than novices.

Table 3.

Average flight distance.

condition G	$_{1}G_{2}$	$G_{1}B_{2}$	$B_{1}G_{2}$	$B_{1}B_{2}$
expert average distance (SM)	64.22	30.39	64.22	20.71
novice average distance (SM)	64.22	34.49	42.71	26.47

Note. $_{1}$= reported weather (ATIS), $_{2}$= actual weather.

These results may indicate that expert pilots rely more on what they are experiencing in real time on their flight, while novices tend to rely more heavily on information provided to them via external sources. This supports the findings of Wiegmann, Goh, and O'Hare (2002), who found a negative correlation between flight experience and distance flown into adverse weather. It seems that when comparing experts and novices, experts trust internal sources (i.e., their own flight experience) more heavily than external sources (i.e., weather reports in ATIS transmissions), while the opposite is true for novices. In these conditions, the novices' decision to trust external sources is either dangerous or expensive. In the bad weather condition with a good ATIS report, novices are flying into adverse weather that they are not equipped to handle, and could end up as one of the statistics in the Nall Report. In the condition where the weather is good but the ATIS is reporting bad weather, novices make a decision to divert which is very possibly an expensive one (e.g., more fuel and Hobbs time, hotel, different form of travel, etc.) when the true weather condition is actually safe to fly in. It is possible that these findings also provide opportunity for better training in weather aeronautical decision-making; further investigation should take place.

Conclusion and Discussion

We presented results from two experiments that studied how expert and novice pilots resolved conflict among inconsistent weather information during pre-flight planning and in-flight decisions. Experiment 1 shows that expert pilots, compared to novices, are in general more likely to make a "go" decision in bad weather when presented with inconsistent reports. This could indicate a dangerous over-confidence effect in more experienced general aviation pilots, and should be investigated further.

One important result we found was that expert and novice pilots differed the most when the major and minor sets of weather reports were inconsistent, and when the present and forecast reports were inconsistent. In particular, when the minor set or the forecast sets of weather reports indicated bad weather, expert pilots tended to put less weight on them, making them more likely to make a "go" decision when the major or present weather reports indicated good weather. On the other hand, novice pilots tended to be more "conservative", as they seemed to put equal weights on all reports and would less likely to make a "go" decision when some reports indicated bad weather.

Interestingly, in Experiment 2, when making in-flight decisions, novice pilots were more likely to fly further into bad weather before making a decision to divert. This suggested that while novice pilots were relatively more conservative when making a "go" decision compared to experts, they were less likely to detect bad weather and decide to divert. It was possible that, for example, novice pilots had fewer attentional resources while flying, and were less experienced in handling changes in weather, supporting Wiegmann, Goh, and O'Hare's results (2002). It is possible that even though expert pilots made more "go" decisions even when there were reports indicating bad weather, they were more prepared to change in-flight, and thus were more ready to detect changing a weather situation. Our current results provide direction for future research to investigate these issues further to determine training techniques to prevent low-time pilots from continuing flight into dangerous weather situations.

References

Kenney, D. J., and Deres, P., Peterson, B., &Vasconcelos, K. (2009). *2009 Nall Report*. AOPA.

Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules into adverse weather. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*(189). Retrieved January 15, 2011, from <http://hfs.sagepub.com /content/44/2/189>.

Acknowledgements

We would like to thank the staff and faculty of the Institute of Aviation and the Human Factors Division at the University of Illinois. Specifically, thanks to Sybil Philips and Laura Gerhold from the Institute of Aviation for allowing us to post flyers and send e-mails to recruit participants. We would like to thank Karen Ayers for helping with reception of participants and acquiring our funds for remuneration. We would like to thank Ron Carbonari for helping with the computer side of data collection.

WEATHER INFORMATION PRIORITIES FOR COMMERCIAL PILOTS AND DISPATCHERS

Russell J. Branaghan
Roger W. Schvaneveldt
Cognitive Science and Engineering Program
Arizona State University, Mesa, AZ

The Next Generation Air Traffic System's (NextGen) goal is to increase capacity three-fold (JPDO, 2007). Given that approximately 70 percent of system delays can be attributed to weather, planning is focused on reducing weather-related delays by at least fifty percent (Leader, 2007). NextGen plans to integrate information from multiple sources, providing the same information to pilots, controllers, and dispatchers. However, different stakeholders may require different information at different times. This research identifies information needed by dispatchers and commercial pilots for pre-flight and in-flight planning and decision-making.

Thirty dispatchers and sixteen commercial pilots prioritized weather concepts. Results illustrate the information elements that were important to each group during each phase. Results show that pilots pay attention to fewer information elements pre-flight, however the two groups rely on similar information once the flight is airborne. User interface implications are presented.

The United States Air Transport System is at capacity, and the cost of operations and maintenance are too high for the revenue being generated. The solution to this problem will not be found in evolving the current air transportation system, but instead it will require an entirely new design (Arbuckle, et al, 2006). This new system, the Next Generation Air Transportation System (NextGen) has the goal of safely increasing capacity three fold by 2025 (JPDO, 2007). This goal will be achieved by reducing spacing among aircraft, improving departure and arrival arrangements, and potentially running simultaneous operations on a single runway. Scheduling will be enhanced through data driven processes that facilitate planning, make information available to multiple communities of interest, and distribute decision making effectively. Additionally, weather information will be better assimilated into decision making, providing a common weather picture to multiple stakeholders, integrating multiple sources of weather information, feeding multiple weather forecasts, and reducing the need for human interpretation.

This focus on improving weather information is prudent, since approximately 70 percent of national airspace system delays are attributed to weather (Leader, 2007). Indeed, NextGen has the goal of reducing weather-related delays by at least fifty percent (Leader, 2007). An important piece of the effort will be to integrate disparate weather data so that information can be shared among all NextGen users. It is important, however, to recognize that different user groups have different responsibilities, and work in different environments. Consequently, the various user groups will likely require different information at different times.

Whereas general aviation (GA) pilots plan their own flights and conduct their own weather research, commercial pilots rely on a substantial support system including dispatchers and meteorologists. The information needed by pilots and dispatchers may be different, though obviously not contradictory. Pilots and dispatchers are likely to have different information needs and place different priorities on pieces of information at different phases of flight. Showing them the exact same display at all times may not be wise.

Our research focuses on determining what weather information should be shown to pilots and dispatchers and when. We identify the pieces of information that are critical to both groups at all times, the pieces of information that are needed rarely, and the similarities and differences in the requirements of the two groups. Although the Airline Operations Center (AOC) includes many roles, such as supervisors, coordinators and meteorologists, the majority of the AOC personnel work in dispatch. As a result, to simplify exposition we will use the term dispatcher to refer to these research participants, even though they sometimes included professionals in the other roles as well. According to Heuwinkel (1993) dispatchers are responsible for conducting the following activities:

- Develop and file flight plan
- Gather weather information
- Provide weather information to the pilot
- Respond to pilot requests for weather information and rerouting
- Distribute information on changing weather to the pilot
- Reroute aircraft
- Develop strategic flight schemes for group of flights according to weather conditions.

Broadly, dispatch involves two components: flight planning and flight following. Flight planning occupies the most time and is proactive in nature. Flight following is usually uneventful, but can also be the most critical and time-compressed activity. This is especially true when unexpected weather occurs. In such situations, dispatchers need to make decisions quickly, taking into account the location of the airplane, proximity to various airports, details about the airports, airplane configuration, and fuel status. Further, because weather can be geographically broad, it may affect many aircraft at the same time. Generally, the process of flight following involves looking for changes. The dispatchers track the weather, looking for deviations from the forecast. If adverse weather develops, the dispatcher may ask the pilot to hold for a while (while the weather dissipates) before attempting to land. If the adverse weather is not likely to disperse, the dispatcher advises the pilot to proceed to an alternate airport.

Ensuring safety typically requires choosing routes that avoid “big weather”, choosing appropriate departure and arrival alternatives, and avoiding turbulence. It is the last of these that is the most demanding, since turbulence can be localized and difficult to detect. Keeping the pilot updated is crucial for dealing with turbulence, ensuring customer safety and comfort. Communicating turbulent conditions to the pilot in a timely manner enables the pilot to change altitude or to warn passengers of impending turbulence.

Several studies have identified important weather factors in aviation (Beringer & Schvaneveldt, 2002; Comerford, 2004; Heuwinkel, 1993; Krozel, Capozzi, Andre, & Smith, 2003). Two of these (Beringer & Schvaneveldt, 2002; Heuwinkel, 1993) have also identified priorities associated with the factors. Schvaneveldt, Branaghan, Lamonica, and Beringer (2008) reviewed these studies and selected the factors shown in Figure 1 as representative of the weather factors identified in all of the studies. The diagram shows that weather factors tend to cluster around five central nodes: precipitation, clouds, wind, visibility and temperature. These informational factors were used in the priority ratings obtained in the present investigation.

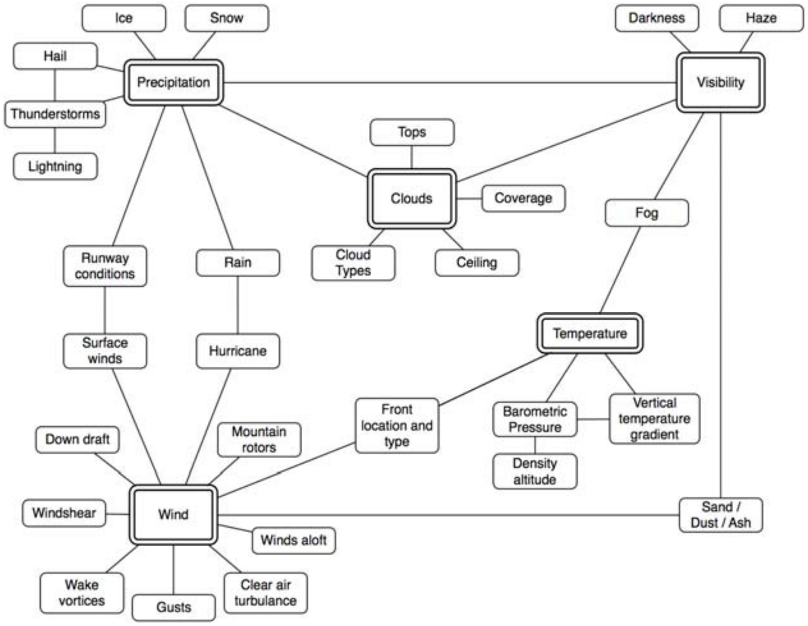


Figure 1. Weather factors and relations among them.

Method

Participants

Thirty dispatchers employed in the AOCs of four airlines based in the United States and sixteen commercial pilots served as research participants. The airlines included two large operations with substantial international service, one smaller operation with national service, and a small regional operation. Although the general procedures and goals are similar across the various AOCs, there are several differences in the specific weather products being used and in the extent to which individual airlines create their own software systems for the use of their dispatchers and meteorologists.

Materials

A website, Survs.com, was used to collect priority rating data. Participants accessed the survey via a link sent in an invitation email. In an earlier study of GA pilots (Schvaneveldt, et al., 2008), we included three phases of flight (departure, en route, and arrival) in addition to planning. For dispatchers, it made more sense to consider only two phases, planning and flight following; this is how they organize their activities. Consequently, the data from commercial pilots were collapsed across in-flight phases for comparison by taking the maximum priority for each factor across the flight phases.

Procedure

The survey presented an introductory page providing an overview of the study and estimating the time it would take to complete. Next, the software displayed instructions for how to complete the survey, and a demographic questionnaire. Participants then rated the of each information element for the activities of flight planning and flight following on a 4 point scale, with 1 representing least important and 4 representing most important. Completion took approximately 10 minutes. Data collection took place over a three-month period. The importance ratings were converted to priorities by subtracting them from 5. Thus, 1 becomes the highest priority and 4 is the lowest.

Results and Discussion

Table 1 shows the median priority ratings for dispatchers and commercial pilots. Results suggest that, certain weather information related to visibility (e.g., fog, haze, sand and dust), dangerous precipitation (e.g., ice, and freezing rain), dangerous wind (e.g., hurricanes, tornadoes, and windshear) and runway conditions should always be salient. These high priority items are highlighted in the table.

Other informational needs depend on the stakeholder's role. Specifically, during both flight planning and flight following, pilots seem to depend on dispatchers to identify some information about clouds, precipitation, visibility and wind. This may be due to the fact that dispatchers are generally monitoring certain routes to begin with, and likely already have this information for several flights. It also may be due to the fact that dispatchers have substantially more room for informational displays.

Table 2 illustrates the correlations among the priorities assigned for phases and roles. Again, it highlights that dispatchers and pilots require somewhat different information. Indeed, while pilots require slightly different information during flight planning and flight following, dispatchers seem to require pretty much the same information the entire time. This suggests that in-flight weather systems designed for the pilot could layer information, placing the most critical information on the top layer and less critical information at the next layer.
Table 1.

Median Priority Ratings for Dispatchers and Commercial Pilots.

Information Element	Planning		Flight Following	
	Dispatch	121	Dispatch	121
Barometric pressure	3 3		3 2	
Clouds/ceiling	1 2		1.5 2	
Clouds/coverage	1 3		2 2	
Clouds/tops	2 3		2 2	.5
Clouds/types	2 2		2 2	
Density altitude	3 2		3 2	
Front location & type	2 2		2 2	
Precipitation/Ice/freezing rain/sleet	1	1	1	1
Precipitation/Rain	1.5 2		2 2	
Precipitation/Snow	1 2		1 2	
Present/Forecast Temperature	2 2	.5	2.5 2	
Runway conditions	1	1	1	1
Thunderstorms/Hail/Lightning	1	1	1	1
Vertical temperature gradient	3 3		3 3	
Visibility	1	1	1	1
Visibility/Fog (dew point)	1 2		1	1
Visibility/Haze	1 2		1.5 2	
Visibility/Sand/Dust/Ash	1 2		1	1
Wind/Clear air turbulence	1.25 2		1 2	
Wind/Downdraft	1.25 2		1	1
Wind/Gusts	1.25 2		1.25 2	
Wind/Hurricanes	1	1	1	1
Wind/mountain rotors	1.25	1	1.5	1
Wind/Surface winds	1 2		1.25	1
Wind/Tornadoes	1	1	1	1
Wind/Wake vortices	2 2		2	1
Wind/Winds aloft	1.5 2		1.75 3	
Wind/Windshear	1	1	1	1

Note. Shaded items indicate median priority of 1.5 or below.

Table 2.

Correlations of Priority Ratings Among Pilots and Dispatchers and Phases

		Planning		Flight Following	
		Dispatch	Pilot	Dispatch	Pilot
Planning	Dispatch	-			
	Pilot	0.58	-		
Flight Following	Dispatch	0.92	0.69	-	
	Pilot	0.55	0.71	0.64	-

Ideally, for the pilot, information would be tied to the 4-D profile of the flight. This would provide information about the factors relevant to their particular flight and the appropriate time. The situation for dispatchers is quite different because they are simultaneously involved with several flights with various 4-D profiles. Dispatchers need a bigger picture of the weather to accomplish their tasks. Of course, they have the luxury of having multiple large displays that can be customized to their needs. The limits of the cockpit place more constraints on displays for pilots.

ACKNOWLEDGMENTS

The authors thank the Federal Aviation Administration, Delta, American, Mesa, and Frontier Airlines for their participation.

References

Arbuckle, D., Rhodes, D., Andrews, M., Roberts, D., Hallowell, S., Baker, D., Burleson, C. Howell, J., & Anderegg, A. (2006). U.S. vision for 2025 Air Transportation. ATCA Journal of Air Traffic Control, (1), http://www.atca.org/index.asp.

Beringer, D. B, & Schvaneveldt, R. W. (2002). Priorities of weather information in various phases of flight. *Proceeding of the Meetings of the Human Factors and Ergonomics Society*, Baltimore, October.

Comerford, D. (2004). Recommendations for a Cockpit Display that Integrates Weather Information with Traffic Information. NASA/TM-2004-212830.

FAA NextGen Website: http://www.faa.gov/about/initiatives/nextgen/ accessed 10/17/2009

Heuwinkel, R. J. (1993). Strawman of industry needs and priorities. Proceedings of National Aviation Weather Users' Forum, DOT/FAA, November 30 – December 2, Reston, Virginia.

JPDO (2007) Concept of Operations of the Next Generation Air Transportation System, http://www. faa. Gov.

Krozel, J., Capozzi, B., Andre, T., & Smith, P. (2003). Future national airspace system - design requirements imposed by weather constraints. *2003 AIAA guidance, navigation, and control conference and exhibit*; Austin, TX; Aug. 11-14, 2003, AIAA Paper 2003-5769.

Leader, C. (2007). Statement of Charles Leader, Director, Joint Planning and Development Office, before the House Committee on Science and Technology, Subcommittee on Space and Aeronautics on JPDO and the Next Generation Air Transport System: Status and Issues, March 29, 2007.

Schvaneveldt, R. W., Beringer, D. B., & Lamonica, J. (2001). Priority and organization of information accessed by pilots in various phases of flight. *International Journal of Aviation Psychology,* 11(3), 253-280.

Schvaneveldt, R. W., Branaghan, R. J., Lamonica, J., & Beringer, D. B (2008). Weather information priorities in various phases of flight with special attention to weather information in NextGen. Report on Tasks 1, 2, and 3 of FAA contract DTFAAC-08-P-00873.

COMPLEX DATA INTEGRATION FOR TRAINING IN TECHNOLOGICALLY-ADVANCED AIRCRAFT

Michael S. Nolan
Department of Aviation Technology, Purdue University
West Lafayette, IN, USA

Erin E. Bowen
Department of Technology Leadership and Innovation, Purdue University
West Lafayette, IN, USA

Michael W. Suckow
Department of Aviation Technology, Purdue University
West Lafayette, IN, USA

Morgan E. Hall
Department of Aviation Technology, Purdue University
West Lafayette, IN, USA

Brent D. Bowen
Department of Aviation Technology, Purdue University
West Lafayette, IN, USA

Flight data collection systems have been used for many years to help monitor and trend flight parameters to better understand aircraft performance and maintenance. However, this information is rarely used to understand and enhance cognitive/behavioral factors impacting flight and maintenance training. This conceptual paper outlines a strategic research model in the collegiate aviation setting, designed to manage the successful integration of data from technologically-advanced aircraft (TAA) into flight, maintenance, and dispatch training and operations. A critical review of the key issues and approach framework should be discussed in the research community. The model will address issues of organizational change management, training and curriculum modification research to accompany data utilization in conjunction with TAA aircraft integration, safety and human factors issues, and perceptual/attitudinal factors. The model presents a framework for assessing and guiding the potentially revolutionary change inherent when advanced aircraft and their data are integrated into training, maintenance, and operations.

The purpose of this article is to lay the foundation for aviation data integration not only throughout collegiate aviation education, but in the global training context. Such integration not only has the potential to improve training efficiency and performance but also improve overall safety. The persons who engage in the early phase of this innovation (early adopters themselves) will be those who provide leadership as they look to change and enhance aviation education technology and safety. The insertion of technologically advanced aircraft into the existing collegiate aviation setting is a spur to drive a paradigm shift from reactive to proactive/predictive aviation systems, and the authors hope that the present paper provides a framework for managing this fundamental shift in education, technology, and innovation.

Flight data recorders were first mandated by The Civil Aeronautics Administration in 1958 to record the actual flight conditions of the aircraft including; heading, altitude, airspeed, vertical accelerations, and time (Haas, Walker, & Kough, 2008). As the years and technology progressed the amount of information collected from the aircraft increased dramatically and interest grew in the application of this data to improve safety and performance. Today, these data gathering systems have evolved into what is now known as Flight Operational Quality Assurance (FOQA) programs (Haas, Walker, & Kough, 2008). FOQA programs were first implemented by scheduled airlines to increase safety, but economic value was quickly realized as the airlines became more familiar with the data collection process. FOQA-style programs moved through the aviation community from scheduled airline operators to helicopters to commercial Part-135 operators and eventually to the military. The data collection programs were producing results from reduced engine removals and increased fuel management to increased crew efficiency and reduced insurance premiums (Haas, Walker, & Kough, 2008).

Modern flight data collection systems have the capacity to gather information on a wide array of both aircraft system and operator performance components. Today flight and engine parameters are being used to increase the maintenance cycles of engines by monitoring internal temperatures and vibrations and extending the life of airframes by monitoring stresses during flight. One study done on the Red Arrow Flight team of the UK showed that wingman were typically experiencing structural stresses double and triple that of the lead aircraft, something we may never have known without structural recorders (Brooks, 1999). While this is a significant gain from previous reactive maintenance systems, there is limited evidence demonstrating that such data are being used to refine training, design, or display information (Stephens, 2004).

The introduction of flight data collection systems via the conduit of Technologically-Advanced Aircraft (TAA) into general aviation presents a revolutionary shift in the information accessibility of GA operations. The basic definition of a Technologically-Advanced Aircraft is an aircraft containing, at minimum, a moving-map display, an IFR-approved GPS navigator, and an autopilot (AOPA Air Safety Foundation, 2005). In many TA aircraft the aforementioned equipment is accompanied by new-generation avionics, traffic, weather, and terrain avoidance information all with system redundancies. The TA aircraft has certainly made an impact in regards to aircraft production. According to the Aircraft Owners and Pilots Association (AOPA, 2005), in 2004 there were 1,758 light GA pistons manufactured by General Aviation Manufacturers Association (GAMA) member companies, of which ninety-two percent were TAA or contained TAA-like equipment. Today, even though most general aviation aircraft still utilize analog or “steam” gauges, new production of general aviation transportation aircraft such as those manufactured by Cirrus and Diamond are virtually all TAA (AOPA Air Safety Foundation, 2005).

Challenges of TAA Integration

The substantial growth in TAA production for the general aviation sector presents significant training challenges, particularly in the collegiate aviation environment. Initially, financial considerations for integrating advanced aircraft into existing training programs are greater. This includes not only the initial higher purchase price for TA aircraft as compared to their non-TAA counterparts, but maintenance costs for the TA aircraft run at a higher level over the long term. In addition, TAA software and tools often require subscription services with annual costs to maintain accuracy of charts and functionality of databases for TAA data storage. Retraining of the flight instructor workforce to include TAA technology and concomitant revisions to training curricula are also expenditures a collegiate training program must consider as part of the TAA evolution.

While financial considerations of TAA integration may be one challenge facing collegiate aviation, of equal (or perhaps greater) concern are the myriad strategic and conceptual programmatic changes that such technology bring to training. TA aircraft cannot simply be thought of as a new tool to fit into an existing structure, but should rather serve as an impetus for a substantial and critical look at the role of information and information capacity in all aspects of aviation. The defining feature of TA aircraft is that they are "technologically advanced"; that is, students training in such aircraft are required to learn both the technical aspects of flying as well as the utilization of the technology integrated into the aircraft.

Previous research demonstrates that this combination of technology familiarization and flight skill training provide significant initial stress and complexity for the beginner pilot, leading to decreased situational awareness and an increased risk of human factors-related errors (e.g., Wickens et al., 2004). As part of the "SAFER SKIES" initiative the FAA conducted a General Aviation TAA safety study to research the early observed safety issues with TA aircraft. Safety problems were identified in the accidents that the team studied that were typical problems found after an introduction of new technology. The safety problems were also all characteristic of general aviation pilot judgment errors, (Fiduccia et al., 2003) these two factors combined contribute to the large training concern of new TAA pilots; "However, the existing training infrastructure currently is not able to provide the needed training in TAAs," (p.4). One program to identify and develop mitigation strategies for these issues is the SAFER project, a collaboration between Middle Tennessee State University and the NASA Langley Research Center. The SAFER project's goals are to discover improvements and adaptations of the traditional pilot training methods to complement technologically-advanced aircraft. This is researched by taking beginner pilots through their private pilot license and instrument rating in TA aircraft (Craig, Bertrand, Dornan, Gossett, & Thorsby, 2005).

One of the key features of collegiate aviation programs that distinguish them from other, commercially- or professionally-focused programs are their foundation within the academic, university setting. This setting presupposes that collegiate aviation students are present not only to earn the required certifications for their desired aviation career, but also to earn an academic degree with a broad-level base of knowledge. In addition, collegiate instructors possess not only technical expertise and certification, but are immersed in a foundation of scholarship and the continued contributions to aviation as an academic field. Aviation training in the collegiate setting should thus be set within a context of the academic discipline of aviation; the integration of TA aircraft to training makes this context of the highest necessity.

Given the potentially revolutionary nature of TAA inclusion in collegiate aviation, the authors argue that complete integration of the full potential of TAA technology cannot be achieved without the implementation of a strategic research model that provides a framework and context for programmatic evolution. The authors have developed such a research model and are currently working to implement it in a collegiate aviation program at a major research university. While based in a classic input-process-output (IPO) framework with multiple feedback loops, the TAA integration model provides guidance on full integration of TAA capabilities into a collegiate system. Such models have been similarly put forth to assess and improve aviation system performance (e.g., Patankar, Bigda-Peyton, Sabin, Brown, & Kelly, 2005) safety integration among flight crews (e.g., Block, Sabin, & Patankar, 2007) and other subsystems in the aviation industry. Failure to utilize such a model not only limits the research and scientific potential inherent when such data-gathering is an ongoing part of the flight, maintenance, and operations programs, but increases the likelihood that human errors based in the challenging combination of high-technology learning and new skill acquisition (Wickens et al., 2004) will occur.

Technologically Advanced Aircraft Data Integration Model

The TAA data integration model was developed at Purdue University as a multidisciplinary collaboration between the departments of Aviation Technology and Technology Leadership & Innovation. This collaborative development is itself a result of the inclusion of TA aircraft into the existing collegiate aviation setting, and it can be seen in the model's context that further technology integration necessitates an increasingly interdisciplinary interaction such as this. Model development followed the methodological guidelines of the Policy Research Construct (PRC) developed by Bowen and Lu (2004) and utilized many of the strategic collaboration approaches proposed by Bowen, Block, and Patankar (2009).

The TAA model utilizes the PRC flow chart basis for the systematic portrayal of an evidence-based systems-theory approach for integration of TAA into a collegiate aviation program. The matrix portrayed graphically in the conference presentation (and available on requested by contacting the corresponding author, Dr. Erin Bowen) provides an X axis that is based on Objective Performance (as measured by the TAA aircraft) and a Y axis that represents Standards (as codified by the Federal Aviation Regulations (FARs)). This mirrors the policy research construct of Data acquisition feeding Policy analysis as represented by the Practical Test Standards (PTS). With a growing body of data collected from the TAA aircraft, meta-analysis can be conducted to recommend policy actions.

This proposed TAA integration model is presented to provision a baseline for deployment in collegiate aviation programs. While not yet fully defined, the intent is a starting point for a new line of research enabled by TAA deployment in a rapidly increasing capacity. Such a research frame work could allow a consistent evaluation over time similar to that provided in the PRC.

While TAA inclusion is a common spur to the application of the integration model, it does not have to be the only or the initial mechanism for model application in a collegiate setting. What is essential in early data integration is the digitization of key data inputs into the model. Until the majority of data inputs can be placed into a digitally accessible format, the integration potentiality of a single piece of technologically advanced equipment is severely limited. For example, collaborators on the present article are developing methodology for digitizing flight operations, dispatch, maintenance records, and even curriculum data. As with other systems-based models, model input data must be accessible in a common format before connections and feedback loops may evolve.

Interoperability of input data in a shared center follows inherently on the heels of the digitization process. Presently, TAA data integration model developers at Purdue have participated in the creation of an Aviation Data Center (ADC), designed to serve as a common physical and electronic destination that centralizes and maintains interoperable databases of all digitized data. This interoperability is a significant challenge in the TAA data model, as it requires both substantial technical and evaluative capability. However, until the various data components can be related to one another in a meaningful way, the system-level interactions and patterns will remain beyond the view of system managers. It is precisely this goal to observe, track, and modify micro- and macro-level changes to the system that is a key goal of the TAA integration model. This goal shifts the collegiate aviation education setting from a reactive system to a proactive/predictive system; much as maintenance researchers advocate the use of enhanced technological tools to conduct predictive maintenance protocols rather than engaging reactive maintenance behaviors (e.g., Stephens, 2010).

Data digitization and interoperability leads to a growth in collaboration and the establishment of formalized feedback loops in the TAA integration model. These collaborations and loops represent a profound shift in the way collegiate aviation perceives itself, from being a relatively semi-autonomous program to a small piece of a globally complex system. Collaboration may begin on a relatively micro level (as evidenced in the intra-university research collaboration that developed the model itself), but through strategic feedback development should expand to interact with other collegiate programs, industry and government partners, and a global aviation community. The establishment of such collaborative linkages promotes innovation and will require feedback to the model's inputs; for example, growth from collaboration may point to necessary changes/evolutions in educational curriculum or professional training needs.

The collegiate aviation setting has the potential to serve as an opportune test bed for the refinement of the TAA data integration model. Currently, Purdue University researchers are working to implement this model and evaluate system components in order to provide a baseline for scaling to other educational and industry settings. The Purdue Aviation Data Center provides a physical foundation for the digitization and interoperability of the multi-level data inputs into this new technology-based aviation education system. Researchers are identifying and implementing mechanisms for formalizing feedback among system components.

Driving this model in part is the utilization and enhancement of multiple feedback loops at various stages throughout the integration process. While it is easy to encourage such feedback on an informal level, true data integration cannot occur until such reciprocal loops are made explicit and a fully evolved part of the daily function of the collegiate aviation system. In addition, the level of active collaboration within and across the collegiate setting of necessity increases as the feedback loops solidify; this in turn evolves into a model that promotes global interoperability.

The effective organization is likely to be one that demonstrates *alignment* among the different contextual variables in the system; that is, the processes, infrastructure, environment, and rewards systems in place in the organization are designed to support the organization’s values and goals, without being rigidly locked to one another (Semler, 1997; Snow & Hambrick, 1980). For example, an organization may spend several million dollars designing, implementing, and evaluating a training program designed to teach teamwork skills to employees, but if supervisors continue to reward only individual success, strong hands-on managers, and limiting team interaction, the likelihood of that training causing positive organizational change is very slim. Understanding and working to align as many system variables as possible is an essential part of improving organizational effectiveness (Snow & Hambrick, 1980). This logic extends to collegiate aviation training organizations, and the inclusion of TA aircraft makes this logic a mandate for successful integration. If true value is to be achieved from the investment in such technology, then a model that explicitly identifies, tracks, and continuously monitors and evolves the alignment and feedback requirements is required.

Currently integration of TA aircraft in this setting is in the initial phases. Development of a data center research facility at Purdue University provides secured data storage and access for students and research collaborators. An ongoing challenge is creating an integrated database in which system models can be visually created and simulations run on various input scenarios. The purpose of this TAA Integration Model is to provision a model that can be applied universally in the application of a changing collegiate flight education paradigm.

Conclusion

The paradigm shift with the incorporation of TAA in collegiate aviation is the move from treating the collegiate program as a series of connected yet semi-autonomous parts to a fully-integrated system. The research model visibly demonstrates to students, faculty/staff instructors, and relevant stakeholders that shifts in one aspect of the model have the potential to cause ripples throughout the program; ripples that may have safety or performance consequences on the entire system. Collegiate aviation is an ideal setting to develop, test, and refine such an integration model because it benefits from a level of structure and control other aviation industry settings, subject to the whims of commercial pressures, do not have. The authors predict that continued inclusion of TAA and other such tools will spur a drive toward more tightly interconnected program components and require a higher level of academic and professional scholarship, both from students and faculty. TAA data serve as a starting point and impetus for perceiving, treating, and responding to aviation as the complex system that it is, which in turn better prepares TAA-trained students to function in a global aviation setting. The authors present this paper as a proposal to apply rigorous methodology as a foundation for revolutionary change in collegiate aviation. In result of this research for improvement in collegiate aviation education, the first global endeavor for deployment of the model is being presented by coauthors to potential collaborators at Moi University (Kenya) and Qatar Aeronautical College (State of Qatar) prior to the publication of these Proceedings.

References

AOPA Air Safety Foundation. (2005). *Technically advanced aircraft safety and training*. Retrieved from http://www.aopa.org/asf/publications/taa_1_6.pdf

Block*, E. E., Sabin, E. J., & Patankar, M. S. (2007). The structure of safety climate for accident free flight crews. *International Journal of Applied Aviation Studies, 7,* 46-59.

Bowen, B. D., Block*, E. E. & Patankar, M. (2009). A Network Collaborative Design Construct for the Dissemination of Aviation Safety Research. *Proceedings of the International Symposium on Aviation Psychology. Dayton, OH.*

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.

Brooks, J. L. (1999, May). *Smiths industries flight data/cockpit voice recorders.* Paper presented at the International Symposium on Transportation Recorders, Arlington, Virginia.

Craig, P. A., Bertrand, J. E., Dornan, W., Gossett, S., & Thorsby, K. K. (2005). *Ab Initio training in the glass cockpit era: New technology meets new pilots.* International Symposium on Aviation Psychology.

Fiduccia et al. (2003). *General aviation technically advanced aircraft FAA-industry safety study, final report*. Retrieved February 11, 2011, from: http://www.faa.gov/training_testing/training/fits/research/media/TAA%20Final%20Report.pdf

Haas, D., Walker, J., & Kough, L. (2008). *Using flight data to improve operational readiness in naval aviation.*

Patankar, M. S., Bigda-Peyton, T., Sabin, E., Brown, J., & Kelly, T. (2005). A comparative review of safety cultures. *Federal Aviation Administration: ATO-P Human Factors Research and Engineering Division*.

Semler, S. W. (1997). Systematic agreement: A theory of organizational alignment. *Human Resource Development Quarterly, 8*, 23-40.

Snow, C. C. & Hambrick, D. C. (1980). Measuring organizational strategies: Some theoretical and methodological problems. *The Academy of Management Review, 5,* 527-538.

Stephens, M. P. (2010). Productivity and Reliability-Based Maintenance Management. West Lafayette, IN: Purdue University Press.

Wickens, C., Lee, J., Liu, Y, & Becker, S. (2004). *An Introduction to Human Factors Engineering (2nd edition).* Upper Saddle River, NJ: Pearson.

---

[*] E. E. Block is now E. E. Bowen and has published scholarly articles in the aviation and psychology literature under both names.

USING MICROSOFT FLIGHT SIMULATOR IN THE CLASSROOM TO IMPROVE STUDENT PILOT AERONAUTICAL DECISION-MAKING SKILLS

Wendy Beckman
Middle Tennessee State University
Murfreesboro, Tennessee

In the Aerospace Department at Middle Tennessee State University, Microsoft Flight Simulator (MFS) has been utilized in the classroom for several semesters in an effort to develop student aeronautical decision-making (ADM) skills. This software is used to create realistic scenarios which are experienced in class. Two Private Pilot ground school classes were evaluated to determine if experiencing these MFS scenarios had an impact on student development of ADM skills. At the beginning of the semester, each student completed a baseline evaluation of their ADM skills. One class was taught incorporating MFS scenario-based training, while the other class discussed the same scenarios in traditional case study format. At course completion, students completed a second evaluation of their ADM skills. It was found that while both classes made gains in their ADM abilities over the course of the semester, the class taught using MFS demonstrated significantly higher gains in these skills.

Aeronautical decision-making (ADM) has long been identified as a skill that should be taught to aspiring pilots, beginning 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). Since that time, many efforts have been made to identify how best to teach student pilots effective ADM. When the FAA Industry Training Standards (FITS) approach was embraced in 2004, the scenario-based approach to flight training entered the general aviation training paradigm (FAA, 2004; Glista, 2003). Scenario-based flight training attempts to utilize realistic scenarios in training to provide pilots the opportunity to make decisions, and experience the results 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 training exceeds that of students taught using conventional methods (Ayers, 2006; Beckman, et al, 2008; Craig, et al, 2005a, 2005b; Dornan, et al, 2007a, 2007b, 2006). The success of scenario-based flight training led to the addition of scenario-based training concepts in the Private Pilot ground school classes in the Professional Pilot degree program at Middle Tennessee State University (MTSU). Given the physical constraints of the classroom environment, the use of Microsoft Flight Simulator (MFS) as a method of bringing realistic scenarios into class was identified as an innovative solution.

The MFS software series was first made available in 1980, and over the past 30 years there have been ten editions released (Grupping, 2007). In the early editions of the software, the graphics and processing capabilities of computers and the level of sophistication of the software resulted in the program not portraying flight very realistically. This caused certificated pilots to view the software as solely a game; an entertaining and fun diversion, but not useful for training or proficiency purposes. However, over the last several years, 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 MFS package by certificated pilots both for training and proficiency purposes. While use of the program for practicing specific flight maneuvers or procedures is the most common application of the program, MTSU has found that the software is very well-received by students when teaching the concepts of ADM in Private Pilot ground school classes. A series of MFS scenarios have been developed, and are used throughout the Private Pilot ground school course (Beckman, et al., 2009). These scenarios revolve around aerodynamic concepts (load factors, stalls, and spins), aircraft system malfunctions (electrical system, vacuum system), weather, and cross country decision making. Anecdotal evidence from students suggested that this approach was preferable to simply discussing case studies involving the same situations (Beckman, et al.), but there was no data to substantiate whether student ADM skills were improving more than they would utilizing conventional teaching methods. The purpose of this study was to measure the impact of using MFS based scenarios to teach ADM, versus traditional case study methodologies.

Methodology

In 1998 the report, “Evaluating the Decision-Making Skills of General Aviation Pilots” (Driskill, Weissmuller, Quebe, Hand, & Hunter) was published by the FAA. The authors of this report developed an inventory of 51 items, designed to assess general aviation pilots’ ADM skills. The inventory was a set of multiple choice questions, with each question stem consisting of a scenario describing a situation. There were four alternative answer choices the participant could select from to resolve the presented situation. The survey items were developed from both a survey of experienced pilots about “lessons learned” and from NTSB accident and incident reports. The inventory items were divided into five categories: mechanical, weather, biological, sociological, and organizational. A group of expert pilots was utilized to rank the alternative answer choices for each scenario, as a means of determining the most favorable course of action for a typical 500 hour pilot. The results of a survey of approximately 250 general aviation pilots found that, for all but seven items, the surveyed pilots were in agreement with the experts on the best alternative for a given scenario. While the study also found that there was a large variation in the individual rankings of correct alternatives, the high level of correlation between the number one choice of alternatives by both the experts and the actual 500 hour general aviation pilots indicated that the questions used in the inventory are good indicators of the ADM skills of the survey taker.

Forty items were selected from the fifty-one items in the Driskell et al. (1998) inventory. The seven items that were found to not have a strong correlation between the experts’ choice of the most favorable alternative and the average 500 hour pilots’ choice of most favorable alternative were discarded, as were four additional items selected at random. The forty remaining items were then assembled into a twenty question Initial ADM Assessment and a twenty question Final ADM Assessment. Care was taken to keep an equal number of scenario questions involving weather, maintenance, sociological, biological, and organizational factors in both the Initial and Final Assessments.

In the fall of 2009, two Private Pilot ground school classes at MTSU were utilized to conduct this study. MTSU Institutional Review Board approval was received to conduct this human subject research. The same faculty member was assigned to teach both sections of the course. The section randomly selected to be the Case Study Section had 19 students enrolled, while the MFS Section had 36 students enrolled. On the first day of class, the Initial ADM Assessment was administered to the students in each section. The two classes were then taught in the exact same manner for the duration of the semester, with the exception that the Case Study Section discussed, in traditional case study format, seven ADM situations at various points throughout the course. The MFS section also considered seven ADM situations over the course of the semester, but the scenarios were presented using the MFS software package. The cockpit view (including instrument panel and windscreen) was projected onto a large screen in the classroom, and the scenarios were experienced in real-time while a student volunteer “flew” the software package. Examples of the scenarios that were either discussed by case study or presented using MFS included: 1) The effect of load factor and uncoordinated flight on the stall characteristics of an aircraft, as experienced when a pilot is distracted by passengers in turning flight (sightseeing to look for a passenger’s house); 2) An alternator failure at night in a glass cockpit aircraft, 3) A vacuum pump failure on a dark, moonless night in a conventional aircraft, 4) Flight into deteriorating visibility, 5) Cross country navigation with greater than forecast winds aloft, 6) Cross country planning with fuel and weight limitation considerations, and 7) The diversion decision making process due to an ill passenger. The emphasis of the lesson involving each case study or MFS scenario was making appropriate flight management decisions.

On the last day of the course, a Final ADM Assessment was administered to the students in each section. Seventeen students completed the inventory in the Case Study Section (2 students had dropped the course) and 33 students completed the inventory in the MFS Section (3 students had dropped the course). The results of the Initial and Final Assessments were then compared for each section.

Results

For each Assessment that was administered, the students’ responses indicating their best solution to the presented situation were compared to the alternative selected by the experts in the Driskell et al. inventory. The percentage of questions in which the student chose the same best alternative as the experts was taken as the student’s score on the Assessment. The results of both the Initial and Final Assessment for each section of the Private Pilot class can be seen in Table 1 below. The students completed the assessments anonymously, so the results are not listed by individual student (i.e., there is no comparison between initial and final assessment scores by student, only as an aggregate).

Table 1

Listing of Student Scores on Initial and Final ADM Assessments

Case Study Section		**MFS Section**	
Score on Initial Assessment	Score on Final Assessment	Score on Initial Assessment	Score on Final Assessment
20	60	25	40
40	50	50	50
55	30	60	40
30	30	15	70
60	50	60	70
45	60	40	60
25	50	15	50
45	40	30	45
55	55	25	45
50	65	15	55
15	50	55	45
30	30	45	65
30	70	40	85
40	80	55	70
30	50	55	80
15	45	25	80
50	40	20	70
45		30	75
25		50	65
		40	80
		35	70
		15	45
		45	65
		45	45
		10	80
		20	60
		30	60
		45	70
		25	55
		25	75
		25	45
		55	40
		30	60
		50	
		55	
		35	
M= 37.11 **SD = 13.87**	**M = 50.29** **SD = 14.08**	**M = 35.97** **SD = 14.87**	**M = 60.91** **SD = 13.83**

An alpha level of .05 was used for all statistical tests in this study. The first statistical test performed was to evaluate whether there was a significant difference between the two class sections at the beginning of the semester. This was important, as there was the potential that the two classes were different in some way prior to experiencing the course. As can be seen in Table 1, the Initial ADM Assessment mean score for the Case Study Section was 37.11 (SD=13.97), while the Initial ADM Assessment mean score for the MFS section was 35.97 (SD=14.87). An independent samples t-test was conducted to compare the Initial ADM Assessment mean scores of the two sections. There was no significant difference found in the scores of the two sections, t(39) = .281, p = .780. This meant the two sections were statistically performing at the same level regarding ADM skills prior to taking the course.

Next, an independent samples t-test was performed to determine if there was a significant difference between the mean initial score and the mean final score for each section. For the Case Study Section, it was found that there was a significant difference between the initial scores (M=37.11, SD=13.97) and final scores (M=50.29, SD=14.08), t(33) = 2.83, p = .008. For the MFS section, it was also found that there was a significant difference between the initial scores (M=35.97, SD=14.87) and the final scores (M=60.91, SD=13.83), t(67) = 7.22, p < .001. Thus, both classes did show a statistically significant improvement in their ability to interpret and choose correct actions for situations requiring ADM from the beginning to the end of the semester.

Finally, an independent samples t-test was performed to determine whether the final scores of the two classes were significantly different. The final scores of the Case Study Section (M=50.29, SD=14.08) were compared to the final scores of the MFS Section (M=60.91, SD=13.83), and it was found that there was a significant difference between the two sets of scores, t(32) = 2.54, p = .016. Therefore, the MFS Section demonstrated a statistically significant higher level of performance on the Final ADM Assessment than did the Case Study Section.

The data was then grouped by question category. On both the Initial and Final ADM Assessments, there were two biological questions, five maintenance questions, four organizational questions, four sociological questions, and five weather questions. The mean percentage correct score for each category of question was tabulated for both sections, along with the amount of improvement experienced. These results can be seen in Table 2.

Table 2

Mean Initial and Final ADM Assessment Scores For Each Category of Question

	Case Study Initial	Case Study Final	Improvement	MFS Initial	MFS Final	Improvement
Weather	30	38	+8	28	52	+23
Biological	26	42	+16	37	50	+13
Organizational	53	67	+14	47	81	+33
Sociological	49	61	+12	47	70	+22
Maintenance	26	45	+19	25	51	+25

Based on independent sample t-tests, there was no significant difference found between the initial Case Study scores by category and the MFS initial scores by category in any of the five categories. It can be seen that both groups experienced improvement in each of the question categories. The MFS Section showed greater improvement than the Case Study Section in all of the subject areas except the biological questions. The largest gain for the Case Study Section was seen in the maintenance category, while the largest gain for the MFS Section was seen in the organizational category. T-tests were conducted to determine if the Case Study Section Final Assessment scores per category were significantly different from the MFS Section Final Assessment scores by category. These t-tests revealed that only two categories showed a statistically significant difference. When the Case Study Section Final Assessment weather scores (M=37.8, SD=8.90) were compared to the MFS Section Final Assessment weather scores (M=52.6, SD=7.77), a significant difference was found, t(32) = 2.80, p = .023. Likewise, when the Case

Study Section Final Assessment organizational factors scores (M=66.75, SD=6.40) were compared to the MFS Section Final Assessment organizational factors score (M=81.5, SD=7.05), a significant difference was found, t(32) = 3.10, p=.021. Thus, only the weather and organizational factors categories were affected significantly by the use of MFS for teaching ADM.

Conclusions

The results of the study indicate that the MFS software package is of benefit in teaching student pilots ADM skills while in private pilot ground training, as demonstrated by the significant difference in Final Assessment scores between the Case Study Section and the MFS Section. The two areas where the use of MFS appeared to be the most helpful were in weather factors and organizational factors, as a significant difference was experienced in the mean scores on these two categories of questions. Although there was not a significant difference between the biological category scores of the two groups, it was interesting to note that the Case Study Section improved slightly more than the MFS group in that category; this was the only category where this was the case. In looking at the data, this change appears to be the result of the Case Study Section obtaining a rather low score initially in the biological area. However, it should also be noted that there were only two biological questions on both the Initial and Final Assessments, making it by far the smallest category. As such, it was more sensitive than the other categories that were analyzed.

Both the Case Study method and the MFS method of teaching ADM resulted in Final Assessment scores improving significantly from their original level, as evidenced by the significant difference in initial scores and final scores for both groups of students. However, it was of concern that the ADM assessment scores remained quite low, even after a semester of discussing ADM concepts in either format. Clearly, more work needs to be done to effectively teach ADM to student pilots. Since using MFS does appear to be helpful, perhaps greater use of MFS-based scenarios throughout the semester would further increase student ADM skills. In addition, targeting the scenarios to a particular ADM category, and explaining specifically what that category involved to the students, may assist them in developing ADM skills. The difficulty with this approach is that it requires additional time in class, and private pilot ground schools are already very full of required course material.

Student reaction to the use of the MFS scenarios in the classroom was overwhelmingly positive. Even students who otherwise appeared bored or disinterested became engaged in the class activities when the software package was utilized. As a group, the current generation of students is very visually oriented, and being able to provide an almost “real life” image to view was enthusiastically received. This study was obviously quite small, and should be replicated in future Private Pilot ground schools, and at other institutions, before any final conclusions are drawn about the utility of the MFS package for teaching ADM. But, based on the evidence thus far, it appears that further scenario development and further testing of the efficacy of MFS scenarios for teaching ADM is both warranted.

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., Craig, P.A., Gossett, S., Dornan, W. (2009). Using Microsoft Flight Simulator X to Develop Aeronautical Decision Making Skills in the Classroom. *Proceedings of the 2009 International Symposium on Aviation Psychology*, 613-618.

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, Dayton, Ohio.

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. *Proceedings of the 2007 International Symposium on Aviation Psychology*, 177-181.

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.

Driskill, W.E., Weissmuller, J.J., Quebe, J.C., Hand, D.K., & Hunter, D.R. (1998). *Evaluating the Decision-Making Skills of General Aviation Pilots.* (DOT/FAA/AM-98/7). Washington, DC: United States Department of Transportation.

Federal Aviation Administration. (2004). *FAA-Industry Training Standards (FITS) Program 2004*. Retrieved from: 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 Retrieved from http://fshistory.simflight.com/fsh/timeline.htm

COLLEGIATE AVIATION SAFETY REPORTING SYSTEMS

Beth M. Beaudin-Seiler, MPA
Western Michigan University, College of Aviation
Battle Creek, Michigan

The Federal Aviation Administration has paid close attention to the safety reporting systems of the airline industry over the last thirty years. The Aviation Safety Action Programs, housed at NASA, allow pilots and crews to report safety issues into a central database that tracks these reports and provides valuable knowledge to the industry on safety related issues. However extrapolating information that is pertinent to general aviation from these reports, specifically collegiate aviation, is difficult. One of the barriers to data collection is having a commonly understood language among reporters in order to ensure accurate information is reported. The goals of this project include steps to better understand the hurdles that have impeded safety initiatives at the collegiate level; identification of a common language with operational definitions that would be used in tracking safety information; and to conduct initial testing of the common language in a currently used reporting system.

The current Aviation Safety Reporting System ("ASRS") created by NASA has been successful because of the Federal Aviation Administration's ("FAA") conviction for "identifying deficiencies and discrepancies in the national aviation system to provide a knowledgeable basis for improving the current aviation system; and providing data for planning and improvements to future systems." (Corrie, 1997)

The ASRS program provides for limited immunity protection to the reporter. The immunity protection only applies if a) a violation was inadvertent; b) the incident did not involve a criminal offense, accident or action disclosing a lack of qualification or competency; c) the reporter was not previously found to have committed a regulatory violation within 5 years; and d) the reporter proves that the ASRS report was filed within 10 days of the incident (Corrie, 1997).

The ASRS program provides confidentiality and protects the identity of the reporter and all other parties involved in an occurrence. Once the report is thoroughly screened, the reporter identification strip is removed from the report and returned to the reporter (Corrie, 1997).

The ASRS program requires that incident reports be compiled daily and screened by analysts. Analysts look for potential time-critical issues that require immediate attention of the FAA and industry. Any report meeting certain alerting criteria are flagged and processed through a two pronged alert message system. The two types of alert messages are the Alert Bulletin and the For Your Information. Alert Bulletins are issued when a hazardous condition has been well documented and involve serious safety concerns. The For Your Information bulletins are issued when a problem is not well documented and involve less serious conditions. Finally, the ASRS program utilizes computer databases to assign unique numbers to reports and search through the information by using coded data and narratives (Corrie, 1997).

The ASRS program has been extremely successful. The FAA along with industry has created an aviation system that is the safest in the world, and getting in front of information flow regarding safety and operation issues will keep it that way (Fiorino, 2003). The ASRS purpose is to collect; analyze and respond to the voluntarily submitted safety reports in order to lessen the likelihood of aviation accidents. It collects data from pilots, controllers and others and includes the general aviation arena (ASRS, 2010).

The ASRS database allows inquires of the general aviation pool however it does not provide a breakout of issues facing collegiate aviation programs. While queries can be narrowed to "training" or filtered through "pilot schools", it is not clear if baccalaureate or associate degree seeking programs submit these reports, or another type of pilot school. Secondly, standardization on the use of the ASRS database is lacking and an understanding of the definitions and terms of use has not been established. Therefore general aviation pilots across the country could potentially be providing descriptions in their reports that spread in to multiple areas across the system without consistency in terminology, which, in turn, could cause analysts to classify the incidents into categories ultimately that are not correct.

During a query of the accident/incident database, narrative text including landing, situational awareness, weather, and pilot error were searched for the calendar year of 2010. Out of 93,450 reports searched, 0 reports came back from the query. The near midair collisions database was queried using the keywords of air congestion, air traffic control and situational awareness for the calendar year of 2010. Out of 6,633 reports searched, 0 reports came back with those keywords. Finally, in the Aviation Safety Reporting System database the keywords searched were landings, weather, air traffic control, and air congestion. Out of 632,677 records 0 were queried by those keywords (ASRS, 2010). It is difficult to use the system and be confident that what is being searched for is actually what is being retrieved. In the description of the ASRS database by the FAA

> "The data received in an ASRS report represents what reporters communicate they saw or experienced. Except through the alert message part of the program, ASRS reports are not investigated, and therefore the accuracy of the report information is not verified. The reporter's experience, visibility conditions, duration of the event, trauma experienced by the reporter or other factors can influence the accuracy of the data. Many factors can influence the decision to file a report, such as, lack of awareness of ASRS, motivation to report which can differ considerably between different segments of the aviation community, and the perceived severity of an incident may influence the decision to report. The cumulative effect of these and other factors is that ASRS reports submitted to NASA represent a portion of the total number of similar events that may and could be reported. For these reasons ASRS information should not generally be used to determine distributions or trends but may be very effective for identifying hazards, accident precursors and safety issues for further analysis." (ASRS, 2010)

The gap in the understanding and the knowledge on safety information that is the foundation of this project is 1) that aviation training in collegiate programs is significantly different enough to warrant a system that can break down the casual factors of risk pursuant to this particular environment; 2) that the terminology be defined and trained to the participating pilots so as to reduce the number of incidents that are misclassified. By addressing these issues, the reports submitted to the safety system can be reliably analyzed in order to provide foundations for more robust methodologies to create change to curriculums, program procedures, attitudes and behaviors and the overall culture of safety at participating institutions.

Current Reporting System

Western Michigan University's College of Aviation has created a Collegiate Aviation Safety Reporting System ("CASRS") to address the gap in knowledge and understanding that is created by co-mingling general aviation safety reporting documents together.

CASRS is a web-based, non-punitive safety event reporting system that employs a process to identify event types and causal factors in a manner that facilitates data analysis. At key points in the process, e-mails are generated to key individuals in order to provide timely notification of the event. When e-mails are generated, what information (i.e., data fields) is included and to whom the e-mails are to be sent are all selectable. Causal factors, as many as two per event, are fixed and based on the work of Dr's Krokos and Baker of the American Institute for Research (2005). They include:

- Air traffic congestion
- Conflicting ATC clearance
- Frequency congestion
- Hear back/read back
- Incorrect ATC clearance
- Late ATC clearance
- Unclear ATC clearance
- Uncontrolled airport, Non-standard procedures
- Aircraft damage
- Aircraft equipment malfunction
- Equipment limitation
- Ground equipment inoperative or malfunctioned
- Inadvertent or intentional disregard for policy or procedure
- Misapplication of flight controls
- Attention to detail
- CRM – Communication
- CRM – Leadership and command
- Experience level
- Fatigue
- High workload/task saturation
- Interruption/distraction
- Personal attitudes towards safety

- Self-induced time pressure
- Situational awareness
- Inadequate training
- Conflicting policies or procedures
- Confusing policies or procedures
- Inaccurate policies or procedures
- Lack of policy or procedure
- Animal/bird strike
- Excessive cold
- Icing
- Low visibility/low ceiling
- Ground surface contamination

Extensive use is made of drop-down menus (e.g., aircraft type, aircraft registration, phase of operation, etc.) where possible and is also amendable. CASRS is housed in a server that is accessible from on or off site for the submission of reports and administration of the system via the internet. The server automatically removes identifying information and stores both identified and de-identified data for retrieval. Only de-identified data is available for sharing. Identified data remains the province of the unit. The server also collates the data on a weekly basis, in two matrices, one by event type and the other by causal factors. The collated reports are color coded by the number of reports of a type in a week and individual reports are selectable directly from the matrix (Jones, 2009).

Access to CASRS is limited to students, faculty and staff members of the College of Aviation plus invited guests of the college. With the exception of members of the College of Aviation Safety Committee, access to CASRS is limited to the submission of reports (Jones, 2009).

When an individual submits a report, an e-mail is generated to key individuals including the Director of Safety. The Director of Safety randomly assigns the report to two members of the Safety Committee for their independent assignment of as many as two causal factors to the event. This moves the report from an "open" status to "pending review." If the assigned causal factors are identical from both individuals, the report moves from "pending review" to "reviewed." If the members don't agree, CASRS so advises the members and they get another opportunity to submit. If they still don't agree, the Director of Safety will assign causal factor(s). At the next biweekly Safety Committee meeting, all reports submitted during the previous two weeks are reviewed by all the members for general consensus. After the meeting, the Director of Safety makes any modifications mandated by the committee and closes the reports. This moves the report status from "reviewed" to "closed." (Jones, 2009)

The Director of Safety currently has the authority to edit reports, questions, users, causal factors, display and e-mail settings. As the system evolves and expands, protocols will be needed to structure some of those functions. The Director of Safety can also designate which reports will not be included in the data base in the event of duplicate reports. (Jones, 2009)

This system has collected over 600 safety reports at WMU. Initial research on the system examined the hurdles that have impeded safety initiatives at the collegiate level; and identification of a common language with operational definitions that would be used in tracking safety information.

Methods

In 2010 an internal research development award was given to the author by WMU. Objectives included preliminary steps towards the establishment of a common safety language and an initial look at how the current safety reporting system at WMU would need to change.

Collaborators from the Historically Black College Consortium, which include Delaware State University, Hampton University and Florida Memorial University, came together with WMU to provide insights to the project. The use of these institutions provides strong academic and flight standards, coupled with a diverse way of completing flight training and unique safety concerns. WMU owns and maintains over 40 aircraft in-house, instructing over 400 professional pilot students, while Delaware State University also owns and maintains their aircraft it is on a smaller scale and therefore have different safety concerns. Hampton University and Florida Memorial University both contract their flight training to outside flight schools. This makes for an interesting position in terms of safety reporting and brings a unique perspective to the discussion as a whole. All the participating programs have interesting weather concerns that will also provide for interesting discussions.

All the collaborating institutions provided subject matter experts to answer an electronic survey to identify the most appropriate operational definitions to the thirty-five causal factors currently listed in the safety reporting system.

The survey was designed to provide the operational definition of all causal factors currently listed in the safety reporting system as well as two alternatives. There was also an opportunity to write in comments or notes on each question. Subject matter experts from each of the collaborating institutions were then asked which operational definition was most appropriate. If they thought none were, they were asked to provide one of their own, or at least make comments as to why they did not consider any of the options appropriate.

Findings

Twenty-five out of the thirty-five causal factors had a majority agreement (>52%) to an appropriate operational definition, leaving ten for further discussion and analysis. Those ten include:

Air traffic congestion
Conflicting ATC clearance
Incorrect ATC clearance
Uncontrolled airport non-standard procedures
Inadvertent disregard for policy or procedure
Intentional disregard for policy or procedure

CRM-Leadership and Command
High workload/Task saturation
Conflicting policies or procedures
Lack of policy or procedure

A positive working foundation for a common safety language has been established with this initial research.

Preliminary discussions with subject matter experts on the user interface of the current safety reporting system revealed the need for even more drop down menus, for easier query functions and for more delineated categories. This feedback allowed WMU to begin understanding the necessary workload needed to introduce these enhancements to the current reporting system. A proposal being developed for the submission to the FAA looks to advance both of these initiatives to the point of usability.

References

Corrie, S. (1997). The US Aviation Safety Reporting System, *American Institute of Aeronautics and Astronautics,* Report 975562.

Federal Aviation Administration. (2010). 2009-2013 FAA Flight Plan. Retrieved from: sdhttp://www.faa.gov/about/plans_reports/media/flight_plan_2009-2013.pdf.

Fiorino, F. (2003). FAA's 'Stock' in Safety. *Air Safety, 2*(159), 48.

Jones, S. (2009). Collegiate Aviation Reporting System (CASRS), *Internal WMU Document unpublished.*

Krokos, K. & Baker, D. (2005). Development of a Taxonomy of Causal Contributors for Use with ASAP Reporting Systems. *FAA Grant#99-G-048*.

National Aeronautics and Space Administration. (2010). Aviation Safety Reporting System. Retrieved from http://asrs.arc.nasa.gov

Acknowledgements

The author would like to thank the collaborating organizations and their respective points of contact for the initial work on this project. Dr. Donna Wilt, Hampton University; Capt. Steve Speed, Delaware State University and Dr./Capt. AJ Tolbert, Florida Memorial University.

And thanks to Western Michigan University's internal Research Development Award for the initial research work.

Perceptions of a Flight Operations Quality Assurance Programs in a Collegiate Environment

Brian G. Dillman
Department of Aviation Technology, Purdue University
West Lafayette, Indiana, USA

Lauren M. Vala
Department of Aviation Technology, Purdue University
West Lafayette, Indiana, USA

Erin E. Bowen
Department of Technology Leadership & Innovation, Purdue University
West Lafayette, Indiana, USA

Michael S. Nolan
Department of Aviation Technology, Purdue University
West Lafayette, Indiana, USA

The use of Flight Operations Quality Assurance (FOQA) programs has become more feasible for flight training in small single-engine aircraft. What has still to be determined is the total impact on pilot training quality through the utilization of digital data recording systems. With the introduction of FOQA, the standard instructor now has the capability to determine flight skills and abilities through digital data collection, and the effects on pilots' perceptions needs to be identified. It is suspected that pilot's awareness of FOQA data collection occurring during flight will have a similar effect on their performance. Survey analysis of student perceptions of the FOQA implementation process and the knowledge of the purpose and functionality of FOQA programs within the aviation industry. Surveys will be administered to students and instructors in the flight training program of a Part 61 training school.

The concept of a Flight Operations Quality Assurance (FOQA) program has roots in previous quantitative and qualitative aviation recording programs such as flight data recorders (FDRs), the Aviation Safety Action Program (ASAP), and the NASA Aviation Safety Reporting System. As indicated by program success at the airline level, a FOQA program must be accompanied by safety management systems (SMS) and a sound safety culture (Wiley 2007; FAA, 2006b). Management must fully support the FOQA program initiatives and strong communication channels through all levels of the flight entity must be in place. Confidentiality and data protection issues remain the largest barrier to FOQA program implementation (FAA, 2004; FSF, 1998) and are also discussed in the FOQA context. Airlines have realized much success from FOQA programs, though no efforts have yet been made to tailor these programs to the unique needs of the university flight training setting. With relative inexperience in the collegiate market lies an increased potential for misperceptions and unique challenges which must be assessed. This paper outlines an initial process for evaluating student-level perceptions of a potential FOQA program under consideration for implementation at a large U.S. university.

Current Aviation Reporting Systems

Information systems intended to promote and encourage safe operations are not a new concept in the aviation industry. Though a few have captured quantitative data, most systems have relied on qualitative pilot reports for such data collection (Wiley, 2007; FAA, 1997). Pilot reports gather subjective information, while information from flight data recorders and quick access recorders provide objective information which provides a different view of events. As aviation has progressed and advanced as a science, reporting methods have as well. Specifically, NASA's Aviation Safety Reporting System (ASRS) can be identified as influencing FOQA program initiatives in the de-identified, non-punitive reporting styles that are characteristic of each (FSF, 1998).

There are other qualitative reporting systems that are currently operated in the aviation industry. The Aviation Safety Action Program (ASAP) is a qualitative, airline specific pilot-initiated reporting system. Self-reporting systems of this sort are non-punitive and the best way to keep abreast of potential hazards and risks in the operation (Corrie, 1997). Wiley (2007) states that these reports are beneficial in acknowledging the existence of discrepancies, but usually fall short of addressing the real problems at hand, since all information gathered is subjective and biased from pilot recounts of actual flight scenarios. Though information collected from ASRS reports has occasionally assisted operators in finding problems and safety-compromising conditions in the past, there is still a large amount of relevant qualitative safety information that operators miss from events due to this subjective reporting style. A potential conflict arises with the utilization of equipment to feed a FOQA program in that it is no longer voluntary. Data is being recorded at all times the aircraft is being utilized. This dichotomy between voluntary reporting and mandatory data collection has the potential to create negative perceptions of a flight operation.

FOQA Program Development

FOQA is a significantly different program than all previous safety programs discussed. Unlike the ASRS or various FAA Aviation Safety Action Programs (ASAPs), FOQA uses quantitative, objective data from flights to enhance trend monitoring and address operational risk issues (FAA, 2004; FSF, 1998). FOQA programs can lead to the development of advanced training programs such as Advanced Qualification Programs (AQPs). Specifically, FOQA data can accurately verify pilot learning outcomes required by AQPs (FAA, 2006a). Historically the only individuals that knew the true events concerning a given flight were those that were in the actual cockpit. The pilot and sometimes first officer or flight instructor would be the only individuals that could recount the events of the entire flight. With FOQA data and the ability to verify the pilot's aircraft manipulation ability there is a new input that many industry and educational professionals don't know how to effectively utilize.

The first workshop attempting to identify the benefits, utilization, and to encourage adoption worldwide of FOQA programs was by the Flight Safety Foundation (FSF) in Taiwan in 1989 (FSF, 1998). According to the Foundation (1998), their blueprint for FOQA has been the backbone for FOQA progress in the United States, though there is much more work to be done. The FAA took initiative to development a formal FOQA program in 1990 by hosting a FSF workshop in Washington, DC, and in 2001 developed a rulemaking committee to further work in this area (FAA, 2003; FSF, 1998).

Before FOQA received full support from the FAA, a demonstration project was carried out to assess the costs, benefits, and safety enhancement associated with the program (FSF, 1998). During this project, the FAA provided hardware and software to four airlines which agreed to implement FOQA programs and share data with the FAA. As a result of the project, the FAA determined that FOQA programs would be made voluntary, as data collection and use for advanced FOQA programs were still in primitive form. The project demonstrated that the FOQA concept was a success for airlines by allowing enhanced trend monitoring and the identification of operational risks (FSF, 1998). The FAA did not attempt to create a

FOQA program for non-commercial use during their three year demonstration project (FSF, 1998), though a FOQA program for the general aviation sector, including collegiate flight operations, would improve safety and operational performance and assist in the training of new pilots (Mitchell, Sholy, & Stolzer, 2007).

An airline FOQA program development guideline is available in Advisory Circular 120-82, which discusses the benefits, set up, and maintenance of such a program (FAA, 2004). This document also provides a template for the Implementation and Operations (I & O) plan set-up as well as key definitions that must be addressed during program establishment (FAA, 2004). In order to be fully operational in a university flight school setting, a FOQA program must fit into the safety program goals and be supported by the university flight department. A safety culture must exist if additional programs, such as FOQA, are to be successful (Wiley, 2007).

Before a FOQA program or further safety management system can be launched at a university flight school, it must be determined if the cultural environment is in place to support it (Wiley, 2007). The FAA (2006b) states that, "the principles that make up the [Safety Management System] functions will not achieve their goals unless the people that make up that organization function together in a manner that promotes safe operation" (p. 4). This organizational aspect is termed a safety culture (Block, Sabin, & Patankar, 2007; FAA, 2006b; Wiley, 2007). A safety culture is composed of psychological, behavioral, and organizational elements (FAA, 2006b). Organizational elements are ones that management has the most control over, and it has been discovered that if this element does not exist and thrive, a safety culture will likely fail (Wood, Dannatt, & Marshall, 2006).

An important aid to the development and sustainability of a safety culture is to hold regular safety meetings with personnel from a wide range of departments and levels (Wood et al., 2006). Wood et al. explains the goal of such meetings is to share information, highlight and discuss any known threats, and make sure that all personnel have the same perspective on the threats. This assists in developing the feeling of safety within operations being a shared responsibility within the company (Wood et al., 2006).

Airline officials, pilot union representatives and the FAA recognized that data protection issues were the biggest roadblock for FOQA program implementation (FSF, 1998). Initially, pilot unions were reluctant to sign FOQA agreements with airlines as they feared a lack of protection for collected FOQA data. FSF (1998) highlights three concerns airline pilot unions had with program implementation:

> "[first,] that the information may be used in enforcement/discipline actions; [second,] that such data in the possession of the federal government may be obtained by the public and the media through the provisions of FOIA; and [third] that the information may be obtained in civil litigation through the discovery process" (FSF, 1998, p. 7).

To address these concerns, 14 CFR Part 13 Section 13.401 was created. This document mandates FOQA data be stripped of any information that may identify the submitting airline before the data is passed to the FAA (FAA, 2004). The FAA ensures that "aggregate data that is provided to the FAA will be kept confidential and the identity of reporting pilots or airlines will remain anonymous as allowed by law" (FAA, 2004, p. 1). It is believed that relatively little exposure or experience with FOQA programs in any context will directly impact the perceptions of the individual within the flight program utilizing FOQA.

The possibilities FOQA programs offer are too beneficial to be ignored by university flight school operations. However, the process of adapting FOQA programs to university flight needs proves daunting and cumbersome for traditional operators. Guidance from previous systems may assist with collegiate FOQA development, but attention must be paid to the legalities of data collection which relate to collection of student data. With support from management and a solid safety culture in place, a data collection system can be developed and standardized. Hopefully, university flight schools would provide similar benefits that airlines have realized from FOQA programs while at the same time preventing a culture of fear or retribution from being developed either in reality or in perception. It is this issue that the survey that was developed and delivered attempted to identify.

Student Perceptions Survey

The authors developed a web-based survey to assess the perceptions of flight instructors and students in the flight program at a large Midwestern university. The survey was designed to determine the current and proposed methods of upset recovery training in each flight program and to use the results to foster dialogue between institutions to determine the most effective method of upset recovery training. The online survey was conducted during the first quarter of 2011. The authors were able to obtain survey information from 67 of the 208 potential respondents for a 32.2% response rate.

Respondents were primarily individuals that have been pilots from between 1-4 years (59%) and 27% having been pilots for more than 4 years. The amount of flight time of the participants was almost split evenly between less than 200 flight hours (49%) and more than 200 hours (51%) with the highest level of respondents being above 300 hours (36%).

The survey was broken down into three distinct areas. The first section attempted to discover the perceptions of the respondents in regards to the current policies and procedures of the University's flight program. The second section attempted to discover the knowledge level and depth of understanding that the respondents had of FOQA programs both in collegiate aviation and in the aviation industry. The final section attempted to determine the perceptions of the respondents in regards to the existence of an ongoing FOQA implementation within the academic program.

	Statement Wording	Mean	Standard Deviation
Perceptions of Current Policies and Procedures 5-Point Likert Scale Strongly Disagree - 1 Strongly Agree - 5	The University's Flight Operation policies and procedures make sense to me	3.77 0.	97
	The University's Flight Operation policies and procedures are too strict	3.28 1.	13
	The University's Flight Operation policies and procedures are based upon legitimate safety concerns	4.22 0.	93
	I feel secure about my flight performance at the University	4.58 0.	68
	The Flight Department has a code of professional conduct that Students and Instructors are expected to follow	4.22 1.	07
Knowledge of FOQA Programs 4-Point Likert Type Scale Very Knowledgeable -1 Not Knowledgeable at all - 4	I am knowledgeable about FOQA programs at the airlines/aviation industry	2.55 0.	94
	I am knowledgeable about FOQA programs in collegiate aviation	2.94 0.	83
	I am knowledgeable about the University's Implementation of a FOQA program	2.66 1.	02

	Statement Wording	Mean	Standard Deviation
Perceptions of FOQA Program Implementation 5-Point Likert Scale Strongly Disagree - 1 Strongly Agree - 5	The feedback that I receive about my performance from my flight instructor is sufficient and I do not need digital information from the FOQA program to make improvements	3.57 1.	02
	Personal input from flight instructors and students should be utilized more for evaluation than FOQA digital data	4.06 0.	95
	Performance information that the FOQA digital data provides is more objective than personal input from flight instructors/students	3.3 1.	07
	I see the implementation of FOQA data as a threat to my freedom as a pilot	3.48 1.	19
	The Administration at the University believes that Students and Instructors are the most important asset of our flight program	3.21 1.	31
	With the implementation of a FOQA program at the University, pilots will be constantly watched to assure that rules and procedures are followed	3.76 1.	13
	FOQA programs discourage creativity, innovation, and continuous improvement	3.19 1.	16
	With the implementation of FOQA digital data I will become a number/statistic rather than a person/pilot	3.42 1.	23
Much more Positive than Negative - 1 Much more Negative than Positive - 5	Overall, when I think about the Flight Operations Quality Assurance Program at the University, I feel…	3.03 1.	03

Conclusion

It is possible that the knowledge of FOQA programs in general played a part in the low response rate to the survey. Despite this level of response some initial conclusions can be drawn. As the program becomes more mature additional data will be taken to assess the change in perceptions over time and as more comfort is developed with the system. Trust in a new and unknown system must occur over a period of time and it is not surprising that these results show the same conclusion.

In general the students and instructors believe that the policies and procedures at the university are based upon reasonable safety concerns and that a code of professional conduct is required. The students also believe that the rules are too strict to some degree, but are willing to abide by the policies and procedures by virtue of them relating to safety concerns and that they generally make sense to each of the pilots. Overall there was moderate to strong support for the policies and procedures established and a willingness to follow the prescribed requirements established by the flight administration.

In general the students and instructors responded that they were not very knowledgeable with FOQA programs either within collegiate aviation, within the aviation industry, or in regards to the implementation process at the university. There was a kickoff meeting at the end of the 2010 calendar year in which the basic facets of a FOQA program were covered, but since the students and instructors have yet to actually engage with the program and get direct feedback there is a deficit in the knowledge and understanding level at this point.

In general the students and instructors didn't have strong feelings toward the negative or positive side in regards to their perceptions of the FOQA program implementation. There was a sense of being unsure as to what it would do and how it would be used and a concern that too much emphasis would be placed on the aspect of data utilization. The students and instructors are well versed in the way feedback is accomplished currently in a very personal and individual manner. All respondents felt that that current system of feedback is sufficient and that the implementation of a data collection system wasn't necessary. Much like the airlines, pilot unions, and FAA had to learn what benefits and drawbacks there was to the implementation of a FOQA program within their own operations, it will be necessary for students and instructors eyes to be opened to the possibilities within the FOQA program.

References

Aviation Safety Reporting System. (2010). *Program briefing*. Retrieved from http://asrs.arc.nasa.gov/overview/summary.html.

Block, E.E., Sabin, E.J., & Patankar, M.S. (2007). The structure of safety climate for accident free flight crews. *International Journal of Applied Aviation Studies, 7*(1), 46-59.

Corrie, S. J. (1997, October 13-16). The US aviation safety reporting system. Proceedings from AIAA/SAE International Symposium: *1997 World Aviation Congress*. Anaheim, CA.

Federal Aviation Administration. (2006b). *Introduction to safety management systems for air operators.* (DOT Advisory Circular No. 120-92). Washington, DC: U.S. Government Printing Office.

Federal Aviation Administration. (2004). *Flight operational quality assurance.* (DOT Advisory Circular No. 120-82). Washington, DC: U.S. Government Printing Office.

Federal Aviation Administration. (2003). *Flight operational quality assurance aviation rulemaking committee*. (DOT Order No. 1110.131A). Washington, DC: U.S. Government Printing Office.

Federal Aviation Administration. (1997). Aviation Safety Reporting Program. (DOT Advisory Circular No. 00-46D). Washington, DC: U.S. Government Printing Office.

Flight Safety Foundation. (1998). Aviation safety: U.S. efforts to implement flight operational quality assurance programs. *Flight Safety Digest, 17*(7-9), 1-54.

Mitchell, K., Sholy, B., & Stolzer, A. (2006, December 21). General aviation aircraft flight operations quality assurance: Overcoming the obstacles. *IEEE Aerospace and Electronic Systems Magazine, 22*(6), 9-15

Wiley, J. (2007, June). C-FOQA: Has its time arrived? *Business & Commercial Aviation, 100*(6), 76-82.

Wood, M., Dannatt, R., & Marshall, V. (2006, June). *Assessing institutional resilience: A useful guide for airline safety managers?* (Australian Transport Safety Bureau No. B2004/0240). Canberra City: Australian Capital Territory.

HIGH-FIDELITY SIMULATION AND TRAINING TO IMPROVE COORDINATION BETWEEN AEROSPACE SPECIALIZATIONS

Paul A. Craig
Richard G. Moffett, Glenn E. Littlepage, Michael B. Hein
Andrea M. Georgiou, Gerald L. Hill, Jennifer A. Henslee
Paul R. Carlson, Nora A. Cole, Joseph H. Cooper
Durant Bridges, Justice Amankwah, Donald J. Tipton, Alan Waid
Middle Tennessee State University
Murfreesboro, TN, USA

This Symposium contains the first reports of research being conducted at Middle Tennessee State University using a scenario-based teaching methodology with students across multiple disciplines of aviation training. The MTSU Center for Research on Aviation Training is a NASA funded project that had built a replica of an airline's flight operations center. The basic research of the project involves the interactions between aviation professionals. These interactions can produce smooth and safe operations for passengers, profits for employers and economic benefits for the national and the world – or these interactions can yield chaos, frustration and loss of revenue. This research brings together pilots, flight dispatchers, controllers, maintenance technicians, weather forecasters and managers into one simulation. The research aims to learn the "best practices" for using this real-world scenario teaching method and to send graduates into the workforce better prepared for the interactions that they will face.

Today, students are trained in various disciplines of aviation in isolated clusters. The pilots train with pilots, the maintenance technicians with other technicians, dispatchers with dispatchers, controllers with controllers, and so forth. Prospective employees are coming to the job market through independent "silos" of training, but this training does not always reflect the way operations run in the real world. Once students enter the job market, they realize that success and efficiency depend on cross-disciplinary communications and understanding. By dismantling the silos, we want to prepare the next generation of aviation professionals in a real-world environment and enable employees to perform better on the job from the first day of hire.

Imagine a situation where students from all aviation and certain business disciplines come together in a laboratory. Instead of attending a typical classroom lecture, they are immersed in a practical, hands-on experience as they "work a shift." They enter a room with a bank of screens on each wall. The screens project real-time weather, aircraft tracking maps, aircraft status boards, crew schedules, aircraft parts inventories and any other information required to run an airline for that shift. This is an airline's Flight Operations Center – sometimes called the "war room." Located in the Business & Aerospace Building, MTSU Aerospace has a replica of such an Operations Center that allows students the opportunity to learn in a real-world scenario based environment. The students' shift in the Center might begin with incoming flights that have maintenance issues. Pilots in the scenario would have to troubleshoot the problem while in flight and communicate to flight dispatch and maintenance technicians the nature of the problems. When the airplane lands, the technicians would go to work on the problem and soon a decision would have to be made about that airplane's availability. Can the problem be repaired before it is time for the airplane to be loaded and dispatched on its next flight, or will another airplane be required? Do we even have another airplane that can do the job? When it was time for the push of departing aircraft to begin, the action would shift to an air traffic control ramp/tower simulation. The simulation reproduces the size and layout of a ramp tower with "out the window" visual systems, allowing students to look out of a virtual window onto the flight line. Ramp/Ground controllers orchestrate the entire departure sequence from push back to takeoff. At the time of departure, a scenario could be presented of deteriorating weather. Thunderstorms in Nashville might delay departures, and fog in Florida would delay arrivals—all of this must be worked out using a total employee scenario-based approach. Solutions to the problems proposed by one group might create even more difficult problems for others. To avoid complicating the problem, students would have to learn the other group's concerns and issues—just like in the real world.

In creating the Center for Research on Aviation Training, the Aerospace Program at MTSU is testing a new method for preparing students to work in the aviation industry. We are not aware of any other university or training center that is using this approach. The center itself goes by the name Flight Operations Center – Unified Simulation or FOCUS. The FOCUS lab consists of three work areas: the ramp tower, the aircraft flight deck and the operations center. The operations center has separate workstations for the Flight Operations Coordinator, a Flight Data position, Maintenance Control, Maintenance Scheduling, Crew Scheduling and a weather station. These workstations are strategically located within the center to facilitate cross communications.

Several subcontractors supply the hardware and software to make the virtual airline operation. The Computer Sciences Corporation (CSC) is the contractor for the ramp tower and radar displays. Talon Systems is the provider of crew, aircraft and maintenance scheduling software. Next year Frasca International will be the contractor for a CRJ 200 Regional Jet simulator which will represent one airliner in the fleet. The virtual airline is named Universal E-Lines and operates to fourteen cities in the Southeastern United States with hubs in Nashville, Tennessee and Jacksonville, Florida.

In the early 2000's, the MTSU Aerospace Department became a research leader in the area of Scenario Based Training for pilots. The Federal Aviation Administration has cited this research in the Federal Register as the seminal work that has changed the way pilots are being trained today. Much of that early research centered on an individual's situational awareness (SA) and the high quality decisions made because of that awareness. This new research takes the gains made with pilots in scenario training and broadens the reach to all aviation functions. Again the primary focus is on situational awareness, but now it is about Group Situational Awareness (GSA) or "shared mental models." The research will produce a series of "best practices" for real-world aviation training and should produce students who are better prepared for the work force.

Someday, a businessperson or vacationer will make a connection from one airplane to another without any problems or cares. The arriving and departing flights will be on time. The aircraft will have all its maintenance performed and will be completely safe. The flight crews will be well rested and alert. The airline will earn a profit. Even the baggage will make the correct connection. The system will work well that day because employees of that airline made competent and informed decisions that allowed the system to work smoothly - Decisions they first learned to make at MTSU's Center for Research on Aviation Training.

The Symposium

The work of the NASA FOCUS lab is a multi-year project, 2010-2011 being the first academic year. Additional opportunities for data collection, observations and curriculum development are still to come, but this symposia presents the first deliverables from the project.

The first presentation, delivered by Dr. Paul A. Craig, describes the development and use of the simulation lab. It involves how the NASA FOCUS lab accomplishes the simultaneous operation of multiple flights from multiple airports and the various professional specializations required to operate the virtual airline and deal with unplanned situations.

The second presentation, by Dr. Mike Hein details the affects that stereotypes play with regard to effective interaction among aviation professionals. Psychology professors will present Using a Measure of Occupational Stereotype to Assess Ingroup-Outgroup Bias among Aerospace Specializations.

Ms. Andrea Georgiou presents the third paper on the Development of Criterion Measures to Assess Interpositional Knowledge and Task Mental Models.

The fourth presentation, lead by Ms. Jennifer Henslee is a report of research on Multi-team Coordination within the simulated airline operation. The research focuses on Interpositional Knowledge and Task Mental Models.

The final presentation by Mr. Gerald Hill discusses the application of these concepts into the aerospace college curriculum. One goal of the overall research effort is to develop a list of “best practices” that would guide collegiate aviation away from the “silo-only” model and toward a more realistic student preparation model.

USING A MEASURE OF OCCUPATIONAL STEREOTYPE TO ASSESS INGROUP-OUTGROUP BIAS AMONG AEROSPACE SPECIALIZATIONS

Richard G. Moffett III
Michael B. Hein
Glenn L. Littlepage
Middle Tennessee State University
Murfreesboro, TN, USA

Safe and efficient flight operations require the effective intergroup coordination across multiple aerospace specializations. Various factors can impact this coordination including attitudes such as occupational stereotypes. This study reports on the development of a measure to assess ingroup-outgroup bias among aerospace specializations. Students from six aerospace specializations (Administration, Aircraft Maintenance, Air Traffic Control, Flight Dispatch and Scheduling, Professional Pilot, and Technology) at a southern university completed a questionnaire designed to assess stereotypes using adjectives to describe the members of ingroups and outgroups. Results indicate that students who had identified themselves as pilots or dispatchers exhibited ingroup-outgroup bias but in different ways. Implications of the findings for educating and training students in aerospace specializations are provided.

Safe and efficient flight operations require the effective coordination across multiple aerospace specializations such as Pilot, Flight Dispatch, Air Traffic Control, and Aircraft Maintenance. Various factors can impact this coordination including attitudes. According to Gregorich, Helmreich, and Wilhelm (1990) “considerable evidence exists which suggests that attitudes about the management of flight deck resources do influence the quality of crew coordination” (p. 682). Such findings have been applied to situations found in flight operations centers as noted in the evolutionary shift from cockpit to crew resource management (CRM) perspective (Helmreich, Merritt, & Wilhelm, 1999).

O’Conner, Campbell, Newton, Melton, Salas, and Wilson (2008) state that the primary measure used to assess attitudes that might impact crew coordination, especially in CRM training programs, is the cockpit management attitude questionnaire (CMAQ) developed by Helmreich (1984). According to Helmreich the CMAQ is a self-report survey that measures attitudes related to communication, coordination, command responsibility, and recognition of stressor effects. However, one attitude that has not been investigated as a possible impact on CRM is occupational stereotypes represented by aerospace specializations such as Pilots, Flight Dispatchers, and Maintenance Technicians.

Stereotypes can be defined as a “socially shared set of cognitions (e.g., beliefs, expectations) about the qualities and characteristics of the members of a particular group or social category” (Forsyth, 2010, p. 426). Stereotypes about occupations have been studied for sometime (e.g., Walker, 1958) but typically in the context of gender (e.g., White & White, 2006). This contextual limitation holds true for aerospace specializations (e.g., Davey & Davidson, 2000). However, some researchers have focused on stereotypes between occupations such as medical researchers and biomedical scientists (Lewitt, Ehrenborg, Scheja, & Brauner, 2010), and between engineers and managers (Jemielniak, 2007). Occupational stereotypes about aerospace specializations might impact the performance of individuals in a flight operations center through the process of ingroup-outgroup bias.

Ingroup-outgroup bias is the “tendency to view (one’s own group) and its members more favorably than other groups” (Forsyth, 2010, p. 82). Forsyth noted that ingroup-outgroup bias can increase conflict and result in a breakdown in the ability of groups to work effectively towards a common goal. Furthermore, he states that when intergroup conflict occurs, typically the “ingroup favoritism is stronger than outgroup rejection” (p. 423).

One way to assess ingroup-outgroup bias among occupations is to identify the likeableness of the words used to describe the members of one’s ingroup and to describe the members of the outgroups. If the words chosen for one’s ingroup are significantly more favorable than those chosen for the outgroups, evidence for possible ingroup-outgroup bias could be said to exist. By having members of aerospace specializations (Dispatch, Maintenance, and Pilots) choose words that they think generally describes the members of their own specialization and other specializations, we can begin to assess if ingroup-outgroup bias exists.

The Present Research

Development of a Measure of Stereotypes about Aerospace Specialization

Dumas, Johnson, and Lynch (2002) developed a list of 844 person-descriptive words (adjectives) and established ratings of likeableness and familiarity of those words. The authors reported that their list expands on earlier word lists that researchers frequently use to “describe of qualify an individual’s personality, dispositions, or behavior” (p. 523). Mean values and standard deviations for both likeableness and familiarity are available for each word. The results from Dumas et al. indicate that the ratings are reliable and correlate well with ratings found by previous researchers.

For the present research, we selected words from the list of 844 adjectives established by Dumas et al. in which both likeableness and familiarity were rated on 6-point scales. We initially selected items that met two criteria: high familiarity (mean rating of 5 or above) and consistency of rated likableness (*SD* < 1). Each of the authors examined the resulting list of 314 adjectives and independently selected items that represented personal characteristics relevant to work teams. For example, the word “competent” was used but not “amorous.” From these items, 139 adjectives were selected by all three authors. These items were then alphabetized.

Hypotheses and Analyses

H1: Pilots will select adjectives with a higher mean likeableness rating to describe pilots than they will to describe maintenance personnel.

H2: Pilots will select adjectives with a higher mean likeableness rating to describe pilots than they will to describe dispatch personnel.

H3: Dispatchers will select adjectives with a higher mean likeableness rating to describe dispatchers than they will to describe maintenance personnel.

H4: Dispatchers will select adjectives with a higher mean likeableness rating to describe dispatchers than they will to describe pilots.

Each of the four hypotheses will be tested with a series of dependent sample t-tests. These analyses will be conducted separately for ratings by students in the Professional Pilot, Flight Dispatch and Scheduling, and Aircraft Maintenance programs and will contrast the mean rating of the adjectives selected to describe each specialty area.

Method

Participants and Procedure

Researchers at Middle Tennessee State University, working through a NASA-funded project titled MTSU Center for Research on Aviation Training, built a lab which is a replica of an airline’s flight operations center. The simulation lab is referred to as NASA FOCUS, (flight operations center-unified simulation). Senior-level MTSU students from six aerospace specializations participate in high-fidelity simulations of a regional airline. Students interactively participate in a simulated work shift playing roles of aircraft dispatchers, pilots, ramp controllers, maintenance technicians, crew schedulers, and weather briefers. Combining the efforts of aerospace and psychology professors and graduate assistants, there are a number of research efforts currently underway including concepts of teamwork, shared task mental models, interpositional knowledge, attitudes, and future curriculum revisions.

As part of this larger research project, 60 students from six aerospace specializations (Administration, Aircraft Maintenance, Air Traffic Control, Flight Dispatch and Scheduling, Professional Pilot, and Technology) completed the “Perceptions of Aerospace Professionals” questionnaire. For the purposes of the present research only 39 subjects were used: 11 subjects identified themselves as Dispatch and 28 subjects identified themselves as Pilot. The remaining aerospace specializations had too few subjects to be included in the analyses.

The “Perceptions of Aerospace Professionals” questionnaire consisted of three copies of the list of adjectives described above; one for each of the target specializations to be described (i.e., one for pilots, one for flight dispatchers, and one for aircraft maintenance personnel). Each participant was asked to circle the specialization that best reflected his/her specialization. The specializations included Air Traffic Control, Flight Dispatch & Scheduling, Aircraft Maintenance (training or management), Professional Pilot, Administration, and Technology. Next, participants were asked to circle the adjectives that are generally descriptive of members of each of the specified aerospace specializations (Dispatch, Maintenance, and Pilot).

For each participant, a likeableness score for each adjective chosen was determined by using the mean established for that word by Dumas et al. These scores were then averaged within a specialization to yield a mean likeableness score for that specialization.

Results and Discussion

Overall, partial support was provided for the hypotheses.

Figure 1 presents the mean likeableness ratings of those who identified themselves as Pilots. As is shown in this figure, these subjects viewed themselves and the members of the other groups in a generally favorable light.

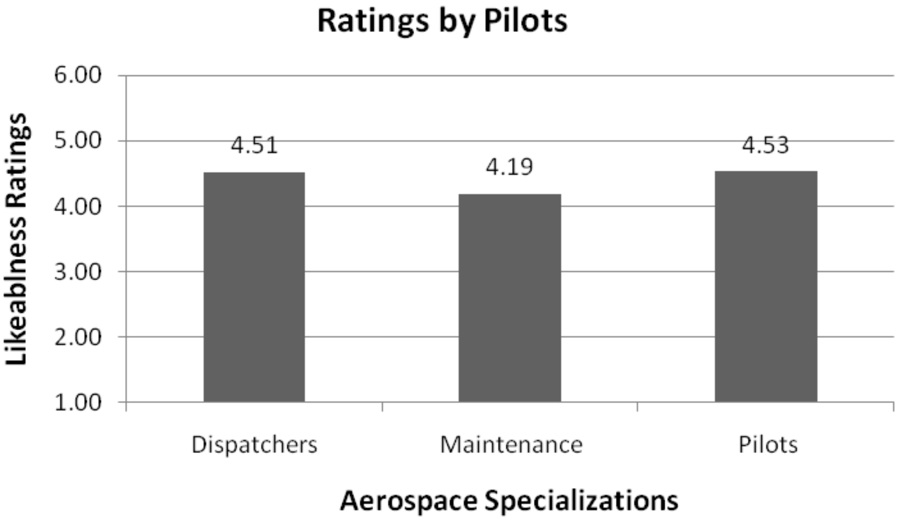


Figure 1. Mean likebleness ratings by pilots.

The results of a dependent sample t-test, comparing the mean rating of the adjectives selected to describe each specialty area by Pilots supported Hypothesis 1. There was a significant effect for specialty area, $t(27) = 2.31$, $p < .05$, with pilots (M=4.53, SD=.60) receiving higher scores than maintenance personnel (M=4.19, SD=.91) from pilots. However, Hypothesis 2 was not supported. There was a not a significant effect for group, $t(27) = .11$, $p > .05$, with pilots (M=4.53, SD=.60) receiving scores almost identical to dispatchers (M=4.51, SD=.69) from pilots.

Figure 2 presents the mean likeableness ratings of those who identified themselves as Dispatchers. As is shown in this figure, these subjects viewed themselves and the members of the other groups in a generally favorable light.

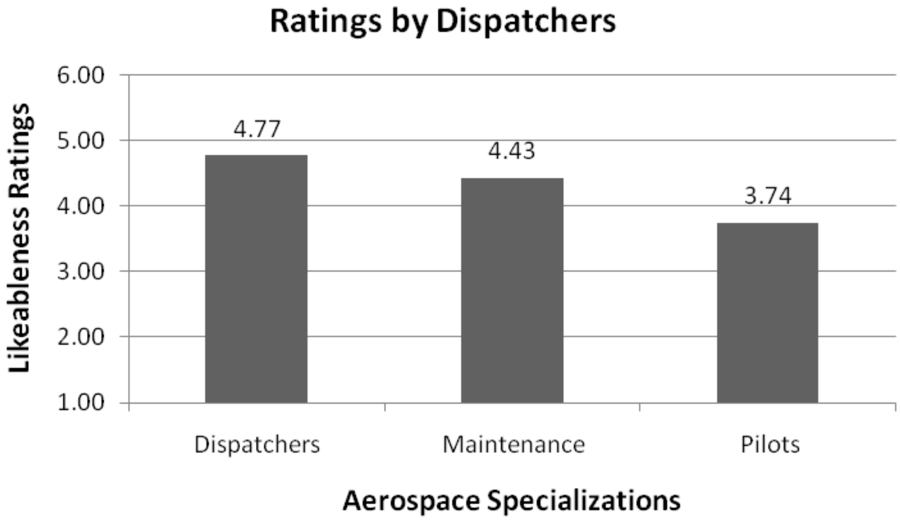


Figure 2. Mean likeableness ratings by dispatchers.

The results of a dependent sample t-test, comparing the mean rating of the adjectives selected to describe each specialty area by Dispatchers did not support Hypothesis 3. There was a not a significant effect for group, $t(10) = .12$, $p > .05$, with dispatchers (M=4.77, SD=.50) receiving scores not significantly different than maintenance personnel (M=4.43, SD=.63) from dispatchers. However, Hypothesis 4 was supported. There was a significant effect for specialty area, $t(10) = 2.77$, $p < .05$, with dispatchers (M=4.77, SD=.50) receiving higher scores than pilots (M=3.74, SD=1.12) from dispatchers.

The above results indicate that both Pilots and Dispatchers may have ingroup-outgroup bias towards other aerospace specializations, but this effect is manifested toward different groups. Pilots exhibit ingroup-outgroup bias towards Maintenance personnel but not Dispatchers who they view very similarly to themselves. On the other hand, Dispatchers exhibit ingroup-outgroup bias towards Pilots but not Maintenance personnel.

It is beyond the scope of this study to speculate as to the reasons for this difference in ingroup-outgroup bias between Pilots and Dispatchers. However, looking at the trends of the differences it appears that the size of the ingroup-outgroup bias held by Dispatchers towards Pilots indicates that the stereotypes held by Dispatchers could be more problematic and might impact effective coordination between these groups.

Some limitations of this study involve the sample used. The overall sample size was small which reduces the generalizability of the findings. Additionally, if data from other aerospace specializations could have been collected the nature of the observed trends might be better understood.

As mentioned earlier, the present study is part of a larger research effort. The focus of this research effort is to understand the effects of training aerospace specializations in an intensively interdependent simulated environment called the NASA Flight Operations Center – Unified Simulation (FOCUS). This simulation replicates a flight operations center of regional airlines. This simulation requires students from six aerospace specializations to perform their jobs and coordinate their efforts to deal with normal and unplanned situations. The simulation requires face-to-face contact in the pursuit of a superordinate goal of safely and efficiently operating the flight operations center. These intergroup opportunities provide the foundation for reducing possible aerospace specialization stereotypes and thus reducing ingroup-outgroup bias (Pettigrew & Troop, 2006).

To test the effectiveness of the simulation to reduce this bias, the measure used in the present research will be re-administered as a post-test once students have received training in the simulation. If the experience of working with other aerospace specializations in the intensively interdependent simulated environment can be shown to reduce stereotypes about aerospace specializations, then suggestions about education and training can be developed to increase coordination and reduce conflict, which may facilitate increasing airline safety and efficiency.

Acknowledgements

This research was supported in part by a contract (NNX09AAU52G) from NASA awarded to the Middle Tennessee State University Center for Research on Aviation Training. We are grateful to Paul A. Craig and Andrea M. Georgiou at the Department of Aerospace and the Center for Research on Aviation Training at Middle Tennessee State University for their helpful advice. We thank Jennifer Henslee and Ryan Lillard for their help with data collection, entry, and coding. Correspondence concerning this article should be addressed to Richard G. Moffett III, Department of Psychology, MTSU Box 87, Middle Tennessee State University, Murfreesboro, TN 37132. E-mail: rmoffett@mtsu.edu.

References

Davey, C. L.,& Davidson, M. J. (2000). The right of passage? The experiences of female pilots in commercial aviation. *Feminism and Psychology, 10,* 195–225.

Dumas, J. E., Johnson, M., & Lynch, A. M. (2002). Likableness, familiarity, and frequency of 844 person-descriptive words. Personality and Individual Differences, 32, 523-531.

Forsyth, D.R. (2010). Group dynamics (5th ed.). Belmont, CA: Wadsworth.

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. (1984). Cockpit management attitudes. Human Factors, 26, 583-589.

Helmreich, R.L., Merritt, A.C., & Wilhelm, J.A. (1999). The evolution of crew resource management training in commercial aviation. International Journal of Aviation Psychology, 9, 19-32.

Jemielniak, D. (2007). Managers as lazy, stupid careerists? Contestation and stereotypes among software engineers. Journal of Organizational Change Management, 20, 491-508.

Lewitt, M.S.; Ehrenborg, E.; Scheja, M.; & Brauner, A. (2010). Stereotyping at the undergraduate level revealed during interprofessional learning between future doctors and biomedical scientists. Journal of Interprofessional Care, 24, 53-62.

O'Connor, P., Campbell, J., Newton, J. Melton, J. Salas, E., & Wilson, K. (2008). Crew resource management training effectiveness: A meta-analysis and some critical needs. International Journal of Aviation Psychology, 18, 353-368.

Pettigrew, T.F., & Troop, L.R. (2006). A meta-analytic test of intergroup contact theory. Journal of Personality and Social Psychology, 90, 751-783.

Walker, K.F. (1958). A study of occupational stereotypes. Journal of Applied Psychology, 42, 122-124.

White, M.J., & White, G.B. (2006). Implicit and explicit occupational gender stereotypes. Sex Roles, 55, 259-266.

DEVELOPMENT OF CRITERION MEASURES TO ASSESS INTERPOSITIONAL KNOWLEDGE AND TASK MENTAL MODELS

Andrea M. Georgiou
Glenn E. Littlepage
Jennifer A. Henslee
Middle Tennessee State University
Murfreesboro, TN, USA

Of utmost importance in the aviation industry is the ability for professionals to work well in a team and understand the intersections of positions for safe operations. In an effort to enhance the understanding of teamwork and communication, senior-level undergraduate aerospace students are currently participating in a NASA funded replica of an airline Flight Operations Center of a regional airline. Students from six aerospace specializations interactively complete a simulated work shift playing roles of aircraft dispatchers, pilots, ramp controllers, maintenance technicians, crew schedulers, and weather briefers. Surveys were collected from Subject matter Experts (SME), and the data statistically analyzed to determine areas of significant agreement. As a result, criterion measures were developed to assess the degree of accuracy and similarity of tasks mental models, as well as positional and interpositional knowledge.

Within high-risk environments such as aviation, it is vital individuals understand the importance of teamwork and the presence of interdependencies in order to ensure safe and efficient operation. Particularly with the ever-increasing technology present with the Next Gen Initiative, every person working within an airline must be competent in a wide range of complex tasks from flying sophisticated aircraft, ensuring safe separation, properly dispatching the flight, and maintaining the aircraft in an airworthy condition. All these tasks, and much more, require every worker to understand how their position affects the entire organization and to take into consideration ways in which their decisions may affect daily operations. At the individual level, this type of knowledge of the basic functioning of tasks which allows individuals to form predictions and expectations about future circumstances is referred to as task mental models (DeChurch & Mesmer-Magnus, 2010). Undoubtedly, the absence of task mental models would be detrimental to the safety of airplane operations as individuals would view their actions in a sort of “tunnel-vision” and lack the knowledge of the overall organizational structure present in aviation.

Resounding throughout the federal regulations, advisory circulars, and National Transportation of Safety Board (NTSB) safety recommendations, the Federal Aviation Administration (FAA) and NTSB emphasize concepts of teamwork, communication, and interdependence of expertise from all levels within an organization with a familiar term coined crew resource management (CRM) (FAA, 2004; FAA, 2005; FAA 2010, NTSB, 2010). According to the FAA Advisory Circular 120-51E, crew resource management training focuses on the importance of crewmembers, dispatchers, mechanics, Air Traffic Controllers, flight attendants, and others involved to operate as one unit or team and to understand their actions and attitudes affect safety (2004). In an effort to better prepare aviation collegiate students to enter the workforce with a deeper understanding of CRM, senior-level undergraduate aerospace students from six aerospace specializations participate in high-fidelity simulations of a regional airline. Funded through a NASA grant, students interactively participate in a simulated work shift playing roles of aircraft dispatchers, pilots, ramp controllers, maintenance technicians, crew schedulers, and weather briefers. Combining the efforts of aerospace and psychology professors and graduate assistants, the simulation lab, referred to as NASA FOCUS, (flight operations center-unified simulation) has a number of research efforts currently underway including concepts of teamwork, shared task mental models, interpositional knowledge, attitudes, and future curriculum revisions. The purpose of this paper is to present the types of criterion measures developed to assess positional and interpositional knowledge and the degree of similarity and accuracy of task mental models, as well as report the findings from the subject matter expert (SME) surveys. The SME data were collected and statistically analyzed to determine areas of significant agreement.

Review of Related Literature

Generally, the collegiate environment offers very few opportunities for cross-training across various disciplines for aerospace students. As it stands, students pursuing a career in aircraft dispatch take a majority of their major courses with other aircraft dispatch students, pilots take courses with other pilot classmates, and so forth. In this type of educational environment a student has very little exposure to the responsibilities and tasks required of other positions and as a result have limited interpositional knowledge. Some of the earliest research in team member effectiveness indicated that cross-training better equips individuals to handle high workload conditions and positively affects the overall team member effectiveness. (Cannon-Bowers, Salas, Blickensderfer, & Bowers, 1998; Volpe, Cannon-Bowers, Salas, & Spector, 1996). Cross-training was positively correlated with improvements in team accuracy, speed, interpositional knowledge, and volunteering information. Further illustrating the importance of cross-team processes, Marks, DeChurch, Mathieu, Panzer and Alonso (2005) discovered that for organizations that utilize two or more teams which interact directly and interdependently, the cross-team processes were more important than within-team processes for predicting effective team performance. In other words, the action processes of the cross-teams within a system are most valuable when operating in a unique environment of high interdependence.

Another important aspect of team functionality towards reaching a common goal is the accuracy and similarity of individual and shared mental models. Shared mental models "allow team members to draw on their own well-structured knowledge as a basis for selecting actions that are consistent and coordinated with those of their teammates" (Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000, p. 274). Mathieu et al.,(2000) discovered cognitive ability was important for the initial emergence of mental model accuracy and similarity in teams. In other words, since new teams lack the experience necessary to develop shared mental models necessary to predict behaviors and actions necessary for team effectiveness, each team member must have a foundational level of cognitive ability to be aware of basic tasks necessary to complete the mission. Although mental model similarity was positively related to team agreeableness, it was not significantly related to goal accomplishment. A study of Air Traffic Controller's interactions among shared mental models proves it is less important to agree on a strategy and more important to consistently judge the likelihood of success of potential strategies under different conditions (Smith-Jentsch, Mathieu, & Kraiger, 2005). Taking into consideration the ever-changing environment unique to aviation, it is of utmost importance for individuals to be able to quickly, and effectively develop strategies while being mindful of all other factors affecting operations.

More important than similar shared mental models was the degree of accuracy as there was a small positive relationship with mental model accuracy and goal accomplishment (Mathieu et al., 2000). A team may develop a similar mental model yet this does not help towards accomplishing a goal as the entire team may have developed an inaccurate model of thinking. This type of flawed phenomena is often referred to as groupthink, in which an entire group of people conform to the same ideas and beliefs despite the fact they are not accurate. The priority becomes seeking agreement among the group rather than accurate task models or appropriate decisions (Kassin, Fein, & Markus, 2008). Proving similar mental models does not necessary translate to accomplishing a task, Resick et al., (2010) discovered mental model similarity does not always correlate with goal accomplishment, as the entire team may have developed similar yet inaccurate task mental models. DeChurch and Mesmer-Magnus (2010) point out that regardless of accuracy or similarity, shared mental models positively relate to team performance.

Method

Participants

The participants for this study were voluntary and purposefully selected based on their area of expertise. According to Leedy & Ormrod (2010), purposive sampling is appropriate when specific groups of people are needed for a particular purpose, which aligns with this study's focus on aircraft dispatchers, mechanics, and pilots. Two groups of Subject Matter Experts (SME) were utilized for the development of the criterion-measures. The first group of SME's consisted of MTSU Aerospace faculty members who were responsible for the design of the assessment tool survey; the second group of SME's was working aviation professionals who provided answers to the surveys. Responses were received from 10 professional pilots (ranging from 4-25 years experience, 1,000 to 16, 900 logged flight time), 10 aircraft mechanics (8-25 years experience), and 7 aircraft dispatchers (9-32 years experience). With the exception of one individual, all SMEs that completed the survey work at Part 121 Air Carrier operations.

Criterion Measures for Interpositional Knowledge and Task Mental Models

The MTSU Aerospace faculty members developed 38 questions to measure knowledge of various aerospace specializations and seven scenarios. These questions were later rated by industry SMEs to develop measures of positional knowledge (PK) which is knowledge of one's area of specialization, and interpositional knowledge (IPK) which is knowledge of other team member's roles and responsibilities. Each of the scenarios described some type of dilemma occurring either on the ground or in flight, followed by strategies to resolve the issue. The industry SMEs were asked to rate the likelihood that each strategy would be effective in resolving the issue. Ratings were made using a likert scale from 1-11 with 1 representing 0% likelihood of success and 11 representing 100% likelihood of success. The purpose of the scenarios was to develop criterion to measure the accuracy of task mental models.

Selected quiz items and scenarios were included in questionnaires that were emailed to the industry SMEs. Three different questionnaire packets were developed and emailed to individuals working within the specific area, each containing quiz items and scenarios relevant to one of the following specializations: aircraft dispatchers, aircraft mechanics, and professional pilots. To clarify, professional pilots were emailed the professional pilot questionnaire, dispatchers the dispatch questionnaire, and aircraft mechanics the appropriate questionnaire. Individuals participated on a voluntary, anonymous basis and completed demographic information to ensure they worked in the appropriate area and had sufficient experience to be able to contribute to the study. Each questionnaire contained the following three sections: 1) information about the extent and nature of professional experience, 2) multiple choice questions relevant to the specialization, and 3) scenarios of a non-routine nature relevant to the specialization.

Results

For the multiple choice questions, the frequency distribution was calculated to determine the percentage of SME agreement on the correct answer. For aircraft dispatch, all 7 items showed acceptable agreement ranging from 71% to 100% agreement. For aircraft maintenance 5 items showed acceptable agreement (ranging from 80% to 100% agreement) and two question items had significant disagreement. With agreement ranging from 67% to 100%, 5 of 7 items indicated agreement for professional pilot experts. Any items with significant disagreement will not be used as criterion to assess accuracy of task mental models.

For the scenarios, we calculated the mean effectiveness rating for each option for each of the seven scenarios. Because some scenarios were relevant to only some specializations, Scenarios 1-4 were rated only by pilots (N=10) while Scenarios 5 & 7 were rated by all 3 specializations (N =27) and Scenario 6 was rated by pilots and dispatch (N = 17). Within each scenario, we examined the mean SME rating of effectiveness for each option.

Across the seven scenarios, the mean difference between the most and least effective options to resolve a specific scenario was 5.43. When looking at each scenario separately, the difference between the most and least effective options ranged from a low of 2.5 to a high of 7.41. This indicates that although some scenarios provided options with only a limited range of effectiveness, in general, SMEs perceived substantial differences in effectiveness between various options. One-way repeated measures ANOVAs for each of the seven scenarios further supported the conclusion that SMEs ratings differed across various responses to the scenario. Effectiveness ratings differed for six of the seven scenarios ($p < .05$); this was true even though 4 of the scenarios were rated only by pilots ($N = 10$).

For each scenario, SME agreement was assessed by examining the similarity between the effectiveness ratings of the SMEs. The average correlations between SMEs varied considerably. For two scenarios, they were very low (-.01 and .03), for one scenario the mean correlation was .34, and for the remaining four scenarios the mean correlation was higher (ranging from .53 to .71). Despite the varied correlations, a composite measure derived from the set of SME ratings yielded a highly reliable measure for 5 of the 7 scenarios ($\alpha > 84$). The set of mean ratings of SMEs to each of the options serves as the criterion needed in later studies to evaluate the accuracy of the task mental models.

Discussion

There was excellent agreement between aircraft dispatchers for the multiple choice items. All SMEs selected the same answer for all 7 items, which indicates accuracy among the dispatcher's task mental models. Agreement among maintenance and pilot SMEs was lower, but still resulted in a pool of items with acceptable levels of agreement. Aircraft maintenance had at least 80% agreement on the answers for 5 out of 7 items, and the pilots had at least 66% agreement, which still indicates a fairly accurate mental model. The pool of items with acceptable SME agreement provides a way to measure accuracy of knowledge relevant to three professional specializations, dispatch, maintenance, and pilots. These items allow for the measurement of positional and interpositional knowledge of students in each of these disciplines. In later phases of this research program, this will facilitate the examination of the following questions related to positional knowledge (PK) and interpositional knowledge (IPK).

1. Will students have a greater level of their own specialization (PK) than of other specializations (IPK)?
2. Will a student's knowledge of their own area (PK) improve after participating in the NASA FOCUS lab simulations?
3. Will the NASA FOCUS training increase student interpositional knowledge?

SME evaluations of the effectiveness of various options indicate that some options are more effective than others. The mean SME evaluations establish criteria that can be used to evaluate the accuracy of task mental models. For future purposes, the data collected and analyzed from the SME measures will be used as the criterion to research the following questions:

4. Do traditionally trained aerospace students have accurate task mental models?
5. Does the training in the NASA FOCUS lab cause the student's mental models to become more similar?
6. Does the training in the NASA FOCUS lab cause the students mental models to more accurate?

A limitation to this study was in the small sample size. Future studies of similar nature with a larger sample size of subject matter experts would supplement and solidify the development of criterion measures to assess mental models, positional and interpositional knowledge as it correlates with effective teamwork. Another limitation is the small number of quiz items. Only 5-7 items are available to measure knowledge of each of the three focal aerospace specializations. Likewise, only two scenarios are related to maintenance and only three are related to dispatch, while all seven involve pilots. This imbalance may make it more difficult to adequately access the mental model similarity and accuracy for maintenance and dispatch. A larger pool of quiz items and a larger and more varied pool of scenarios, as well as a larger sample of SMEs could provide for more effective measurement of PK, IPK, and similarity and accuracy of task mental models. Nevertheless, the current work seems to provide an initial foundation that can be used to examine PK, IPK, and task mental models among aerospace students and professionals.

While our primary focus has been on the use of SMEs to establish criteria for measuring knowledge and mental models, the program of research utilizes additional measures that will be used to supplement our understanding of the effects of coordination training in the NASA FOCUS lab. Examples of these measures include:

- Perceptions of Aerospace Professionals (Occupational Stereotypes). For each of three specializations (dispatch, maintenance, pilot) measures perceptions of own versus other specializations. Measures are computed from the favorability ratings of adjectives used to describe members of various aviation specializations.
- Interdependence Questionnaire. Measures the extent to which participants think their specialization needs to rely on other specializations.
- Beliefs about Groups. Measures reactions to working in teams vs. alone.
- Communication Requirements. Measures perceptions of information flow requirements across various specializations.
- Communication Patterns. Measures extent of communication between various positions in the simulation.
- Teamwork Questionnaire. Measures the extent of teamwork during the simulation.

Although our program of research is clearly a work in progress, it provides a foundation to evaluate efforts such as the coordination training provided in the NASA FOCUS lab. Learning how to facilitate effective coordination across aerospace specializations holds much promise for increasing airline safety and efficiency.

Acknowledgements

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

References

Cannon-Bowers, J. A., Salas, E., Blickensderfer, E., & Bowers, C. A. (1998). The impact of cross-training and workload on team functioning: a replication and extension of initial findings. *Human Factors, 40*(1), 92-102.

DeChurch, L.A., & Mesmer-Mangus, J. R. (2010). Measuring team mental models: A meta analysis. *Group Dynamics: Theory, Research, and Practice, 14*(1), 1-14. doi 10.1037//a0017455

Kassin, S., Fein, S., & Markus, H.R. (2008). Social Psychology (7th ed.) Boston, MA: Houghton Mifflin Co

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

Marks, M. A., DeChurch, L. A., Mathieu, J. E., Pranzer, F. J., & Alonso, A. (2005). Teamwork in multiteam systems. *Journal of Applied Psychology, 90*(5), 964-971.

Mathieu, J.E., Heffner, T.S., Goodwin G.F., Salas, E., & Cannon-Bowers, J.A. (2000). The influence of shared shared mental models on team process and performance. *Journal of Applied Psychology, 85*(2), 273-283. doi: 10.1037//0021-9010.85.2.273

National Transportation Safety Board [NTSB] (2010). *Most wanted transportation safety improvements: Aviation issue areas.* Retrieved on February 16, 2011 from http://ntsb.gov/Recs/mostwanted/aviation_issues.htm

Resick, C.J., Dickson, M.W., Mitchelson, J.K., Allison, L.K., & Clark, M.A. (2010). Team composition, cognition, and effectiveness: Examining mental model similarity and accuracy. *Group Dynamics: Theory, Research, and Practice, 14*(2), 174-191. doi: 10.1037/a0018444.

Smith-Jentsch, K. A., & Mathieu, J. E. (2005). Investigating linear and interactive effects of shared mental models on safety and efficiency in a field setting. *Journal of Applied Psychology, 90*(3), 523-535 doi: 10.1037/0021-9010.90.3.523

United States Department of Transportation. Federal Aviation Administration [FAA] (2004). *Crew resource management training*. (Advisory Circular 120-51E). Retrieved on February 25, 2011 from http://rgl.faa.gov/Regulatory_and_Guidance_Library/rgAdvisoryCircular.nsf/list/AC%20120-51E/$FILE/AC120-51e.pdf

United States Department of Transportation. Federal Aviation Administration [FAA] (2005). *Dispatch resource management training.*(Advisory Circular 121-32A). Retrieved on February 25, 2011 from http://rgl.faa.gov/Regulatory_and_Guidance_Library/rgAdvisoryCircular.nsf/list/AC%20121-32A/$FILE/AC%20121-32a.pdf

United States Department of Transportation. Federal Aviation Administration [FAA]. (2010). *Federal Aviation Regulations/ Aeronautical Information Manual*. 14 CFR, Part 121. New Castle, WA: Aviation Supplies & Academics

Volpe, C. E., Cannon-Bowers, J. A., & Salas, E. (1996). The impact of cross-training on team functioning: An empirical investigation. *Human Factors,* 3(1)8, 87-100.

.

MULTITEAM COORDINATION IN SIMULATED AIRLINE OPERATIONS: ASSESSMENT OF INTERPOSITIONAL KNOWLEDGE AND TASK MENTAL MODELS

Glenn E. Littlepage
Middle Tennessee State University
Murfreesboro, TN
Jennifer A. Henslee
Middle Tennessee State University
Murfreesboro, TN

Effective airline operations require coordination among various specializations such as pilot, flight dispatch, and maintenance. Interpositional knowledge (IPK) and task mental models are emergent cognitive states that can facilitate effective coordination. This study examined the extent of IPK and similarity and accuracy of task mental models among aerospace students. Results indicated relatively low levels of IPK and mental model similarity and moderate levels of mental model accuracy. Training activities that enhance IPK and task mental models have the potential to improve coordination and performance of airline personnel.

Flight cancellations and delays inconvenience passengers, disrupt business activities, and cause operational problems for airlines. The overall economic impact of delays and cancelations in the U.S. exceeds $31 Billion annually (NEXTOR, 2010). While disruptions cannot be eliminated entirely, more effective coordination among differing specializations of aviation professionals offers to reduce their frequency and duration. When unexpected problems arise (e.g. weather conditions, mechanical difficulties, passenger illness or incidents) effective communication, similar mental models, effective coordination, and proactive action can help avoid or limit possible disruption. Currently, we are conducting a multi-year project to study coordination between various aviation specializations. In order to examine group processes and emergent states that impact multiteam performance, we utilized a high-fidelity simulation that incorporates both routine and non-routine work situations. Recent theory and research on multiteam systems provides a perspective from which to view the coordination required to maintain efficient airline performance.

A multiteam system is composed of two or more teams that must work interdependently to reach one or more collective goals (DeChurch & Marks, 2006; Mathieu, Marks, & Zaccaro, 2002). Just as members of cross-functional teams may have mixed motives, component teams may have proximal goals that are not fully synchronized, nevertheless they must coordinate to achieve critical distal goals. The effective and efficient operation of an airline depends on the coordinated actions of persons in various aviation specializations such as pilot, flight dispatch, maintenance, and others. Although these specializations may have differing proximal goals (e.g., thorough maintenance inspection vs. on-time departure), they share common superordinate goals of safety and operational efficiency.

Research suggests that coordination among persons or teams with different types of expertise is critical to effectiveness (Marks, Mathieu, & Zacarro, 2001; Salas, Sims, & Burke, 2005). This may be especially true for teams that operate in dynamic environments, such as emergency response, military operations, and commercial aviation. Shared cognitive states are important factors that facilitate effective, coordinated team performance (DeChurch & Mesmer-Magnus, 2010; Mathieu, Hefner, Goodwin, Salas, & Cannon-Bowers, 2000; van Ginkel, Tindale, & van Knippenberg, 2009). One cognitive state that relates to effective team performance is interpositional knowledge (IPK). IPK represents a team member's knowledge of the tasks, roles, and behaviors required of other team members. A related cognitive state that enhances team performance is a shared mental model. A shared mental model exists when team members share a common view of the task or teamwork requirements. Team coordination, viability, and performance are enhanced when members share accurate mental models (Resick, Dickson, Mitchelson, Allison, & Clark, 2010; Smith-Jentsch & Mathieu, 2005).

Traditionally, aviation students are trained primarily in their specific specialization and have limited knowledge of other specializations. Upon entering professional employment, aviation personnel must develop the ability to effectively coordinate with other specializations. We are currently engaged in a multi-year project directed at two goals: to understand emergent states and processes that affect effective performance, and to develop and evaluate a training program to enhance coordination among aviation professionals. This report describes only results of the first of three phases of this project. Before presenting the results of this initial phase, an overview of the entire project is presented.

For this project, a high-fidelity simulation of an airline was constructed. The simulation involves coordination among several workstations at three separate locations: local airport operations (Ramp Tower), cockpit operations, and airline flight operations. Together these locations control the operation of a simulated airline with a fleet of 30 aircraft. The simulation requires coordinated action of 10-12 persons working in different aviation specializations. The local airport operations center controls airport functions such as gate departure and arrivals, taxiway clearances, and takeoff and landing clearances. This location houses up to three participants and provides a panoramic view of the airline's gates. Persons working at these stations control all aspects of airplane movement from the gate to takeoff. The airport area and all plane movements are displayed on three large video screens. A pilot/captain and first officer work in a second location where they control a flight simulator. Currently, they operate a low-fidelity simulator from a computer screen, but in the near future, they will work from a fully functional flight simulator designed to mirror the aircraft used by the virtual airline.

The third location, the flight operations center is the most complex. It contains six workstations: flight operations coordinator, flight dispatch data, maintenance control, maintenance scheduling, crew scheduling, and weather monitoring. Each station contains a computer with station-relevant information. Seven large video screens display information such as flight schedules, radar views of flights in progress, and a weather map. The two flight dispatch positions monitor flights and make scheduling adjustments. Maintenance control engages in real-time conversations with pilots in flight to evaluate maintenance issues that arise. Maintenance scheduling oversees routine and non-routine maintenance activities and is aware of planes that may be available for service. Crew scheduling has data about crew service limitations and availability of other personnel for backup duty. Weather monitoring is aware of weather conditions that may affect airline operations.

In Phase One, we examine the IPK and task mental models of traditionally trained aerospace students—students trained almost exclusively in their area of specialization. These students are not exposed to the simulation lab. In Phase Two, we examine the effects of an extremely low-fidelity, talk-through simulation. In this phase, participants learn the functions of the various workstations and participate in guided discussions of the coordinated actions required to collectively deal with various scenarios such as a bird strike or a temporary disruption of the fuel distribution system. In Phase Three, participants complete the high-fidelity simulation. Participants in this phase manage their workstations and respond to normal and non-routine situations.

Phase One, the focus of this paper, provides baseline data for traditionally trained aerospace students. Because these students have had only few opportunities for coordinated work with students in complementary aerospace specializations, we expect them to have low levels of IPK. Thus, we expect pilot, flight dispatch, and maintenance management students to have higher levels of knowledge specific to their respective specializations than to complementary specializations. Likewise, we expect that mental models of members of different specializations will not be highly similar. We anticipate that interactive coordination training among specializations will enhance IPK and lead to the development of more accurate shared mental models, although tests of this hypothesis await collection of data in Phases Two and Three.

Method

Participants in Phase One consisted of 63 students enrolled in a capstone course in the Aerospace Department of a large university in the Southeastern United States. The specializations represented were Professional Pilot ($N = 29$), Airport Administration ($N = 22$), Flight Dispatch ($N = 8$), and Maintenance Management ($N = 8$).

With the assistance of Subject Matter Experts (SMEs), we developed a number of measures. A quiz covering several aviation specializations was used to assess IPK. Specific questions were designed to reflect job-related knowledge related to each specialization. SMEs from three specializations (pilots, flight dispatchers, maintenance technicians) answered the questions related to their area of expertise and verified the previously identified correct answers. Also, with the assistance of SMEs, following Smith-Jentsch and colleagues (2005), another measure containing seven scenarios was developed. The scenarios included the following problems: alternator failure at night, runway incursion, nighttime runway incursion, communication failure, bird strike, unruly passenger, and an equipment problem complicated by weather. Each scenario included four to six alternative possible responses to the problem. To successfully deal with the problem presented in a scenario, action is required by one or more aerospace specialization(s). All scenarios required responses from pilots, while the last three also required responses from flight dispatchers, and scenarios 5 and 7 also required responses from maintenance personnel. SMEs, from the three specializations mentioned above, also rated the effectiveness of various responses to scenarios that were relevant to their specialization. Their ratings were used to develop indices of task mental model accuracy. Participants in this study completed the quiz and scenario instruments during the final meeting of the course.

Results

Interpositional Knowledge

We evaluated the hypothesis that positional knowledge would be greater than IPK by examining the accuracy scores of responses to quiz items that did and did not reflect the participant’s aviation specialization. To control for difficulty differences across quiz items, scores for each item were standardized prior to analysis. Overall, positional knowledge was somewhat higher than IPK. Flight dispatch students were more accurate on dispatch items ($z = .97$) than non-dispatch items ($z = -.81$), $t\ (7) = 4.49$, $p < .01$. Likewise, maintenance students were more accurate on maintenance items ($z = .88$) than non-maintenance items ($z = -.50$), $t\ (3) = 2.45$, $p < .05$, one-tailed. Pilots in training did not differ in accuracy between pilot and non-pilot items.

We also examined this hypothesis by examining quiz items related to a specific specialization and comparing the responses of students trained in that specialization with students trained in other specializations. For items related to flight dispatch, dispatch students displayed greater accuracy than students in other specializations, $F\ (1, 61) = 9.87$, $p = .003$. For maintenance items, maintenance students displayed a significant tendency for greater accuracy than students from other specializations, $F\ (1, 61) = 3.46$, $p = .034$, one-tailed. For items related to pilots, pilots in training were slightly more accurate than other specializations, but the difference was not statistically reliable, $F\ (1, 61) = 1.18$, $p = .28$. Both sets of analyses provide partial support for this hypothesis; positional knowledge was higher than IPK for both dispatch and maintenance students, but not for pilots in training.

Mental Models

We conducted a mixed factorial ANOVA for each specialization (pilot, flight dispatch, maintenance, aerospace administration) by response alternative for all 7 scenarios. For each scenario, a highly significant ($p < .001$) and strong ($\eta^2 > .45$) main effect for response alternative was observed. This indicates that each scenario contained response options that varied in perceived effectiveness. Only one of the scenarios (bird strike) yielded a significant specialization effect ($p < .05$, $\eta^2 = .15$). For this scenario (that required action by all three specializations), the overall level of effectiveness of the response options varied across specializations. For the other scenarios, effectiveness ratings did not differ across specializations. Based on these analyses, Intraclass Correlations (ICCs) were computed to determine if members of the same aerospace specialization held views about the effectiveness of various response options that were more similar than those of all aerospace specializations. Most of the ICCs were very small (< .044), but for one scenario (bird strike), there was a moderate tendency for members of the same specialization to show greater similarity (ICC = .14). This pattern of results suggests that, among traditionally trained aerospace students, there is not a large degree of differentiation between the mental models of students representing various specializations.

Next, we examined the similarity of evaluations of the effectiveness of response options. This provides evidence about the degree of mental model similarity of aerospace students in general and additional evidence about mental model similarity within specific specializations. For each scenario, we examined the consistency across participants of effectiveness ratings of the various response options. This involved computing the mean of the correlations between all possible pairs of participants. This provided an index of agreement about the relative effectiveness of the various response options to a scenario across all aerospace specializations. For each scenario, we also computed the mean correlation between participants within each specialization that would need to respond to the scenario. In four scenarios only pilots were critical, but for three scenarios multiple specializations were critical. For these scenarios, correlations were computed separately for each critical specialization and then a sample size-weighted average was computed. For all participants, across scenarios correlations ranged from .27 to .56 with a mean of .38. When only participants in critical specializations were examined, correlations ranged from .29 to .55 with a mean of .40. This pattern of results indicates that participants showed moderate agreement about the relative utility of various responses to the problems presented in the scenarios. That is, there is some degree of similarity among the task mental models of participants. Since the level of agreement between members of the same specialization did not differ appreciably from those of all participants, it appears that the mental models of members of the same specialization are not markedly more similar than the models of aerospace students in general.

Mental model accuracy was examined for each scenario. For each response option in a given scenario, each participant's effectiveness rating was compared to the previously established SME mean. For each scenario, these discrepancies were averaged across the response options to yield an error score for the scenario. These individual error scores were averaged across all participants. This procedure was repeated for each scenario. Lower error scores represent higher levels of mental model accuracy. The mean error scores ranged from 2.02 to 2.85 points on the 11-point effectiveness scale. Based on the effectiveness rating of the SMEs, the maximum average discrepancy for a typical scenario was 6.08. The mean error score for each scenario expressed as a percentage of possible error ranged from 25.7% to 42.4% and the mean error across all scenarios was 35.4%. This suggests that participants did not have extremely accurate mental models. Accuracy scores were examined using a specialization (4) by scenario (7) mixed factorial ANOVA. This analysis yielded only a main effect for specialization, $F(3, 59) = 6.31$, $p = .001$, $\eta^2 = .243$. Follow-up LSD tests indicated that maintenance students had higher levels of error ($M = 3.45$) than all other specializations (pilot = 2.28, flight dispatch = 2.65, administration = 2.41). Pilots in training had the lowest error scores on each of the seven scenarios, although these differences were only occasionally significant. Across scenarios, error scores for pilots in training were marginally lower than for flight dispatch students ($p = .082$). These results indicate that mental model accuracy varies across specializations. Maintenance students have less accurate task mental models than all other specializations and pilots in training seem to have the most accurate task mental models.

Discussion

Findings from analyses of IPK suggest that both flight dispatch and maintenance students have a greater knowledge of their respective specializations than of other specializations. However, pilots in training did not display greater knowledge of their specialization relative to knowledge of other specializations. While the findings for pilots in training are puzzling, findings for flight dispatch and maintenance students are consistent with our hypothesis concerning IPK.

Examination of the scenarios revealed that across disciplines, students showed some degree of mental model similarity. Contrary to expectations, mental models were not markedly more similar within disciplines than across disciplines. Comparison of student and SME responses to scenarios indicates that students showed only modest levels of mental model accuracy. Analyses suggest that mental model accuracy is highest for pilots in training. One potential explanation is that all scenarios involved situations requiring action from pilots. Maintenance students have the least accurate task mental models. This may partially reflect the fact that only two scenarios require action by maintenance personnel. However, the pattern of less accurate mental models for maintenance students was also found in the two scenarios that involve actions by maintenance. Thus, it seems that the maintenance function may be more isolated and maintenance students (and perhaps airline maintenance personnel) are less aware of the big picture of airline operations. Because some situations require close coordination between maintenance and other functions, this may be problematic.

This study has a number of limitations that should be addressed in subsequent studies. The sample involved all students in the capstone course across two semesters, but it contained only a small number of flight dispatch and maintenance students. Data collection over a longer period of time or across aerospace and aviation programs would provide for a larger sample and more stable baseline measures. Utilization of data from various programs would provide evidence concerning the extent of generalization of our current findings. The quiz could be expanded to include more items from each specialization and the set of scenarios could be expanded. Of particular importance, problem scenarios that do not require responses from pilots should be developed.

Despite these limitations, current findings provide evidence concerning the state of IPK, mental model similarity, and mental model accuracy among aerospace students. They suggest that traditionally trained students do not have extensive awareness of knowledge relevant to job demands of other aviation specializations. Results also suggest that task mental models of aerospace students are moderately similar, but the level of similarity is not much greater within specializations than across specializations. Finally, results indicate that aerospace students did not have highly accurate task mental models and that accuracy was lower for maintenance students than for students in other specializations. Because mental models facilitate coordination and effective performance, steps to increase IPK, mental model similarity and accuracy may enhance effective airline performance and safety. The results of this study can provide a baseline that can be used to evaluate the effectiveness of training programs designed to enhance coordination among aviation students.

References

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

DeChurch, L. A. & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology, 95*, 32-53. doi:10.1037/a0017328

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review, 26*, 356-376. doi:10.2307/259182

Mathieu, J. E., Hefner, T. S., Goodwin, G. F. Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*, 273-283. doi:10.I037t/0021-9010.85.2.273

Mathieu, J. E., Marks, M. A., & Zaccaro, S. J. (2002). Multiteam systems. In N. Anderson, D. S. Ones, H. K., Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology* (Vol. 2, pp. 289-313). Thousand Oaks, CA: Sage.

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

Resick, C. J., Dickson, M. W., Mitchelson, J. K., Allison, L. K., & Clark, M. A. (2010). Team composition, cognition, and effectiveness: Examining mental model similarity and accuracy. *Group Dynamics: Theory, Research, and Practice, 14*, 174-191. doi:10.1037/a0018444

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

Smith-Jentsch, K. A., Mathieu, J. E., & Kraiger, K. (2005). Investigating linear and interactive effects of shared mental models on safety and efficiency in a field setting. *Journal of Applied Psychology, 90*, 523-535. doi:10.1037/0021-9010.90.3.523

van Ginkel, W., Tindale, R. S., & van Knippenberg, D. (2009). Team reflexivity, development of shared task representations, and the use of distributed information in group decision making. *Group Dynamics: Theory, Research, and Practice, 13*, 265-280. doi:10.1037/a0016045

Acknowledgements

This research was supported by NASA Contract NNX09AAU52G, Paul Craig principal investigator. Correspondence concerning this article should be addressed to Glenn Littlepage, glittlepage@mtsu.edu.

USABILITY OF COCKPIT AUTOMATION IN MANAGING NORMAL AND ABNORMAL AIRCRAFT OPERATIONS

Ioana Koglbauer
Brightline Avionics
Kalsdorf, Austria

Reinhard Braunstingl
Institute of Mechanics, University of Technology
Graz, Austria

Klaus Fruehwirth
Brightline Avionics
Kalsdorf, Austria

Erich Grubmueller
Brightline Avionics
Kalsdorf, Austria

Siegfried Loesch
Institute of Mechanics, University of Technology
Graz, Austria

In the present study we assessed pilots` usability and acceptance of an advanced glass cockpit system during normal and abnormal operation. Two matched groups of pilots (female and male) attended a series of tasks which required the use of automated and manual cockpit functions. The cockpit was implemented in the research flight simulator at Graz University of Technology. Pilots` performance, psychophysiological state, workload and situation awareness were evaluated in a repeated measures design. Benefits and limitations of the automated and manual cockpit functions are discussed. Conclusions for the optimization of glass cockpit functions in normal and abnormal operation were drawn out in order to facilitate better and safer flight performance. Results were implemented to optimize a glass cockpit system developed by Brightline Avionics, which shall be installed in light aircraft.

According to the National Transportation Safety Board (2010) the introduction of glass cockpit avionics into light aircraft did not yet result in the expected increase of safety. Generally, pilots` attitudes towards different glass cockpit systems such as Primary Flight Display (PFD), Moving Map, Autopilot, GPS are positive, but there are also concerns about complexity and learnable quality of the systems, and about becoming over-reliant on them (Casner, 2008). As Fennell and Pruchnicki (2009) showed, there are concrete demands of the International Airline Pilots Association to be considered in the evaluation of new cockpit systems and of automation. Main requirements refer to human-centered-design of automatic functions, feedback and control of automation, as well as the evaluation of devices with pilots in normal and abnormal operations.

Focus of this research was the formative evaluation of a new glass cockpit system which provides electronic checklists for aircraft configuration in each phase of the flight, allowing also automatic configuring actions of the system. Results of this preliminary evaluation were used to optimize the system.

Method

Participants

Ten women pilots aged between 21 and 62 years (M = 41.90, SD = 14.54) and ten men pilots aged between 18 and 54 years (M= 37.20, SD = 12.50) were recruited from flight clubs and pilot associations from Austria, Germany and Switzerland. All participants were in possession of a current private pilot license, 4 pilots held an IFR rating and 2 pilots held an airline pilot license. Women and men pilots in the study were yoked according to their total flight time and ratings. As the medians showed, more than half of the pilots had little or no experience with glass cockpits and certified flight simulators. Mean flight experience of women pilots was 407 total flight hours (SD = 564.62), whereas that of men pilots was 469 total flight hours (SD = 853). All participants owned a computer and used it weekly for private purposes or for work. Differences between the groups in respect to total flight, glass cockpit, certified simulator and computer usage time calculated by means of Wilcoxon-Test did not reach statistical significance.

Procedure

After reading and signing an informed consent form that described the experiment as a cockpit evaluation and optimization study, each pilot received a written description, hands-on training in using the device and a briefing.

The test procedure consisted in configuration of an aircraft for all phases of flight, including a missed approach and a go-around. Therefore, the test sequence had a fixed order of configurations: before start, after start, taxi, engine run up, line-up, after takeoff, cruise, approach, cruise, approach, final, go-around, after takeoff, cruise, approach, final, after landing, and parking. For abnormal operation scenarios an alternator failure was induced by the experimenter at the time when pilots completed the first cruise configuration.

Aircraft configuration during normal operation, as well as aircraft configuration and management of the electrical system during alternator failure was performed in both manual and automatic mode, resulting in following blocks: manual-normal, manual-failure, automatic-normal, automatic-failure. The presentation orders were randomized across participants, but yoked across genders.

Equipment

The new glass cockpit system was implemented in a generic aircraft simulator at Graz University of Technology <www.flightsimulation.tugraz.at> . The simulation included a Primary Flight Display (PFD) and manual switches of devices (lights, flaps, trims, transponder, radio and so on). An instructor station was connected to the simulator allowing the experimenter to induce the alternator failure.

Dependent measures

For performance evaluations we adopted self-ratings of the pilots as a part of NASA-TLX, duration to complete the task and the number of omissions in configuring the aircraft. Workload was rated post-trial by the pilots using NASA-Task Load Index (Hart & Staveland, 1988) using a scale from –5 (very low) to +5 (very high). Mean heart period or inter-beat-interval (IBI) in milliseconds and the heart rate variability (HRV) as root mean square of successive differences (RMSSD) were recorded for each block using the heart monitor Polar RX800C. Physiological data of the pilots were standardized using values of rest measurements (Koglbauer, Kallus, Braunstingl, Wurmitzer & Boucsein, 2010). Situational awareness was evaluated using the SART 3D (Taylor, 1990) with ratings from –5 (low) to +5 (high). By means of the positive and negative affect scales of PANAS (Watson & Clark, 1988) pilots rated the intensity of their actual emotional state for each item using a five-point scale from 1 (very slightly or not at all) to 5 (extremely). A modified version of the usability evaluation form of Chin, Diehl and Norman (1988) was applied at the end of experiment. Acceptance of the device was assessed by means of an evaluation form which was adopted with some modifications from Jensen, Guilkey and Hunter (1988).

Results

Main effects of the control mode (automatic vs. manual) and the interactions between mode and operation were calculated using Greenhouse-Geisser tests. Data analysis adopted an α-level of .05 and results with statistical probabilities up to p =.10 are presented.

Analyses show a significant mode effect on task duration [F(1,18) = 10.88, p = .004], as pilots needed significantly more time to accomplish the task in manual (M = 441.37, SD = 22.58) than in automatic mode (M = 387.69, SD = 21.38). The number of omissions was higher in automatic (M = 2.45, SD= .89) than in manual mode (M = .12, SD = .08), differences reaching statistical significance [F(1,18) = 7.22, p = .01]. Self-ratings of performance did not differ significantly between the manual and automatic mode. However, pilots have spent more effort in manual (M = -.85, SD = .53) as compared to automatic mode (M = -1.47, SD = .50), differences being marginally significant [F(1,18) = 3.71, p = .07]. The demand on attentional resources was higher when using the manual (M = -1.76, SD = 1.22) than during the use of the automatic mode (M = -3.00, SD = 1.23), [F(1,17) = 3.11, p = .09]. Pilots` positive emotions were stronger when using the device in automatic (M = 3.72, SD = .13) than in manual mode (M = 3.62, SD = .15), differences reaching statistical significance [F(1,18) = 5.79, p = .027]. No significant main automation effects were found on the measures of mental, physical, temporal demand, negative emotion and pilots` cardiac activity.

Analyses of the interactions between mode and type of operation show significant interactions between the mode and type of operation on self-ratings of performance [F(1,18) = 5.37, p = .032], mental workload [F(1,18) = 4.39, p = .039] and situation awareness [F(1,18) = 4.59, p = .047]. As Table 1 shows, pilots` self-ratings of performance during normal operation were higher in manual than in automatic mode. However, during abnormal operation pilots` performance was better in automatic than in manual mode.

Mental workload of the pilots during normal operation was higher in automatic than in manual mode. Nevertheless, as illustrated in Figure 1, during abnormal operation mental workload was lower in automatic than in manual mode. Pilots rated their situation awareness higher in manual than in automatic mode during normal operation, but during abnormal operation they have rated their situation awareness lower in manual than in automatic mode (Figure 2).

Table 1.

Means and standard deviations of performance, workload and situation awareness ratings during different modes and types of operation.

Mode	Manual				Automatic			
Operation	Normal		Abnormal		Normal		Abnormal	
Measure	M	SD	M	SD	M	SD	M	SD
Performance	2.85	.39	.90	.62	1.90	.65	1.75	.57
Mental workload	-.90	.64	1.25	.60	.30	.68	.35	.67
Situation awareness	3.09	.38	1.83	.56	2.51	.43	2.28	.46

Note. Ratings on a scale from –5 (very low) to +5 (very high).

No significant interactions between mode and type of operation were found on the measures of physical, temporal demand, effort, emotion, pilots` cardiac activity, duration of the task and omissions of the pilots.

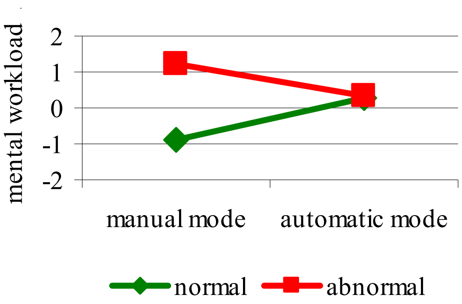


Figure 1. Pilots` ratings of mental workload when using the device in different modes and types of operation (ratings from –5 very low to +5 very high).

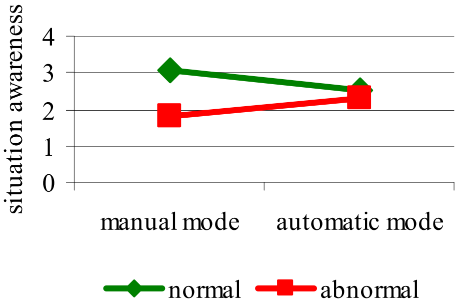


Figure 2. Pilots` ratings of situational awareness when using the device in different modes and types of operation (ratings from –5 very low to +5 very high).

Descriptive analyses of usability data included pilots` ratings on semantic differential scales ranging from -5 to 5 (i.e.: from -5 confusing to +5 easily understood). As the medians (Mdn) show, the majority of pilots rated high the easiness of their understanding of display elements indicating automatic (M = 3.55, SD = 1.66, Mdn = 4) and manual functions (M = 3.85, SD = 1.22, Mdn = 4). Pilots gave also good ratings for the usefulness of automation in configuring the aircraft during normal (M = 3.35, SD = 2.23, Mdn = 4.5) and abnormal operation (M = 3.40, SD = 2.03, Mdn = 4). More than half of the pilots reported a good understanding of both, manual (M = 4.90, SD = .30, Mdn = 5) and automatic modes of control (M = 4.80, SD = .69, Mdn = 5), and of the manual override button (M = 4.55, SD = .88, Mdn = 5). They also rated in average high their awareness concerning the status of the device (M = 3.40, SD

=2.16, Mdn = 4) and knowing what the device was doing (M = 3.20, SD =2.28, Mdn = 4). The majority of pilots rated the reliability of the device extremely high (M = 4.15, SD = 1.26, Mdn = 5).

In respect to acceptance of the device, pilots reported that the device was extremely helpful (35%) and helpful (45%), whereas 15% found the device neutral or not helpful (5%). The majority of pilots reported their definite intention to use the device in the future (65%), 25% might use it, 5% were not sure and 5% reported not to intend using it. With one exception, all pilots reported that they would recommend other pilots to use the device (95%). The great majority of pilots (95%) considered that the device was improving their own personal safety during flight. However, there were pilots (5%) who considered that the device would jeopardize their safety during flight because they might become too reliant on it.

Discussion

Main focus of this research was pilots` evaluation of a new device to be integrated in the glass cockpit of General Aviation (GA) aircraft, which makes possible the use of automation in monitoring and configuring the aircraft for different flight phases.

At first glance the results show significant benefits of automation to reduce the task duration as well as the effort and the demand on attentional resources in pilots. Furthermore, the pilots experienced stronger positive emotion when using automation. However, the number of omissions was higher in automatic than in manual mode pointing to a sub-optimal use of automation.

System evaluations during normal and abnormal operation show further limitations of automation. The use of automation during normal operation was associated with lower performance, lower situation awareness and an increased mental workload in pilots as compared the manual mode. Interestingly, the use of automation during abnormal operation resulted in better performance, better situational awareness and reduced mental workload in pilots as compared to the manual mode of operation. During abnormal operation in manual mode can be noticed a pattern of considerable decrement in performance and situational awareness, associated with an increase in mental workload. A contrary pattern was found when pilots used the automatic mode showing that their levels of performance and situational awareness did not considerably decrease during abnormal operation, but they were below the levels attained in manual mode during normal operation.

The majority of pilots understood easily the automation, the displayed information related to automation and the possibility to control automation using the manual override function. The majority of pilots rated very high the reliability of automation and the feedback provided by the device. Generally, pilots found useful the support of automation in configuring the aircraft for different flight phases during normal and abnormal operation. However, some pilots were concerned about becoming too reliant on the new device and on automation. These findings confirm the results of Casner (2008) who also identified such concerns in pilots.

Results of this research found direct application in optimization of the new glass cockpit system. On one side, the system was improved to provide pilots with more support in managing abnormal operations. On the other side, the automation was optimized to buffer pilots` omissions when using it. Monitoring automation remains an additional task of pilots which certainly increases mental demand of the normal flight operation. However, when pilots are current in using it, automation is a tool that can balance the effort of managing abnormal procedures during flight. This is especially important for GA pilots who merely fly alone and cannot share their tasks with a co-pilot.

Despite the need for further optimization and development of the new glass cockpit system which was detected at this early stage of development, the great majority of pilots considered the device helpful, contributing to improve their own personal safety and they expressed their definite intention to use the device in the future and to recommend it to other pilots too.

Acknowledgement

This research was supported by the Austrian Federal Ministry for Transport, Innovation and Technology and the FEMtech Program (Project No. 561022).

References

Casner, S. (2008). General aviation pilots` attitudes toward advanced cockpit systems. *International Journal of Applied Aviation Studies*, Vol. 8, No. 1, 88-112.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of SIGCHI* , 213-218.

Fennell, K., & Pruchnicki, S. (2009). Aviation automation design and implementation – the need for human factors considerations. In *Proceeding of the International Symposium on Aviation Psychology*, Dayton, Ohio, 299-304.

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,* Amsterdam: North Holland Press, 239-250.

Jeannott, E., Kelly, C., & Thompson, D. (2003). The development of situation awareness measures in ATM systems. *EATMP Report*. Eurocontrol HRS/HSP-005-REP-01.

Jensen, R. S., Guilkey, J. E., Hunter, D. R. (1988). *An evaluation of pilot acceptance of the personal minimums training program for risk management*. Federal Aviation Administration, DOT/FAA/AM-98-6.

Koglbauer, I., Kallus, K. W., Braunstingl, R., Wurmitzer, K., & Boucsein, W. (2010). Evaluation of pilots` performance and psychophysiological state in managing unusual attitudes, full stalls and spins during simulator training and real flight tests. In *Proceedings of the 29th Conference of the European Association of Aviation Psychology*, Budapest, September, 339-344.

National Transportation Safety Board (2010). *Introduction of glass cockpit avionics into light aircraft.* Safety Study NTSB/SS-10/01. Washington, DC.

Taylor, R. M. (1990). Situational awareness rating technique (SART): the development of a tool for aircrew system design. *Situational Awareness in Aerospace Operations*, (AGARD-CP-478), Neuilly Sur Seine, France: NATO-AGARD, 3/1-3/17.

Watson, D., & Clark, L. A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070.

WHY AIR TRAFFIC CONTROLLERS ACCEPT OR REFUSE AUTOMATED TECHNOLOGY

Marek Bekier
Brett R. C Molesworth
Ann M. Williamson
Department of Aviation, University of New South Wales, Sydney, Australia

Increased utilisation of automation is seen as a potential, if not the most likely solution to cope with the forecasted increase in air traffic (SESAR, 2006; FAA, 2010). However, Air Traffic Controller Operators (ATCOs) are very selective about forms of automated assistance (EUROCONTROL, 2000). Automation acceptance is considered crucial for the successful implementation of any new technology within air traffic management and therefore is one of the largest challenges the industry faces (Hilburn & Flynn, 2001). Since traditional predictors of automation acceptance such as trust and job satisfaction appear to be diminishing (Bekier, Molesworth & Williamson, in press), the main aim of the present research was to identify the factors that help to explain ATCOs willingness to accept more mature forms of automated assistance. The results revealed that ATCOs value automation that is user-friendly, removes 'boring' and 'standardized' tasks, and importantly keeps them cognitively challenged in their role.

Air service providers are under pressure to modernisation their Air Traffic Management (ATM) systems in order to deliver the required capacity gains for the forecasted increase in air traffic movements. While it has been acknowledged that this is long overdue (Metzger, 2001), how exactly it is to occur remains a contentious issue. According to Parasuraman and Rovira (2010), one solution is to reduce the size of the traffic sectors controllers are required to manage. However this has a potential negative flow-on effect for the controller in terms of understanding the scheme of traffic and its relationship to the adjoining sectors. Metzger and Parasuraman (2005) propose an alternate which involves sharing the decision-making responsibility between pilots and air traffic controllers. In contrast, Odoni (1999) believes a congestion-based pricing would be the most effective means of delivering the required capacity gains, through a better distribution of the traffic movements. Hilburn (1996) sees technology in the form of automation as a solution to the problem. Specifically, Hilburn contends that efficiency gains are best achieved through increasing the role of automation within the system. It is the latter of these suggestions that appears to have received the most attention (Kirwan, 2001; Agogino & Tumer, 2009), simply because it is viewed that automation currently employed in many Air Traffic Management (ATM) systems is underutilized.

The utilization of automation in many present ATM systems focuses largely on low level cognitive tasks such as data gathering and storage, data compilation, the computation and presentation of summaries of data, the retrieval and updating of data, as well as data synthesis (Hopkin, 1999). However, it is envisaged with the proposed increase in automation utilization (SESAR, 2006; FAA, 2010) tasks other than those involving low level cognition will be considered (e.g., decision-making). Key to the success of these changes is air traffic controllers' acceptance or willingness to use any increase in automated technology.

According to EUROCONTROL, (2000) air traffic controllers as a whole are very selective about forms of computer assistance. History is littered with examples where users' reluctance to embrace new technological advances have seen the implementation of a new system or tool fail. In a simulated study with unmanned air vehicles and various levels of automated decision-aids, Ruff and colleagues found that as the accuracy of the decision-aid decreased and users detected the errors produced, acceptance of the system, in terms of trust dramatically decreased (Ruff, Narayanan, & Draper, 2002). This study also found that as the level of automation increased, performance deteriorated. Similar results have been identified elsewhere (Wickens, Mavor, & McGee, 1997;

Wiener, 1988). For designers of automation, these results are the opposite of what they intended to achieve. According to Ruff et al., (2002) automation does more than replace the task performed by the human; it changes the operational function requirement and as a result often imposes a greater level of demand on the operator. For the end user, this adds to their level of scepticism regarding any new implementation of technology. Nonetheless, since increased utilisation of automation is seen as a viable option to facilitate the forecasted increase of traffic within ATM, the present study sought to determine what factors predict user acceptance within this domain.

Previous research has demonstrated that trust in automation (automated system continually performing its duties; Muir & Moray, 1996; Lee & See, 2003) and job satisfaction – both past and present (Hopkins, 1991; Lee, Rhee, & Dunham, 2009) are significant predictors of automation acceptance. Within air traffic management, Bekier and colleagues have found similar results, in addition to a number of other factors such as age and automation experience (Bekier, Molesworth & Williamson, in press). However, the variance in automation acceptance accounted for by these variables was very low. Therefore, future work is needed to examine other factors that predict automation acceptance. Hence, the main aim of the present study was to investigate the predictors of automation acceptance with air traffic controller operators. Specifically, the research sought to answer the following question.

Research Question

If the traditional predictors of automation acceptance such as trust and job satisfaction explain only a small portion of air traffic controllers' willingness to accept/use automated technology, what other factors account for the variation across controllers?

Method

Participants

20 (16 males) professional air traffic control operators from one centrally located air navigation service provider volunteered for the research. The mean age of participants was 37.70 (SD = 8.36) years. On average the participants had been working as an air traffic controller for 15.4 (SD = 8.07) years. All procedures in this study were approved in advance by the University of New South Wales ethics panel.

Apparatus and Stimuli

Two questionnaires comprised the material, namely a demographics question (e.g., age, gender, experience) and an air traffic control questionnaire. The air traffic control questionnaire consisted of six 'open-ended' questions. The questionnaire was divided into three parts. The first part contained two questions which were specifically designed to examine the tasks and/or components of an ATCOs' role they considered to be most motivating and important in their daily ATC work. Specifically question one read '*Which tasks or aspect of your daily ATC work are the most motivating for you? Give examples of these which you rate positively and negatively*'. Question two read *'what would you describe as the most important task in your role as an ATCO and why?'*

The second part of the questionnaire was designed to determine which existing ATM tools are presently used and ATCOs' view on the characteristics of these tools, both positive and negative. Specifically question three read '*Of all the support tools at your disposal, which features do you rate the most positive and why?'* Question four was similar to question three except it was concerned with which features ATCOs did not like. It read '*Of all the support tools at your disposal, which features do you rate the most negative and why?'*

The third part of the questionnaire was designed to elicit the characteristics of an ideal futuristic support tool, as seen by the ATCO and to obtain insight into the views of ATCOs about the limitations and capabilities of ATM automation. Hence, question five read *'If an automated system was to be introduced in your role, hence to assist you – what design features would you insist on?'*

The final question (question six) read ‘*Do you believe that the core ATC tasks (as stated in response to question 2) can today be performed by automation in a way that is superior to human performance?’*

A tape recorder (Sony IC Recorder, ICD-8500) was used to record all interviews.

Procedure

One central European based air navigation service provider was selected for the research primarily because of the automation-supported environment of its operations and the availability of different ATCO groups (ACC, APP/DEP, TWR) on location. An email providing background information of the research was sent to the workforce of the air navigation service provider seeking voluntary participation in the research. Potential participants expressed interest in the research through email where they were provided additional information in the form of an information sheet. Upon confirmation to assist with the research, a mutually suitable time was arranged to conduct the research/interview. All the interviews were recorded and on average each interview took 30 minutes. At the conclusion of the research participants were thanked for their time.

Results

Data Reduction

In order to analyze participants’ responses to Part 2 and 3 of the questionnaire, the data were grouped into common categories. In total, three categories were formed and were titled: User-Friendliness, Functionality and Quality. These categories were created based on a word or phrase used by participants, which were attributed to the object under examination (i.e., adjectives). Definitions for the following categories were:

- User-friendliness - refers to the design of the machine/user interface and its impact on human performance (e.g. ergonomic, easy to use, and well arranged).
- Functionality – refers to the quality of being suited to serve a purpose (e.g. supportive, facilitating work and useless).
- Quality – refers to the degree of excellence including output of the software and future use (e.g. robust, missing accuracy and producing false alarms).

Following this and to ensure reliability of coding, two industry specialists were asked to code the data independently. In total, 52 independent words or phrases were identified and coded. Agreement rate amongst the three raters was 96%.

Questionnaire

The responses to the six questions were summarized and are described below. Remember the format of the study involved open-ended questions with the objective of obtaining as much information as possible from the viewpoint of the ATCO. This often resulted in the ATCOs providing more than one response to each question. As a result, the below summary includes all responses presented as a percentage. Total responses to each question are reported at the conclusion of each summary.

Part 1 – Motivating in present role. In response to the first question, the majority of the ATCOs mentioned at least one task or component/function of their role which they rated positively in their daily ATC work. In this context, radar and related work was reported to be the most motivating aspect of the daily work (38.5%), with ‘radar work’ cited by 23.1% and ‘conflict detection and resolution’ cited by (15.4%). The second most frequently cited term was ‘challenge’ (23.1%), followed by ‘radio contact with pilots’ or ‘interaction with pilots’ (10.3%). Further mentioned terms were ‘teamwork’ (7.7%), ‘efficiency’ (7.7%), ‘customer service’ (5.1%) and ‘all task’ (5.1%). Safety as a holist construct was only reported once (2.5%). Total stated 39.

The most frequently cited negative aspect of ATM was ‘no traffic’ (37.5%). This was closely followed by ‘nil’ or no aspect is negative (25%). Further mentioned tasks/aspects included ‘regulations’ or ‘restrictions’ (16.7%), ‘monotonous and repetitive tasks’ (8.3%), ‘unnecessary and diverting tasks’ (8.3%) as well as ‘cooperation with military’ (4.2%). Total stated 24.

When asked what would you describe as the most important task/component in the role as an ATCO (question 2), the term ‘separation/safety’ featured highest (65.5%). This was followed by ‘efficiency’ (26.9%), ‘customer service’ (3.8%) and ‘to stay calm’ (3.8%). Total stated 26.

It appears that ATCO find their daily tasks to be motivating, particularly the ATC core tasks such as conflict detection and solution or the radio communication with the pilots. Boredom and monotony on the other hand are very much disliked. Not surprisingly safety and efficiency are seen as the most important ATC functions.

Part 2 – Characteristics of existing tools. Based on the grouping of participants’ responses into the three categories (e.g., user-friendliness, functionality and quality), the majority (58.1%) of responses related to the positive experiences with automation based on the ‘functionality’ of the automation (e.g., practical, supportive) followed by its ‘user-friendliness’ (37.2%; e.g., usable, ergonomic). The remaining (4.7%) responses related to ‘quality’ aspects of the automation such as reliable and fast. Total stated 43.

When asked about the negative features of support tools at their disposal, ‘user-friendliness’ rated highest (45.1%) followed by ‘functionality’ (35.5%) and ‘quality’ (19.4%). Total stated 31.

Part 3 – Ideal futuristic tool. In relation to participants’ responses to the hypothetical question regarding the design features they would insist on with a new automated system, 24 terms were used. Interestingly 52.9% of these related to ‘user-friendliness’ such as ‘easy to use’ or ‘ergonomic’, followed equally by ‘quality’ (25.5%) and ‘functionality’ (21.5%). Total stated 51.

Finally, ATCOs were provided the opportunity to reflect on whether advanced technology could perform the core tasks they personally indentified in the role. The participants responded unanimously on this question with all answering in the negative. In other words, 100% of participants felt that the core ATC tasks could not be performed by automation in a way superior to the human. The limitations to this question are discussed below.

Discussion

Increased utilization of automation within the ATM system is seen as a real and viable solution to the forecasted increase in air traffic movements (FAA, 2010; SESAR 2006). Crucial to the success of such a change is operators’ acceptance of the automation. Since traditional predictors of automation acceptance such as trust and job satisfaction explain only a small portion of user acceptance (Bekier et al., in press), the aim of the present study was to investigate the main drivers that may better account for this acceptance. The results from part 1 of the study revealed that the core-task of conflict detection and resolution is liked and motivating for ATCOs. Similarly, it was clear that ATCOs dislike the monotony aspects of their tasks. This was most evident by the high citations rate of terms such as ‘no traffic’, ‘boredom’ and ‘monotonous and repetitive tasks’. It is noteworthy that a quarter of all the participants were unable to report a ‘negative (most)’ aspect of their work, which can be a good indication that overall ATCOs enjoy their daily tasks, as they exist today. By extension, this may translate into a reluctance to support fundamental changes such as an increase in automation in their work environment. This conclusion is supported by existing literature (Hopkin, 1995; Eurocontrol, 2000) that describes ATCOs as ‘conservative’ and ‘selective’ about forms of computer assistance. In the context of automation implementation, these results suggest that from an end users’ (ATCO) point of view it is probably easier to automate non-core tasks such as ‘flight data management and standardized coordination’s’ and/or ‘hand-offs’ rather than core tasks such as ‘conflict resolution’, since it is precisely these tasks that ATCOs find most motivating and key drivers.

The results from part 2 of the questionnaire that was concerned with the benefits and shortcomings of existing automation and the expectations towards future automation revealed a consistent theme across all three questions. Specifically, 'user-friendliness'- referring to the design of the machine/user interface and its impact on the human performance featured highly in participants' response in all questions (37.2%, 45.1% and 52.9% respectively). A similar, although not as prominent trend was evident (58.1%, 35.5% and 21.5% respectively) with 'functionality' - quality of being suited to serve a purpose. In contrast, quality – the degree of excellence including output of the software and future use featured the least across two questions (4.7%, 19.4% and 25.5% respectively).

When asked about the positive aspects of today's automation, 'functionality' featured highest (58.1%) followed by 'user-friendliness' (37.2%) and 'quality' (4.7%). However, when asked why automation is not liked, the distinction between 'functionality' and 'quality' is less, 35.5% and 19.4% respectively, while 'user-friendliness' increased to 45.1%. When questioned about expectations towards future ATM automation, both 'functionality' and 'quality' were of almost equal importance, while 'user-friendliness remained relatively the same. In other words, existing ATM automation is 'liked' because it is intuitive to use and suited to the purpose. In this sense, it appears that software stability, reliability and accuracy are almost taken for granted. However, when existing automation is not liked, it is predominately due to its lack of lack of user-friendliness (45.1%) and to a lesser degree because of its functionality or quality. This distribution is similar with future automation that is expected to be user-friendly and to a lesser but equally similar degree functional and quality. To put it simple, ATCO enjoy automation because it supports them to do their task and is easy to use, hence allowing them to focus on their core task/s. However, where automation is not liked it is mainly because it is not user-friendly and the handling of the automation is distractive from executing their core tasks.

Limitations and Future Research

The results from the present study need to be interpreted within the context of the research. Specifically, the present study surveyed 20 ATCOs. Due to this relatively small number of participants, caution should be taken to ensure the appropriate weight is applied to the percentages discussed. A caveat should also be noted with the sample. Specifically, participants were all recruited from one air navigation provider, albeit a large and prominent provider. Finally, it could be argued that the results from question six are somewhat expected due to the leading nature of the question. This being the case, future research should focus on investigating if this scepticism surrounding automated technology performing controllers' core task is as prominent as reported in this study.

Conclusion

In summary, the results from the present study suggest that if air navigation service providers are planning to increase the use of automation within the ATM system, they need to choose carefully the tasks they elect to automate if they wish to obtain user support. In addition, they need to ensure that the automation is well designed from a usability perspective as well from a reliability perspective, although the latter appears to occur already. Importantly, any new automation should ensure the ATCOs are cognitively engaged in the task, hence sufficiently challenged in their role. Most importantly, ATCOs would like to see new automation remove the 'boring' and 'standardized' tasks in their role, thereby allowing them to focus on the core task of radar work namely conflict detection and resolution.

References

Agogino, A., & Tumer, K. (2009). *Improving air traffic management through agent suggestions*. Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (pp. 1271-1272). Budapest, Hungary: International Foundation for Autonomous Agents and Multiagent Systems.

Bekier, M., Molesworth, B. R. C., & Williamson, A. (in press). Defining the Drivers for Accepting Decision-Making Automation in Air Traffic Management. *Ergonomics.*

EUROCONTROL. (2000). *ATCO attitudes towards future automation concepts: A literature review*. Brussels, Belgium: Author.

FAA. (2010). *The National Aviation Research Plan 2010*. Washington, DC: Federal Aviation Administration Headquarters

Hilburn, B. G. (1996). *Strategic Decision Aiding: A Comprehensive Bio-Behavioral Assessment of Human Interaction with Real - Time Decision Aids*. Brussels: North Atlantic Treaty Organization - NATO.

Hilburn, B., & Flynn, M, (2001). *Air traffic controller and management attitude towards automation: an empirical investigation*. Paper presented at the 4th USA/Europe Air Traffic Management Seminar, (pp. 1-7). Santa Fe, NM: USA. USA/Europe Air Traffic Management.

Hopkin, V. D. (1991). The Impact of Automation on Air Traffic Control Systems. In J. A. Wise (Ed.), *Automation and Systems Issues in Air Traffic Control*. Berlin: Springerverlag.

Hopkin, V. D. (1995). *Human Factors in Air Traffic Control*. London, England: Taylor & Francis.

Hopkin, V. D. (1999). Air Traffic Control Automation. In D. J. Garland, J. A. Wise & V. D. Hopkin (Eds.), *Handbook of Aviation Human Factors* (pp. 497 - 517). Mahwah, NJ: Lawrence Erlbaum Associates.

Kirwan, B. (2001). The role of the controller in the accelerating industry of air traffic management. *Safety Science, 37*, 151-185.

Lee, J. D., & See, K. A. (2003). Trust in Automation: Designing for appropriate reliance. *Human Factors, 46*, 50-80.

Lee, D., Rhee, Y., & Dunham, R. B. (2009). The role of organizational and individual characteristics in technology acceptance. *International Journal of Human-Computer Interaction, 25*(7), 623-646.

Metzger, U. (2001). *Automated decision aids in future air traffic management: Human performance and mental workload.* Technische Universität Darmstadt, Darmstadt

Metzger, U., & Parasuraman, R. (2005). Automation in future air traffic management: An empirical study of active vs. passive monitoring. *Human Factors, 47*, 35-49.

Muir, B. M., & 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., & Rovira, E. (2010). Transitioning to Future Air Traffic Management: Effects of Imperfect Automation on Controller Attention and Performance. *Human Factors, 52,* 411-425.

Ruff, H. A., Narayanan, S., & Draper, M. H. (2002). Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. *Presence, 11*(4), 335-351.

SESAR. (2006). *SESAR Definition Phase: Air Transport Framework - The Current Situation*. SESAR Consortium for the SESAR Definition Phase, Brussels, Belgium: Author.

Wickens, C. D., Mavor, A., & McGee, J. M. (1997). *Flight to the future: Human factors in air traffic control*. Washington, DC: National Academy Press.

Wiener, E. L. (1988). Cockpit automation. In E. L Weiner & D. C. Nagel (Eds.), *Human Factors in Aviation* (pp. 433-461.). San Diego, CA: Academic Press.

Automation for Human-Robotic Interaction: Modeling and Predicting Operator Performance

Angelia Sebok, Christopher Wickens, Marc Gacy
Alion Science & Technology, MA&D Operation
4949 Pearl East Circle, Suite 200
Boulder, CO 80301
www.alionscience.com

Human-robotic interaction presents numerous challenges to designers and operators. One way to address these challenges is through task automation. However, appropriate application of automation, reducing workload while keeping the operator informed and in control, without causing skill degradation, is not generally understood. In this paper, we describe a human performance modeling and simulation approach to evaluating the effects of automation on operator and system performance. In this research, we identify and combine relevant factors that affect operator performance into operator-robotic system interaction models. The result of this project will be a partially-validated tool to help system designers evaluate potential automation strategies for their expected effects on operator and system performance.

This paper describes a unique research and development project to address human-automation interaction in robotic missions. In this project, we are developing the Function Allocation Simulation Tool (FAST) to help mission planners, automation designers and human performance researchers at NASA Johnson Space Center (JSC) evaluate human-automation interaction in space robotics missions. On the International Space Station (ISS), robotic equipment provides a primary means for performing repairs, and conducting docking, assembly, and maintenance tasks in the hostile environment of space. These robotic tasks offer multiple challenges. In particular, mismatches between the operator's viewpoint and the direction of movement of the robotic arm, and between the direction of control movement and corresponding arm movement adversely impact performance. Further, a limited number of cameras, few lighting options, and multiple potential collision surfaces add to the complexity of the task.

One potential solution for helping astronauts perform these missions is to automate the robotic tasks. Giving control to the automation could alleviate some of the operator's workload, but automation includes its own hazards. In particular, unreliable automation, or even highly-reliable automation that unexpectedly fails, can result in worse performance than continuous manual control (Wickens et al, 2010; Bainbridge, 1983; Parasuraman & Riley, 1997; Endsley & Kiris, 1995). As one example, if the operator expects that collision avoidance automation will prevent any impacts, s/he can become complacent, fail to diligently monitor the arm's position with respect to collision surfaces, lose situation awareness, and be surprised when the arm collides with a structure.

Other challenges with human-automation interaction include specifying how automation is to be implemented and the ways in which automation failures occur. Automation can be implemented so it takes over the "easy to automate" tasks. In these situations, it is typically the manual control tasks that get automated. This can reduce operator workload, but leave the operator removed from the loop. This can contribute to degraded operator situation awareness, and eventually cause the operator to lose their direct control proficiency.

Given the high-stakes missions in space, where collisions can damage expensive equipment, compromise mission completion, or potentially put astronauts' lives in jeopardy, it is imperative that human-automation allocations are carefully evaluated prior to implementation. Our approach to this work is to build human performance models of operators performing typical robotic missions, and have these models interact with an actual robotic simulation, to make predictions of human and system performance in different conditions. The tool we are developing, FAST, will allow planners, researchers, and designers to evaluate potential automation strategies, and identify their predicted effects on human and system performance, before implementing them. It will provide the opportunity to compare the effects of different types of automation by testing them (i.e., running simulations to gather data on predicted performance), evaluate the effects of sub-optimal reliability, and evaluate different types of automation failures.

The modeling and simulation (M&S) approach we describe in this paper has been used successfully in many applications (e.g., Allender, 2000; Foyle & Hooey, 2007). Modeling approaches offer the significant advantages of providing operator and system performance data without requiring human-in-the-loop research. The time and expense of obtaining institutional review board (IRB) approval for research with human subjects, planning and conducting studies, and the common problem of only being able to evaluate a few, limited scenarios is avoided. Models that represent situations being evaluated can be easily modified, allowing analysts to evaluate a wide variety of situations. Performance data can be gathered with a few computer key presses, rather than multiple experimental scenarios.

Human Automation Interaction in Space Robotic Missions

One of our first tasks in this work was to explore the domain of space robotics and identify the relevant human-automation interaction issues. We obtained a NASA JSC robotic simulation to include in the tool and attended a week-long Generic Robotics Training course offered at JSC. Based on the course, interviews with NASA robotics system instructors and a former astronaut, and findings from readings of the space human factors literature (e.g., Cizaire, 2007; Kanas & Manzey, 2008), we identified a set of potential automation concerns to include in the tool. The starting point for our model development was grounded in the stages-levels view of automation (Parasuraman, Sheridan & Wickens, 2000). This model defines stages of automation that correspond to different roles in an information processing / decision making / action framework. In this model, automation can support humans by: (1) gathering information and presenting it, (2) integrating information in such a way as to improve operator understanding, (3) supporting decision making by presenting options, and (4) implementing actions. Within each stage of automation, there are different levels, where the degree of automation increases. The following table shows how we characterize the robotic domain in terms of the stages-levels framework.

Table 1

Examples of Stages and Levels of Automation in Space Robotic Tasks

	Stage			
Level	***Information Acquisition***	***Information Analysis / Integration***	***Choosing / Deciding***	***Executing***
High	Automation highlights the camera view it infers is most valuable.	The automation diagnoses which control axis has excessive deflection	Automation assigns the best 3 views to the 3 viewports (but allows human to override)	Automatic control of XYZ trajectories, to points that are human designated
Inter-mediate	Highlights joints approaching singularities, potential collisions	Auto diagnoses collision state based on trajectory extrapolation	Automation recommends 3 camera views	Manual control of XYZ trajectories but automation control of joint angles
Low	Presents all raw data	None	Automation recommends a set of camera views to choose between	Manual control of joint angles

Following the task of identifying stages and levels, we identified a variety of specific types of automation, relevant to the robotic domain. These include trajectory control, camera control, lighting control, hazard alerting, and rate control. Within these different types of automation, there are different ways in which the automation can fail. Automation can simply fail to alert users, and lead them into an unsafe condition they were expecting to avoid. Automation can provide false alarms, alerting the operator to a hazard when none exists. Further, automation can recommend a suboptimal course of action; in the worst cases, this recommended course of action can be more risky than the one currently being performed by the operator. In the executing stage, automation can fail to implement an action correctly; it could choose and execute an incorrect trajectory for example.

Our current challenges include: determining the breadth of possible changes, identifying those most relevant to NASA, and focusing our research on a highly-relevant subset of those factors. Our main concern is to specify the automation conditions that can affect operator and system performance, and model these appropriately.

The goal of the project is to produce a Function Allocation Simulation Tool (FAST) that allows users at NASA JSC to evaluate predicted operator and system performance in robotic tasks under a variety of different automation situations. Thus, we have been actively identifying and specifying what types of automation are reasonable to consider, how these might fail, and what types of adaptive automation strategies might be included. These are all factors we will include in our software tool and in the operator models.

Tool Development

Software Development: FAST

FAST is a wrapper-based tool that includes a NASA JSC robotic simulation (The Basic Operational Robotic Instructional System, BORIS), a computational model of a human controller, known as MORRIS, and a user interface through which different scenarios can be created and evaluated. The MORRIS operator provides inputs to the BORIS simulation, and BORIS responses provide inputs to the operator model. The user interacts with the tool by creating and evaluating specific scenarios. Figure 1 shows a concept display for the data-entry screen. The user specifies what mission is being performed (e.g., moving to a destination, grappling an object, or performing an extra vehicular activity), what systems will be automated, how automation will be implemented, e.g., if trajectory control is performed manually by the operator using hand controllers or if the operator merely specifies the destination and the automation moves the robotic arm to that location and orientation. The FAST user also specifies if the automation is progressively adaptive, and if the automation is unreliable. Unreliable automation is further specified in terms of degree of reliability and type of failure. FAST users also specify aspects of the environment in which the task is performed, by placing obstructions (i.e., tables and no fly zones) in the area. Finally, users select the simulated operator's level of experience (i.e., novice, experienced, or expert). The tool then provides predictions of operator and system performance. FAST will allow users to evaluate and compare predicted performance across different automation conditions, to identify the best operating situation for the particular mission. Users have the flexibility to define the "best situation" according the parameters of most importance, e.g., minimal time to complete the task, most reliable performance, and/or lowest operator workload.

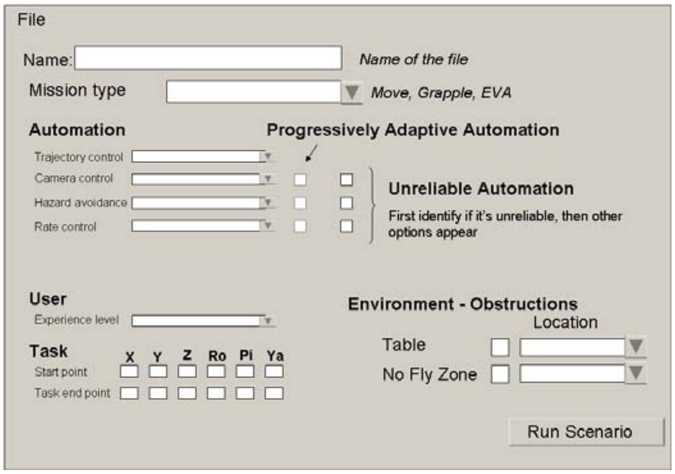


Figure 1. Concept of a FAST data entry screen, through which the user specifies the scenario.

Once users define a scenario, the tool sends parameters to both BORIS and MORRIS to configure the simulation and select the appropriate mission model, and set the human performance modeling parameters. By hitting Run the user initiates the simulation. The visualization of a run shows the following information (Figure 2 below): the BORIS screens of the robotic simulation (on the left) and the graphical user interface (on the right), the name of the scenario and time into the run, the operator performing the task and his SEEV-predicted viewcone (the area where the eyes are looking at any given time), and his/her predicted situation awareness and workload, and major events as identified in the interaction between the operator model and robotic simulation.

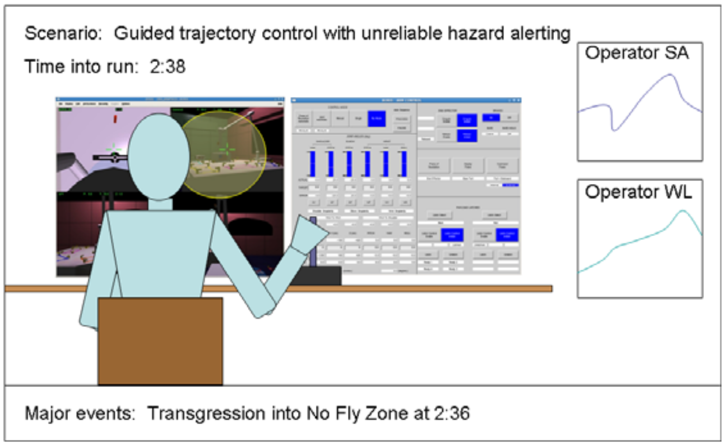


Figure 2. Sample display of the FAST scenario visualization capability, showing predicted operator performance.

BORIS Robotic Simulation

The Basic Operational Robotic Instructional System (BORIS) is a simulation environment consisting of a six-degree of freedom (DoF) generic robot arm in a simulated room with tables and payload latch-points. BORIS is the primary instructional aide in the General Robotics Training (GRT) program at NASA JSC, used for training general robotic arm control concepts and camera manipulations. BORIS provides a generic environment in which to encounter and practice many of the issues with robotic arm control. The BORIS training environment simulates a 15m x 30m x 15m room with a six DoF arm attached to one of two wall mounts. The room includes grids and distinguishing features on the walls. BORIS offers the capability to insert a large table, either for payload placement or as an obstacle, into the room (see Figure 3). Finally, BORIS has seven camera positions: a “window” view, four cameras mounted in the room corners, one at the end of the robot arm, and one mounted on the arm joint. The operator controls the arm through two hand controllers and receives feedback on arm position and certain features such as self-collisions and arm singularities (positions where automated arm movement fails) on a set of displays. In our simulation, a virtual operator will be controlling the arm, giving commands and receiving the update information.

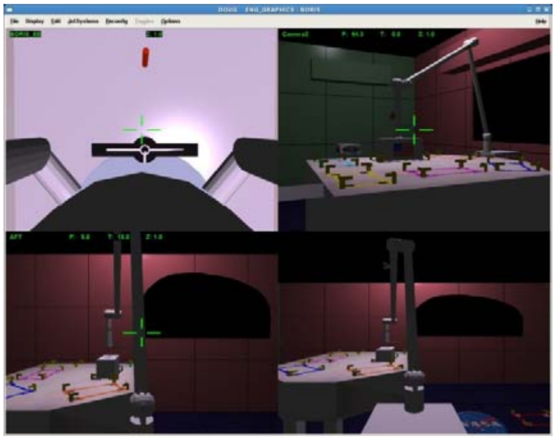


Figure 3. The BORIS training simulation, showing the robotic arm and operating environment

Operator Model Development

MORRIS: the computational model of the robotic operator

The goal of MORRIS was to create a software model of a robotics controller, working with BORIS, in a manner that generates the same sorts of errors, makes the same kinds of decisions, and experiences the same sorts of workload and situation awareness profiles (across a mission) as would the actual human controller. Initially, the model will be validated against human performance data from live operators controlling the equivalent systems. Then, to the extent that MORRIS is valid, it can be used to predict the consequences of different design, automation and mission changes that a mission planner would wish to predict. MORRIS attempts to incorporate the cognitive and physical aspects of the robotics task. Based on the inputs from our data collection efforts described previously, we refined and integrated three cognitive models, (Figure 4). The models represent decision making, spatial transformation (the Frame of Reference Transformation, FORT; Wickens, Keller, Small, 2010), and eye movements (Salience, Expectancy, Effort, and Value; the SEEV attention model; Wickens et al, 2009). These affect the critical operator behaviors of trajectory control, control mode selection, and camera view assignment, and contribute to predictions of workload and situation awareness.

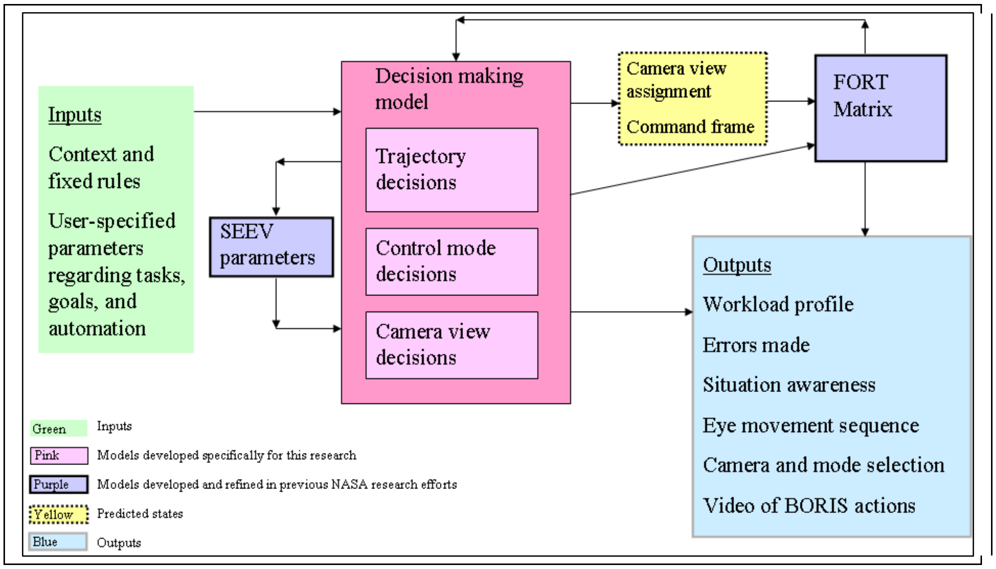


Figure 4. The simulated operator cognitive model, showing decision making, trajectory selection and attention.

The **decision model** makes three types of decisions: 1) ***trajectory decisions*** - how to move the robotic arm and whether to halt it in mid trajectory (e.g., if it should be approaching a hazard), 2) ***control mode decisions*** – which type of movement should be selected (e.g., rate of movement, arm-referenced or room-referenced movements), and 3***) camera view decisions*** – which cameras should be selected for display in the three windows available to the operator. All of these decisions are based in part upon certain fixed if-then rules, where the “if” defines a context. For example if the end-effector is within 1.6 m of the target, then the slow, Vernier rate control mode should be selected. Importantly however, the decisions are also based on continuously varying **utility** of different choices, determined by continuous and changing variables as the arm moves through the workspace. Most of these variables are represented, cognitively, by the **frame of reference** with which the end-effectors is viewed in the cameras, and by which the control movement governs the arm movement. Thus a second model, the **(2) frame-of-reference transformation**, or FORT model provides critical inputs to all three classes of decisions.

The FORT model was developed for a broader class of spatial manipulations (Wickens, Keller & Small, 2010), and based on extensive empirical data from spatial cognitive operations (see Wickens, Vincow & Yeh, 2005). In the current project, we are modifying it to account for particular costs (cognitive and perceptual-motor challenges) imposed on the operator in the robotic environment, including line-of-sight ambiguity and control-display compatibility. FORT calculations are used to assess the quality of various camera views, and influence the modeled operator’s decision to select a different camera view.

The **model of visual attention** across the workspace is **SEEV** (Wickens & McCarley, 2008; Steelman-Allen et al, 2009). The model is particularly important because it can predict attentional tunneling and **areas of neglect**, such as when an operator becomes so focused on a window guiding precise movement to a target, s/he fails to monitor another display that portrays the proximity of the arm’s elbow to colliding with an object in the workspace. Essentially SEEV predicts the moment to moment scan between the three different camera windows and the master system monitor display (i.e., the GUI), based upon the **Salience** or each display, the **Effort** (proportional to distance) required to move the scan from one area to another, the **Expectancy** that information will change in an area and, most importantly the **Value** of each display to the robotics subtask in question. The FORT and SEEV models in FAST interact with one another, in that displays returning high FORT penalties are of low value, whereas those of low FORT penalties are of high value, and will be looked at much of the time.

Discussion

This paper describes the initial development of a model and simulation-based tool for predicting operator and system performance in robotic missions. The efforts described in this paper are being conducted to develop the FAST tool, to support researchers in evaluating human performance in potential robotic automation strategies, and help ensure that the design of new automation concepts does indeed support better operator and system performance.

Acknowledgements

The team would like to thank Barbara Woolford, the scientific and technical monitor, and Janis Connolly and Mary Kaiser for their support, interest, and inputs to our effort. This work was sponsored by a grant from NASA Johnson Space Center, NNX09AM81G.

References

Allender, L. (2000). Modeling Human Performance: Impacting System Design, Performance, and Cost. In M. Chinni (Ed.) *Proc of the Military, Govt and Aerospace Simulation Symp, 2000 Adv Sim Tech Conf,* 139-144.

Bainbridge, L. (1983). Ironies of Automation. *Automatica, 19,* 776-779.

Cizaire, C. (2007). *Effect of 2 module docked spacecraft configurations on spatial orientation* (Master’s thesis). Cambridge, MA: MIT

Endsley, M. R. & Kiris, E. O. (1995). The out-of-the-loop performance problem level of control in automation. *Human Factors, 37*(2), 387-394.

Foyle, D.C. & Hooey, B.L. (2007). *Human Performance Modeling in Aviation*. Boca Raton, FL: CRC Press.

Kanas & Manzey, D.(2008) *Space Psychology and Psychology*, El Segundo CA: Microcosm Press.

Parasuraman, R. & Riley, V. (1997). Humans and Automation: Use, misuse, disuse and 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: 30*, 286-297

Steelman-Allen, K.S., McCarley, J.S., Wickens, C., Sebok, A., & Bzostek, J. (2009) N-SEEV: A Computational Model of Attention and Noticing. In *Proc of the Human Factors & Ergonomics Society Conference*, 774-778.

Wickens, C.D. & McCarley, J.S. (2008). *Applied Attention Theory.* Boca Raton, FL: CRC Press.

Wickens, C.D., Keller, J.W. & Small, R.L. (2010). Left, No, Right! Development of the Frame of Reference Transformation Tool (FORT). In *Proc of the Human Factors & Ergonomics Society Conference,* 1022-1026.

Wickens C.D., Li H., Santamaria, A., Sebok, A. & Sarter, N. (2010) Stages & Levels of automation, An integrated meta-analysis. In *Proceedings of the Human Factors & Ergonomics Society Conference,* 389-393.

Wickens, C.D., Vincow, M., & Yeh, M. (2005). Design applications of visual spatial thinking: The importance of frame of reference. In A. Miyaki & P. Shah (Eds.), *Hndbk of Vis Spat thinking*. Oxford, UK: Oxford Univ Press.

DEVELOPING ASAP (ANTICIPATION SUPPORT FOR AERONAUTICAL PLANNING)

Sami Lini[1,2]
Pierre-Alexandre Favier[1]
Xavier Servantie[2]
Bruno Vallespir[3]
Sylvain Hourlier[2]

[1] Equipe CIH / ISCC — UMR-CNRS 5218 — HEAL / ENSC — IPB
Bordeaux, France
[2] Human Engineering for Aerospace Lab. (HEAL) — THALES / ENSC
Bordeaux, France
[3] IMS, Université de Bordeaux, CNRS
Talence, France

The 2009 A320 ditching on the Hudson River revealed to the public that critical decision-making ability was a key asset for pilots. Situation awareness and workload management are two key elements as well as the ability to anticipate how the situation will evolve. Thales Avionics is funding research through its HF research Lab (HEAL) on HMI facilitating pilot anticipation. This presentation covers a literature review on anticipation so as to develop a model of anticipation in a realistic flying task. Analyzing the concept of anticipation leads us to consider it as a metacognitive process relying on cognitive resource management, situation awareness and time management, using abstract representations. After reviewing these concepts, we focus on time management and anticipation. It reveals a closed loop mechanism inspired from the reafference principle. This will be the basis for a HMI based on a model of anticipation in flight.

The feat of ditching on the Hudson in January 2009 has revealed the criticality of certain decisions under high pressure. Metaknowledge and anticipation are the cornerstones of human ability to manage these types of situations: helping pilots to better anticipate situations is therefore a major challenge that the ASAP project (Anticipation Support for Aeronautical Planning) intends to address. This presentation is an anticipation-oriented literature review, with the intention to develop a model of anticipation in the context of a simulated flight task. This work will enable the design of a user-centered HMI prototype dedicated to assisting in-flight anticipation.

First, we will compare different definitions of the "*anticipation*" concept in diverse disciplinary fields. This will lead us to consider anticipation as a metacognitive process based on cognitive resource management, situation awareness and time management, using abstract representation. These themes are discussed with regard to their contribution to the process of anticipation. Second, we concentrate on time management within the scope of cognitive psychology. Emphasis is placed on the relationship with time and the environment: workload as a function of the perceived distance to the goal is specifically addressed. Finally, work on anticipation is presented, highlighting a closed-loop mechanism similar to the reafference principle.

Anticipation as a metacognitive process

Definition

Even though the term anticipate is commonly used, it is nevertheless still difficult to define precisely. We confuse it and often use *predict, foresee* and *plan ahead* instead. There is one idea common to these concepts; it is a process that is both in the present and the future, as suggested in its Latin etymology: "*anticipare*" means "*take in advance, take the initiative, take the lead*".

In psychology, Sutter (1983) defines anticipation as a "*movement by which man carries his entire being beyond the present into a future, near or far, that is essentially his future*". Even though this definition does not exclusively consider taking action, it elicits the idea of thinking ahead: to anticipate is to represent ourselves and our environment in a process of evolution and adaptation, it is a "*transplant of the future in the present*" (Minkowski (1968)). A new idea emerges from the definition of an anticipatory system in computer science provided by Rosen (1985): an anticipatory system is *"a system containing a predictive model of itself and/or of its environment, which allows it to state at an instant in accord with the model's predictions pertaining to a later instant".* In computer science, to anticipate entails two phases: a prediction phase and a phase of using the prediction. It is therefore a system that has a kind of "future memory", a database enabling it to infer the evolution of the situation as a function of similar situations already encountered.

In cognitive psychology, Cellier (1996) gives the following definition: an "*activity consisting of evaluating the future state of a dynamic process, determining the type and timing of actions to undertake on the basis of a representation of the process in the future, and, finally, mentally evaluating the possibilities of these actions. It is dependent on the overall goal assigned to an operator in a dynamic situation, which is to keep the process, physical or otherwise, within acceptable limits, and therefore avoid the propagation of disturbances. It is also governed by a logic aimed at reducing the complexity of a situation. Finally, it is a way of managing individual resources".* All of the preceding ideas are included in this definition: assessment of the evolution of the situation, mental simulation and encoding in both temporality and the action. However, a few additional points are presented: the teleonomic aspect, anticipation only makes sense in light of an overall goal; from a cognitive point of view, the process also meets the requirement to reduce the load and the complexity of the environment.

For this bibliographic work, we will therefore focus on anticipation as a metacognitive process enshrined in a dynamic temporality: a process aimed at conserving cognitive resources, requiring awareness of oneself and of the situation, as well as time management ability.

Anticipation as a means of cognitive economy: impact of cognitive resource management

In the management of the high-pressure situation presented earlier, it is essential for the pilot to avoid any cognitive overload. Anticipation is presented as a process that makes it possible to overcome the limits placed on his resources.

One strategy consists of spreading cognitive processing out through time: Amalberti (1995) gives the example of preparing a response in advance to a probable combination of events. This example is itself a continuation of the SRK model (Rasmussen (1983)): in the face of a complex situation requiring the use of knowledge to develop a response, anticipation enables the construction of a routine that is ready to be used when needed. Amalberti (1996) further clarifies this idea: the operator avoids complex situations as much as possible in line with his expertise, and prepares for those situations that he cannot avoid by preparing his responses in advance.

Within a complex system, such as a cockpit, the operator may be faced with a situation where several distinct temporal dynamics coexist (inertia of the airplane, transmitting information via the radio, etc.), Leplat & Rocher (1985) emphasize that, in these cases, anticipation exploits the temporal tolerance induced by these dynamics. We will discuss this later in more detail.

With regards to an activated plan, anticipation also consists of mentally simulating the evolution of a situation: this is the last level of Endsley's situation awareness model (Endsley (1995)), which we will also discuss later. When the task requirements increase, anticipation allows assumptions to be made on the evolution of the situation as well as to test them (Amalberti (1996)). The operator imagines the consequences of these situations and consequently changes his conduct by adapting his plan of action if necessary. This attitude allows him to keep the situation within controllable limits while managing any deviations to the plan. To illustrate this, in aviation education, it is said that "*the pilot must learn to be in front of his plane*" and should be encouraged to "*make permanent assumptions about the future situation in order to actively adapt to this situation and not wait for it to occur*". Thus, according to Amalberti (2001), anticipation enables the subject to integrate his own ability in order to impact one's environment and protect oneself from feared events by adjusting the plan.

Anticipation and projection of the evolution of the situation: impact of situation awareness

Even though we emphasized the importance of the representation of the situation during a crisis, it is still correlated to having a good awareness of the situation. The idea of anticipation is included in the definition given by Endsley (1987), "*situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future."* The last part of the definition corresponds to level 3 of this model: awareness of the status and dynamic of the elements is made possible by the first two levels of the situation analysis and allow the evolution to be assessed.

Baumann & Krems (2009) add the following to this analysis: the information collected activates the associated knowledge stored in long-term memory. From this knowledge network, a coherent representation is constructed by following a constraint-satisfaction process: the compatible elements are activated, the incompatible elements deleted. The subject's expertise allows him to add expectations regarding the evolution of the situation to this representation. These expectations are related to schemas that are activated when certain environmental patterns are encountered; more details will be provided for this later.

Comprehension of the situation consists not only of giving coherence to the facts but also explaining them and predicting their evolution. This is therefore a continuous mechanism that is based on a dual relationship with temporality: backwards to explain the cause and forwards to predict the future. All comprehension is teleonomic: it therefore involves permanently constructing an adapted representation, at a given moment, depending on the specific purpose.

Time management

The concepts addressed earlier have shown that the idea of anticipation is inconceivable outside of a temporal framework. The issue of cognitive processes allowing the appropriation of this framework cannot be ignored.

The psychology of time

Michon (1985) defines time as "*the conscious experiential product of the processes that allow the (human) organism to adaptively organize itself so that its behavior remains tuned to the sequential (i.e., order) relations in its environment*." From this perspective, time is no longer a concept for which attributes need to be defined, nor can we still consider it as an intrinsic property of our world, but as a co-occurrence of processes that allows synchronization with the evolution of the environment. This implies that the cognition of temporality belongs in the field of declarative knowledge; the temporal representations would then be a form of high-level cognition.

Of all the theoretical concepts dealing with representations of time, we adopt one principal idea based on recurring events. Among farmers-ranchers, Valax (1986) establishes that a plan is determined around repeated daily pivotal tasks and time-based goals, which are irrelevant both before and after the appointed time. However, two systems provide margins of tolerance: tasks are always planned with some built-in flexibility to make up for a predecessor's possible delay, and open tasks, which are not constrained by a "*when to act*" and may be inserted in moments of calm. As briefly mentioned earlier, the process of anticipation plays a significant role in the situation described here: by simulating the evolution of his future cognitive load, the operator plays with the margins that time gives him to smooth out and avoid peaks in the cognitive load, i.e. cognitive overload. This means that the operator must be able to assess his cognitive investment based on the distance to the action.

Time and cognitive control

In managing the time available for the action, there is a fundamental difficulty related to the anticipation span. There is a proportional relationship between the distance to an expected event and the uncertainty around it. Hollnagel (1998) distinguished four types of control, functions of the perception of time available for the action and the familiarity of the situation: strategic, tactical, reactive and disorganized, which define different degrees of mastery of the situation and behavior types. This is closer to the SRK model (Rasmussen (1983)), mentioned above.

Reynolds (2006) addresses this question in the aeronautical field. He states that, in the short term, we place ourselves in a persistent area, where the environment will not change very much. A little further in the future, models, such as physical laws, can be applied to get a relatively accurate idea of the evolution of the situation; we have now positioned ourselves in a deterministic area. Beyond a certain distance in time, the rules for assessing the evolution of a situation are subject to a combinatorial explosion linked to the set of variables to be considered. Here, we are in a probabilistic area, in which the slope of the uncertainty curve increases exponentially. This raises the question of the relationship between gain associated with anticipation and cost associated with the uncertainty generated by the distance to the event and thus, the importance of planning.

Planning

Hoc (1987) defined planning as "*the development or implementation of plans*". He attributes two fundamental properties to the plan: it is both schematic and orientated towards anticipation. Several cognitive psychology concepts related to the plan express this second characteristic. Thus, the notion of scheme introduced by Piaget (1952) continues in this direction: any action in reality would be made with a previously established scheme in mind and then adapted to the current situation. Therefore, this involves a mental representation that contains a component of expectations about the evolution of the situation that is action-orientated with an identified purpose.

The concept of schema (Bartlett (1932)) provides a second element in this analysis. Bobrow & Norman (1975) hypothesize that this structure, which is not fully specified, represents the relationships between the variables and the constraints on these variables. They will be distinguished during the implementation through environmental cues, which determine its application. Anticipation operates on this idea.

Anticipation

To illustrate how anticipation works, Tanida & Pöppel (2006) propose a wider application of the reafference principle proposed by von Holst & Mittelstaedt (1950) . Mundutéguy & Darses (2007) confirm this functioning in the context of driving a car: a schema of the situation is activated from a set of sensory cues detected in the environment. An efferent copy of the expectation component is made in order to be compared with the objective reality; this is called a corollary discharge. In order to validate the chosen representation, an oriented search for cues is carried out. The schema constitutes an active means of recognition in itself: once activated, it guides and orientates the search for information in order to validate itself (Amalberti (1996)). The schema is validated when all of its contents in the situation are determined, but the mechanisms provide an opportunity to address any existing gaps by values that are, by default, seemingly realistic.

The teleonomic component of the schema is submitted to two forms of monitoring (Amalberti (1996)): an external monitoring, involving the physical process and the situation, and an internal monitoring, involving the cognitive actor of the process. In case of a problem (negative self-evaluation of the performance), the internal supervision makes it possible to increase the cognitive load in order to adjust the selected mental model: enrichment, adjustment or even construction of a new solution. This metacognitive monitoring is also in charge of arbitrating the following processing: intensity, priority, stopping.

On a more global scale, the Cognitive Architecture of Dynamic Control model described by Hoc & Amalberti (1994) illustrates the possibility for the operator to open several cognitive loops at the same time, which project to different levels of temporal depth: a knowledge base is partially activated through the current representation, and partially mobilized at an unconscious level by an activation lattice controlled by this representation. The performance, models and anticipations are continually self-evaluated. The operator's cognitive limits results in compromises with regard to the possible corrections, function of the available resources and different requirements related to the task. Three loops take place within three different temporalities, three distinct ambitions to make corrections, from automatic control to the complete reconsideration of the current representation. As presented earlier, the attentional supervisor is responsible for both filtering and weighting the sensory inputs as well as a potential local change of the level of control of the action.

In line with the work presented earlier, two alternatives are available to design an efficient tool: help to choose the right plan, or help to validate it. A fundamental paradox appears at this stage: helping the operator to anticipate basically makes sense in the case where a representation of the situation is too terse. The paradox resides in the fact that to automatically provide help would result in giving him precisely this representation. In fact, this is another argument in favor of the direction followed until now: the operator is at the center of the decision loop. The potential additional information that we could give him would help him to construct or complete his representation but is not, in any case, intended to replace it.

Conclusion

In this bibliographical work, we have established that anticipation is a metacognitive process aimed at optimizing cognitive resources, based on situation awareness and time management. Its operation in a closed loop opens opportunities for constructing an HMI aide for anticipation. In the ASAP project, the design method is centered on the final user: the understanding of its cognitive functioning in terms of anticipation makes it possible to envisage the design of the tool while keeping the end-user in mind.

This bibliographical review opens up two development directions: first, contextualization in the aviation industry; this will involve breaking down the tasks constituting a defined flight phase based on their relationship with anticipation. This formalization will allow the creation of a knowledge base in order to fuel future HMI prototypes. Second, the definition of the optimal span of anticipation in a flying task: this will involve defining the window of time within which the help information given for anticipation will improve the performance at a locally minimal cognitive price.

References

Amalberti, R. (1995). Anticipation. In M. de Montmollin (Ed.), Vocabulaire de l'Ergonomie: Octares.

Amalberti, R. (1996). La conduite de systèmes à risques. Paris: Presses Universitaires de France.

Amalberti, R. (2001). La maîtrise des situations dynamiques. Psychologie française, 46(2), 107-118.

Bartlett, F. C. (1932). Remembering : A Study in Experimental and Social Psychology: Cambridge University Press.

Baumann, M., & Krems, J. (2009). A Comprehension Based Cognitive Model of Situation Awareness. Digital Human Modeling, 192-201.

Bobrow, D., & Norman, D. (1975). Some Principles of Memory Schemata. In Representation and Understanding : Studies in Cognitive Science (Bobrow Daniel, Collins Allen H. ed., pp. 131-150): New York : Academic Press.

Cellier, J. (1996). Exigences et gestion temporelle dans les environnements dynamiques. In La gestion du temps dans les environnements dynamiques (pp. 20-48). Paris: Presses Universitaires de France.

Endsley, M. (1987). SAGAT: A methodology for the measurement of situation awareness. NOR DOC, 87-83.

Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors: The Journal of the Human Factors and Ergonomics Society, 37(1), 32-64.

Hoc, J. (1987). Psychologie cognitive de la planification: Presses Universitaires de Grenoble.

Hoc, J., & Amalberti, R. (1994). Diagnostic et prise de décision dans les situations dynamiques. Psychologie française, 39(2), 177-192.

Hollnagel, E. (1998). Context, cognition and control. In Co-operative process management, cognition and information technology (pp. 27-52).

Leplat, J., & Rocher, M. (1985). Ergonomie du contrôle de processus en marche simultanée : un cas dans l'industrie biochimique. Psychologie et Education, 9(1-2), 5-26.

Michon, J. A. (1985). The complete time experience time experience. In J. A. Michon & J. L. Jackson (Eds.), Time, mind and behavior (pp. 20-52). Heidelberger: Springer Verlag.

Minkowski, E. (1968). Le temps vécu: Delachaux et Niestlé.

Mundutéguy, C., & Darses, F. (2007). Perception et anticipation du comportement d'autrui en situation simulée de conduite automobile. In Le travail humain (Vol. 70, pp. 32). Paris: Presses Universitaires de France.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE transactions on systems, man, and cybernetics, 13(3), 257-266.

Reynolds, H. (2006). Modeling the Air Traffic Controller’s Cognitive Projection Process. PhD Thesis, Cambridge, MA: Massachusetts Institute of Technology.

Rosen, R. (1985). Anticipatory systems: philosophical, mathematical, and methodological foundations: Pergamon.

Sutter, J. (1983). L’anticipation: Presses Universitaires de France, Paris.

Tanida, K., & Pöppel, E. (2006). A hierarchical model of operational anticipation windows in driving an automobile. Cognitive Processing, 7(4), 275-287.

Valax, M. (1986). Cadre temporel et planification des tâches quotidiennes: étude de la structure des plans journaliers chez les agriculteurs. Thèse de Doctorat, Université de Toulouse Le Mirail.

von Holst, E., & Mittelstaedt, H. (1950). Das reafferenzprinzip. Naturwissenschaften, 37, 464-476.

CONFLICT DETECTION IN AIR TRAFFIC CONROL: DISTUINGUISHING BETWEEN JUDGMENTS OF CONFLICT RISK AND INTERVENTION DECISIONS

Stéphanie Stankovic
Université Paris Descartes
Paris, France
Esa Rantanen
Rochester Institute of Technology
Rochester, NY, USA
Nicolas Ponomarenko
French Civil Aviation Authority
Toulouse, France

This paper seeks to make a distinction between cognitive processes involved in conflict risk judgment and those involved in conflict avoidance decisions (controllers' interventions for separation assurance). First, we conducted a systematic review of the conflict detection literature to identify studies that focused on conflict risk assessments and studies that focused on conflict avoidance (intervention). We then report empirical data pertaining to controller intervention judgments. Studies on conflict avoidance have rarely described the intervention decision making process and its interaction with the conflict risk assessment process, whereas our data indicated differences in terms of information processing between judgments of conflict risk and intervention judgments. We provide recommendation for future studies on conflict detection and conflict avoidance. These findings also have implications for the development of automated conflict detection tools.

Air traffic controllers (ATCos) assure the safe, orderly, and expeditious flow of controlled aircraft between a departure point and a destination. Detecting and resolving potential conflicts between aircraft are the most important tasks that ATCos perform. To do so, ATCos have to systematically scan the display (radar screen) and check the trajectories of the aircraft to assess whether a minimum separation is maintained between pairs of aircraft, and if not, develop and implement a solution that will assure sufficient separation. Conflict detection, which is the first tasks to achieve to keep the air traffic safe, is a complex and dynamic task that is cognitively very demanding. This task has been the topic of several studies in the field (for example, Bisseret, 1981; Remington, Johnston, Ruthruff, Gold, & Romera, 2000; Rantanen & Nunes, 2005; Boag, Neal, Loft, & Halford, 2006; Stankovic, Raufaste, & Averty, 2008; Loft, Bolland, Humphreys, & Neal, 2009). Several models of conflict detection have also been proposed (among the most recent are Rantanen & Nunes, 2005; Stankovic et al., 2008; Loft et al., 2009). More general models of separation assurance which describe both conflict detection and conflict resolution processes have also been developed but they are less numerous than conflict detection models (see for example Niessen, Eyferth, & Bierwagen, 1999). We argue in this paper that there are two distinct processes that come to play after initial conflict detection, one that pertains to judgments of conflict risk and another regarding conflict avoidance (intervention) decisions, and that great care should be taken not to confound them in experimental tasks or in models of conflict detection and avoidance.

Conflict Detection

The focus of current theoretical models of conflict detection concerns how controllers determine whether specific aircraft pairs that have been selectively attended will be in conflict or not. We define conflict as the potential loss of separation between two aircraft, or a situation where two aircraft would lose separation at some time in the future should they continue on their present trajectories. Consider two aircraft flying at the same altitude on converging courses. Determining whether these aircraft will violate lateral separation requires the integration of speed and distance information to predict the distance or time between the aircraft at the point of intersection of their trajectories (Law et al., 1993; Loft, Neal, & Humphreys, 2007; Neal & Kwantes, 2009). When aircraft are also changing altitude (especially in approach control), the prediction of loss of separation in the future also requires the integration of the vertical speeds and altitudes of aircraft, and the subsequent computation of whether the difference between aircraft altitudes at the time of the position overlap on the lateral plane are below a minimum separation (Stankovic, Loft, Rantanen, & Ponomarenko, in press; Xu & Rantanen, 2003). Several theoretical accounts of how controllers detect conflicts have been published (Loft, Bolland, Humphreys & Neal, 2009; Rantanen & Nunes, 2005; Stankovic, Raufaste, & Averty, 2008).

In empirical studies of conflict detection, researchers have typically either asked controllers to assess the risk of conflict between aircraft or to estimate the likelihood that they would intervene on a given aircraft pair to assure separation, but there has been little consideration of whether these two types of judgment are tapping into the similar decision process or not.Bisseret (1981) made a distinction between those two cognitive processes thirty years ago. In his study on decision making made by ATCos, he reported several studies that investigated conflict detection made by expert ATCos and trainees. In this study, Bisseret distinguished the information processing (risk of loss of separation) from the decision (intervention) process. Bisseret showed that trainees made more accurate separation estimations than expert controllers. This particular result has been explained by the safety margins that experienced controllers apply. Bisseret further hypothesized that experienced controllers are more concerned about succeeding in their overall control task (which also includes orderly and expeditious flow of traffic) than in the accuracy of their conflict risk assessments. On the basis of this hypothesis, Bisseret suggested that experienced controllers' operating decision process should not be entirely based upon their diagnosis of the air traffic situation. In an earlier study Bouju (1978) had asked experienced controllers to make a diagnosis (conflict or not) for twelve (six conflict and six non-conflit) air traffic situations, and to express their intention for intervention (action or not) for the same situations. Results showed that controllers decided to take actions in nine situations diagnosed as non-conflict, lending support to Bisseret‘s hypothesis that conflict risk judgments and decisions to intervene are separable.

Stankovic, Neal and Hasenbosch (2010) proposed in a recent study a model of separation assurance that described five, instead of three, main cognitive processes: (1) information gathering (selection of information), (2) trajectory anticipation (risk judgments of loss of separation), (3) intervention (decision making about future intervention), (4) solution choice (selection of conflict solution strategy), and (5) planning (resolution course of actions). These separation assurance processes are influenced by changes in ATCos' environment and their strategies. This study showed the importance of consider all separation assurance processes and their interactions when studying conflict detection, and in doing so makes a distinction between the *conflict risk process* (which corresponds to trajectories anticipation in their model) and the *intervention process.* Conflict risk consists of a judgment ATCos make about the loss of separation between aircraft, whereas intervention requires ATCos to decide whether to intervene or not on a particular pair of aircraft. Recent studies have reported quite different results using the two questioning methods (risk versus intervention).

Judgment of Conflict Risk

Several experimental studies have asked the participants to judge the risk of conflict between two or more aircraft. Bisseret (1981) did so when he investigated controllers' diagnosis process. More recently, Stankovic et al. (2008) used this method and three specific variables to predict controller judgment of conflict risk, (1) the distance between the crossing point of the aircraft pair trajectories and the closest aircraft to that point, (2) the distance between the two aircraft when they are laterally closest, and (3) the lateral distance between the two aircraft when their growing vertical distance reached 1,000 feet. These variables accounted for up to 50% of the variance in conflict judgments made by expert controllers. Stankovic et al. asked expert ATCos to judge the risk of conflict between two aircraft using an 8-pt scale (from "assured airprox" to "no conflict"). Large individual differences in controllers' judgments were evident. One group of controllers seemed to be more influenced by a distance that takes into account vertical separation between aircraft (the lateral distance between the two aircraft when their growing vertical distance reaches 1,000 feet) when judging the risk of conflict between aircraft than the other group of controllers.

In a recent study, Stankovic, Loft, Rantanen and Ponomarenko (in press) also reported individual differences in the effect of vertical separation on conflict risk judgments. In this study, we asked fourteen expert controllers to judge the risk of conflict between aircraft for situation were four variables (environmental cues) were manipulated, (1) lateral conflict geometry, (2) vertical separation between aircraft, (3), time to lateral separation threshold (3 nm), and (4) groundspeed difference between the aircraft. The question used to collect conflict risk judgment was the same used in the Stankovic, et al. (2008) study, but the scale was different (a 12-pt scale, from "no risk of conflict" to "extreme risk of conflict"). Results showed that an important effect of vertical separation between aircraft on controllers' judgments of conflict risk. Overall, the conflict risk judgments increased as the vertical separation decreased. Most importantly, we found individual differences in the effect of the vertical separation on conflict risk judgments. One group of controllers (group 1, $N = 7$) made lower conflict risk ratings than the other group of controllers (group 2, $N = 7$), and the effect of vertical separation on conflict risk judgment was greater for controllers in group 1 than for the controllers of group 2. Controllers in group1were more experienced (i.e., total experience, sector specific experience, older) than controllers in group 2, indicating that experienced controllers were less conservative and took vertical separation more into consideration than their less experienced counterparts.

Decisions of Intervention

Another experimental task consists in asking participant about their intention to intervene in a situation to assure separation between aircraft. This was one of the methods used by Bouju (1978). More recently, Loft et al. (2009) used this method for their study on conflict detection. Like Bisseret (1981), Loft et al. proposed that controllers use ‘safety margins’ to assure separation between aircraft. These safety margins reflect expectations regarding likely variation in aircraft trajectory, and also the degree to which controllers are biased to favor safety over accuracy. Depending on the magnitude of safety margins, controller predictions of aircraft position at specific points in the future will be some distance closer or further (or higher or lower in the vertical plane) than the positions predicted by aircraft state values. To test their the theory, Loft et al. (2009) presented to their participants pairs of aircraft and asked them to provide intervention judgments on a four point scale. A two-parameter computational model that emulated how controllers approximate aircraft trajectory closely predicted the conflict risk judgments made by controllers. A key finding reported by Loft et al. (2009) was that there was no variability in risk judgment as a function of the vertical separation between aircraft. Instead, risk judgment only varied with changes in aircraft lateral separation. To account for these data, the Loft et al. (2009) model was simplified to assume that controllers always consider aircraft pairs to be in vertical conflict when aircraft are descending or climbing through the levels of other aircraft. Loft et al. argued that controllers prefer to intervene to assure aircraft separation when aircraft are climbing through the levels of other aircraft in order to manage their own workload (Loft, Sanderson, Neal & Mooij, 2007), and thus that their computational model should indeed be able to predict risk judgments without the setting of a vertical separation safety margin parameter. This result contradicts the effect of vertical separation on conflict detection reported by Stankovic et al. (in press).

Once ATCos have detected a conflict risk between two aircraft, then they will intervene by instructing one of the aircraft to maneuver to avoid the conflict (Stankovic, Neal & Hasenbosch, 2010). This risk-intervention process has been rarely investigated as a unified process except by Bisseret (1981). Researchers have used risk judgment to examine conflict detection or conflict resolution, but to our knowledge only Bisseret’s study examined cognitive processes involved in intervention decision. Despite the close relationship between judgment about the risk of conflict and intervention decision, it is worthwhile to specify the cognitive processes involved in each operation. This question is particularly urgent as new conflict detection tools that automate the conflict detection process are to be implemented in the air traffic management (ATM) systems worldwide (for a review see:Neal, Flach, Mooij, Lehman, Stankovic et al., 2011). Moreover, as showed by Bisseret (1981) intervention actions may occur even if a situation is diagnosed as non-conflict (cf. Loft et al., 2009).

Some studies on conflict resolution, however, have focused on the type of intervention actions that are applied by ATCos, once they have decided to intervene on a particular aircraft. For instance, Leroux (1999) defined three different policies that controllers apply when resolving conflicts according to task load level: (1) “be elegant first”, (2) “be efficient first”, and (3) “be safe and nothing more”. A good description of how ATCos adapt their control strategies to anticipated workload and task demands is also provided in Loft et al. (2007).

Recent studies have reported quite different results using the different experimental tasks (judging the risk of conflict or the likelihood of intervention). For example, Loft, et al. (2009) claimed that there was no variability in risk judgment as a function of the vertical separation between aircraft, whereas Stankovic et al. (in press) reported an effect of vertical separation on conflict risk judgments made by expert controllers. The latter showed that controllers’ judgments of conflict risk increased as vertical separation decreased for the half of the participants. In the same study and in addition to the risk of conflict judgments, Stankovic et al. asked fourteen licensed air traffic controllers (12 men and 2 women) to judge the likelihood by which they will intervene to assure separation between aircraft. Three questions were asked about three other judgments relating to strategies used to ensure separation, (1) will you intervene by assigning a new level to the descending aircraft? (2) will you intervene by assigning a new route to the descending aircraft, or (3) will you intervene by assigning both a new level and a new route to the descending aircraft. These data on intervention judgments have not been reported before.

First, we compared the three types of intervention (level, heading or both level and heading) to each other and we found a preference for the level solution compared to the heading solution ($t(1,13) = 3.11$, $p = .002$) and to the both level and heading solution ($t(1,13) = 3.23$, $p = .001$). This result is not surprising since in approach control controllers sequence aircraft for landing and takeoff by using mainly level solution. For this reason we decided to put together all intervention judgments (level, heading and both level and heading) by keeping the higher scores among all intervention judgments. We obtained thus just one intervention judgment variable. We used the *Tukey* for Post hoc tests and the values for small, medium and large effect sizes of.10, .25 and .40 respectively (Cohen, 1988). One important result was the significant difference between judgments of conflict risk and intervention judgments, $t(1,13) = 2.85$, $p = .004$. Post hoc tests showed that controllers made higher intervention judgments ($M = 9.27$, $SD = .68$) than conflict of risk judgments ($M = 7.30$, $SD = .53$). This result confirmed Bisseret (1981) hypothesis, expert controllers intervene on pair of aircraft that have been diagnosed not to be in conflict. We also found an effect of vertical separation on intervention judgments for the Same heading scenarios, $F(2, 26) = 5.41$, $p = .011$, $\eta_p^2 = 0.29$; for Opposite heading scenarios, $F(2, 26) = 5.54$, $p = .010$, $\eta_p^2 = 0.30$; and for the Cross Heading scenarios, $F(2, 26) = 2.94$, $p = .071$, $\eta_p^2 = 0.18$. These patterns are similar to those found for risk of conflict judgments; however this effect of vertical separation on intervention judgments was less than the effect found for conflict risk judgments. Only four controllers were sensitive to vertical separation in making intervention assessments. This result showed that the effect of vertical separation is less important on intervention judgments than on conflict risk judgments. This is consistent with Loft et al. (2009) findings and may attest to the difficulty of vertical speed calculations or estimations in conflict situations.

Discussion

There are empirical results that support the need for distinguishing between judgments of conflict risk and intervention decisions. First, we reported at least two recent studies on conflict detection that presented results in contradiction but that also confirm our hypothesis about of two sub-processes in ATC conflict detection and avoidance. Loft et al. (2009) showed that conflict detection (intervention decisions) was not affected by variation in vertical separation between aircraft. Hence, Loft et al. concluded that the controllers' made no calculations about vertical distance between aircraft but rather applied safety margins for detecting conflicts. On the other hand, Stankovic et al. (in press) showed that experienced controllers made calculations of vertical separation for judging the risk of conflict between aircraft. These two different results are, however, in line with Bisseret's (1981) conclusions. Conflict detection implies a conflict risk process which consists of assessment of the future separation between aircraft mainly based on calculations, or at least more than the intervention decision process which is mainly based on the conflict risk assessment.

Moreover, the comparison of results reported on conflict risk judgments reported in Stankovic et al. (in press) with the results on intervention decisions reported above also confirm the existence of two conflict detection sub-processes. Hence, results on conflict risk judgments showed that controllers took into account vertical separation between aircraft to judge the risk of conflict between aircraft. Results on intervention judgments showed also an effect of vertical separation on controller's intervention ratings; however this effect was reported for only 4 controllers. Controllers also made higher intervention judgments than conflict risk judgments for the same situation. Latter result shows that controllers made their intervention decision based on conflict risk assessment, but it also shows that controllers are more cautious when deciding to intervene than when assessing the risk of conflict between aircraft.

In conclusion, the empirical results reported above confirm Bisseret's conclusions that controllers are more interested in overall control task than in accuracy of individual conflict risk judgments, and that experienced controllers' operative decision process are not be entirely based upon his diagnosis (conflict risk judgment) of the situation but it is also include the application of safety margins such as Loft et al. (2009) showed. Conflict detection process implies two sub-processes: (1) a conflict risk judgment process and, (2) an intervention decision process.

Conclusion

This research highlighted the differences between cognitive processes involved in intervention and those involved in conflict detection. Controllers make more calculations in estimating the risk of conflict between aircraft than for deciding to intervene in a given situation. As for conflict detection, it seems that intervention operation is affected by vertical separation between aircraft, thus conflict detection tools which displays pair of aircraft in conflict and which suggest intervention solution should integrate such differences. This particular result on differences between conflict detection and intervention processes should be considered for the account of the reason why ATCos very often do not trust conflict detection tools when completing conflict resolution task.

It may also be argued that conflict avoidance, that is, decisions to intervene in potential conflicts early on, without careful assessment of the actual conflict risk is part of strategic ATC and an effective means to manage workload and maintain an accurate picture of the traffic situation. Accurate conflict risk assessment requires considerable amounts of cognitive (especially attentional) resources that ATCos working busy traffic can ill afford to trade off against strategic advantages. That experienced controllers exhibit such traits supports this hypothesis.

The aim of this paper was to clarify the conflict detection process rather than claiming that one method (conflict risk versus intervention question) is better than another. On the contrary, what it is claimed here is that both conflict risk judgments and intervention decisions should be considered when studying the conflict detection processes. For the design of future experimental protocols, distinguishing conflict risk judgments and intervention decisions is crucial since this clarification should guide researchers for choosing an appropriate method (question about the risk of conflict or about intervention) for the particular research questions they investigate. Understanding how expert controllers make conflict detection decisions in air traffic control is also crucial for the design and the evaluation of the future Air Traffic Management (ATM) systems (Stankovic, Neal & Hasenbosch, 2010). In particular, when conflict detection tools envisaged in the future ATM systems allocate conflict risk assessments to an automated system and as a consequence separate conflict risk judgments from intervention decision.

References

Bisseret, A. (1981). Application of Signal Detection Theory to Decision Making in Supervisory Control: the effect of the Operator's Experience. *Ergonomics, 24*, 81-94.

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.

Bouju, F. (1978). Analyse *de la prévision de séparation entre deux avions évolutifs convergents* [Analysis of the separation forecast between two converging aircraft] (No. C.O.7802-R54). Toulouse, France: Centre d'Etudes de la Navigation Aérienne.

Leroux, M. (2000). *Cognitive aspects and automation.* In N. B. Sarter & R. Amalberti (Eds.), Cognitive engineering in the aviation domain. (pp. 99-130). Mahwah, NJ: Lawrence Erlbaum Associates.

Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (2009). A theory and model of conflict detection in air traffic control: Incorporating environmental constraints. *Journal of Experimental Psychology: Applied, 15,* 106-124.

Loft, S., Neal., A., & Humphreys, M. S. (2007). The development of a general associative learning account of skill acquisition in a relative arrival-time judgment task. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 938-959.

Nunes, A., & Mogford, R. H. (2003). Indentifying controller strategies that support the 'picture'. Paper presented at the *47th Annual Meeting of the Human Factors and Ergonomics Society*, santa Monica, CA.

Neal, A & Kwantes, P. (2009). A random walk model of conflict detection performance in a simulated air traffic control task. *Human Factors, 51*, 164-180.

Niessen, C., Eyferth, K. and Bierwagen, T. (1999). Modelling cognitive processes of experienced air traffic controllers. Ergonomics, 42, 1507-1520.

Rantanen, E. M., & Nunes, A. (2005). Hierarchical conflict detection in air traffic control. *International Journal of Aviation Psychology, 15*, 339-362. doi: 10.1207/s15327108ijap1504_3.

Remington, R. W., Johnston, J. C., Ruthruff, E., Gold, M., & Romera, M. (2000). Visual search in complex displays: Factors affecting conflict detection by air traffic controllers. *Human Factors*, 42, 349-366.

Stankovic, S., Loft, S., Rantanen., E. M., & Ponomarenko, N., (in press). Individual differences in the effect of vertical separation on conflict detection in air traffic control. *International Journal of Aviation Psychology.*

Stankovic, S., Neal, A., & Hasenbosch (2010). A separation assurance model for automation design and evaluation. Paper presented at *HCI-Aero 2010*, 3-5 november, Cape Canaveral, FL.

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)*. Maisons-Alfort, France: Author. (www.statsoft.fr)

CHARACTERIZING AIR TRAFFIC CONTROLLER SECTOR SPECIFIC KNOWLEDGE: AN ABSTRACTION-BASED ANALYSIS

Rahul Bhagat[1]
[1]Department of Systems Design Engineering, University of Waterloo
Waterloo, Ontario, Canada
Jonathan Histon[1]
Emilio Alverne Falcão de Albuquerque Filho[2],
R. John Hansman[2]
[2]Department of Aeronautics & Astronautics, Massachusetts Institute of Technology
Cambridge, Massachusetts

The need to train air traffic controllers on sector-specific operational traffic patterns and procedures creates staffing inflexibility and training inefficiency across the National Airspace System (NAS). The deployment of generic airspace, or air traffic control sectors with similar operational characteristics, is one means of addressing these challenges in next generation (NextGen) ATC operational concepts. Based on prior work, local, sector specific knowledge, is thought to be captured in part through abstractions, or simplifications of a controller's mental model. This paper describes a methodology used to identify key properties characterizing three distinct abstraction types (standard flows, handoffs, and merges). This categorization provides a useful basis for assessing the relative importance of differences in abstractions between sectors, a key step in assessing operations similarities required for the generic airspace concept.

Recent years have seen accelerated rates of retirement rate amongst United States air traffic controllers (FAA, 2010A). In addition, controllers maintain proficiency on only a limited number of sectors. In combination with the retirement pressures, this is creating the possibility of localized and national shortages of air traffic controllers. This creates a need for greater staffing flexibility: an effective response is to transfer experienced controllers to provide coverage for sectors experiencing shortfalls. However, efficient transfer requires minimizing the amount of retraining an experienced controller needs.

A key factor affecting controller training is airspace structure. Airspace structure is defined as the physical and informational elements that organize and arrange the air traffic control environment (Histon and Hansman, 2008). It plays an important role in developing air traffic controller mental models and strategies. However, airspace structure can vary considerably between sectors and across facilities necessitating site-specific training. Air traffic controller training includes a considerable amount of time devoted to on the job (OJT) training where controllers learn relevant airspace structures and internalize the mental models and strategies that help them safely control traffic. This training develops localized sector-specific knowledge that has to be learned when even experienced controllers transfer to a new sector.

One strategy for mitigating these training needs is the development of generic airspace with similar structure such that controllers only require training on the minimal differences between sectors (FAA, 2010B). This approach requires assessing the applicability of a controller's sector-specific knowledge to other airspace sectors and identifying the cognitive differences amongst sectors. . In order to provide a framework for conducting these assessments, this paper uses previously identified knowledge of how controller's use structure to reduce complexity as a basis for determining the similarity of one or more airspace sectors. The sector abstraction binder provides a comprehensive tool for assessing generic airspace sector groupings for cognitive similarity.

Background

Characterizing Airspace Sectors

The generic airspaces concept identifies opportunities to standardize airspace in an attempt to increase air traffic controller training efficiency. In the short to mid-term, the goal is to identify similarities across existing airspace sectors and produce sector groupings based on minimizing training differences within each group. In the longer term, the factors used to asses these similarities can be used as heuristics for sector redesign with the goal of reducing overall NAS-wide differences in training.

Previous attempts at characterizing airspace sectors have mostly looked at aggregate complexity measures based on a combination of air traffic and structural considerations. Christien (2003) proposed a set of complexity factors, or a complexity index, which could then be evaluated and compared across airspace sectors. Goldman et al. (2006) similarly proposed a set of sector factors which were independent of specific air traffic situations. Yousefi (2003) proposed metrics for measuring airspace density and transit time. These works show promise for characterizing airspace sectors and are used as a basis for deriving factors that characterize abstractions within the SAB. However, the factors presented in these works lack a strong association to structure based abstractions which are shown to greatly influence air traffic controller mental models (Histon and Hansman, 2008).

Structure-Based Abstractions

Figure 2 is a representation of an air traffic controller's mental model (Histon and Hansman, 2008). A key component of this representation is the working mental model which supports the generation and maintenance of situation awareness along with decision-making and implementation processes of the controller's task. The working mental model is a result of the specific air-traffic situation, or operational environment, that the controller is managing and the mental models and abstractions that the controller has knowledge of within their long term memory.

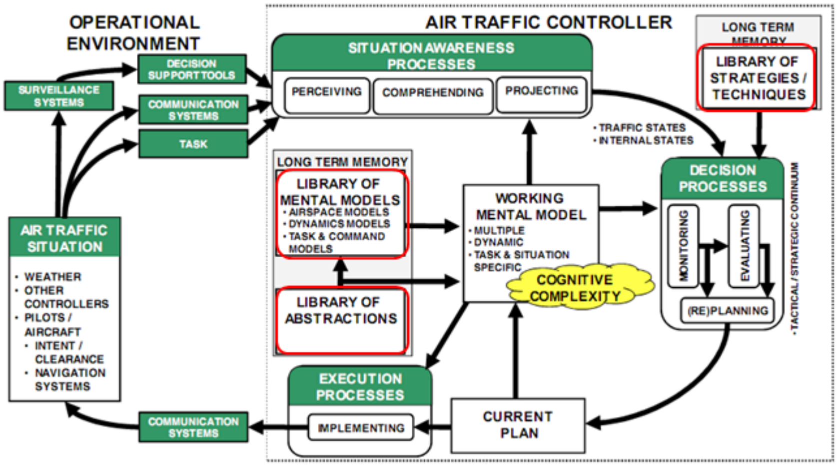


Figure 2. Representation of an air traffic controller mental model.

Over the course of OJT, controllers build up their libraries of knowledge as generalized abstractions and are thought to use sector-specific instantiations of those abstractions when they are being utilized. For example, a controller may develop an abstraction for a merge pattern that involves a consolidation of flows within heuristic operating limits such as maximum and minimum flow separation angles. However, the specifications regarding the specific map location of the merge, spacing, and velocity requirements may differ between airspace sectors. If the controller is accustomed to utilizing a certain abstraction then a transfer to a different sector that also requires a similar abstraction can be accomplished with reduced training because only the discrete specifications need to be relearned.

The working mental model is also influenced by the operational environment or context under which the abstraction is used. This incorporates the effects of other structure-based abstractions on the abstraction of interest. For example, a military operations area may project certain constraints on a merge abstraction if they are in close proximity. Learning the operational environment and context of a specific sector is also an important part of OJT.

The presence and context under which structure based abstractions are utilized across airspace sectors can be used as a method of clustering for the purposes of generic airspaces. The underlying hypothesis of this research is that controller transfers between airspace sectors should involve minimal training if the needed abstractions exist and are "similar". The challenge is to assess the similarity of these abstractions by determining the specification and context based factors that influence them, evaluating these factors and comparing them across sectors. This document presents the Sector Abstraction Binder (SAB) which is a bottom up method for identifying and evaluating the similarities in abstractions across airspace sectors.

Sector Abstraction Binder

The Sector Abstraction Binder (SAB) is a bottom up methodology for assessing cognitive similarities across airspace sectors by leveraging the importance of structure based abstractions. To limit scope, the analysis is limited to four commonly used abstractions with a focus on a high-altitude enroute airspace sectors. However, additional abstractions can be easily incorporated. The abstractions include merges, inbound and outbound handoffs and, standard flow segments. Table 1 provides a working definition for each of these abstractions and reasoning for inclusion into the SAB.

Table 1

Abstractions chosen for the Sector Abstraction Binder (SAB)

Abstraction	*Definition*	*Reason for Selection*
Std. Flow Segment	This abstraction is the presence of densely organized air traffic that is generally but not exclusively associated with jet routes.	It is a very common occurrence and tends adjoin other abstractions such as merges and handoffs.
Merge	This abstraction involves the consolidations of n flow segments to n-1 or fewer segments while resolving sequencing conflicts.	Merges were selected because they are commonly found and involve some amount of traffic sequencing by air traffic controllers.
Inbound Handoff	Inbound and outbound handoffs are the process of giving away control versus receiving control of an aircraft. They are treated as separate abstractions because there are significant procedural differences between the two.	These abstractions were selected because they are very common and are an example of an air traffic situation where coordination with another controller is required.
Outbound Handoff		

Bottom Up Process

Figure 3 provides an overview of the bottom up nature of the SAB. First, each *abstraction instance* within a sector is evaluated. This involves the assessment of characteristic factors related to that abstraction type. The factors capture key properties of the abstraction determined from an assessment of how it fits in to a controller's mental model. There are two distinct types of factors: specifications and context (Table 2). Specifications represent the core parameters required to describe an abstraction and distinguish it from another instance of the abstraction; for example, the frequency of an adjoining sector is a key specification of a handoff abstraction. Context captures the relationship between an instance of the abstraction and features in the airspace. For example, the same handoff abstraction may occur at a different distance from key confliction points in the sector, leading to different abstraction instances. Generally, specifications tend to be discrete bits of information while context tends encapsulate the operational environment and behaviours of abstractions.

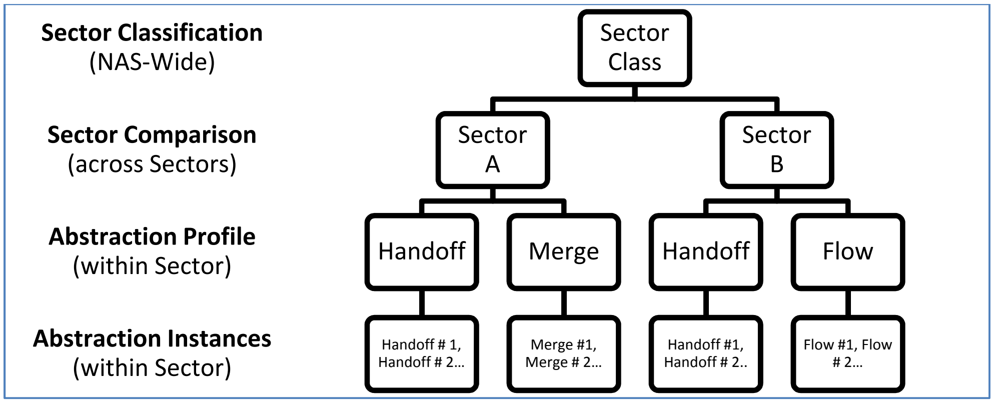


Figure 3. Bottom-up approach used by the SAB.

Table 2

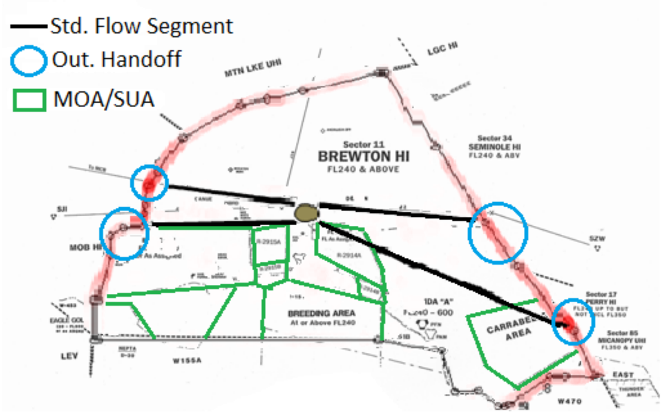
Examples of abstraction characterization factors

		Charac. Factor	*Definition*
Handoffs (Inbound/Outbound)	*Specs.*	Position/Location	- sector boundary location of handoff instance
		# Interacting Sector(s)	- # sectors that feed (inbound)/accept (outbound) aircraft to/from instance
		Adjacent Sector Freq.	- adjacent sector primary/backup radio frequencies
		.....	
	Context	Dist. to Internal Critical Points (nm)	- dist. from handoff to critical points within sector - distance from handoff to critical points within receiving sector
		Handoff Angle (deg)	- angle between handoff flow and sector boundary
		.....	
Merges	*Specs*	Position/Location	- location of merge instance
		Entry/Exit Headings	- headings of n entry flow segments and n-1 or less exist flow segments
		....	
	Context	Dist. to Internal Critical Points (nm)	- distance from merge point to other critical points within sector
		Nearby Elements	- list of nearby airspace elements not including MOA/SUA
		.....	
Std. Flow Segments	*Specs*	Position/Location	- heading required to maintain flow segment track
		Flight Levels	- flight levels available to aircraft that will use flow instance
		.....	
	Con.	Segment Length	- length of flow segment
		Terminal Elements	- elements that establish the end point elements of flow segment
		.....	

Consolidating the results of these abstraction instances provides an *abstraction profile* which is a summary of the range of both specifications and context factors found for each abstraction type within a sector. The abstraction profiles represent the types of specification and contextual environments an abstraction operates under for a specific sector. This can be seen in Table 3 which illustrates a partial abstraction profile of outbound handoffs for the Brewton high level sector in Jacksonville Center. Table 3 also shows a visualization of some of the contextual features. These profiles can be compared across different sectors to establish cognitive similarities. Such profiles can be used to create groupings of cognitively similar sectors or *sector classes*. Since the analysis begins from individual structural features within airspace sectors and progressively makes generalizations, this method leverages benefits of bottom up methodologies. It provides transparency into the causes of cross sector dissimilarities as any difference can be traced to a specific factors and their corresponding data source. Furthermore, the approach allows the consolidation of various qualitative and quantitative data sources at an early stage ensuring a comprehensive analysis.

Table 3

Selected portion of an outbound handoff abstraction profile for Brewton-HL (Bhagat and Histon, 2011)

Specifications		*Visualization of Brewton-HL Outbound Handoffs*
Position/Location	eastern bound.; western bound.	
# Interacting Sector(s)	1-2	
Adjacent Sector Freq.	128.07/307.2, 135.65/291.7, 135.32/380.25...	
...	...	
Context		
Dist. to Internal Critical Points (nm)	*50-120*	
Handoff Angle (deg)	*30-90*	
Nearby MOAs/SUAs	Up to *3*	
... ...		

The following provides a step-by-step breakdown of the SAB process:

1. Identify instances of merges, handoffs (inbound & outbound), and standard flow segments within sector of interest.
2. Characterize each instance by evaluating it against each of the specification and context factors devised in the SAB.
3. Develop an abstraction profile for each of the four abstractions (inbound/outbound handoff, merge, standard flow segment) within the sector. This involves consolidating each instance analyzed in step 2. The consolidation can vary depending on the type of characterization factor.
4. Repeat steps 1-3 for another sector of interest.
5. Compare abstraction profiles across the two sectors to determine emergent differences in abstractions.
6. Expand analysis to more sectors and group by abstraction similarities to create sector classes.

The SAB uses 19 factors to characterize inbound and outbound handoffs, 17 factors for merges and 15 factors for standard flow segments. Examples of these characterization factors are shown in Table 2. The complete list can be found in Bhagat and Histon (2011). These characterization factors were devised from literature reviews of existing complexity factors, qualitative analysis of standard operating procedures, change notices and reviews of first-hand descriptions of air traffic control procedures and processes (Majumdar and Orcheing, 2004; Histon et al., 2001; VATSIM's Jacksonville ARTCC 2011).

Results - Cognitively Similar Groupings

The results of classification using the SAB approach are groupings of cognitively similar airspace sectors with respect to the considered abstraction types. The degree of similarity can be used to asses good candidate sectors for cross sector controller transfers. If two sectors show significant similarities between their abstraction profiles then they have to potential to facilitate a transfer with minimal re-training. This is because controllers in both sectors are used to operating abstractions under similar operating conditions. However, in the presence of differences between the abstraction profiles, further research needs to be performed to assess the criticality of the difference. Since the characterization factors are not ranked or prioritized, the effect of a difference and the magnitude of a significant difference is not known and must be further researched. Initial progress has been made by the specifications and context groupings of the characterization factors. Since the specification-type factors are generally discrete bits of knowledge, such as frequency values, it is expected that they are easier to train then contextual differences such as operating an abstraction in close proximity to another.

Furthermore, upon the analysis of a larger set of sectors, groups of cognitively similar airspace sectors can be used to create sector classes. An analysis of the emergent differences across these classes, together with the high visibility provided by the bottom up approach, can be used to develop difference mitigation strategies and establish the specific sources of cognitive differences. This is a preliminary but essential step towards realizing the NextGen Generic Airspaces concept. It is expected that supporting the SAB analysis on a NAS-wide scale would be infeasible to perform manually due to its resource intensity. This may be addressed either through automated evaluation of characterization factors or through a complementary top-down classification approach that performs initial classification which is then further analyzed using the Sector Abstraction Binder.

Further Research

This paper presented preliminary developments of the Sector Abstraction Binder - a structural abstraction based framework for establishing cognitive similarities across airspace sectors. The next step is to apply the SAB framework to a comparison of two sectors thereby establishing methods for identifying abstraction instances and evaluating the characterizing factors. Upon identifying emergent differences between the two sectors, the differences need to be verified and significance can be established through controller questionnaires and field interviews. The effects of certain prominent characterizing factors may also be studied in greater detail through experimental evaluation. Finally, methods of automating the SAB evolution such that it can be scaled NAS-wide should be explored or an alternative top-down classification approach should be explored if automation seems infeasible.

Acknowledgements

The authors acknowledge the Federal Aviation Administration for funding this research through FAA Grant 06-G-006.

References

Christien, R. (2003). *Air Traffic Complexity Indicators & ATC Sectors Classification*, Eurocontrol

Bhagat, R.V. & Histon, J.M. (2011). *The Sector Abstraction Binder: A Comparison of the Brewton & Geneva high level airspace sectors*. University of Waterloo: Humans in Complex Systems Lab.

Federal Aviation Administration. (2010A). A Plan for the Future: 10-Year Strategy for the Air Traffic Control Workforce. Retrieved from http://www.faa.gov/air_traffic/publications/controller_staffing/media/CWP_2010.pdf

Federal Aviation Administration. (2010B). FAA's NextGen Appendix B: Implementation Plan. Retrieved from http://www.faa.gov/nextgen/media/appendix_b.pdf

Goldman, S. et al. (2006). *Static Sector Characteristics and Operational Errors*, Federal Aviation Administration, DOT/FAA/AM-06/4, Office of Aerospace Medicine

Histon, J.M. et al. (2001). *Introducing Structural Considerations into Complexity Metrics*, 4th USA/Europe Air Traffic Management R&D Seminar, Sante Fe: New Mexico

Histon, J.M. & Hansman, R.J. (2008). *Mitigating Complexity in Air Traffic Control: The Role of Structure-based Abstractions*. Massachusetts Institute of Technology: International Center for Air Transportation.

Majumdar, A., Orcheing, W.Y. (2004) Factors affecting air traffic controller workload: multivariate analysis based on simulation modeling of controller workload. Transp. Res. Rec 1788, 58–69

VATSIM's Jacksonville ARTCC. (2011). Jacksonville Air Route Traffic Center Forum. Retrieved from http://www.zjxartcc.org/PNphpBB2.html

Yousefi, A. et al. (2003). Investigation of Airspace Metrics for Design and Evaluation of New ATM Concepts, Air Transportation Systems Lab, George Mason University

THE COMPARISON OF AIR TRAFFIC CONTROLLERS' TO COLLEGE STUDENTS' MEMORY PERFORMANCE AND BRAIN ACTIVITIES

Sehchang Hah[1], Ben Willems[1], Hasan Ayaz[2], Scott Bunce[3], Kurtulus Izzetoglu[2], & Atul Deshmukh[4]
[1] Federal Aviation Administration, Atlantic City International Airport, NJ; [2] Drexel University, Philadelphia, PA;
[3] Penn State University, Hershey, PA; [4] Hi-Tec Systems, Inc., Egg Harbor Township, NJ

We compared controllers and college students memory task performance and brain activities while performing the task. The purpose of these comparisons was to test the hypothesis that controllers must have acquired special memorization skills from many years of air traffic control and that their brains would respond differently from college students' brains. To perform the task, they must maintain a sequence of characters in their working memory and manipulate the characters. We compared controllers' brain activities to college students' recorded at the prefrontal cortex with functional near infrared (fNIR) spectroscopy while performing the task. Our results showed that controllers performed significantly better than college students. Controllers and college students also showed distinctly different brain activity patterns. Controllers used the areas of the prefrontal cortex more evenly than college students. We discuss the implications of the group difference.

To become fully certified, air traffic controllers receive training that usually takes more than three years. As they control air traffic, they monitor and collect the relevant information such as aircraft positions, speeds, altitudes, distances to nearby aircraft, fixes, sector boundaries, and aircraft destination. They must maintain all the information current and integrate them to make safe and efficient decisions. We hypothesized that through many years of air traffic control, air traffic controllers developed skills in memorization, manipulation of memorized information, decision making, and control. The prefrontal cortex executes these functions, and the functional Near InfraRed (fNIR) pad is able to record its activities.

In this paper, we compared controllers of Federal Aviation Administration (FAA) Air Route Traffic Control Centers (ARTCCs; also called en route centers) and college students in their memory task performances and brain activities measured by the fNIR technology. As a performance task, we used a standard memory task, the nBack task (Wikipedia, 2011a). In this memory task, the participants viewed a character presented one by one on the display sequentially and identified a target. In the 0Back condition, the target was always X. In the 1Back condition, the target was the character that was shown previously. In the 2Back task, the participant's target was the character shown two characters previously. In the 3Back condition, the target character was shown three characters previously. The target changed constantly, and the participants needed to hold one, two, and three characters current in their working memory in 1Back, 2Back, and 3Back task conditions, respectively.

Researchers have used fNIR technology to study brain activities for about two decades (Cope, 1991; Ayaz, Izzetoglu, Platek, Bunce, Izzetoglu, Pourrezaei, & Onaral, 2006).The technology uses a pad covering the brain, and the pad has two critical parts: light source and light detector. It sends out light to the cortex of the brain and collects reflected light. From this, it calculates oxygenated and deoxygenated hemoglobin levels in the blood, and these levels are directly related to neuronal activities (Bunce, 2006). Specifically, for neurons to function, they need energy, and this comes from glucose. But they must metabolize glucose with oxygen, which is delivered by hemoglobin molecules in red blood cells. As hemoglobin molecules deliver the oxygen and absorb carbon dioxide, they become deoxygenated and change colors. Although most biological tissues are relatively transparent to light in the red and near-infrared range of the electromagnetic spectrum between 700 nm and 1,000 nm, hemoglobin is a strong absorber of light in this range. By emitting light of two wavelengths (one more sensitive to oxygenated hemoglobin and another more sensitive to deoxygenated hemoglobin) and collecting the reflected light, the fNIR system can calculate changes of their concentrations in the blood over time. From this data, scientists deduce neuronal activities.

Even though we used a standardized simple memory task, nBack, not an air traffic control related task, we predicted controllers would perform better in the task. We predicted that both college students' and controllers' oxygenation levels would increase with the difficulty of NBack tasks. We also predicted that controllers' brains would respond differently from college students', and we would see the different patterns of fNIR data by the different levels of nBack tasks (0back vs. higher level nBack tasks) and by different prefrontal cortex areas. For instance, D'Espito (2001) stated that based on fMRI research results, both ventral and dorsal areas of prefrontal cortex would engage in a simple maintenance tasks, but for high load and complex tasks, the dorsal area would engage more heavily. Accordingly, we hypothesized that fNIR data of the 0Back task where participants would need to maintain the target in working memory and recognize it if displayed would be different from those of more difficult nBack tasks where they must not only maintain characters in their working memory but also switch characters and select the target among them in their working memory. We assumed that for the 0Back task, both dorsal and ventral areas would be involved but for the more complex nBack tasks, the dorsal area would be involved more heavily.

We tested our predictions with memory task performance and fNIR data and discussed the results focusing on the group differences and the fNIR technology as a measure of working memory task performance.

Method

Ayaz, Bunce (now at Penn State), and Izzetoglu ran the experiment at Drexel University and Hah, Willems, and Deshmukh (Hi-Tec contractor for the FAA) ran it at the William J. Hughes Technical Center. We used the same task, materials, software, and procedure.

Participants

Nine college students volunteered to participate in the experiment at Drexel University. Their age range was between 21 and 30 years. Twenty-eight Certified Professional Controllers (CPCs) from en route centers participated as volunteers at William J. Hughes Technical Center, but we analyzed data from 24 participants only because we had a system problem with the first group of four participants. The participants had medical certificates that were current within 30 days prior to their participation in the experiment. Their average age was 44 with a range between 24 and 55 years. They had worked as controllers for 20 years on average with a range between 3 and 30 years.

Materials

fNIR pad.

The fNIR pad (see Figure 1; 6 cm x 16.5 cm x .5 cm; 30 grams) covered the participant's forehead only and had four small light emitting diodes (LEDs) in the middle generating peak wavelengths at 730 nm and 850 nm and 10 light detectors (or sensors) surrounding the LEDs at 2.5 cm apart. Low power light from the diodes shone through the skin of the forehead onto the brain, and sensors recorded the reflection. With this arrangement, we collected data from 16 different places called voxels in the prefrontal cortex at the sampling rate of 2Hz.

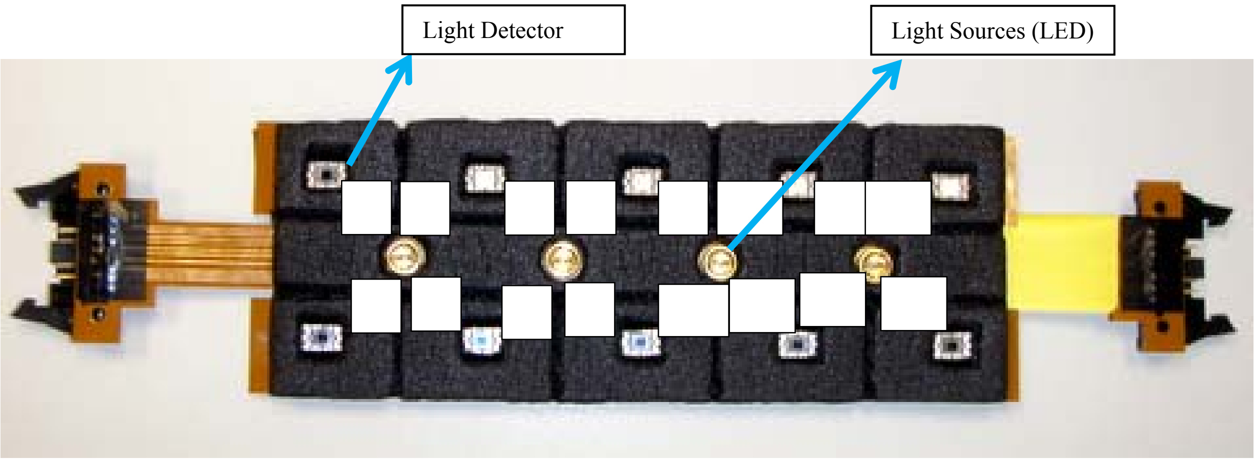


Figure 1. The fNIR pad we used. The numbers represent the schematic locations of voxels in the cortex when it was worn.

The frontal lobe covers about a third of the brain, and the prefrontal cortex encompasses half of the frontal lobe. The prefrontal cortex is responsible for working memory, decision making, strategy generation, and executive functions (Gazzaniga, Ivry, & Mangun, 2002; Gazzaniga, 2004; Saper, Iversen, & Frackowiak, 2000). It connects to perceptual, motor, and limbic areas. It receives projections from wide areas of the brain including the occipital cortex. Even the substructures of the cortex are connected indirectly through thalamic connections. Based on these connections, it controls various brain functions through excitation and inhibition. The fNIR pad is light, small, not-invasive, and portable unlike other technologies such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG).

Because everyone has different sizes and shapes of the forehead, it is difficult to pin-point which voxel data of the fNIR pad would correspond to which areas of the prefrontal cortex. However, in general we assumed it covered Broadmann areas of 9, 46, 10, 11, 47, 45, and 44 (see Figure 2).

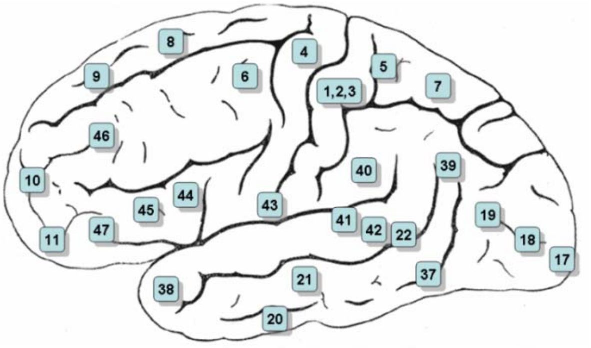


Figure 2. Broadmann map from Wikipedia (Wikipedia, 2011b) (the lateral view of the brain)

nBack memory tasks.

For the memory task, we used four nBack tasks: 0Back, 1Back, 2Back, and 3Back tasks (see Figure 3). The participants viewed a stream of characters displayed on a liquid crystal display (LCD) one by one and responded by pressing the left mouse button for ‘yes’ if the displayed character was the target. In all nBack tasks except the 0Back task, the target changed as the trial progressed, and the participants must refresh their working memory of the characters. For the 0Back condition, the target was always X, and the participants responded if the displayed character was an X. For the 1Back condition, they needed to hold one character in working memory and decide if the current character was the same as the previous one. For the 2Back condition, they must hold two characters in working memory and decide if the current character was the same as the one shown two characters before. After their decision, they must abandon the oldest one and hold the new two characters in their working memory for the upcoming trial. In the 3Back condition, the target character was three characters before. The participants must always hold the newest three characters in their working memory.

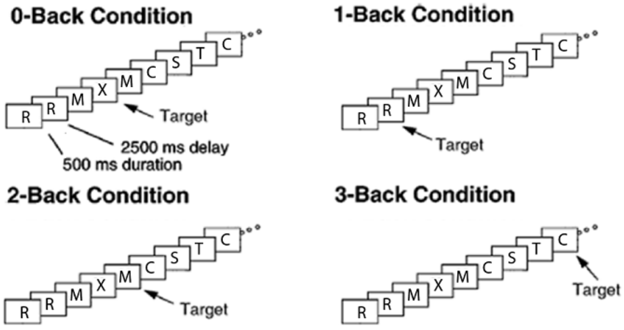


Figure 3. Four nBack tasks showing the target in the sequence.

Procedure

The participants received a short training for all types of nBack tasks before the data collection runs. While they performed the nBack tasks, the participants wore an fNIR pad. They had 28 blocks, that is, 4 nBack types x 7 repetitions that were presented randomly. Each block had 20 trials. For each trial, the participants decided if the character on the display was a target or not. The participants pressed a left mouse button for a target. They did not respond for characters that were not targets. Thus, each participant made 560 decisions. At the start of each block, ‘relax’ was shown on the display for 7.5 sec. Then, which type of nBack trials to be followed such as ‘3back’ was shown on the display for 2.5 sec. Following that display, a stream of characters followed. Each character was displayed for .5 sec with a 2.5 sec interval between characters. This was repeated 20 times. Then, the next run started. The total length of the experiment was just over an hour depending on the speed of the participants’ responses. The sequences of the characters and nBack types were pseudo-randomly presented and were the same for all participants.

Results

We used non-parametric Mann-Whitney U tests using ranks to compare college student and controller nBack performance and brain activities shown in the fNIR data. The controllers performed better than college students with a higher hit rate: 0Back (z = -.4.881, p < .001), 1Back (z =-3.058, p = .002), and 2Back (z = -2.008, p = .045). There was no group performance difference for the 3Back. For the false alarm rate, there was no performance difference between them for any of the nBack tasks. Hit rates declined and false alarm rates rose as the number of characters to be remembered increased. The error bar in Figures 4 and 5 represents one standard error.

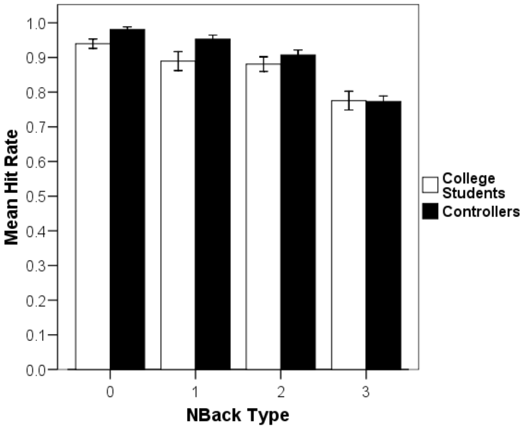


Figure 4. Hit rates by nBack types.

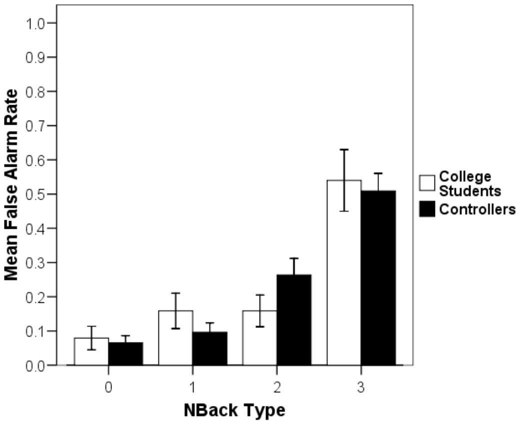


Figure 5. False alarm rates by nBack types.

Controller brain activities shown in the fNIR data were more even across voxels for all NBack tasks than college student brain activities (see Figure 6). Also, as shown in the 0Back figure, controller fNIR levels were generally higher than college student levels. We considered 0Back task as the baseline task.

Since controllers had more heightened oxygenation levels for the baseline (0Back) than college students, it was not meaningful to compare their oxygenation levels with college students’ levels directly at different nBack task conditions. We decided to compare college students’ to controllers’ incremental effort as the task became more difficult. We subtracted 0Back oxygenation levels from 1Back, 2Back, and 3Back oxygenation levels for each group, respectively. In general, college students’ incremental effort was larger than controllers’ in all nBack tasks. To test this statistically, we averaged increases of individuals at each voxel for each group. Then, we compared two groups across 16 voxels for each of (*NBack – 0Back*) separately using Mann-Whitney U tests. The results showed college students exerted more effort than controllers for *1Back-0Back* (z = -3.279, p = .001) and *2Back-0Back* (z = -4.033, p < .001). The difference at *3Back-0Back* between the two groups was not significant.

Because the fMRI research results showed the different roles of dorsal and ventral areas of prefrontal cortex (D’Espito, 2001), we examined if there would be different oxygenation patterns collected from the upper and lower parts of the pad . The upper part represented voxels 1, 3, 5, 7, 9, 11, 13, and 15. The lower part represented the rest of the voxels. We assumed the upper part would receive responses from the dorsal areas mostly and the lower part would receive responses from the ventral areas mostly. Figure 7 shows the differential oxygenation levels at the upper and at the lower parts of the pad. The results did not show the expected trend, that is, there was no trend of higher oxygenation at the dorsal area for more difficult nBack tasks.

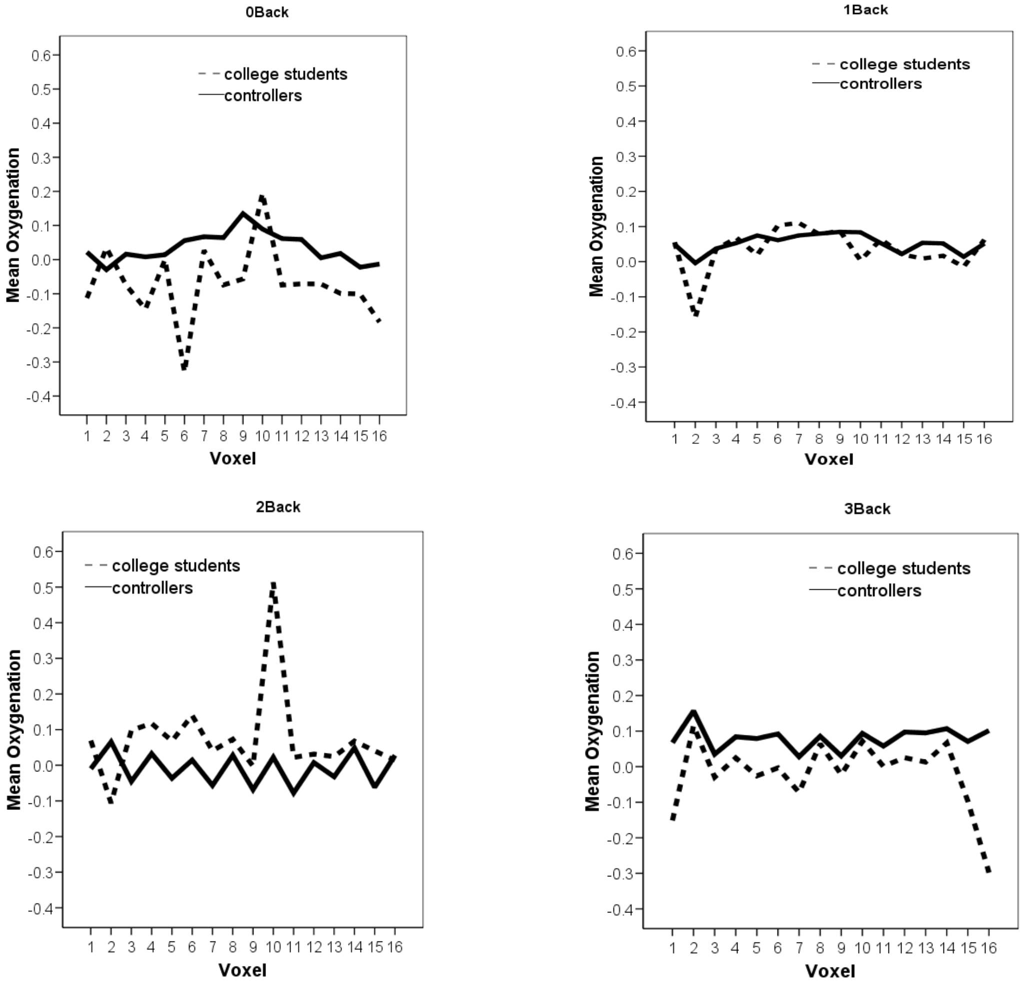


Figure 6. Mean oxygenations at voxels by different NBack tasks of college students and controllers.

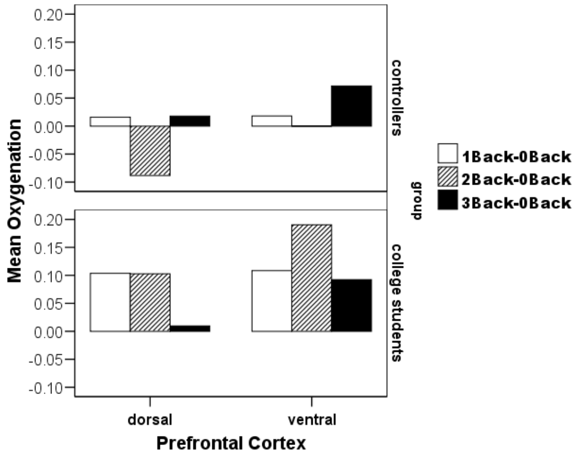


Figure 7. Oxygenation levels at dorsal and ventral areas for (1Back-0Back), (2Back-0Back), and (3Back-0Back) tasks of college students and controllers.

Discussion

We compared controllers and college students memory task performance and their brain activities. Our experimental results showed the controllers performed better in all nBack tasks except the 3Back task. We assumed the controllers must have acquired skills relevant to memory tasks in their job. In the 3Back task, both groups did not perform well. Their hit rates were below 80%, and noise in the data might have contributed to the failure of revealing the groups' difference. Their false alarm rates were also close to 50% in the 3Back task.

We expected that oxygenation measured by fNIR would be higher with more difficult NBack tasks. But our results did not support it. We also expected that the contribution of the dorsolateral prefrontal cortex would increase with the difficulty of the nBack tasks because the participants must manipulate characters in their working memory in addition to maintaining them in the high-level nBack tasks. Our results did not support this either. One possible reason is that to perform the nBack tasks, the participants might have used other parts of the brain extensively, which our fNIR pad did not cover, and the role of the prefrontal cortex may not be that significant. However, our fNIR data showed controllers spent less incremental effort than college students as the difficulty of the nBack tasks increased. As Posner and Raichle (1997, p. 244) mentioned, more repetitions (that is, higher skills) lead to less effort. This particular result is in line with the performance results that controllers performed better than college students.

It was intriguing to find in our fNIR data that the controllers responded evenly across the prefrontal cortex in contrast to college students. This needs further research such as the effect of aging, but it may imply that controllers have used all parts of the prefrontal cortex for all of the nBack tasks. This may implicate that their frontal cortex had developed a tight and well-connected organization, and one part may be closely linked to the other parts of the prefrontal cortex. Since they exhibited superior performance over college students, we conjecture that the presumed tightly organized prefrontal cortex like our controllers may be a developed form of the prefrontal cortex in performing memory tasks such as nBack tasks. This has a very significant implication in training and system designs. When researchers devise training methods and systems, they use results of the past research that used ordinary adults as the participants, not necessarily the skilled and domain specific adults such as air traffic controllers. We conjecture that some assumptions of human capabilities and limitations known to us may not be applicable to skilled populations such as air traffic controllers.

Acknowledgement

We express our appreciation to the 28 participant controllers and nine college students who volunteered for this study. We thank Dr. Mike McAnulty and Dr. Eric Neiderman for their reviews and suggestions for the paper.

References

Ayaz, H., Izzetoglu,M., Platek, S. M., Bunce, S., Izzetoglu, K., Pourrezaei, K., & Onaral, B. (2006). Registering fNIR data to Brain Surface Image using MRI templates. *Proceedings of the 28th IEEE EMBS Annual International Conference*. 2671-2674.

Bunce, S. C., Izzetoglu, M., Izzetoglu, K., Onaral, B., & Pourrezaei, K. (2006). Functional Near-Infrared Spectroscopy; An Emerging Neuroimagining Modality. *IEEE Engineering in Medicine and Biology Magazine*, July/August.

Cope, M. (1991). *The development of a near-infrared spectroscopy system and its application for noninvasive monitoring of cerebral blood and tissue oxygenation in the newborn infant*. London: University College.

D'Esposito, M. (2001). Functional Neuroimagining of working memory. In Cabeza, R. & Kingstone, A. (Eds.), *Handbook of Functional Neuroimaging of Cognition*. Cambridge, MA: The MIT Press.

Gazzaniga, M. S. (2004). *The cognitive neurosciences.* (3rd ed.). Cambridge, MA: The MIT Press.

Gazzaniga, M. S., Ivry, R. B., & Magnun, G. R. (2002). *Cognitive neuroscience: The biology of the mind.* (2nd Ed.). NY: W. W. Norton Co.

Posner, M. I. & Raichle, M. E. (1997). *Images of Mind.* NY: Scientific American Library.

Saper, C. B., Iversen, S., Frackowiak, R. (2000). Integration of sensory and motor functions: The association areas of the cerebral cortex and the cognitive capabilities of the brain. In Kandel, E. R., Schwartz, J. H., Jessell, T. M. (Eds.), *Principles of Neural Science.* (4th ed.). NY: McGraw-Hill.

Wikipedia. (2011a). *n-back.* Retrieved from http://en.wikipedia.org/wiki/N-back

Wikipedia. (2011b). *Broadmann map*. Retrieved from http://en.wikipedia.org/wiki/File:Gray726-Brodman.png

FACTORS INFLUENCING FEMALE AVIATION PROFESSIONALS' CHOICE OF CAREER

Gail Zlotky and Wendy S. Beckman
Middle Tennessee State University
Murfreesboro, Tennessee

Despite much attention, there continues to be a low level of female participation in the aviation industry. In an earlier qualitative study of female aviation students, it was found that having parents who were supportive of their daughter's career choice was the most influential factor in student career decision making. As a follow up to this earlier study, a Likert-scale type survey was distributed electronically to both active female aviation professionals and to current female aviation students. The survey allowed for participant identification of the factors that were most influential for them both choosing and remaining in an aviation career. Factors such as influence from parents and other significant adults, previous exposure to the career field, education and training requirements, and lifestyle factors were all evaluated. Supporting the previous study, parental support was found to be the most influential factor in a female's choice of an aviation career.

The attraction of female students into the traditionally male-dominated fields of science, technology, engineering, and math (STEM) is 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). This bill was subsequently reauthorized in January of 2011 (America Competes Reauthorization, 2011), signifying that progress needs to continue in this area. 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 is 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, 2009a, 2009b). For example, there were only 5,636 female certificated Airline Transport Pilots in the United States in 2009, out of a total of 144,600 Airline Transport Pilots. This means that only 3.9% of the pilots at the highest level of certification are women. The Commercial rated pilot numbers are slightly better, but even then only 6.6% of the Commercial pilots are female; similarly, only 6.7% of the nation's Certified Flight Instructors (CFIs) are women. The aviation maintenance area is even more dismal, with only 2.1% of the certificated A&P Mechanics being women. One area that is slightly higher is certificated Dispatchers, where 16.8% of those certified are women; but even this is obviously far below the level of parity with men.

Perhaps even more sobering than today's statistics is the realization that little improvement has been made in the last decade. In 2000, 3.1% of the nation's ATP's, 4.7% of Commercial pilots, 6.4% of the CFI's, 1.5 % of A&P Mechanics, and 12.6% of Dispatchers were female (Federal Aviation Administration, 2009a, 2009b). When these figures are compared to the 2009 data, it is obvious that efforts to date to increase female participation in aviation careers have not been highly successful.

Numerous studies have been conducted to attempt to determine how to improve female representation in traditionally male fields. Some of these studies 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. In a 2009 study regarding female flight students' perceptions of gender biases at their institutions, Depperschmidt & Bliss found that female flight students felt there were not enough female professionals employed at their institution to serve as role models for female students. In another 2009 study, the factors that either encouraged or discouraged female students from enrolling in the Aerospace Department at Middle Tennessee State University (MTSU) were explored (Zlotky & Beckman, 2009). This study found that the majority of the study participants had parents that strongly encouraged education and supported their daughter's career choice. It was also found that many female students had someone outside the family that encouraged them to pursue higher education, although not specifically an aviation career. In fact, most of the students interviewed did not know anyone working in aviation prior to pursuing a collegiate aviation degree. Finally, it was found that about half of the students interviewed had encountered others with negative stereotypes about females in aviation. This was consistent with the findings of the Depperschmidt & Bliss study.

Methodology

Since the research that has been done to date on female attraction to and retention in aviation careers has focused primarily on students, it was decided that it would be a logical next step to determine what factors influenced currently employed female aviation professionals to enter their career field. This data could then be compared and contrasted with the factors that influenced female aviation students to pursue their career field. To carry out this study, approval was granted from the MTSU Institutional Review Board (IRB) to conduct a human subject research study of female aviation professionals and students. A survey was created using the online survey website, SurveyMonkey. The survey was distributed to aviation professionals via two methods. First, the organization Women in Aviation, International provided a link to the survey on their February 2009 website, and also posted the survey link on their blog. Second, the electronic aviation newsletter, "Flight Safety Information", distributed by Curt Lewis, provided a link to the survey in the February 15th edition of the newsletter. There were a total of 30 female aviation professionals who responded to the survey from these two sources. An identical survey was distributed via e-mail to current female aviation students at MTSU. There were 24 students who responded to the survey.

The survey contained several demographic questions to determine the basic backgrounds of the individuals completing the survey, including the field of aviation in which they are currently employed or studying, and (for the professionals) their number of years of experience in their field. The rest of the survey consisted primarily of Likert-scale questions. The first set of questions dealt with the influence of the participants' parents (or primary care-givers) on their choice of career. Items such as the participant's relationship with their parents as a child and adolescent, their parent's personality type, and how their parents felt about their choice of aviation as a career field were all explored. Next, the influence of other individuals on the participant's choice of career field was examined, including that individual's relationship to the participant, if that person was involved in the aviation industry, and if their influence was positive or negative. Then, a series of questions for the professional participants attempted to discover the level of difficulty females experienced in entering their chosen aviation career. Items such as the level of difficulty they experienced in training, how well their education prepared them for their career, how well they feel they have been accepted by their peers, the level of difficulty they have experienced in being promoted, the challenge of balancing a career and family, and the pressure they feel to conform to traditional roles were all evaluated. While the student surveys also contained this section, their responses were not used in the data evaluation, since they largely have no industry experience upon which to base their responses.

Data Analysis

As indicated above, there were 30 female aviation professionals and 24 female aviation students who completed the survey. The first set of questions addressed the participant's relationship with their parents or primary caregivers. The answer choices for these questions were "excellent," "good," "neutral," "poor," or "bad." As can be seen in Table 1, a high percentage of both aviation professionals and aviation students indicated that their relationship with both of their parents (or primary caregivers) was either “excellent” or “good”.

Table 1

Percentage of Participants Indicating Their Relationship With Their Parents as Children and Adolescents Were Either “Excellent” or “Good”

	Professionals	**Students**
Question Stem	Percent "Excellent" or "Good"	Percent "Excellent" or "Good"
Relationship with mother as child	83.4	81.9
Relationship mother as adolescent	60.0	77.3
Relationship with father as child	75.8	86.4
Relationship with father as adolescent	73.4	72.7

The participants were then asked to categorize both their mother and father’s personality as either very strong, strong, average, mild, or meek, and to then indicate whether or not they felt each parent’s personality had any effect on their choice of career. The percentages of participants that indicated their parents’ personalities as either very strong or strong can be seen in Table 2.

Table 2

Participant Indications of Parent’s Personality Type

	Professionals	**Students**
Question Stem	Percent "Very Strong" or "Strong"	Percent "Very Strong" or "Strong"
Mother’s personality type	73.3	72.7
Father’s personality type	44.8	68.2

When asked about the influence of their mother’s personality on their career choice, the aviation professionals most frequent response was “not at all” (40.0%) while for aviation students the most frequent response was either “somewhat” (27.3%) or “not at all” (27.3%). Regarding the influence of their father’s personality on their career choice, professionals’ most frequent response was “strongly” (33.3%), and students’ most frequent response was “strongly” (40.9%) as well. When asked about parental support of their career choice, the majority of respondents indicated that their parents were either strongly supportive or supportive of their career choice, as can be seen in Table 3.

Table 3

Participant Indication of Parental Support of Their Career Field

	Professionals	**Students**
Question Stem	Percent "Strongly Supportive" or "Supportive"	Percent "Strongly Supportive" or "Supportive"
Mother's support	80.0	86.3
Father's support	76.6	77.3

A majority (63.3%) of the professionals indicated there had not been a person outside their immediate family who was influential in their growth and development, while 54.4% of the students indicated the same. However, for those that did have an outside influence, this individual was largely indicated as being either "very positive" or "positive" in encouraging the participant to choose an aviation career, with 100% of the professionals indicating this to be the case, and 90% of the students indicating the same. Both groups indicated there was at least one family member (most often an extended family member) who did not support their career choice, with 54.5% of the professionals experiencing this, and 80.0% of the students experiencing it. Since this was one of the few areas of difference between the two groups of participants, a chi square test was performed to see if the difference was significant, and it was found to not be.

Most of the participants did not choose an aviation career based on having a family member in the industry, with 73.3% of the professionals and 68.2% of the students not basing their decision on this factor. However, for those that did have a family member in the industry, those family members were a positive influence on their decision to enter the aviation career field. The professionals indicated that 87.5% of the family members in the industry were either "strongly positive" or "positive" in terms of influence, while 100% of the students indicated the same. For both groups, this individual was likely to be their father (71.4% of professionals and 83.3% of students). The most frequent response of age range at which interest in aviation first occurred was 11-15 for the professional participants, while it was the 16-20 age range for the student participants.

The last series of questions was analyzed only for the professional participants, as students who have not yet started their careers could not answer the questions in an informed manner. A strong majority of professionals felt that their education had prepared them for their career either "very well," "well," or "adequately" (93.3%), while 96.6% of the professionals felt either "very well," "well," or "adequately" accepted in their career field. However, 60% of participants found it was either "very difficult" or "difficult" to enter their career field, and 63.4% experienced either "quite a bit" or "some" difficulty becoming accepted in their field. When asked about the level of difficulty experienced in being promoted, 40.0% of respondents indicated promotion was either "very difficult" or "somewhat difficult." The most frequent response to a query regarding the challenge of balancing their career and family life was "neutral" (43.3%), while the most frequent response to a question regarding the pressure felt to conform to the aviation field's traditional roles was "somewhat." The majority of respondents (56.7%) felt either "very strongly," "strongly" or "somewhat" that the expectations for personnel in aviation careers should change. The mode for the range of years of experience for the professional respondents was 16-20 years.

The open responses provided by participants in the "comments" section provided insights into the experiences of some of the participants. Such comments included: "One employer no longer wanted to work with me when I became pregnant because he did not believe women with children should work outside the home;" "As a woman, I have to prove myself beforehand, but men get promoted on potential and then show they can handle it afterwards;" and, " Career can consume you. It is easy to become overloaded. Very hard to put priorities in order sometimes." But, one participant acknowledged, "As hard as it is to be promoted as a woman, there is the flip side, that it is much more accepted for women to balance their career/family lives than it is for men. It is always frowned on when my husband has to take time off for the kids."

Discussion

It was interesting that very few differences were found between student and professional responses. For both groups, the factor that most influenced their choice of aviation career was having parents that were supportive of this career choice. Most participants reported having positive relationships with both of their parents, and it was noticed that the majority of respondents indicated that their mother's personality was stronger than their father's. While the majority of participants had no significant influence outside their parents in choosing their career, there were still a number of participants that did have this outside influence (36.7% of professionals and 45.5% of students). When there was an outside influence, this individual was very influential in causing the participant to choose an aviation career. Interestingly, it was found that a far greater percentage of professionals (80.0%) than students (54.4%) had experienced a family member who was not supportive of their career choice. Since the students are obviously of a younger generation than the professionals, this does seem to demonstrate a shift in attitudes for the better. While most participants did not have a family member in the aviation industry, those that did reported that they received positive influence from this individual. These findings are all in support of the earlier qualitative study that was conducted utilizing MTSU undergraduate aviation students, and indicate that supportive parents and other influential individuals in young females' lives are the single most important factor in influencing the choice of an aviation career.

The most serious limitation of this study was the low response rate. It had been hoped that at least 100 aviation professional participants would be obtained, along with 50 aviation student participants. However, the point of this study was to extend the understanding of the strongest influences on females to entering the aviation industry, beyond what was previously learned from a limited qualitative (interview-based) study of MTSU female students. This study did indeed support the findings of the previous study. As a next step, the responses of the female aviation professionals in this study will be compared to those of male aviation professionals, in an attempt to identify the differences that exist between these two groups.

References

America COMPETES (Creating opportunities to meaningfully promote excellence in technology, education, and science) Act of 2007, Pub. L. No. 110-69, (2007).

America COMPETES Reauthorization Act of 2010, Pub. L. No. 111-358, (2011).

Chavanne, B. (2008). Sea change. *Aviation Week and Space Technology, 169(7)*, 81-82.

Depperschmidt, C.L. & Bliss, T.J. (2009). Female flight students: perceptions of barriers and gender biases within collegiate flight programs. *Collegiate Aviation Review, 27(2),* 18-27.

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.

Federal Aviation Administration. (2009a). Table 1. Estimated active airmen certificates held: December 31, 2000-2009. Retrieved from http://www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics/2009/

Federal Aviation Administration. (2009b). Table 2. Estimated active women airmen certificates held: December 31, 2000-2009. Retrieved from http://www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics/2009/

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.

Zlotky, G. and Beckman, W. (2009). Family factors influencing female aerospace student's choice of major. *Proceedings of the 2009 International Symposium on Aviation Psychology,* 414-419.

PERSONALITY TYPE PREFERENCE ASSESSMENT AS A COMPONENT OF PILOT TRAINING

Gil Sinclair
Western Michigan University
Kalamazoo, MI
Tracey Moon
Western Michigan University
Kalamazoo, MI

In aviation education and training, students experience a number of tests to determine their technical abilities and theoretical knowledge. When piloting an aircraft, many stage checks and examinations must be completed before a license is issued. Even then, the testing does not stop; pilots in commercial aviation are subject to a whole range of periodic checks in simulators and in the air. During these evaluations, shortcomings in technical abilities can be identified and corrective action taken before retesting. However, during testing, candidates are especially careful to minimize any tendencies they may have under normal circumstances to, for example, make a short cut in a checklist. Hazardous thought patterns are unlikely to surface under test conditions. In the past, numerous personality tests have been administered to pilots in attempts to determine the likelihood of success. Conversely, very few studies have been done to use such tests to attempt to determine the likelihood of failure or, even further, to determine the most likely type of failure. This paper discusses the possibility of using a personality type preference assessment (Myers-Briggs Type Indicator (MBTI®)) to determine the most likely type preferences conducive to failure (e.g. impulsiveness) of individuals and using the information to develop interventions at an early stage in training.

Flying can be a dangerous game in which human error, if not managed successfully, can lead to disastrous consequences. Crew Resource Management (CRM) uses what is already known about human error and combines it with accident analyses and critical incident reports to develop methods to reduce the incidence and consequences of such errors to a minimum. CRM advocates “error management” to avoid errors, trap errors and mitigate errors. CRM assumes that human error is inevitable, but the first line of defense is avoidance. In a perfect world, the only errors that cannot be avoided are those that cannot be predicted. Unfortunately, it is not a perfect world and humans will still occasionally commit those errors, intentional or not, that we do know are possible.

Many studies have been conducted regarding human error and although this has resulted in extensive knowledge of contributing factors, total elimination is still not possible and probably never will be. Several approaches to human error have been developed, including the “**person model**”, which is a traditional approach to human error that "names, blames, and shames" one or more individuals as "causing" the accident. The underlying assumption is that mistakes and errors are the result of faulty, negative mental processes such as negligence or lack of skill. This model uses fear and discipline to attempt to improve safety and considers errors a "moral" issue that "bad things happen to bad people” (Reason, 2000). Another model is the “**system model”** which recognizes that there are systemic contributions to and causes of error. It acknowledges that organizational culture, human-to-system interface design, and environmental elements can create "latent failure" conditions which contribute to human error. The model recognizes human limitations and accepts that human error is inevitable, and thus concludes that systems should be designed to anticipate human error and to mitigate the consequences (Reason, 2000).

If every effort is made to prevent the occurrence of human error, then inevitability can be reduced somewhat. Current CRM training seeks to encourage the avoidance of human error by some simple and common sense measures, such as complete pre-flight and in-flight crew briefings or strict adherence to Standard Operating Procedures (SOPs). Adherence to these CRM principles helps to maintain situational awareness on the part of the whole crew (FAA), which also reduces the inevitability of the errors that may be likely.

The International Civil Aviation Organization (ICAO) Safety Management Manual cites the need for intervention strategies in safety management systems (SMS). Safety risk management in SMS operations builds upon a system design in which appropriate safety risk controls intended to eliminate or mitigate the consequences of anticipated hazards are embedded in the system. Identification of these is therefore the first step in a formal process of collecting, recording, acting on and generating feedback about hazards and safety risks in operations (ICAO, 2009). This paper is concerned about dangers from individuals who, although seemingly professional in their work attitudes, may nevertheless be prone to particular types of hazardous thought patterns due to their fundamental personality type preference. If correlations can be made between personality type preferences and types of errors, then intervention during training to bring awareness of these correlations may reduce the number of occurrences.

Personality profiles concept

Profiles administered in an effort to determine psychological type concentrate on identifying *traits* and measuring the strength of these traits, and are often used as a predictor of suitability for certain career paths. However, traits are the result of quantitative analysis, using continuous data, and have scores in relation to a comparative group. When graphed, the results of such measurements usually form a bell curve and often the greatest importance is given to how close or far an individual is from the mean. These tests are often used diagnostically and results are described in terms of means and standard deviations. In comparison, instruments such as the Myers-Briggs Type Indicator (MBTI®) assess *type preferences* – a description of an individual's *preferred* way of behaving or acting. These assessments do not measure; instead, they *sort* into categories. Furthermore, they do not indicate a strength of a particular preference but, instead, a *clarity* which merely indicates how certain a person is of his or her preference at the time he or she took the test. Type assessments are qualitative and results should never be used diagnostically (Briggs-Myers, McCaulley, Quenck, & Hammer, 1998) (Briggs-Myers, McCaulley, Quenck, & Hammer, 1998).

The MBTI® assessment is based on a theory of personality type developed by Carl G. Jung. Jung's theory is intended to explain the normal differences between healthy people and, based on his observations, led him to conclude that differences in behavior result from inborn tendencies to use one's mind in different ways. More importantly, Jung theorized that as people act on these tendencies, they develop predictable patterns of behavior. Jung observed that active minds are involved in two mental activities: taking in information (perceiving) and organizing information and reaching conclusions (judging). Jung defined two opposite ways that people perceive (sensation and intuition, the S-N dichotomy) and two ways of judging (thinking and feeling, the T-F dichotomy). He also observed two opposite ways in which people focused their energy, basically outward (extraversion) or inward (introversion) or the E-I dichotomy. Briggs and Myers further developed Jung's theory and added another dichotomy describing how people orient themselves to the outer world, or organize – the judging-perceiving dichotomy (J-P) (Myers, 1998).

The MBTI® instrument has undergone decades of reliability and validity testing. Test-retest reliabilities show consistency over time, with levels of agreement much greater than by chance. Changes, although rare, are most likely to occur in only one preference pair, usually in cases where the original clarity was low. Evidence exists for the validity of the four preference scales along with evidence for the validity of whole types. Detailed descriptions of research into reliability and validity may be found in the MBTI® manual (Briggs-Myers, McCaulley, Quenck, & Hammer, 1998) (Briggs-Myers, McCaulley, Quenck, & Hammer, 1998).

Personality tests may be administered at various stages of career progression – training, pre-employment, concurrent and post-employment and are often used for selection and placement. However, the MBTI® tool is bound by a strict code of ethics which prohibits its use for evaluation of performance and suitability for employment. Administration of the assessment must be purely voluntary and results are confidential between the responder and the practitioner. Only those properly certified as MBTI® practitioners may administer the instrument and conduct the interpretation. Furthermore, individuals are subject to type development as they mature. This is not to say that their four-letter type is going to change as they grow older, but merely that they develop their less preferred side and learn when to use them. In fact, people of the same four-letter type may well exhibit different behaviors.

Use of personality studies in aviation

Personality characteristics have historically been used in attempts to predict those who would be successful and safe pilots. Bearing in mind that personality is the characteristic way in which a person normally thinks, feels and behaves (American Psychiatric Association, 1980), it is difficult to apply a personality characteristic to success in aviation due to the stressful nature of the field. In fact, several studies have failed to find a relationship between pilot personalities and success in pilot training programs (Dillinger, Wiegmann, & Taneja, 2003). The Army Air Force's aviation psychology program conducted one of the earliest studies to identify personality characteristics that would predict aviation performance (Guildford, 1947). In general, the personality measures used did not predict success in primary flight training. By far the majority of similar studies have been done using data on military pilots and very few studies have been conducted using data on commercial pilots. Furthermore, a study completed in 1983 (Ramachandran, Wadhawan, Kumar, Chandramohan, & Rao, 1983) found that the personality profiles of commercial aviation pilots differ from those of military pilots. However, a study carried out by a team from the University of Illinois, Human Factors Division (Dillinger, Wiegmann, & Taneja, 2003) concluded that civil aviation pilots have different personality characteristics than non-pilots.

Do some people cope better with stressful situations than do others, and can behaviors linked to certain personality traits actually be changed through intervention and training? Indeed, many intervention strategies have been applied in the field of aviation to try and reduce the occurrence of human error, mostly in the form of improved ergonomic design, improved training and hazard awareness. The FAA has long used awareness training of five hazardous attitudes – anti-authority, invulnerability, impulsiveness, machoism and resignation – in its aeronautical decision making (ADM) training guidelines. However, few studies have investigated whether or not any links exist between personality characteristics and a tendency to display these behavioral characteristics.

Dr. Ganesh A and Dr. Catherine Joseph provide an excellent overview of personality studies in aircrew (A & Joseph, 2005), using Raymond Cattell's definition of personality as "that which permits a prediction of what a person will do in a given situation". When discussing the historical background of aviation psychology, they point out the results of two studies as being contradictive. One described successful pilots as "high-spirited, happy-go-lucky sportsmen" while another described the best aviators as "quiet and methodical men" (Hunter & Burke, 1995). Later tests, using different methods, portrayed aviators as dominant, confident, outgoing and stable. A 2002 study of military trainees (Berg, Moore, Retzlaff, & King, 2002) identified three different types of military pilot and classified them as "typical" (achievement oriented, dominant, affable and stable), "the right stuff" (similar but also aggressive, self-aggrandizing and exhibitionistic) and "the wrong stuff" (cautious, compulsive and socially retiring). A further study conducted on experienced military pilots (Picano, 1991) identified three very similar groups. These two studies do not indicate any stereotype of successful military pilot personality.

The cost of training pilots is very high, thus the military has spent a great deal of time and money in developing and researching methods to select those candidates likely to be successful. However, the studies just described show that this is difficult since no two successful pilots are the same. With the increasing population of female aviators, a recognition of gender differences is also needed. Studies have been conducted on female pilots and one (Novello & Youssef, 1974) concluded that the personality profile of female pilots was more similar to the male pilot profile than to any norms established for U.S. adult males or females. Other studies showed that female pilot profiles differed only slightly from male pilots but differed greatly from non-pilot females. It is generally accepted that selection can be done with personality testing, but only to reject the obvious, such as those with emotional instability, high anxiety or extreme impulsiveness.

Companies in the U.S. and in Europe that offer sponsorship or scholarships for commercial pilot training generally use selection tests to determine aptitude rather than personality. However, the requirements for being accepted into most commercial pilot training programs in the U.S. are medical, as regulated by the FAA. For example, at the institution where the authors of this paper are employed, the only requirement for enrollment is a FAA 2nd class medical and a suitable GPA and ACT score. In the authors' experiences, a whole variety of personality types have found success whereas others, with the same types, have not.

Studies have been conducted to determine relationships between personality and involvement in mishaps. Evidence indicates that many "accident-prone" pilots do indeed share particular traits such as inadequate stress coping ability and blaming of others (Frank, 1981), (Alkov, Giaynor, & Borowsky, 1985). Although the five hazardous attitudes have been found to correlate with certain personality characteristics (Stokes & Kite, 1994) opinions are divided as to whether or not these attitudes are changeable or are traits that will resist change (A & Joseph, 2005).

In aviation training, a method for identification of personality traits or type preferences of those pilots who are most likely to be involved in any kind of mishap would be desirable, allowing early intervention. The goal would be to make individuals aware of their own potential pitfalls and recommend actions to enable them to avoid errors. Some factors already identified in the failing aviator are excessive aggressiveness and resentment of authority (Voge, 1989).

Hypothesis

This paper hypothesizes that certain behaviors, in particular inadvertent or intentional disregard for policies or procedures (rules!) and attention to detail, are more often observed among those with a particular personality type preference. Evidence of this theory would allow intervention at an early stage of pilot training.

Proposed study method

Currently, students in the College of Aviation are asked to take the MBTI® assessment as part of their Crew Resource Management class. Thus far, no student has refused to take the assessment and depersonalized data has been recorded; types are recorded in relation to gender and program only, although the names are also stored for use only by the instructor for class assignment assessment purposes. In addition, a comprehensive safety reporting system has been developed which records details of accidents and incidents categorized by major cause. These two sets of data may be processed to produce a single data set of pilot personality type preference versus mishap cause, in particular policy disregard and attention to detail.

Preliminary results

A preliminary collation of existing data has shown signs of correlation between the results of student pilots' MBTI® assessments and the College safety records. However, this data has also revealed some other interesting tendencies which may necessitate a different approach toward data analysis. Table 1 shows that 62.32% of the 337 students (210) were in the flight program whilst 37.39% (127) were in the management program. Since only around 10% of the management students complete actual flight training, the investigators are considering basing the study only on those in flight.

Table 1

Initial Data Analysis by Type Preference

Total sample 337 Flight 210
Admin 126

	Flight Ad	min		Flight only
ISTJ 8.	31%	5.93%		13.33%
ISTP 8.	90%	4.15%		14.29%
ESTP 8.	61%	3.86%		13.81%
ESTJ 4.	45%	6.23%		7.14%
ISFJ 2.	67%	2.37%		4.29%
ISFP 4.	15%	1.19%		6.67%
ESFP 5.	04%	2.08%		8.10%
ESFJ 2.	08%	1.78%		3.33%
INFJ 0.	59%	0.30%		0.95%
INFP 2.	97%	1.48%		4.76%
ENFP 3.	86%	3.26%		6.19%
ENFJ 0.	89%	0.00%		1.43%
INTJ 1.	19%	0.30%		1.90%
INTP 2.	97%	1.19%		4.76%
ENTP 4.	15%	2.67%		6.67%
ENTJ 1.	48%	0.59%		2.38%
	62.31% 37.	39%		100.00%

Table 1 shows that, of these, 14.29%, 13.81% and 13.33% respectively are ISTP, ESTP and ISTJ. The next largest group is the ESTJs at 7.14%. Thus these types are more likely to be the subject of a safety report purely by being in the majority in the program. Initial data gathered from safety reports revealed only 52 incidents with a the relevant factor(s). However, data collection only began recently and the available sample is not yet large enough to be able to draw any valid conclusions to prove or disprove the hypothesis.

Discussion

The theory to be applied here is that type preference is different from behavior because attitudes shape behaviors. The behaviors that one's type preference tends to produce needs to be examined to determine whether or not it is appropriate, so that the individual can make appropriate behavioral adjustments to function properly and cooperatively in society and in the workplace. Before behavior can be changed, individuals need to be aware of why it needs to be changed, what the benefits are and how changes are made. The MBTI® instrument is not the final answer, but it is a start and could be a useful tool at an early stage. The MBTI® assessment is often used in career advising to determine suitable careers. It can also be used to raise awareness of how individuals need to change their outward behavior and possibly their inner thought processes in order to be successful in their chosen field.

The authors are already using the MBTI® tool in the classroom to raise awareness in aviation students of their type profile and what it means. The students are taught how it may affect their interaction with others and how they can make an assessment of the preference types of others and compensate accordingly. They are also asked to analyze their results and determine what kind of hazardous thought patterns their type may encourage and learn to detect their onset and apply their own antidote. This fits with CRM philosophy in error management by providing students with tools to avoid errors they may commit.

Conclusions

At this point, the authors have concluded that there is insufficient data to either prove or disprove the original hypothesis, however some interesting possibilities for other studies have been revealed. The intention is to continue to gather date related to the original hypothesis whilst exploring other areas of study using data already available.

Bibliography

A, G., & Joseph, C. (2005). Personality studies in aircrew; an overview. *Indian Journal of Aerospace medicine , 49* (1).

Alkov, R., Giaynor, J., & Borowsky, M. (1985). Pilot error as a symptom of inadequate stress coping. *Aviatin, Space and Environmental medicine , 56* (3), 244-47.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: American Psychiatric Associaition.

Berg, J., Moore, J. L., Retzlaff, P., & King, R. (2002). Assessment of personality nand crew interaction skills in successful naval aviators. *Aviation, Space & Environmental Medicine , 73* (6), 575-59.

Briggs-Myers, I., McCaulley, M. H., Quenck, N. L., & Hammer, A. L. (1998). *MBTI®* (Third ed.). Mountain View, CA, USA: CPP, Inc.

Dillinger, T. G., Wiegmann, D. A., & Taneja, N. (2003). Relating Personality with Stress Coping Strategies Among Student Pilots in a Collegiate Flight Training Program. *12th International Symposium on Aviation Psychology.* Dayton, OH.

FAA. (n.d.). *FAA Human Factors Awareness Web Course*. Retrieved August 28, 2009, from Team Performance Module>Error Management - 2 of 4: http://www.hf.faa.gov/webtraining/TeamPerform/TeamCRM014.htm

Frank, L. (1981, February). The myth of the accident prone. *Flying Safety* , pp. 26-27.

Guildford, J. (1947). *Printed classification tests: Army Air Forces aviation psychology program research project (Report no. 5).* Washington, DC: U.S. Government Printing Press.

Henley, I. M. (2003). Aviation Education and Training. In I. M. Henley, *Aviation Education and Training* (pp. 106-112). Burlington: Ashgate.

Hunter, D., & Burke, E. (1995). *Handbook of pilot selection* (1st ed.). Aldershot, Hants., England: Avebury Aviation.

ICAO. (2009). *ICAO Safety Management Manual (SMM).* Retrieved August 29, 2009, from International Civil Aviation Organization (ICAO): http://www.icao.int/anb/safetymanagement/DOC_9859_FULL_EN_V2.pdf

Myers, I. B. (1998). *Introduction to Type.* Palo Alto, CA: Consulting Psychologists Press, Inc.

Novello, J., & Youssef, Z. (1974). Psycho-social studies in general aviation: II. Personality profile of female pilots. *Aerospace Medicine , 45* (6), 630-33.

Picano, J. (1991). Personality types among experienced military pilots. *Aviation, Space and Environmental Medicine, 62* (6), 517-20.

Ramachandran, N., Wadhawan, J., Kumar, V., Chandramohan, V., & Rao, P. (1983). Personality profile os an IAF Pilot; Its usefulness in pilot selection. *Aviation Medicine , 21* (2), 131-139.

Reason, J. (2000). Human error models and management. *BMJ* , 320:768-770.

Stokes, A., & Kite, K. (1994). *Flight Stress: stress, fatigue and performance in aviation* (1st ed.). Aldershot, Hants., UK: Avebury Aviation.

Voge, V. (1989). Failing aviator syndrome: a case history. *Aviation, Space and Environmental Medicine , 60* (7,Suppl.), A89-91.

SCENARIO-BASED FLIGHT SIMULATION TRAINING: A HUMAN FACTORS ANALYSIS OF ITS DEVELOPMENT AND SUGGESTIONS FOR BETTER DESIGN

Christopher M. Johnson
University of Wisconsin – Madison, USA

Douglas A. Wiegmann
University of Wisconsin – Madison, USA

The Federal Aviation Administration (FAA) funded a study aimed at ameliorating issues with visual flight rules (VFR) pilots flying into instrument weather conditions (IMC). Scenario-based simulation training (SBST) was developed to examine threat and error management (TEM) skills among private pilots. SBST was modeled after line-oriented flight training (LOFT), and new technologies were developed to improve weather simulation. This paper includes a Human Factors analysis of flight simulation development, and it details weather simulation improvements. Historical weather data was used for accurate recreation of pre-flight briefings, simulated weather parameters, and in-flight weather reports, and the technologies were tested among distinctly different pilot populations. The first experiment examined career-track aviation students, and the second experiment involved VFR-only pilots with no career aspirations. The technologies developed for this project revealed interesting findings related to inadequate training, and results indicate that SBST can effectively fill holes in *ab intio* flight training and foster higher simulation fidelity at all levels of flight training.

Less than a decade after the Wright brothers conceived powered flight, the utility of flight simulation became a realization. “The invention, therefore, of a device which will enable the novice to obtain a clear conception of the workings of an aeroplane and conditions existent in the air without any risk personally or otherwise is to be welcomed without doubt” (Haward, 1910, p. 1005; from Page, 2000). This 1910 quote in *Flight*, the first aviation journal, turned out to be less accurate than projected. Certainly, flight simulation development has endured engineering designs that were not “welcomed without doubt”; however, in the past century, flight simulation development has grown into a multi-billion dollar industry.

Over the last 100 years, more and more realism has found its way into the simulated cockpit. The first flight simulation designs focused on basic structure for cockpit control of flying, and they were utilized for assessing aptitude and reaction times of eligible airmen during the First World War (Page, 2000). After WWI, instrument flight was conceived to support mail delivery services, which proposed the need for accurate recreation of cockpit instrumentation (Page, 2000), and fix-based, instrument-only simulators are still used today for training instrument skills.

Some early simulators incorporated archaic movable platforms and visual systems, but after a century of development, the advent of affordable computing, high-definition audio-visual systems, and numerous electrical and mechanical advances have made flight simulation development a highly refined science. As flight simulation fidelity increased so did its training utility, and transfer of training evaluations became necessary to determine the extent to which flight simulation could effectively replace actual flight to achieve training objectives (Roscoe & Williges, 1980). For example, Taylor et al. (1999) evaluated the effectiveness of personal computer-based aviation training devices (PCATDs) and flight training devices (FTDs) for training technical instrument procedures. Not only did this work show that combining flight simulation with actual flight training can facilitate the development of technical skills faster than using only airplanes, but this work also paved the way for the development of FAA-approved, desktop simulators, ultimately reducing the cost of flight training.

High-fidelity simulation has fostered the development of scenario-based training (SBT). For over 3 decades, SBT has been conducted in simulators to train commercial pilots for non-technical skills such as aeronautical decision-making (ADM), crew resource management (CRM) and threat and error management (TEM). Commercial pilots learn these higher-order thinking skills (HOTS) through line-oriented simulation (LOS), which allows instructors to evaluate abnormal and emergency procedures that cannot be safely recreated in flight (Butler, 1993). LOS scenarios can be used for training (aka line-oriented flight training, or LOFT) or evaluation (aka line-oriented evaluation, or LOE), and they can be full-mission simulations (e.g. LOFT and LOE) or part-task training (aka special purpose operational training, or SPOT; Bowers & Jentsch, 2004). To maintain fairness in evaluation, LOS scenarios are scripted and contain highly structured event sets that are algorithmically determined from accident data to realistically evaluate technical and non-technical skills (Bowers & Jentsch, 2004). The LOS training model has set the stage for general aviation (GA); however, GA's adherence to the commercial model is not as strict as it should be.

GA cockpits have experienced a rapid influx of automation in the past decade, bridging the automation gap between airliners and GA cockpits, and to ensure that training evolves in parallel with cockpit advances, the FAA-Industry Training Standard (FITS) was developed (Chaparro et al., 2004). FITS syllabi utilize a mixture of simulation and flight training to combine maneuver-based training (MBT) with hypothetical scenarios so that training for automation, navigation, communication, and decision-making (aka single-pilot resource management, or SRM) coincides with psychomotor skill development. FITS uses SBT to embed maneuvers within hypothetical missions that relate MBT to real-life contexts to provide students with a global perspective of why specific maneuvers are important, and SBT incorporates automation training at the beginning stages to teach SRM skills. Although there is a steep learning curve, SBT charges students with taking ownership of each flight, which will ultimately induce a positive change in safety culture; however, FITS lessons often require plan revisions to accommodate unforeseen circumstances in the air, which can be good for teaching ADM skills, but diversions can also hinder task completion and encourage reversion to non-contextual MBT. In short, FITS lessons are necessarily semi-structured to offer latitude to adapt lessons to the inconsistencies of the aeronautical environment. Unfortunately, however, this training mantra has carried over into FITS simulator sessions, which, according to LOS, should be more highly structured in a computed environment with engineered threats.

Today's flight simulators undergo stringent fidelity assessments of performance criteria, and physics data derived from *actual* flight tests have catapulted simulation development. At the highest levels of simulation fidelity, cockpit layouts are exact, and the depiction of virtual aeronautical environments is as realistic as 2 dimensions can offer. History shows a strong positive correlation between simulation fidelity and training utility, and high-fidelity simulation gave rise to SBT. FITS and LOS illustrate that MBT can be accomplished osmotically through SBT, and this study analyzed simulation for training weather-related decision skills. It revealed significant potential for improvement that can benefit all levels of flight training.

Methods

This study began with an analysis of weather-related training and testing for private pilots (PVT). Wiegmann, Talleur & Johnson (2008) found that the PVT written exam primarily tests lower-order intellect, and suggestions toward improve were offered. The analysis also found that the PVT oral exam supplemented the written by testing for higher-order intellect, or HOTS. In short, the written and oral exams were found to be adequately supplemental for evaluating pre-flight decision-making, but these findings lead to a review of in-flight training and testing procedures, which were found to be inadequate for evaluating pilots' ability to dynamically integrate in-flight weather data to update pre-flight weather data.

A review of Federal Aviation Regulations (FARs) revealed that the training environment is less risky than the operational environment (Johnson, Wiegmann & von Thaden, 2009). Specifically, student pilots are trained and tested in relatively conservative weather conditions, and since student pilots are not allowed to carry passengers or travel cross-country on personal "trips", the training environment is generally void of social pressures that may influence the acceptance of weather risk.

Only one PVT practical testing standards (PTS) standard exists that examines PVT applicants' ability to divert from an intended route of flight, but the diversion procedure is not necessarily related to weather (US DOT, 2002). Examiners that *do* wish to examine diversion procedures related to weather, however, must do so under highly unrealistic conditions, and since check rides are conducted in fair weather, PVT applicants lack the visual and non-visual weather information (e.g. radio reports) on which to base their simulated diversion procedures. Similarly, the IFR PTS (US DOT, 2010) includes no requirement for IFR applicants to integrate in-flight weather data during the checkride.

The above findings suggested the need to train and evaluate weather-related ADM skills via simulation, so 3 software platforms were reviewed. The analysis targeted simulated weather functions and certified flight instructors' (CFIs') work-arounds to overcome limitations of weather simulation. The overall goal of this analysis was to identify the technological and social elements that shape simulated flight training and to improve weather simulation to facilitate an examination of VFR into IMC behavior.

Findings

Existing Flight Simulation Technologies

Three major findings were derived from our analysis of GA flight simulators:

1) Programming weather data into the simulator is labor-intensive and unrealistic. Weather parameters generally remain static until CFIs make changes, taking their attention away from the lesson. Some software randomly sets weather, but this prevents instructors from knowing exact weather parameters, which is often desired to predict behavior. Rendering visual weather for IFR training can be argued as unnecessary because IFR students *should be* focused on instruments; however, the results from this study indicate that visual simulation of clouds and haze can and should be used to train IFR pilots to transitions from visual to instrument flight and vice versa, which are vital skills that mitigate potential spatial disorientation (Johnson & Wiegmann, 2011).

2) Radio-based weather reports are generally delivered by CFIs that typically report only partial information by saying exactly the weather parameters that they have programmed. This is problematic because in-flight reports exist so that pilots can understand the weather differential between their current location and a distant region. Also, the abbreviated reports delivered in CFIs' natural voice are more intelligible and less information-saturated than synthetic reports in the operational environment. This creates a signal detection issue that can hinder students' abilities to interpret weather information, and during simulation, CFIs must also pay close attention to simulated radios to cue their delivery of weather reports. If done accurately, this absorbs CFIs' attention, leaving nothing for the lesson, and students often must prompt instructors for weather reports, which diminishes the suspension of a student's disbelief in the scenario's realism.

3) The pre-flight weather experience is greatly simplified amidst the time constraints of a paid service. A simulated, pre-flight briefing typically consists of a short discussion between the CFI and the student about the weather parameters that the CFI will program, yet standardized weather reports for short cross-country trips can range between 10 and 20 pages of coded text with dozens of images. The simplified weather experience hinders skill development, and even commercial developers admit that although "recent research has demonstrated that weather is a prime consideration in aircrew decision making…the creation of weather paperwork has frequently been only an afterthought in traditional LOE development" (Bowers & Jentsch, 2004).

The above findings suggest that there needs to be a formal shift toward designing realistic weather experiences for pilots, so below is a list of the state-of-the-art weather features in GA flight simulation:

1) "Real-time" weather accesses online meteorological reports (METARs) that are not always up-to-date. This is a database management issue, but the radio-based reports that are tied to these METARs are subsequently inaccurate. Furthermore, weather changes within the simulation software are random, which adds more variability, and since "real time" METARs are updated only periodically, this can create sudden shifts of weather parameters. The ideal functionality would incorporate historical METARs tied to a "game" clock and latitude/longitude coordinates of the reporting stations to facilitate accurate rendering of simulated weather and automated reports.
2) Simulating "real time" weather can be a problem for practicing local procedures. Specifically, simulation is commonly used for training when the weather is unsuitable for flying, which generally means that the simulated, "real time" weather would be equally unsuitable for practicing procedures within the virtual environment of a pilot's "home" airport. Of course, since simulation allows pilots to fly anywhere in the world, it is generally possible to find suitable weather;

however, this takes time, students are forced to train in unfamiliar environments, and weather is still rendered with a fair degree inaccuracy.

3) Simulating historical weather was the primary undertaking of this project. Using historical weather allows pilots to choose weather parameters to fit their desired route, and it prevents weather from shaping training goals. To overcome the rendering issues detailed above, programming solutions incorporated improved calculation strategies, and a link between pre-flight weather information and the associated in-flight weather parameters was accomplished with a database management system (DBMS) and a pre-flight briefing interface.

Advanced Flight Simulation Technologies

Three software elements were developed to improve weather simulation fidelity and to provide a new way of training GA pilots. These technologies replicate complete, historical weather scenarios that maintain continuity in all aspects of weather, both pre-flight and in-flight:

1) Pre-flight briefing interface and archived aviation weather database: The pre-flight interface was originally developed as a menus-based executable program (Johnson, Wiegmann & von Thaden, 2009); however, a Pilot Training System was built as a web application that is more representative of actual pre-flight resources, and it includes the underlying DBMS with added functionality to serve as a holistic training tool for flight schools (Johnson & Wiegmann, 2011).

2) Dynamic control of simulated weather: Historical weather parameters and radio-based weather reports were managed by an internal clock, and linear spatial transitions were used to overcome randomness in weather generation between reporting stations. These automated features reduce CFI workload, and they mitigate programming errors and facilitate rapid training initiation.

3) Post-flight output of aeronautical parameters: A post-flight feedback tool was also developed to facilitate performance analysis. The graph illustrates aircraft location in relation to field height and simulated weather parameters, and regulatory information is represented to facilitate a reduction in IMC violations (Johnson, Wiegmann & von Thaden, 2009).

Discussion

This study took a formal design approach to the simulation of weather, and scenario-based simulation training (SBST) was designed to supplement literary-based methods that train SRM skills to GA pilots. In short, SBST is to SRM as LOFT is to CRM, and although SBST was developed to study VFR flight into IMC, the technologies can be used to bring FITS training to a higher standard.

The technologies developed in this study can be expanded to fulfill complex weather simulation at the commercial level, and an evaluation method was also developed that captures implicit measures of situation awareness for correlating pilots' procedural skills (e.g. aviate, navigate, communicate, and process weather data) with their ability to adhere to VFR regulations (Johnson & Weigmann, 2011). These technologies and methodologies revealed some interesting things about novice pilots' IMC violations.

"The key to effective simulation-based training is achieving suspension of disbelief…Subjects must be made to think and feel as though they are functioning within a real environment and to maintain that suspension of disbelief throughout the scenario." (Halamek et al., 2000, p. 4). This is why seamless fidelity is important; however, suspending disbelief begins with a realistic scenario, so in order to motivate these private pilots to pursue weather risk, 2 scenarios were developed that targeted to distinctly different types of pilots. In the first experiment, a cargo-delivery scenario was used, and it was effective because the experiment involved FAR Part 141 pilots who were mostly career-track aviation students pursuing IFR training (Johnson, Wiegmann & von Thaden, 2009). The second experiment, on the other hand, involved VFR-only pilots who flew for various personal reasons, so the scenario that was found to be more effective for these pilots was one in which they were to imagine that they were still finishing up training and that their CFI was endorsing them to make a solo cross-country in marginal VFR (Johnson & Wiegmann, 2011).

In both experiments, IMC experience was found to be a significant, demographic moderator of pilots' ability to execute proper procedures and maintain VFR (Johnson, Wiegmann & von Thaden, 2009; Johnson & Wiegmann, 2011). These findings support the realism of the simulations, and they add validity to using simulation to study naturalistic behavior. Findings from this study also indicate that automation can foster unsafe behavior. Specifically, pilots' abilities to program navigation equipment was shown to have a significant negative correlation with their ability to maintain VFR, and some of the pilots used the autopilot to commit flagrant IMC violations (Johnson & Wiegmann, 2011). For example, 6 pilots penetrated simulated clouds in the second experiment, and 3 of them used the autopilot while doing so. Interestingly, the 3 pilots that were hand-flying the simulator accumulated only 55 seconds of total IMC penetration over 8 encounters (~7 seconds per encounter), whereas the 3 pilots that used the autopilot penetrated the simulated clouds for 10:33 over 4 encounters (~2:38 per encounter). The prolonged periods of cloud penetration were followed by blind descents below minimum safe altitudes, putting the pilots at risk for controlled flight into terrain (CFIT), and both of these findings indicate that cockpit automation can create a "comfort zone" in which pilots use automation to extend their natural flying skills and pursue weather risks that exceed their certificated capabilities.

SBST may be unsuitable for training visual-only pilots. For example, in the first experiment, one subject displayed overtly, risky behavior, and it seemed as though SBST fostered unwarranted confidence that caused this pilot to make willful violations in pursuit of the simulated mission. This finding was related to the fact that this pilot committed IMC violations and achieved "successful" outcomes without receiving corrective feedback (Johnson, Wiegmann & von Thaden, 2009). This finding led to a design change to the weather profile in the following experiment, and in the second experiment, a manual manipulation of the weather was made to engineer an unexpected IMC encounter on approach to the destination airport. This was done to ensure that the pilots did not "win" the scenario, and it also allowed for the analysis of pilots' behavior during abrupt immersion by unexpected IMC (Johnson & Wiegmann, 2011).

Seven of the 16 pilots the participated in the second experiment pursued the simulated route at length. In doing so, they committing VFR violations, and they encountered unexpected IMC at a critical phase of flight, on approach to landing. None of these pilots executed an immediate course reversal when encountering IMC on final approach, which suggests that the 180° turn is not an effective, one-size-fits-all, IMC escape procedure. In fact, 2 pilots crashed within seconds of being immersed in IMC, and several other pilots demonstrated flight profiles characterized by excessive control inputs, which, in a full-motion aircraft, would have likely lead to vestibular confusion and loss of control. Interestingly, the pilots who were able to maintain control of the aircraft all continued to circle the airport environment in zero visibility at low altitudes, putting them at risk for a CFIT accident (Johnson & Wiegmann, 2011).

Conclusion

Entire flight training programs can be built around the Pilot Training System. It brings together all required information for a "go/no-go" decision, and it provides CFIs with a consolidated resource of aeronautical information to teach ground lessons and provide students with a comprehensive overview of the pre-flight experience. The interface can also be utilized by check pilots to assess pilot applicants' pre-flight knowledge, and when combined with appropriate flight simulation software, it allows pilots to experience realistic, simulated weather parameters that would be too risky to experience in the operational environment.

Flight simulation manufacturers have overlooked the pre-flight experience, and this oversight has undermined the weather experience and placed a heavy burden on CFIs and the designers of SBT (Bowers & Jentsch, 2004). Unfortunately, safe pilots do not just "hop in and go", and the weather experience begins with construction of a mental model from pre-flight information. Pilots, then, update their mental models of the weather situation with new weather data made available in flight. That said, SBST technologies bring that reality to simulation, and the historical database removes weather as an obstacle that has traditionally dictated training goals. In fact, the sky is literally the limit with regards to providing pilots with realistic, simulated, weather experiences of far-off regions, and improved weather simulation also provides accident investigators with a tool to recreate weather-related accident scenarios to help validate their findings.

Weather simulation technologies are useful for exposing and training weather-related ADM skills, and in this study they helped uncover errors related to the misuse of automation, which reveal training weaknesses that are indirectly related to weather. These technologies were developed to test a hypothesis about the inadequacies of the current flight training system, so they were tested on certificated pilots; however, more research is required to understand the utility of this training for student pilots. That said, these technologies will need to undergo training transfer assessments, and to avoid SBST's potential to induce risk among VFR-only pilots, it should first be tested for training within syllabi that combine private and instrument training such as those certified under FITS.

References

Bowers, C., & Jentsch, F. Engineered experiences for scaled worlds: An example from aviation. In Schiflett, S. G., Elliott, L. R., Salas, E., & Coovert, M. D. (Eds.), *Scaled Worlds: Development, Validation and Applications*. Burlington, VT: Ashgate.

Butler, Roy E. (1993). LOFT: Full-mission simulation as crew resource management training. In Wiener, E. L., Kanki, B. G. & Helmreich, R. L. (Eds.), *Cockpit Resource Management*. San Diego, CA: Academic Press.

Chaparro, A., Roger, B. L., & Hamblin, C. J. (2004). *Review of the FITS program: Program tasks, goals, and pilot training initiatives* (Final Report: DOT/FAA/AR-xx/xx). Washington, D.C.: U.S. Department of Transportation, Federal Aviation Administration.

Halamek, L. P., Kaegi, D. M., Gaba, D., Sowb, Y. A., Smith, B. C., Smith, B. E., Howard, S. K. (2000). Time for a new paradigm in pediatric medical education: Teaching neonatal resuscitation in a simulated delivery room environment. *Pediatrics, 106* (4).

Haward, D. (1910). The Sander's "Teacher." *Flight, 102*, p. 1005. December 10, 1910.

Johnson, C. M., & Wiegmann, D. A. (2011). Pilot error during VFR flight into IMC: A literature review and experiment using advanced simulation and analysis methods. Submitted for presentation at *Human Factors and Ergonomics Society 55th Annual Meeting.* Las Vegas, NV, Sep 19-23, 2011.

Johnson, C. M., Wiegmann, D. A., von Thaden T. (2009) *Introduction to Single-Pilot Resource Management Scenario-based Simulation Training: A Proof of Concept.* Technical Report AHFD-09-01/FAA-07-G-13. Savoy, IL: University of Illinois, Human Factors Division.

Page, R. L. Brief history of flight simulation. *Proceedings of SimTechT 2000*. Sydney: The SimtechT 2000 Organizing and Technical Committee.

Roscoe, S. N., & Williges, B. H. (1980). Measurement of transfer of training. In Roscoe, S. N. (Ed.), *Aviation Psychology*. Ames: Iowa State University Press.

Taylor, H. L., Lintern, G., Hulin, C. L., Talleur, D. A., Emanuel, T. W., Jr., Phillips, S. I. (1999). Transfer of training effectiveness of a personal computer aviation training device. *International Journal of Aviation Psychology, 9(4)*, 319-334.

U.S. Department of Transportation, Federal Aviation Administration. (2002). *Private Pilot Practical Test Standard for Airplane* (FAA-S-8081-14A).

U.S. Department of Transportation, Federal Aviation Administration. (2010). *Instrument Rating Practical Test Standards* (FAA-S-8081-4E).

Wiegmann, D., Talleur, D., Johnson, C. (2008). Redesigning weather-related training and testing of general aviation pilots: Appling traditional curriculum evaluation and advanced simulation-based methods. *University of Illinois Human Factors Division Technical Report HFD-08-01/FAA-08-01.*

IMPACT OF SIMULATOR OUT-THE-WINDOW VISUAL DISPLAY RESOLUTION ON AIR-TO-GROUND SKILL PERFORMANCE

Jamie L. Estock
Melinda K. Seibert, Dr. Elliot E. Entin
Aptima, Inc.
Woburn, MA

Previous research revealed no difference in air-to-air skill performance between instructor pilots who flew a simulator with a narrow out-the-window visual display field-of-view and instructor pilots who flew a simulator with a wide field-of-view. To evaluate the generalizability of these results to a different fidelity dimension, mission, and pilot type, the current study assessed the impact of out-the-window visual display *resolution* on *air-to-ground* skill performance of *less experienced pilots*. In the current study, 18 F/A-18 Fleet Replacement Squadron pilots flew air-to-ground training missions in two simulators that differed in their visual display resolution. F/A-18 subject matter experts assessed pilot performance during the missions using three observer-based instruments. Results revealed a difference in performance between pilots who flew the simulator with a lower-resolution display and pilots who flew the simulator with a higher-resolution display for two out of 12 air-to-ground skills. We discuss the implications of our findings for simulator acquisition.

Introduction

Employing the appropriate level of simulator fidelity ensures better training results and reduces costs by eliminating investments in unnecessary training and technology. However, simulator fidelity tradeoff decisions are difficult to make because of the lack of available objective data to support these decisions. As a result, we have been conducting simulator fidelity research in operational training environments to collect objective data regarding the level of fidelity necessary for effective simulator-based training. Our previous research revealed no difference in air-to-air (A/A) skill performance between instructor pilots who flew a simulator with a narrow out-the-window (OTW) visual display field-of-view (FOV) and pilots who flew a simulator with a wide FOV (Estock, Alexander, Stelzer, & Baughman, 2007; Estock, Baughman, Stelzer, & Alexander, 2008). The results indicated that instructor pilots rated the simulator with the narrower FOV less effective for training air-to-air skills largely dependent on visual information than the simulator with the wider FOV, yet the in-simulator performance and training effectiveness results showed no difference between the two simulator conditions. In post-study interviews, pilot subject matter experts (SMEs) suggested that we might not have found differences between the two simulators because of the experience level of our pilot participants and the operational mission that we used in our previous study.

The purpose of the research reported in this paper is to examine the generalizability of our previous results to a different fidelity dimension, operational mission, and participant type. Specifically, the current study assessed the impact of OTW visual display *resolution* on *air-to-ground (A/G)* skill performance of *pilots who were just learning to fly the F/A-18 aircraft.* In this study, F/A-18 Fleet Replacement Squadron (FRS) pilots flew A/G training missions in two different simulators—the Weapons Tactics Trainer (WTT) or the Tactical Operational Flight Trainer (TOFT). The primary difference between the two simulators was the resolution of the OTW visual display, with WTT having a lower-resolution visual display (20/80 visual acuity) and the TOFT having a higher-resolution visual display (20/40 visual acuity).

Prior to the study, we administered a survey to six F/A-18 SMEs to identify which A/G skills might be negatively impacted by a lower-resolution OTW visual display in a simulator. We extracted the A/G skills from the U.S. Navy's F/A-18 Competency-Based Training & Readiness Matrix. The SMEs indicated that a lower-resolution visual display could negatively impact a simulator's effectiveness for training 12 A/G skills that are largely dependent on information provided by the depiction of the outside world. Table 1 provides the names and definitions of the 12 A/G skills identified by the SMEs.

Table 1. A/G Skills Largely Dependent on Information Provided by the Depiction of the Outside World.

Skill Name	**Definition**
A/A Target Acquisition	Ability to locate and identify A/A targets
A/G Target Acquisition	Ability to locate and identify A/G targets
Abort Awareness	Ability to recognize when an abort is required
Basic Air Work	Ability to establish and maintain proper altitude, airspeed, and heading during flight
Delivery Parameters (Guided Other)	Ability to execute appropriate delivery parameters to effectively deliver munitions and achieve desired weapons effects
Delivery Parameters (Unguided)	Ability to execute appropriate delivery parameters to effectively deliver ordnance and achieve desired weapons effects
Maintains Formation	Ability to maintain a briefed formation
Sensor Employment	Ability to employ the optimal sensor mechanization to collect and integrate information
Threat Awareness	Ability to recognize that an A/A and/or A/G threat exists within range of a friendly aircraft
Threat Counter-maneuver	Ability to perform the appropriate the maneuver to mitigate the threat.
Threat Counter-measures	Ability to employ the appropriate counter-measures to counter the threat
Threat Identification	Ability to identify the type of threat(s)

As a result, we hypothesized that:

> Pilots flying the WTT would show *poorer performance* on measures related to each of the 12 A/G skills than pilots flying the TOFT.

Methods

Participants

Eighteen U.S. Navy F/A-18 FRS pilots at Naval Air Station (NAS) Lemoore participated in the study. All 18 participants were male. The participants included 14 Lieutenants, Junior Grade (O-2) and four Lieutenants (O-3). The participants had a mean of 4.93 months flying the F/A-18 aircraft (*SD* = 3.67), and a mean of 46.33 flight hours in the F/A-18 aircraft (*SD* = 51.93).

Simulators

The WTT consists of an F/A-18E/F Super Hornet aircraft cockpit running the F/A-18E/F aircraft's Operational Flight Program (OFP). The WTT has a 360-degree horizontal FOV visual display with a 20/80 visual resolution, and contains the actual F/A-18 E/F aircraft controls and displays.

The TOFT consists of an F/A-18E/F Super Hornet aircraft cockpit running the F/A-18E/F aircraft's OFP. The TOFT has a 360-degree horizontal FOV visual display with a 20/40 visual resolution, and contains the actual F/A-18 E/F aircraft controls and displays.

Experimental Design

The between-subjects experimental design compared pilots who flew the WTT with pilots who flew the TOFT on their in-simulator performance. The design focused on the impact of OTW visual display resolution on pilot in-simulator performance related to 12 A/G skills largely dependent on information provided by the depiction of the outside world.

F/A-18 SMEs used the following three questionnaires to rate pilot in-simulator performance during each A/G training mission in the simulator:

Strike SPOTLITE questionnaire. The Strike SPOTLITE Questionnaire was developed to capture detailed data regarding pilot A/G skill performance. The Strike SPOTLITE Questionnaire consisted of 76 individual performance measures and was used to capture pilot performance using a variety of question types (e.g., Yes/No, Likert scales, checklists). Prior to the study, F/A-18 SMEs identified the individual Strike SPOTLITE measures that capture performance related to each of the 12 A/G skills. For example, the F/A-18 SMEs identified four Strike SPOTLITE measures that assess the skill of *abort awareness*, including: (1) employs ordnance within correct parameters, (2) obtains clearance prior to weapons release, (3) aborts when the attack meets abort criteria, and (4) follows established abort procedures.

A/G skill questionnaire. The A/G Skill Questionnaire was developed to capture high-level data regarding pilot performance related to the 12 A/G skills. The A/G Skill Questionnaire was used to capture pilot performance using a five-point Likert scale where one referred to Extremely Poor and five referred to Excellent. A sample A/G Skill Questionnaire statement is “Please rate the pilot’s ability to *recognize when an abort is required [abort awareness]*.”

Instructor gradesheets. The Instructor Gradesheets are standard performance evaluations used by Instructor Pilots to rate pilot performance during each training event. The Instructor Gradesheets were used to rate pilot performance on a four-point scale where one referred to Unsatisfactory and four referred to Above Average. Prior to the study, F/A-18 SMEs identified the individual Instructor Gradesheet items that capture performance related to each of the 12 A/G skills. For example, the F/A-18 SMEs identified two gradesheet items that assess the skill of *abort awareness*, including: (1) navigation system setup/usage, and (2) time-on-target.

Procedures

U.S. Navy F/A-18 FRS pilots reported to regularly-scheduled simulator-based training missions at NAS Lemoore and were recruited to participate in the fidelity study. Upon consent to participate, pilots completed a demographic questionnaire. Then, pilots were assigned to either the WTT or the TOFT by their squadron. Pilots flew the training missions in their assigned simulators while F/A-18 SMEs rated pilot performance using the Strike SPOTLITE Questionnaire, A/G Skill Questionnaire, and Instructor Gradesheet during that mission.

Results and Discussion

To test our hypothesis, we compared the in-simulator performance of pilots who flew the WTT to pilots who flew the TOFT using three expert observer questionnaires. We conducted all of our in-simulator performance analysis at the A/G skill level. As a result, we computed A/G skill performance scores by averaging the ratings of pilot participant performance on the individual Strike SPOTLITE measures that captured performance on each A/G skill. We used the same procedure to compute the A/G skill performance scores using Instructor Gradesheets. We did not need to aggregate items on the A/G Skill Questionnaire because they provided performance at the A/G skill level.

We used independent samples *t* tests to determine if there were statistically significant differences between the in-simulator performance of pilots who flew the WTT and pilots who flew the TOFT. A *p* value of ≤ 0.05 was considered to be a statistically significant difference. A *p* value of ≤ 0.10 was considered to be a marginally significant difference. We also calculated effect size to determine whether a statistically or marginally significant difference has some practical significance, and is not just a statistical artifact. We used Cohen's *d* as a measure of effect size. Cohen (1988) refers to $d = 0.20$ as a small effect, $d = 0.50$ as a medium effect; and $d >= 0.80$ as a large effect.

Strike SPOTLITE questionnaire

An independent samples *t* test revealed that pilots who flew the WTT showed poorer performance on their *ability to recognize when an abort is required (abort awareness)* ($M = 2.53$, $SD = 0.45$) than pilots who flew the TOFT ($M = 2.97$, $SD = 0.08$), $t(15) = -3.53$, $p < .05$, $d = 0.98$. Moreover, the large effect size of 0.98 suggests that there may be practical significance for this finding. Figure 1 presents the mean Strike SPOTLITE rating of in-simulator performance related to the *ability to recognize when an abort is required [abort awareness]* for pilots who flew the WTT and pilots who flew the TOFT.

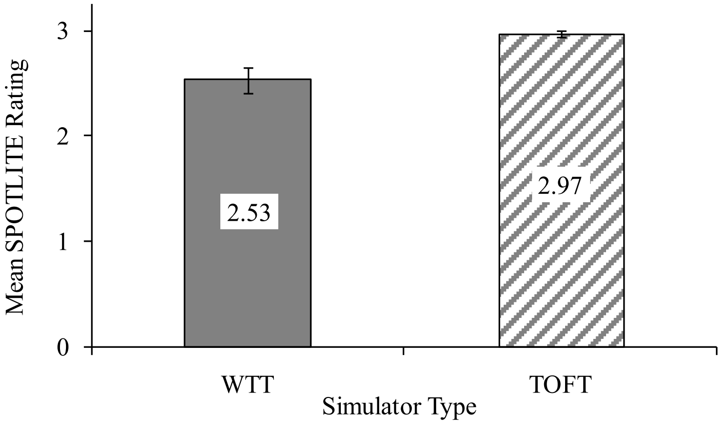


Figure 1. Mean Strike SPOTLITE ratings of in-simulator performance related to *abort awareness* by simulator type.

Independent samples *t* tests revealed no significant in-simulator performance differences between pilots who flew the WTT and pilots who flew the TOFT for the remaining 11 A/G skills.

A/G skill questionnaire

An independent samples *t* test revealed that pilots who flew the WTT showed poorer performance on their *ability to execute appropriate delivery parameters to effectively deliver ordnance and achieve desired weapons effects [delivery parameters – unguided]* (*M* = 2.89, *SD* = 0.68) than pilots who flew the TOFT (*M* = 3.33, *SD* = 0.65), *t*(28) = -1.79, *p* = .08, *d* = 0.66. While this finding is not statistically significant, the medium effect size of 0.66 suggests that there may be practical significance for this finding. Figure 2 presents the mean A/G Skill Questionnaire rating of in-simulator performance related to the *ability to execute appropriate delivery parameters to effectively deliver ordnance and achieve desired weapons effects [delivery parameters – unguided]* for pilots who flew the WTT and pilots who flew the TOFT.

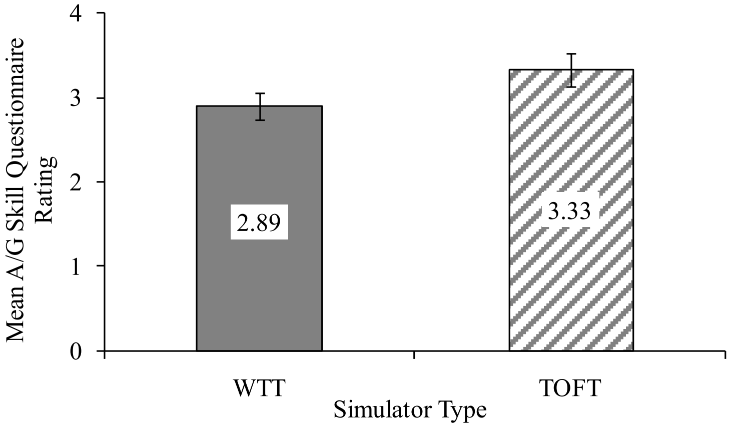


Figure 2. Mean A/G Skill Questionnaire ratings of in-simulator performance related to *delivery parameters* - *unguided* by simulator type.

Independent samples *t* tests revealed no significant in-simulator performance differences between pilots who flew the WTT and pilots who flew the TOFT for the remaining 11 A/G skills.

Instructor gradesheet

Independent samples *t* tests revealed no significant in-simulator performance differences between pilots who flew the WTT and pilots who flew the TOFT for any of the 12 A/G skills.

Conclusions

Our previous research revealed no difference in A/A skill performance between instructor pilots who flew a simulator with a narrow OTW visual display FOV and instructor pilots who flew a simulator with a wide FOV. In this previous research, we conducted post-study interviews with pilot SMEs to obtain an operational explanation of our results. The pilot SMEs suggested that we might not have found differences between the two simulators because of the experience level of our pilot participants and the operational mission that we used in our previous study. Specifically, the pilot SMEs suggested that the instructor pilot participants in our previous study would have mastered the A/A skills, and as a result, may have been able to adapt readily to lower-fidelity simulators. In addition, the pilot SMEs suggested that the pilot participants in our previous study were mostly engaging enemy aircraft beyond visual range during the A/A missions, lessening the importance of high-fidelity visual systems. As a result, we designed this study to explore three new areas: (1) OTW visual display resolution requirements, (2) fidelity requirements for less experienced pilots, and (3) fidelity requirements for the A/G mission.

In this study, we examined the influence of OTW visual display resolution differences on FRS pilot A/G skill performance. The A/G skills selected for this investigation were extracted from the U.S. Navy's F/A-18 Competency-Based Training & Readiness Matrix. Specifically, we examined visual display resolution differences on pilot performance related to 12 A/G skills that are largely dependent on information provided by the depiction of the outside world. The results of this study indicated that the OTW visual display resolution difference between the WTT (20/80 visual acuity) and the TOFT (20/40 visual acuity) affected pilot performance related to two A/G skills—*abort execution* and *delivery parameters (unguided).* Specifically, the pilots who flew the TOFT performed both skills better than the pilots who flew the WTT.

Unlike our previous research, the findings of this study suggest that simulators with higher-fidelity OTW visual displays may be more effective for training certain air combat skills—at least more effective for training *some A/G skills* to *less experienced pilots*. Yet like our previous research, the findings of this study suggest that end-users tend to overestimate the level of fidelity required for effective training. Of the 12 specific A/G skills that the SMEs suggested could be negatively impacted by a lower-resolution visual display, *only two A/G skills* showed any difference between the two simulator conditions. Based on the results of our research to date, we believe that relying solely on end-user-defined fidelity requirements can result in the acquisition of a simulator that is more expensive than necessary for effective training. Future research should be conducted to collect the necessary objective data to employ the *appropriate level* of fidelity to meet the training objectives. In addition, future research should examine whether this finding is consistent for other fidelity dimensions, such as motion fidelity, and other training events, such as carrier landings.

Acknowledgements

This material is based upon work supported by the Office of Naval Research (ONR) under Contract No. N00014-09-C-0079. All opinions expressed in this article are those of the authors and do not necessarily reflect the official opinion or position of Aptima, Inc., ONR, or the U.S. Navy. We would like to thank Dr. Roy Stripling and LCDR Joseph Cohn at ONR for funding this research, and CDR James Patrey and Dr. Melissa Walwanis for their technical guidance and support on this effort. We would also like to thank Ms. Judity Rice, CDR Brandon Hammond, LCDR Greg Brown, MAJ Jon Rabine, and all of the other engineers and F/A-18 pilots at NAS Lemoore who supported this research with their time and expertise.

References

Estock, J. L., Alexander, A. L., Stelzer, E. M., & Baughman, K. (2007). Impact of visual scene field of view on F-16 pilot performance. Paper presented at the *51st Annual Meeting of the Human Factors and Ergonomic Society,* Baltimore, MD.

Estock, J.L., Baughman, K., Stelzer, E.M., & Alexander, A.L. (2008). Fidelity requirements for effective training: Pilot perceptions versus objective results. *Proceedings of the 30th Annual Interservice/Industry Training, Simulation and Education Conference*, Orlando, FL.

COMBINING BEHAVIORAL AND BIOMETRIC MEASUREMENTS FOR AUTOMATED PERFORMANCE ASSESSMENT

Chris Forsythe, Robert A., Abbott, Susan M. Stevens-Adams, Michael Haass, Laura Matzen,
Kiran Lakkaraju
Sandia National Laboratories
Albuquerque, NM, USA

Technologies are needed enabling more cost-effective military aviation training. Automated performance assessment has been advanced as one approach to enable instructors to make more effective use of simulation-based training systems. Recent experimental research will be reviewed illustrating that automated techniques produce student assessments comparable to human appraisals of student performance and employed within an after-action debrief, resulted in more effective training, as compared to a baseline after-action debrief capability. These studies used the E-2 Enhanced Deployable Readiness Trainer (EDRT), a medium-fidelity simulation trainer employed for training E-2 Hawkeye Naval Flight Officers (NFOs). This paper will summarize further developments to combine behavioral with biometric measures as a basis for automated performance assessment. In particular, speech communications and EEG were assessed as two-person teams of E-2 NFOs completed relatively complex mission scenarios on the EDRT. Using biometric measures, it was possible to distinguish the performance of expert and novice teams, providing a proof-of-principle of feasibility.

As military aviation systems become increasingly complex, training has become a significant cost-driver in the life cycle of aviation platforms. Consequently, there is need for technology innovations that increase the effectiveness of military aviation training, while lessening the demands on human instructors, and their support staff.

Automated performance assessment has been advanced as a technology that should allow reductions in the manpower required to support training operations. This assertion is based on current practices which require instructors observe and grade student performance as students complete missions within simulation-based trainers. For complex operations, the instructor-to-student ration may reach one-to-one, with the need for human role players creating even greater manpower requirements. In theory, by using automated techniques to assess certain facets of student performance, there should be a reduction in the cognitive workload for instructors enabling training to be accomplished with fewer instructors. Furthermore, automated performance assessment is well suited for performance measures that involve continuous attention to detailed facets of student performance (e.g. situationally-dependent duration of radio communications), providing a basis for objectifying such measures.

Previous research has established that the performance assessments obtained using automated techniques are accurate, as compared to equivalent measures obtained through human observation and assessment of student performance (Stevens et al, 2010, Stevens et al, 2010). This research utilized the E-2 Enhanced Deployable Readiness Trainer (EDRT), a medium-fidelity simulation trainer used to train E-2 Hawkeye Naval Flight Officers (NFOs), and focused on key performance measures appropriate for entry-level NFO training (e.g. fleet protection, or preventing enemy aircraft from coming within close enough proximity to pose a threat to a Naval carrier group). Furthermore, when presented as a component of an instructor's after-action debrief, more effective training was achieved with automated performance assessments, as compared to a condition employing a baseline after-action debrief capability (Stevens et al, 2010). For these efforts, the input to the automated performance assessment consisted of readily available data, and specifically, the geometric relationships between entities within the simulation (e.g. relative positions, directions and speeds of enemy and friendly entities) and distinct transactions as students operated the simulator (e.g. labeling of entities, depressing foot pedal actuator for radio). The following paper considers extension of these data sources to include analysis of speech communications and biometric measurement of brain activity, presenting a proof-of-principle demonstration in which speech and EEG serve as inputs to automated performance assessment.

Automated Expert Modeling and Student Evaluation (AEMASE)

AEMASE has been advanced as both an approach to automated performance assessment, and as specific algorithmic instantiations of this approach (Abbott, 2006). As an approach, AEMASE consists of a three-step process. First, an expert demonstrates desired behavior within either a simulator or instrumented environment. Key to this step is the prior identification of key performance parameters (i.e. features) underlying task performance. Based on data from experts, machine learning techniques are employed to derive a model of expert performance. The specific techniques employed may vary depending on the performance measure. In many cases, performance measures have been modeled using a vector-based representation combining different features within a multi-dimensional parameter space. For example, student's performance for fleet protection (i.e. preventing enemy aircraft from posing a threat to a Naval carrier group) may be modeled as a vector that combines the features for each enemy aircraft: (1) distance from carrier group; (2) angle off and (3) velocity. In the third step, during a training exercise, data is fed into the expert model which provides predictions concerning appropriate courses of action. The actual performance of the student is then compared to these predictions and the difference between predictions generated by the expert model and the student's actual behavior provide the basis for assessing the student's performance.

While various approaches have been employed for automated performance assessment, such as intelligent tutoring systems concepts (e.g. Corbett, 2001), there is a distinction worth noting. The expert models used in the AEMASE approach are based on statistical analysis of data produced as experts perform within a representative task environment. One of the costliest elements of most automated performance assessment concepts is knowledge engineering (i.e. expert interviews, task decomposition, etc.) required to derive a detailed model of expert performance. AEMASE does require some degree of knowledge engineering, but this is primarily restricted to steps associated with identifying performance measures, and associated data features, and obtaining sufficient instances of expert performance. Thus, AEMASE provides a more cost-effective approach for system development, and given interface features that allow users to readily modify expert models, AEMASE streamlines the process for later updating the system.

Accuracy and Utility of AEMASE Automated Performance Assessment

To date, the most extensive implementation of the AEMASE approach has been for training E-2 NFOs. To assess the accuracy and utility of AEMASE automated performance assessments, laboratory studies have been conducted using the E-2 EDRT simulation trainer. In these studies, test subjects were recruited from the employee population of Sandia National Laboratories with demographics comparable to entry-level E-2 NFOs. Subjects then underwent a program of training to provide them with the basic skills needed to complete relatively complex, yet entry-level E-2 mission scenarios. This training consisted of an 8-hour classroom session taught by a reservist E-2 NFO and five sessions on the simulation trainer focused on the development and practical application of key skills. Students were then brought back for a final data collection session in which their proficiency was assessed as they completed two missions requiring an integration of the knowledge and skills attained in the earlier training sessions.

In the first of two studies, the objective was to compare automated assessments with those of human instructors. For this study, three performance measures were chosen that were each deemed to be highly relevant to the training objectives for an entry-level E-2 NFO. The first concerned fleet protection, or the effectiveness with which students recognized potential threats (i.e. enemy aircraft) to a Naval carrier group and committed friendly aircraft to intercept approaching enemy aircraft within a timely manner. The second measure involved the timeliness with which commercial aircraft were identified and labeled. The third addressed situation awareness and management of the battlespace and in particular, whether students recognized and responded correctly to a developing gap in their air defenses. With each measure, there was good correspondence between the automated and instructor assessments, with values of 100%, 95% and 83% respectively for the three measures.

A second study compared the performance of students trained using an after-action debrief featuring automated performance assessment to students trained with a baseline after-action review capability (i.e. scenario capture and replay). Subject trained with the AEMASE after-action debrief exhibited superior performance for performance measures that included fleet protection, the accuracy and latency for labeling commercial aircraft and the timeliness with which the warfare commander was informed following successful downing of enemy aircraft. There was no difference between groups for the measure of situation awareness and battlespace management described above, and it was concluded that this skill was too complex given the limited training provided to test subjects.

Incorporation of Voice and Biometric Data

Previous studies focused on automated performance assessment using readily available data concerning location of entities within the simulation scenario and student transactions with the simulation trainer. Also, in these studies, training and student assessments occurred on an individual level, outside the context of team operations. Given the degree to which tasks of the

E-2 crew involve a coordinated team effort, there was a certain artificiality in having students conduct missions individually.

A third study was conducted in which subjects participated as two-person teams. In this study, there were 8 subjects, divided into 4 two-person teams. Two of these teams consisted of subjects from the second study that had received training using the AEMASE after-action debrief tool, and were considered to be novices. The other two teams consisted of reservist E-2 NFOs and were considered to be experts. In addition to the data collected in previous studies, voice communications and dense-array EEG was recorded.

Analysis of both voice communications and EEG allowed expert and novice teams to be distinguished. With the EDRT, to activate the radio, students must depress a foot pedal and continue pressing the foot pedal for the duration of the radio call. Initial analysis of voice communications considered the duration of these pedal presses. As shown in Figure 1, across scenarios, expert teams generally pressed the pedal for shorter periods of time, indicating a greater degree of brevity in their radio communications. This is consistent with observations that a key facet of expert NFO performance involves the efficient use radio channels.

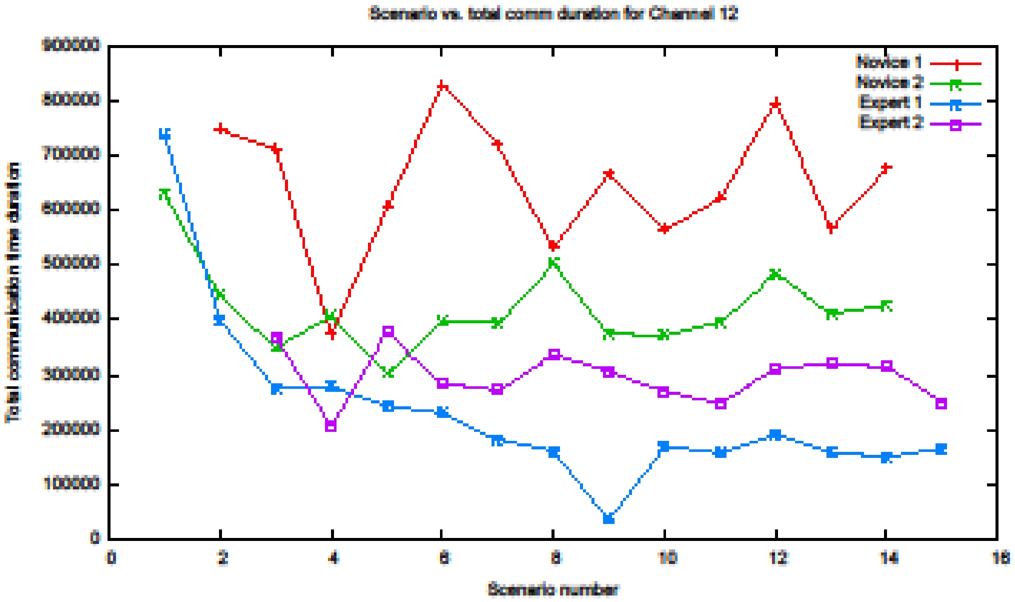


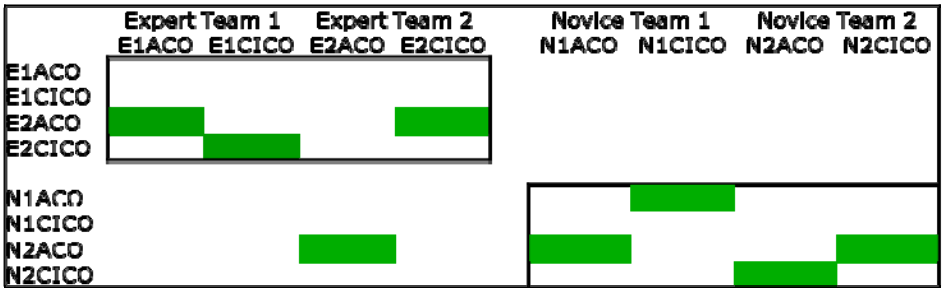
Figure 1. Generally, the duration of radio communications for novice teams was longer than for expert teams.

It was noted that novice subjects often seemed tentative in their radio communications, whereas expert teams tended to be deliberative and concise. Using speech-to-text transcription, the contents of radio communications was assessed. This tentativeness was evident in the use of filler words (e.g. ur, ah), as shown in Table 1. Overall, for five common filler words, their use occurred substantially less with expert, than with novice teams.

Table 1. Use of filler words for an illustrative scenario.

Filler Words	Experts	Novices
ah	1	6
er	4	8
like	5	9
uh	112	307
um	5	28
Total	127	358

Additional analysis of radio communications considered the semantic content of radio communications. For this analysis, transcriptions were indexed using a term frequency-inverse document frequency approach. Each subject's transcript was treated as a separate document and based on cosine similarity, each subject was compared to each of the other subjects to determine who their speech content most closely resembled. As depicted in Figure 2, for three of the four experts, their communications most resembled another expert. Furthermore, all four novices most closely resembled another novice.



	Expert Team 1		Expert Team 2		Novice Team 1		Novice Team 2	
	E1ACO	E1CICO	E2ACO	E2CICO	N1ACO	N1CICO	N2ACO	N2CICO
E1ACO								
E1CICO								
E2ACO	🟩			🟩				
E2CICO		🟩						
N1ACO						🟩		
N1CICO								
N2ACO			🟩		🟩			🟩
N2CICO							🟩	

Figure 2. Semantic comparison of radio communications for expert and novice subjects. Green shaded cells indicate the subject each subject most closely resembled.

Initial analysis of EEG data considered the relative levels of activity in the theta (4-7 Hz) and beta (13-30 Hz) bandwidths. Figure 3 depicts the moment-to-moment transitions of activity for individual electrodes for a sample of data. It was observed that experts exhibited greater variability in the beta bandwidth, whereas novices showed greater variability in the theta bandwidth. This would suggest that the neural processes being engaged by experts were somewhat distinct from those being exercised by novices. Further analysis reported by Dodel et al (in preparation), found higher power correlations and reduced dimensionality with the expert teams, suggesting that the coordination of team performance involves some degree of coordinated activation of neural processes.

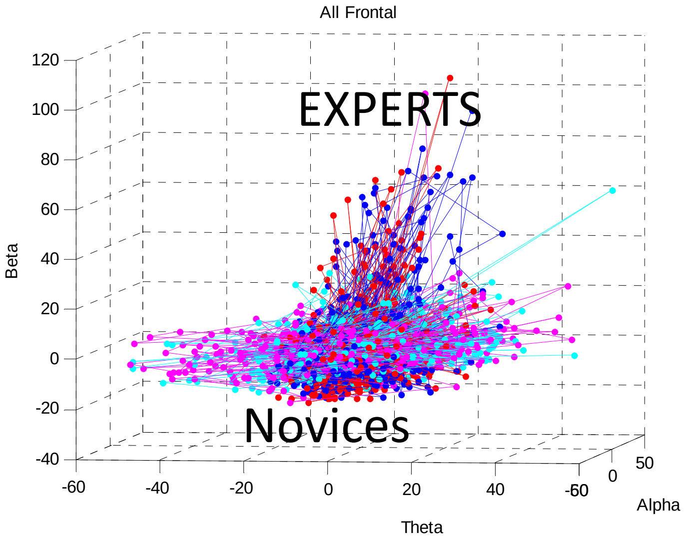


Figure 3. Expert teams showed greater variability in beta (13-20 Hz) bandwidth, while novices showed greater variability in theta (4-7 HZ) bandwidth

Conclusion

Automated performance assessment, as illustrated by the AEMASE approach, provides an opportunity to improve the effectiveness of training, while reducing cost by lessening the workload on instructors and streamlining the development process. Given that the AEMASE approach is based upon expert demonstration of desired performance, concerns may arise regarding the generality from scenarios used to train the expert models to other scenarios that differ in their contents and complexity. However, this same concern exists with expert models developed using traditional approaches based on knowledge engineering, in that resulting models are only valid for the contingencies identified and addressed during domain analysis and model development. A useful conceptualization identifies the key parameters underlying performance and represents those parameters within a multi-dimensional space. Any given scenario, or perhaps mission segment, may be depicted as a point within this parameter space. Ideally, an expert model should generalize to the entire parameter space. However, actual generality will be a function of the extent and care with which the parameter space is sampled in selecting the scenarios utilized to construct the expert model.

It may be noted that in the work summarized in this paper, automated performance assessments were based on comparing student performance to the predictions of an expert model. Often this may not be the most appropriate comparison, and the most appropriate comparison may be to compare a student to a model reflecting performance that is intermediate between the student and an actual expert. This could be readily accomplished by obtaining data reflecting a range of performance such that intermediate levels of performance are represented within the model against which a student's performance is compared. In fact, an important distinction of the AEMASE approach is that the performance of each expert whose data contributes to the model is reflected within the model (i.e. there does not have to be a single correct solution). This accommodates situations in which there are multiple acceptable solutions to a given problem. Thus, in the same way that AEMASE accommodates variation across experts with respect to their performance, varying levels of expertise may similarly be accommodated.

For the most part, development and experimental assessment of AEMASE implementations have only utilized behavioral performance data. The integration of behavioral performance data with biometrics data offers a mechanism by which these capabilities may be extended to provide more thorough assessments of student performance. As illustrated with voice communications, a novice may say all the right words, but do it in a manner that is ineffective (e.g. use of excessive filler words). Similarly, a novice may complete a mission and if their performance is assessed on a behavioral level, they accomplish all their objectives. However, their cognitive and physiological resources may have been taxed nearly to the breaking point, whereas an expert routinely accomplishes the same objectives with ease. This discrepancy may not be readily apparent and the student allowed to progress, despite their capabilities being on the margins, and likely quite brittle if placed in a stressful situation. Biometric measurement should provide a mechanism to not only assess student's behavioral performance, but to additionally assess the levels of mental and physiological exertion required to obtain performance objectives.

Acknowledgements

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. Work described in this paper was performed through contract awards from the Office of Naval Research and DARPA.

References

Abbott, R. G., (2006). Automated Expert Modeling for Automated Student Evaluation. *Intelligent Tutoring Systems*, 1-10.

Corbett, A. T. (2001). Cognitive computer tutors: Solving the two-sigma problem. In UM '01: Proceedings of the 8th International Conference on User Modeling, (pp. 137 – 147). London: Springer-Verlag.

Dodel, S., Jirsa, V. Mersmann, J., Forsythe, C., Wheeler, T.L. & Cohn, J. (in preparation). Expert teams show reduced variability and enhanced coordination of source generators on a complex decision making task. *Proceedings of Human Computer Interaction International*, Orlando, FL.

Stevens, S., Basilico, J., Forsythe, C., Abbott, R. & Gieseler, C. (2010) Experimental assessment of automated knowledge capture. Proceedings of the I/ITSEC, Orlando FL.

Stevens, S., Basilico, J., Forsythe, C., Abbott, R. & Gieseler, C. (2010) Using After-Action Review Based on Automated Performance Assessment to Enhance Training Effectiveness Proceedings of the Human Factors and Ergonomics Society, San Francisco, CA.