

Pilot Subjective Assessments During an Investigation of Separation Function Allocation Using a Human-In-The-Loop Simulation

Dr. Kelly A. Burke¹, David J. Wing² and Timothy Lewis³
NASA Langley Research Center, Hampton, VA 23681, USA

Two human-in-the-loop simulation experiments were conducted to investigate allocation of separation assurance functions between ground and air and between humans and automation. The experiments modeled a mixed-operations concept in which aircraft receiving ground-based separation services shared the airspace with aircraft providing their own separation service (i.e., self-separation). The two experiments, one pilot-focused and the other controller-focused, addressed selected key issues of mixed operations and modeling an emergence of NextGen technologies and procedures. This paper focuses on the results of the subjective assessments of pilots collected during the pilot-focused human-in-the-loop simulation, specifically workload and situation awareness. Generally the results revealed that across all conditions, pilots' perceived workload was low to medium, with the highest reported levels of workload occurring when the pilots experienced a loss of separation during the scenario. Furthermore, the results from the workload data and situation awareness data were complimentary such that when pilots reported lower levels of workload they also experienced higher levels of situation awareness.

I. Introduction

An essential part of developing the Next Generation Air Transportation System (NextGen) is the exploration of new technologies, procedures, and human roles in providing services and functions for the safe and expeditious passage of aircraft. Separation assurance is a key function of Air Traffic Control (ATC) and a core responsibility of air traffic controllers in current-day operations. Due to the safety criticality of separation assurance, a complex system of airspace and route structures, surveillance and communication technologies, and operational controls on aircraft trajectories has evolved to enable a separation assurance environment in which controllers, using voice communication with pilots, can sustain safe operations with manageable workload.

This evolved, complex system is reaching its limits in accommodating new traffic demand and satisfying operator needs for efficiency. A significant constraining factor is the controller's workload in communicating with and separating aircraft. Currently, the controller is responsible for nearly all separation-related functions. Ground automation plays an ancillary role in enhancing controller situation awareness, and the aircraft (i.e., the flight crew and airborne automation) has a passive role with respect to separation, simply obeying trajectory instructions (except in collision avoidance situations where an active role is taken). Given that human workload capacity cannot be substantially increased, other means will be needed to stretch beyond the current limits.

The Concepts and Technology Development Project of the NASA Airspace Systems Program is exploring fundamental changes to separation assurance that are intended to reduce human workload as a limiting factor. Through this "function allocation" research thrust, NASA researchers are testing separation concepts that leverage the extensive use of automation and the untapped, distributed resource of aircraft systems and crews. Concepts for ground-based and airborne separation developed and researched over the last decade and beyond are brought together into a "mixed operations" environment where the maximum use of all resources for separation can be explored.

¹ Human Factors Research Scientist, Crew Systems & Aviation Operations, Mail Stop 152, kelly.a.burke@nasa.gov, AIAA member

² ATM Research Engineer, Crew Systems & Aviation Operations, Mail Stop 152, david.wing@nasa.gov, AIAA member

³ Aerospace Engineer, Crew Systems & Aviation Operations, Mail Stop 152, timothy.a.lewis@nasa.gov, AIAA member

II. Separation Function Allocation Concepts

A. Airborne Separation Concept

The airborne separation concept leverages the attributes of both distribution and automation in its approach to function allocation. In this approach, separation functions for individual aircraft are performed onboard the aircraft (i.e., the “ownship”) to provide separation from all traffic the ownship encounters. The crew (rather than the controller) manages its own trajectory during en route flight, and “self-separates” from all traffic by adjusting the trajectory as needed to resolve conflicts identified by the onboard equipment. With multiple self-separating aircraft in the airspace, the separation “service” is distributed among them and resident onboard each equipped aircraft. Conflict detection and alternative trajectory generation automation onboard the aircraft is heavily leveraged to avoid the flight crew having to provide such capability manually.

Aircraft that manage their own trajectory and separation are referred to as flying under Autonomous Flight Rules (AFR). AFR distinguishes these aircraft from Instrument Flight Rules (IFR) aircraft, which are managed by and receive separation services from ATC. While AFR aircraft optimize their own trajectory through airspace shared with IFR traffic, the mixed operations concept tested in these experiments requires AFR aircraft to yield right-of-way to IFR aircraft in all conflict encounters and to take responsibility for ensuring the separation standard is met. To meet this responsibility, the flight crew uses onboard automation that processes data from ownship avionics and airborne surveillance (Automatic Dependent Surveillance Broadcast, ADS-B) to probe for conflicts and compute resolution maneuvers, and possibly several acceptable alternatives. The crew chooses the desired maneuver and executes it directly. Because the separation function is performed onboard, no ATC approval is needed to maneuver. AFR intent information is electronically available to controllers, but they bear no responsibility for separation between AFR and IFR aircraft. Coordination between AFR pilots and controllers, if needed, is conducted by voice communication.

B. Ground-Based Separation Concept

Separation assurance in the National Airspace System today is ground-based, manual, and limited by how many aircraft the air traffic controller can keep under positive control. Emergent technologies are intended to support air traffic controllers in detecting conflicts and generating solutions and will reduce some of the coordination and communication workload. The current near- and mid-term NextGen plans foresee the introduction of additional decision-support tools to improve operations, but under the current paradigm, the human operator remains responsible for providing separation between all aircraft.¹

This concept of ground-based automated separation assurance utilizes technologies to shift the workload-intensive tasks of monitoring and separating traffic from the controller to the automation. A critical element of this centralized concept makes the ground-side automation, not the controller, responsible for conflict detection. In many cases, the automation, not the controller, is responsible for resolving conflicts as well. However, the controller is responsible for maintaining separation of unequipped aircraft using a conventional voice link and steps in to handle certain off-nominal situations. Thus, under automated separation assurance, air traffic controllers’ roles involve providing services and performing decision-making activities, while the roles of monitoring, providing nominal separation, and back-up solutions in off-nominal situations are allocated to the automation. For a more detailed description of the controller-focused experiment see Wing, et al.²

III. Pilot-Focused Experiment

The pilot-focused experiment addressed the perspective of the AFR pilot in mixed operations. Conducted at the NASA Langley Air Traffic Operations Laboratory (ATOL), the experiment focused on two issues: the ability of AFR aircraft to shoulder the burden of detecting and resolving conflicts with ground-controlled IFR aircraft, and the value of AFR aircraft having access to IFR intent information. The experiment was organized in two parts, a set of “primary” runs to examine the first issue and a set of “exploratory” runs for the second issue.

The primary portion of the pilot-focused experiment was designed to identify the limits under which AFR aircraft can ensure separation from IFR aircraft in normal operations. The parameters of interest included amount of alerting time and conflict geometry. It was hypothesized that achieving adequate separation performance requires some minimum amount of alerting time for pilots, and that conflict geometry does not have an interaction effect with this alerting time. To ensure AFR aircraft can shoulder the separation burden of mixed operations, it will be important that this minimum required alerting time be guaranteed through a combination of ATC restrictions on IFR

maneuvers, IFR intent information sharing, and AFR automation design and parameter settings. An example of the latter is the use of extra separation buffers beyond the minimum required separation standard, a technique initially explored in this experiment. To test the effects of alerting time and conflict geometry, a series of scripted conflicts with a range of carefully controlled IFR maneuver timing and encounter orientation were created for AFR flight crews to resolve using automation tools.

The exploratory portion of the experiment addressed the value of AFR aircraft having access to trajectory intent information from IFR aircraft. Currently, the ADS-B mandate does not require broadcast of intent, either trajectory change points or target state information (e.g., target headings and altitudes). The absence of this intent information may cause conflicts to “pop-up” with shorter notice, as planned IFR maneuvers are executed without forewarning. To explore the issues surrounding intent information exchange, AFR flight crews flew a series of scenarios representing normal en route operations with and without automatic data transmission of IFR intent information. Confederate controllers and IFR aircraft pilots participated in these runs to provide for normal interactions.

C. Experimental Design

The primary experimental test matrix was designed to test the ability of the AFR pilots to resolve conflicts of varying timing and geometry. A test matrix of conflict scenarios was generated using a fractional $[4]X[2X2X2]$ between-subjects design with four categorical factors. The first factor was *Time to Buffer Loss* (TBL), the amount of alerting time given to pilots prior to reaching a buffered protected zone around the IFR aircraft (8 nmi lateral separation, 1000 ft vertical separation). An additional bin for alerting times less than 20 seconds included only vertical encounters, due to the difficulty in creating such short-notice lateral encounters. This bin contained six data points and was not combined with the other bins for statistical analysis.

The $2x2x2$ portion of the design represented three elements of conflict geometry: *encounter angle*, *maneuver dimension*, and *passage orientation*. The first element, *encounter angle*, was divided into convergence angle bins with the two aircraft either roughly aligned (“acute”) or roughly opposed (“obtuse”). *Maneuver dimension* refers to the IFR aircraft approach direction being either “lateral” (same altitude and level with the AFR aircraft) or “vertical” (climbing or descending into separation loss, with the AFR aircraft always level). *Passage orientation* refers to whether the IFR aircraft in the encounter would “pass in front” of the AFR aircraft or “pass behind” if the conflict were left unresolved.

A total of 34 airline pilots participated as 17 flight crews, with each crew flying 12 scenarios in the primary matrix. Each flight crew saw 12 scripted conflicts, but not the same 12 conflicts. A between-subjects design with blocking by groups of 3 crews was used to provide balanced coverage of the fractional test matrix, where not every geometry combination could be tested at every TBL condition. A more detailed description of the primary matrix experiment design and conditions tested is presented in Wing, et. al.²

D. Facility Description

The Langley ATOL was configured with six “team pilot” stations that permitted a flight crew to share a desktop simulator. Referred to as an Aircraft Simulation for Traffic Operations Research (ASTOR), the desktop simulator provides the displays and controls of a modern Boeing-style widebody jet aircraft. Integrated with the avionics system is an automation tool designed to support the flight crew in self-separation operations. The Autonomous Operations Planner (AOP)⁵ provides a full suite of conflict detection, resolution, and prevention tools, using information obtained from ownship systems (aircraft state, autoflight settings, active strategic route, pilot-specified tactical flight targets) and ADS-B (traffic aircraft states, intent information if available), among other sources of information. AOP supports both “strategic” trajectory-based operations, i.e., fully coupled to the Flight Management System (FMS), as well as “tactical” operations using pilot-specified flight targets set on the Mode Control Panel (MCP). The flight crew is alerted to conflicts detected by AOP on a textual display, as well as by audible



Figure 1. Example AOP display of a tactical-urgent conflict.

alerting and graphical depiction on the Navigation Display. Fig. 1 shows an example AOP display for a “tactical urgent” conflict.

IV. Subjective Assessments

Flying an aircraft involves a complex, multidimensional series of behaviors, only some of which can be observed directly. Pilots must communicate, navigate, control, and continuously monitor their environment. They must often accomplish this using a plethora of systems involving human-machine interaction. Cockpit procedures, technology, and instrumentation continue to change and become more complex as researchers design and develop improvements to the human-machine interactions within the cockpit in an effort to enhance safe flight operations. Consequently, there is a continued need to evaluate the potential impact of these changes on pilot workload and situation awareness to understand how well these designs and systems meet their needs, capabilities, and limitations. One primary way to examine these constructs has been the use of subjective assessments. The use of subjective ratings to measure mental workload and situation awareness is a common technique and these are often used in conjunction with performance measures. Both workload and situation awareness have important implications for flight safety and excessive workload and loss of situation awareness are commonly cited as contributing to aviation accidents. Consequently, both are commonly regarded as important design considerations for advanced aerospace systems.

The current experiment explored separation concepts that leverage the extensive use of automation and the untapped, distributed resource of aircraft systems and crews and focused on the ability of AFR aircraft to shoulder the burden of detecting and resolving conflicts with ground-controlled IFR aircraft. The experiment was designed to identify the limits under which AFR aircraft can ensure separation from IFR aircraft in normal operations with the primary parameters of interest being the amount of alerting time and conflict geometry. Clearly it is important to gain an understanding of how these parameters affect pilot workload and situation awareness. In an effort to inform the ongoing design of an airborne separation assurance automation tool, subjective assessments of pilot workload and situation awareness were administered during the experiment.

A. Workload

Pilot mental workload is often high due to the complexity of the flying task and refers to the number of mental and physical tasks a pilot needs to do, the time period in which these tasks must be completed, as well as the complexity of the tasks. Relative increases in pilot workload generally result in subsequent reductions in pilot performance, especially at the cognitive level⁶. Generally, more complex tasks will increase workload more than less complex or less difficult tasks, unless the complex tasks are well rehearsed and become nearly automatic. When workload levels continue to increase, performance decrements will eventually occur. This level will depend on several things, including pilot arousal levels, which are influenced by fatigue and motivation (higher pilot arousal can allow higher workload levels before performance decrements begin⁷), and the commonality of multiple tasks (the more common concurrent tasks are, the more likely task decrements will occur⁸).

Since the advent of concern for human-machine interaction in the cockpit, researchers have been interested in evaluating how well equipment and system designs meet the capabilities, needs, and limitations of the pilots. As technology advances and the systems designed for cockpits become more complex and increasingly reliable, the weak link in the human-machine system is the human operator, whose reliability can often be a function of the load placed on him/her. In aviation in particular, the relationship between safe flight operations and pilot workload could have serious consequences. The evaluation of pilot workload has represented a complex measurement problem since the earliest days of manned flight. A frequently used method has employed post-flight, and when practical, post-task (during flight or simulated flight) questionnaires. Simulation scenarios that impose predictable and objectively determined levels of workload on pilots are an essential element of research on current and future aircraft systems and procedures. In the context of an individual performing a task, workload reflects the degree of congruence between task demands and that individual’s self-appraisal of available resources⁹.

In the current study it was also of interest to evaluate differences in perceived workload between the Pilot Flying and the Pilot Monitoring. Pilots have many tasks to perform and these are normally shared between the Pilot Flying and the Pilot Monitoring. Flight crew workload varies, even during routine flights, from low to high and will rise in the event of abnormal weather conditions or aircraft malfunctions. During high workload situations, the flight crew are especially vulnerable to error if their strategies for effective multi-tasking breaks down.

B. Situation Awareness

The concept of situation awareness (SA) is a construct, not directly observable, used among human-factors specialists to describe and measure the long held intuitive notion among pilots that successful flight results when a pilot has the “big picture,” and conversely when problems arise due to pilot error, it is because some aspect of this picture is missing or incorrect. SA refers to the pilot having an accurate mental representation of the material state of the environment they are operating in at the present time¹⁰. One of the most commonly cited definitions of SA¹¹ defines it 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.” (p. 36). SA involves three stages:

- perception (observing the environment);
- comprehension (how does the state of the perceived world affect me now); and
- projection (how will it affect me in the future)¹¹.

A loss of situation awareness occurs when there is a failure at any one of these stages resulting in the pilot not having an accurate mental representation of the physical and temporal situation.

The situation awareness concept has been extended to other domains such as air traffic control, battlefield management, and medical procedures. These domains share common characteristics; for example: (a) the environment is often dynamic and information rich; (b) the human may sometimes experience high mental workload; (c) extensive training is often required; (d) the problems are often ill-structured; and (e) time is often constrained. However, in the operational setting of aviation, the concept of SA is especially compelling because it involves the operation and control of a complicated system in a dynamic environment. In this environment, the pilot must integrate disparate and sometimes inconsistent inter-sensory input (visual, auditory, tactile, vestibular, etc.) with elaborate cognitive models of the machine and the operating environment to control the movement of an aircraft.

Subjective measures of SA measure situation awareness by either self-assessment ratings or by the assessment of the observer, and are based solely on the opinion of the participant or the observer. For example, on a given scenario or task, a participant might be asked to use a Likert-type scale ranging from “1” to “7” in rating the amount of situation awareness experienced. These measures are useful because they are easy to implement and are also practical because they may be used both in simulations and the actual task environment.

A. NASA-Task Load Index

1. Description

The NASA-Task Load Index (TLX)¹² is a subjective workload assessment tool designed to test workload metrics specifically in the aviation community and allows users to perform subjective workload assessments on operators working with various human-machine systems. It is a multidimensional instrument that provides a reliable index of global workload and also identifies the relative contributions of six sources of workload. Three of those sources reflect the demands that tasks place upon operators (mental, physical, and temporal demand), whereas the remainder characterize the interaction between the operator and the task (performance, effort, and frustration). Aspects of the task, behavior, and the operator are all included in the TLX. To produce a global measure of workload, the scales are traditionally combined using paired-comparison derived weights. Paired-comparison weights for the six scales must be obtained for each task rated. The number of times the dimension is chosen is the derived weight, ranging from 0 (never chosen) to 5 (always chosen). In addition to the paired-comparisons, pilots also make ratings of the six scales which are scored on a 0-100 scale. Each rating scale score is multiplied by its derived weight and the six weighted scores are averaged to obtain a single combined TLX score. This two-pass process of data collection for each task is labor intensive and therefore very time-consuming to administer. Because our participants were active line pilots participating between scheduled trips, we had limited time with them for training and data collection and had to economize time where possible. Consequently, the time available to administer the subjective assessments was limited; therefore the Raw Task Load Index (RTLX)¹³ was evaluated in this experiment. The RTLX is a viable alternative to the NASA-TLX, requires significantly less time to administer and is based upon a sum of the six scales (RTLX = SUM/6). Additionally, Byers, Bittner, and Hill (1989) indicated that RTLX scores can even provide a better account of the workload experienced by the participants than traditional weighted TLX scores. This eliminates the need to collect the paired comparisons portion of the NASA-TLX, which experience and previous

research has shown to be difficult and tedious for the operators to complete, both from a training and administration stand point.

2. Administration Procedure

In the pilot-focused experiment, pilots provided a workload assessment after every scenario by completing the NASA-TLX set of six rating scales. They rated their perceived experience of workload on an 11-point scale, from “Low” to “High” for all factors of workload except performance which was rated from “Good” to “Poor.” The scale and anchor points are consistent with the traditional NASA-TLX. The pilots received the instructions developed by Hart and Staveland for completing the rating scales prior to the start of the experiment. Although two pilots “flew” the scenarios together as a crew, with one pilot assuming the role of Pilot Flying and the other pilot assuming the role of Pilot Monitoring, they were instructed to complete the subjective assessments individually. The pilots alternated roles every scenario so that each pilot assumed each role for half of the scenarios. Additionally, after completing a practice scenario, the pilots were given a practice RTLX so that they could become familiar with the questions and the method of responding. For this experiment, the RTLX was administered electronically via Lime Survey, an online survey system. This method allowed the RTLX to be presented on the computer at each pilot station. After the completion of each scenario, the questionnaire was loaded up on each computer by the researchers so that the pilots could respond individually using their keyboard and mouse. An example presentation of a question is depicted in Figure 2.

* Please select the point on each scale that best indicates your experience of the task you just completed. Each scale has two endpoint descriptors that describe the scale. Note that Performance goes from "Good" on the left to "Poor" on the right. This order has been confusing for some people. Please consider your responses carefully in distinguishing among the different task conditions. Consider each scale individually. Your ratings will play an important role in the evaluation being conducted, thus, your active participation is essential to the success of this experiment and is greatly appreciated by all of us.

Mental Demand

How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?

Low High

○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○

Figure 2. Example question from the NASA-TLX as administered in Lime Survey.

B. Situation Awareness Rating Technique (SART)

1. Description

The Situation Awareness Rating Technique (SART)¹⁴ is a subjective measure of situation awareness that can provide an index of how well operators are able to acquire and integrate information in a complex environment. SART was developed as an evaluation tool for aircrew system design and provides a subjective rating of situation awareness by operators, including pilots and controllers. Operators rate on a series of scales the degree to which they perceive 1) a demand on operator resources, 2) supply of operator resources, and 3) an understanding of the situation. Pilots were asked to rate their situation awareness immediately after completion of each scenario during the experiment. SART for pilots has a total of 14 components: three global ratings of the basic dimensions of Demand, Supply, and Understanding plus ratings of the 10 elements composing each dimension, plus a situation awareness simple rating “How good was your awareness of the situation?”. Pilots rated their situation awareness on each of the ten dimensions using Likert-type scales from 1 (“Low”) to 7 (“High”). The scales are then combined to provide an overall situation awareness score for each pilot after each scenario using the following formula:

$$\text{Situation Awareness (Calculated)} = \text{Understanding} - (\text{Demand} - \text{Supply}).$$

2. Administration Procedure

Pilots provided an assessment of their situation awareness after every scenario by completing the SART. The pilots received the instructions for completing the SART during a briefing prior to the start of the experiment.

Additionally, after completing a practice scenario, the pilots were given a practice SART so that they could become familiar with the questions and the method of responding. The SART was administered electronically via Lime Survey similarly to the RTLX.

V. Results and Discussion

The results presented in this paper represent the data collected from the subjective measures administered over a three-week testing period during the Separation Allocations in Shared Airspace (SALSA) human-in-the-loop (HITL) experiment. Generally, only those results where statistical differences were found are presented in this paper. It was necessary to exclude 16 runs from the system performance analyses due to preemptive pilot maneuvering that invalidated the planned scenario. Essentially some of the pilots began to maneuver their aircraft prior to the start of the intended conflict scenario. This resulted in the pilots experiencing either a different conflict initially or no conflict at all. Because their experience during that scenario was different than their experience with the intended conflict would have been, their perceptions of workload and situation awareness during that run may have also been different. Therefore, those runs were also excluded from the analyses of the subjective measures.

A. NASA-TLX

Participants rated their workload on each of the six NASA-TLX rating subscales following the completion of each primary scenario. Analysis of variance was conducted to assess the impact of the Time to Buffer Loss, Loss of Separation, Buffer Zone Traversal, and the conflict geometry of the intruder aircraft (angle, dimension, and orientation) on the TLX subscale scores. This experiment was executed via a desktop simulation, so the pilots “flew” their aircraft and navigated through the cockpit controls using a mouse. Given that this was not a physically demanding task to begin with and that the physical demands of the task did not change substantially throughout the experiment, changes in ratings of perceived physical demand across conditions were not expected. Consistent with that expectation, there were no significant results for physical demand across all conditions except for pilot role (Pilot Flying vs. Pilot Monitoring), which is most likely a function of the use of the mouse. The desktop simulator utilized a single mouse to operate the cockpit controls, and the Pilot Flying (PF) assumed that responsibility. The Pilot Monitoring (PM) monitored the flight and the radio via a headset and interacted with the PF, but did not use the mouse during the scenarios. The only component of the task that required any amount of physical demand was using the mouse, which was always accomplished by the PF.

Additionally, with the exception of when there was a loss of separation during the scenario, there were no statistically significant differences in the Performance subscale across all conditions. In fact, all the pilots, regardless of whether they were assuming the role of PF or PM, rated their performance very high across all experimental conditions and significantly higher than the ratings of the other subscales. Additionally, because the means of the ratings of Physical Demand were much lower than the means of all of the other subscales and because the means of the ratings of Performance were significantly higher than the means of all the other subscales, potentially affecting the overall workload RTLX rating, all subsequent analyses were conducted on each of the subscales individually to further evaluate specific differences.

B. SART

Participants rated their perceived situation awareness on each of the SART subscales following the completion of each primary scenario. Analysis of variance was conducted to assess the impact of the Time to Buffer Loss, Loss of Separation, Traverse the Buffer Zone, and the Conflict Geometry of the intruder aircraft (angle, dimension, and orientation) on the SART Calculated score.

C. Pilot Role

1. Pilot Workload - NASA-TLX

Participants rated their workload on each of the six NASA-TLX rating scales following the completion of each scenario. Each crew in the experiment was asked to work out their own crew procedures for AFR operations according to their company crew procedures. When the pilots assumed the role of PF, they had more responsibility for flight operations than the PM including “flying” the aircraft using the mouse, monitoring the Navigation Display (ND), monitoring the AOP display for conflict alerts, and decision making about the procedures for resolving the conflicts, while the PM was checking each step and calling out changing information as needed. The PM was engaged and caught potential errors, but the PF had the main work. Therefore, differences in reported workload

between the PF and the PM were expected and a one-way ANOVA was conducted on the subjective ratings of workload by pilot role. As expected, the results indicate that the pilots perceived higher workload when they assumed the role of PF than when they were the PM. However, it is important to note that generally across all conditions, pilots' (PF and PM) perceived workload was low to medium. Specifically, for all conditions tested the mean TLX ratings did not exceed 6.0 on the 11-point Likert scale, with the exception of when there was a loss of separation, where the mean TLX ratings did not exceed 8.0.

There was a significant main effect of pilot role for four of the six subscales, where pilots rated workload higher when they were the PF than when they were the PM. (See Table 1). Although the main effect of pilot role on the subscale of Frustration was not statistically significant, the trends in the data indicate that when pilots were the PF they reported higher levels of frustration than when they were the PM. In light of these results and to further evaluate the differences in perceived workload between the PF and PM, all subsequent analyses for workload were conducted separately for PF and PM. Across all conditions for workload, pilots generally reported higher levels of the TLX scales when they were PF than when they were the PM.

Table 1. TLX ratings as a function of pilot role.

NASA-TLX Sub-scale	Pilot Flying	Pilot Monitoring	
Mental Demand	$M = 5.20, SE = .23$	$M = 4.43, SE = .22$	$F(1, 344) = 5.83, p = .02$
Physical Demand	$M = 3.75, SE = .21$	$M = 2.99, SE = .18$	$F(1, 344) = 7.72, p = .01$
Temporal Demand	$M = 4.86, SE = .24$	$M = 4.17, SE = .22$	$F(1, 344) = 4.41, p = .03$
Effort	$M = 4.25, SE = .20$	$M = 3.68, SE = .19$	$F(1, 344) = 4.10, p = .04$
Frustration	$M = 3.33, SE = .18$	$M = 3.09, SE = .18$	$F(1, 344) = .90, p = .35$
Performance	$M = 9.86, SE = .13$	$M = 9.89, SE = .13$	$F(1, 344) = .02, p = .88$

2. Situation Awareness - SART

There was not a significant main effect of Pilot Role on the SART Calculated scores which suggests that the pilots did not perceive their level of situation awareness differently as a function of when they were assuming the role of PF versus when they were the PM.

3. Discussion

Pilots rated workload significantly higher when they were the PF than when they were the PM. However, pilots did not perceive their level of situation awareness differently as a function of when they were assuming the role of PF versus when they were the PM. These results are consistent with the experimental setup and the responsibilities given to the PF and PM. The test environment was a desktop flight simulator operated via a single mouse under the control of the PF. The PF and PM conferred on all conflicts and decisions regarding maneuvering the aircraft, which supports the finding of similar situation awareness.

D. Time to Buffer Loss

Time to Buffer Loss (TBL) was the amount of alerting time given to pilots prior to reaching a buffered protected zone around the IFR aircraft (8 nmi lateral separation, 1000 ft vertical separation). There were four alerting time bins included in this analysis: 20-60 seconds, 1-2 minutes, 2-4 minutes, and 4-10 minutes.

1. NASA-TLX

Workload data was analyzed to determine whether the TBL affected pilots' perceived ratings of the amount of workload required during that scenario. Significant main effects of TBL were present for Temporal Demand, $F(3, 342) = 5.52, p = .001$; Frustration, $F(3, 342) = 4.00, p = .008$; and Effort, $F(3, 342) = 2.65, p = .05$ (See Figure 3).

To evaluate the differences between alerting time bins further, pairwise comparisons were conducted. For each factor where there was a significant main effect of TBL, there was a significant difference between the 20-60 seconds time bin and all other time bins, with one exception: for the Effort factor there was not a significant difference between the 20-60 seconds bin and the 1-2 minutes time bin. Additionally, for the Temporal Demand factor there was also a statistically significant difference between the 1-2 minutes and 4-10 minutes time bins, $p = .03$.

In general, the results indicate that as the alerting time to buffer loss decreased, pilots reported an increase in temporal demand, frustration, and effort, with higher levels reported for temporal demand, than effort and frustration, respectively. These results are consistent with expectations in that conflict alerting time affected pilots' perceived workload, specifically for temporal demand.

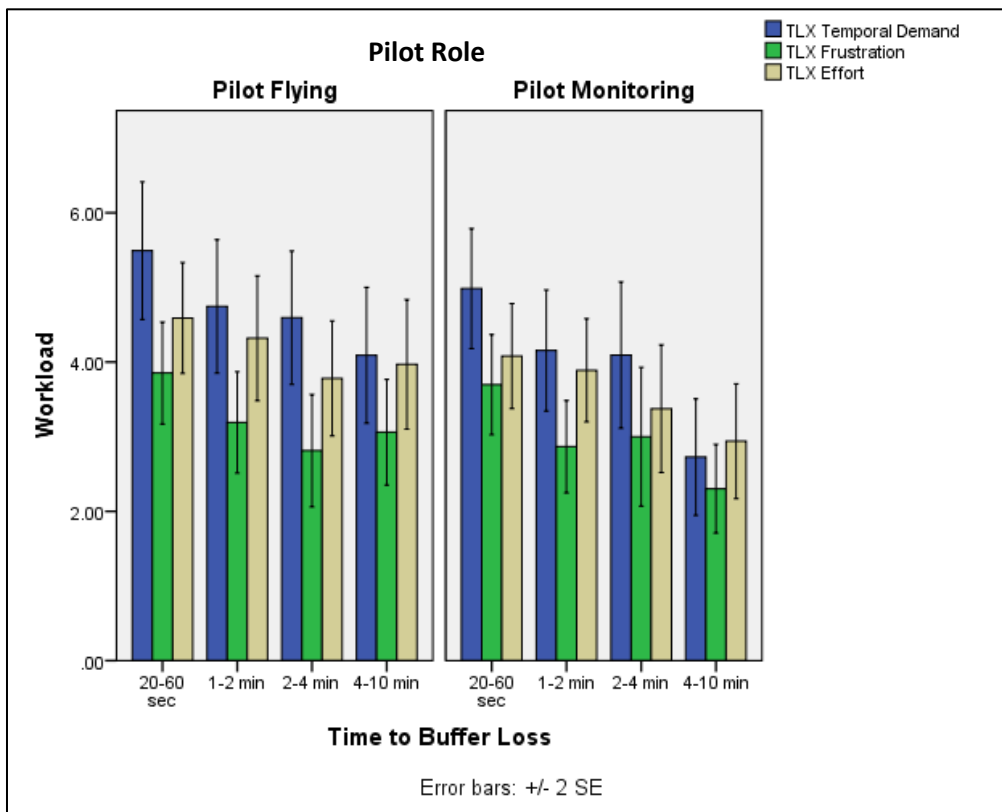


Figure 3. The effect of alerting time to buffer loss on the NASA-TLX subscales of Temporal Demand, Frustration and Effort.

2. SART

The calculated situation awareness score from the SART was analyzed to determine whether the TBL affected pilots' perceived ratings of their level of situation awareness during that scenario. There was not a significant main effect of alerting time to buffer loss on SART scores; however, pilots reported a much lower level of situation awareness when the alerting time to buffer loss was 20-60 seconds than any of the other alerting time bins. Pairwise comparisons revealed a significant difference between 20-60 seconds and the 1-2 minutes and 4-10 minutes time bins (See Figure 4).

3. Discussion

As the alerting time to buffer loss decreased, pilots reported an increase in temporal demand, frustration, and effort, with higher levels reported for temporal demand, than effort and frustration, respectively. Pilots also reported a much lower level of situation awareness when the alerting time to buffer loss was 20-60 seconds than any of the other alerting time bins. The objective of an automated conflict-detection system, such as provided by AOP, is to give an adequate and comfortable amount of warning time to the pilots for conflict situations whenever possible. One of the key purposes of this study was to determine the impact of different alerting times. Specifically timed conflicts were "scripted" by having a nearby aircraft make an unannounced turn towards the ownship. The objective data presented in Wing, et al.,² indicated that all conflicts with alerting times greater than one minute were resolved with no loss of separation (LOS). Supporting this result, the subjective data indicated the pilots had a consistent level of pilot situation awareness for all alerting times greater than one minute. Thus, when conflicts occur with less than one minute's notice, LOS may result not just due to inadequate time to maneuver but also due to a decreased

understanding of the situation. The increase in workload was primarily in temporal demand, but increased effort and frustration were consistent with decreased situation awareness. These results point to the need for improved Human-Machine Interface (HMI) and procedures in time-critical conflict situations or to ensure alerting time always exceeds one minute.

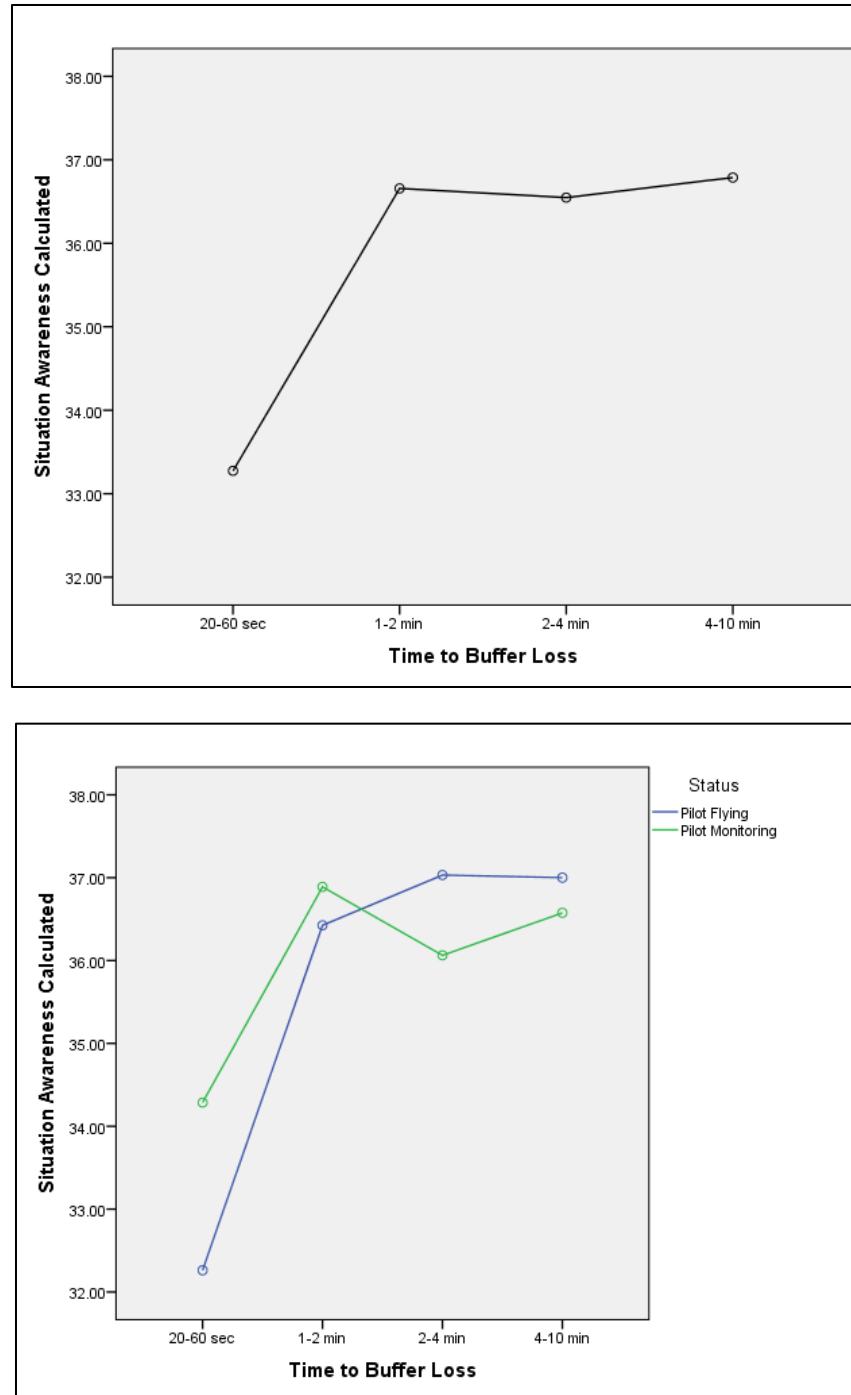


Figure 4. The effect of time to buffer loss on Situation Awareness Calculated

E. Loss of Separation

1. NASA-TLX

The separation standard for the experiment was 5 nmi and 800 ft. Workload data was analyzed to evaluate whether pilots reported higher levels of workload when a loss of separation occurred during the scenario. There was a significant main effect of loss of separation on Mental Demand, $F(1, 342) = 8.79, p = .003$; Temporal Demand, $F(1, 342) = 16.29, p < .001$; Performance, $F(1, 342) = 14.64, p < .001$; Frustration, $F(1, 342) = 22.85, p < .001$; and Effort, $F(1, 342) = 27.53, p = .006$. In general, the results indicate that pilots reported higher levels of workload and a lower level of Performance when a loss of separation occurred during the scenario than when they did not experience a loss of separation. With the exception of Performance and Frustration, there were significant differences between the PF and PM, where the PF reported higher levels of workload than the PM (See Figure 5.)

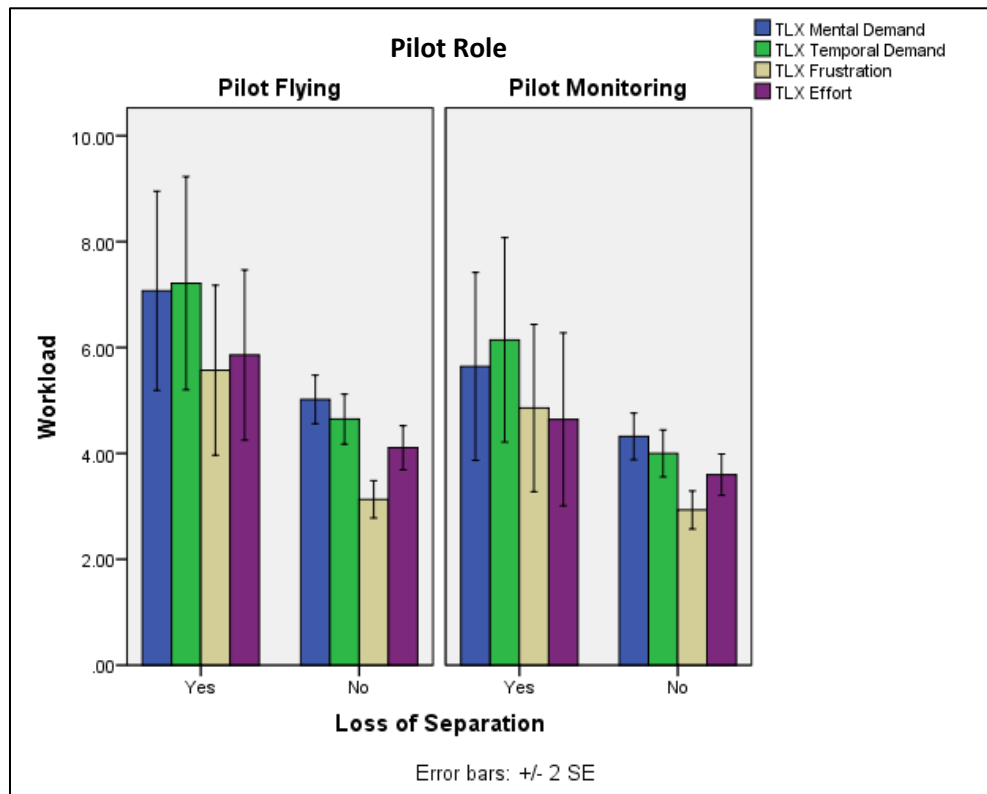


Figure 5. The effect of loss of separation on the NASA-TLX ratings of workload.

2. SART

The calculated situation awareness score from the SART was analyzed to evaluate whether pilots reported lower levels of situation awareness when a loss of separation occurred during the scenario. There was a significant main effect of Loss of Separation on the ratings of situation awareness, $F(1, 342) = 6.41, p = .01$, indicating that when a loss of separation occurred during a scenario, pilots reported lower levels of situation awareness than when a loss of separation did not occur. Additionally, when a loss of separation occurred there was a significant difference between the PF and the PM, where the PF perceived a much lower level of situation awareness than the PM. When a loss of separation did not occur there was no difference in the situation awareness calculated score between the PF and PM (See Figure 6).

3. Discussion

The occurrence of a LOS equates to a failure of the separation assurance system, which in the AFR concept consists primarily of the automation system (AOP) but includes the pilot procedures and other factors, such as the

behavior of the traffic aircraft and whether such behavior was announced or expected. Regardless of the cause, the task for the pilots at this point is to quickly reestablish separation by following AOP instructions. By necessity, it is higher workload because tactical maneuvering is required in short order. An additional element of workload present in this experiment's test environment that would not necessarily impact the flight environment was the mouse-driven controls of the ASTOR flight simulator. Pilot comments during debrief sessions indicated the extra work involved in using the mouse relative to actual hand flying. The reduction in situation awareness is a concern and points to the need for improved HMI and procedures for recovering separation. Algorithmic improvements in the separation recovery logic are also an area for improvement to provide greater stability of guidance during such highly dynamic events.

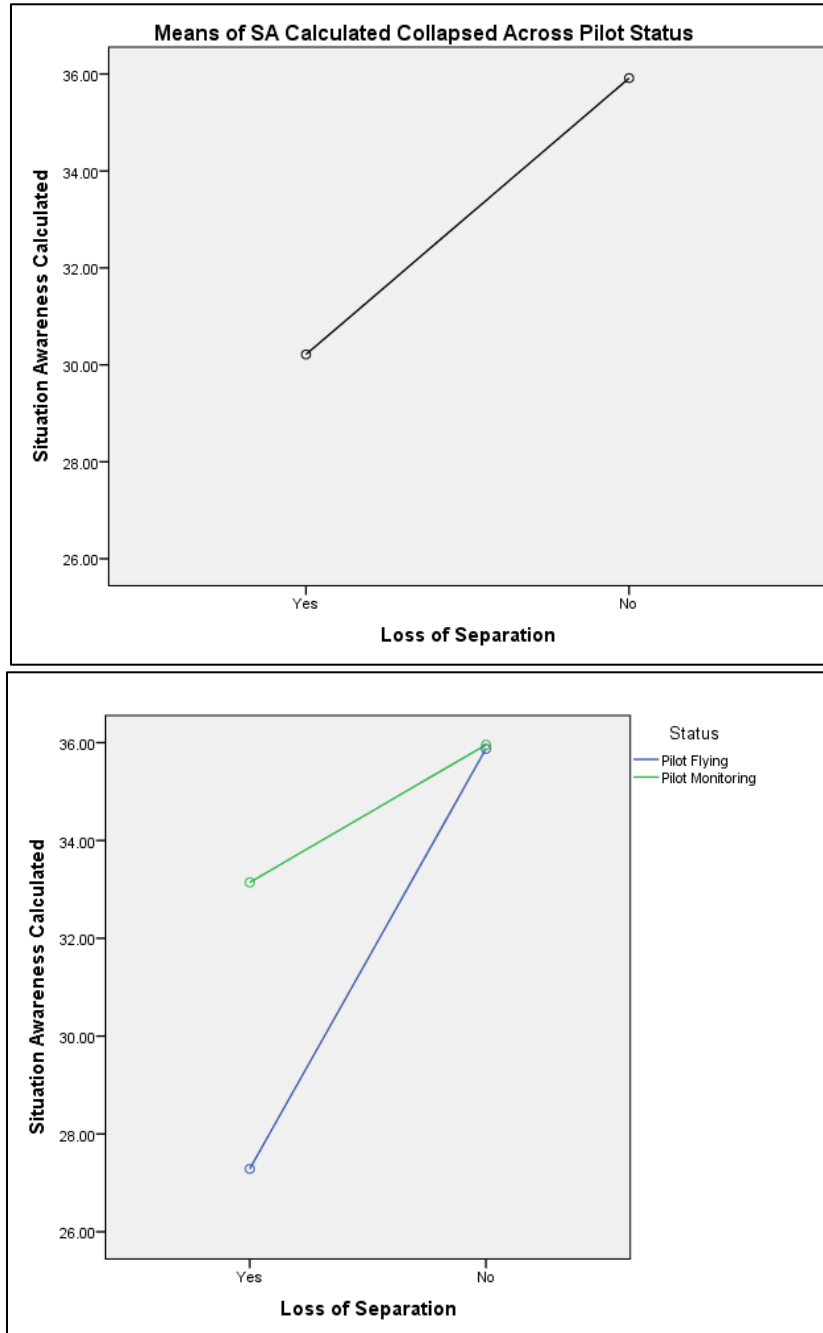


Figure 6. The effect of loss of separation on Situation Awareness Calculated

F. Traverse the Buffer Zone

1. NASA-TLX

Workload data was analyzed to evaluate whether pilots reported higher levels of workload when they traversed into the buffer zone during the scenario. The scenarios analyzed included only those in which the buffer zone was crossed, but not scenarios that also had a loss of separation. There were no significant main effects of traversing the buffer zone on any of the workload subscales.

2. SART

Situation awareness data was analyzed to evaluate whether pilots reported lower levels of situation awareness when they traversed into the buffer zone during the scenario. The scenarios analyzed included only those in which the buffer zone was crossed, but not scenarios that also had a loss of separation. There were no significant main effects of traversing the buffer zone on the calculated score of situation awareness. Additionally, there were no significant differences in perceived situation awareness between the PF and the PM. These results suggest that pilots accepted the addition of the buffer zone to provide additional time to respond to the conflict even when it was occasionally traversed.

3. Discussion

The three-mile lateral buffer was included for the purposes of providing a cushion in detecting and resolving conflicts, particularly in environments without “broadcast intent” of the traffic aircrafts’ planned maneuvers. The buffer was intended to be useful airspace, and subjective results confirm that the pilots were able to use this airspace effectively without effect on workload or situation awareness and that adding an additional buffer zone is an acceptable way to handle unexpected lateral maneuvers of IFR aircraft. Findings from the objective data indicated that a non-circular buffer might be a more effective use of airspace from a flight efficiency perspective. Based on these subjective results, no impact is expected from the pilot’s perspective if a change in buffer geometry is pursued.

G. Conflict Geometry Angle (Acute or obtuse)

The experimental design was composed of three elements of conflict geometry. One element of conflict geometry, *encounter angle*, included two convergence angle bins with the two aircraft either roughly aligned (“*acute*”) or roughly opposed (“*obtuse*”).

1. NASA-TLX

Workload data was analyzed to determine whether the conflict geometry angle affected pilots’ perceived ratings of the amount of workload required during that scenario. There was a significant main effect of conflict geometry angle on Mental Demand, $F(1, 185) = 9.41, p = .002$; Temporal Demand, $F(1, 185) = 8.52, p = .004$; Frustration, $F(1, 185) = 5.52, p = .02$; and Effort, $F(1, 185) = 9.34, p = .002$. Specifically, the pilots reported higher levels of workload when the encounter angle was acute than when it was obtuse (See Figure 7).

2. SART

Situation awareness data was analyzed to determine whether the conflict geometry angle affected pilots’ perceived ratings of their level of situation awareness during that scenario. There was a significant main effect of conflict geometry angle on the situation awareness Calculated score, $F(1, 342) = 7.70, p = .006$. Specifically, the pilots reported higher levels of situation awareness when the encounter angle was obtuse than when it was acute (See Figure 8). Additionally, there were no significant differences in perceived situation awareness between the PF and the PM.

3. Discussion

Pilots reported higher levels of workload and lower levels of situation awareness when the encounter angle was acute than when it was obtuse. Obtuse conflicts (where the intruding aircraft is more “head-on” than not) may appear to have a shorter duration due to the greater closing speed, and they may appear to have a more definitive “end” to the encounter as the aircraft passes “behind the wings.” By comparison, acute conflicts may appear to emerge more slowly and have a more persistent “threat” even after the conflict is resolved, thereby resulting in

higher perceived workload. The subjective results tend to support these two perspectives. While it may not be necessary to address this issue, one simple HMI change would be to eliminate display of the conflict geometry. However this may decrease overall situation awareness below desired levels.

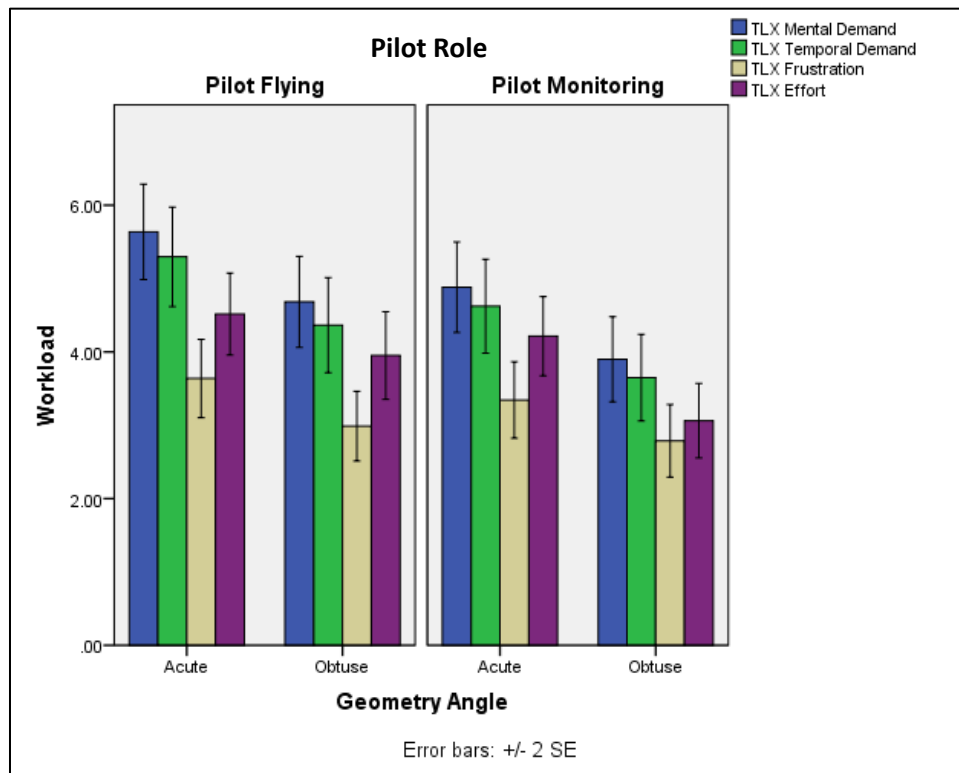


Figure 7. The effect of geometry angle on the NASA-TLX ratings of workload.

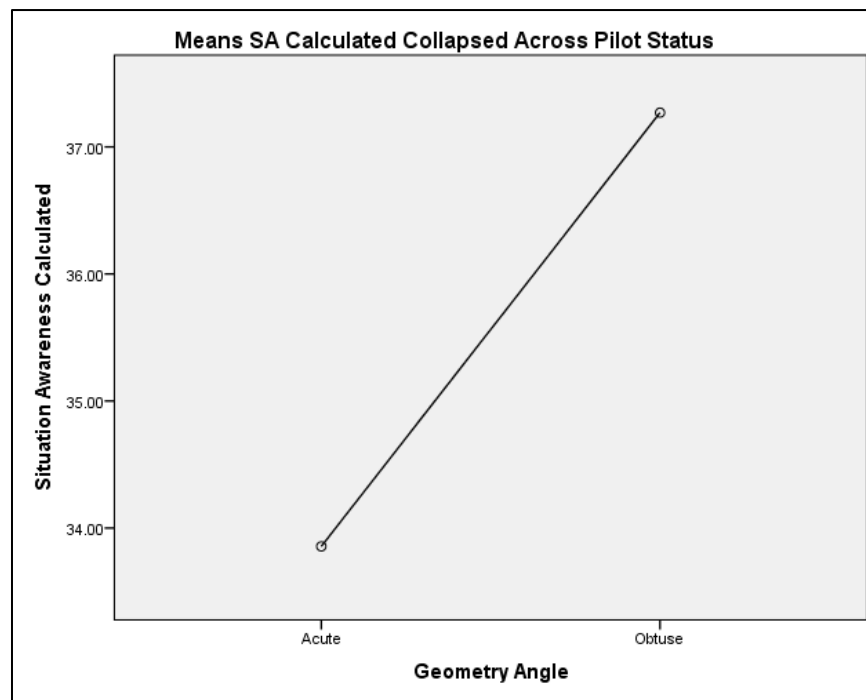


Figure 8. The effect of geometry angle on Situation Awareness Calculated

H. Maneuver Dimension (Lateral or vertical)

A second element of conflict geometry, maneuver dimension, refers to the IFR aircraft approach direction being either “*lateral*” (same altitude and level with the AFR aircraft) or “*vertical*” (climbing or descending into separation loss, with the AFR aircraft always level).

1. NASA-TLX

Workload data was analyzed to determine whether the maneuver dimension affected pilots’ perceived ratings of the amount of workload required during that scenario. There were no significant main effects of maneuver dimension on any of the workload subscales.

2. SART

Situation awareness data was analyzed to determine whether the maneuver dimension affected pilots’ perceived ratings of their level of situation awareness during that scenario. There were no significant main effects of maneuver dimension on the calculated score of situation awareness. Additionally, there were no significant differences in perceived situation awareness between the PF and the PM.

3. Discussion

The maneuver dimension of the conflicts was designed to be either lateral (both aircraft in level flight) or vertical (traffic aircraft climbing or descending). The finding that maneuver dimension did not impact the pilots’ experience is a positive result. The results of the objective data for this experiment¹⁵ indicated that 33 out of the 35 “pop-up” conflicts in the study’s exploratory scenarios (where a pop-up conflict is one that is alerted with less than 5 minute’s notice) were vertical conflicts. Previous research¹⁶ on airspace complexity for air traffic controllers has indicated vertical encounters as a significant contributor to controller workload. The finding here that pilot workload and situation awareness are not affected by maneuver dimension indicates metrics of airspace complexity may be very different between AFR operations and conventional ATC-based separation.

I. Passage Orientation (Front or behind)

A final element of conflict geometry, *passage orientation*, refers to whether the IFR aircraft in the encounter would “*pass in front*” of the AFR aircraft or “*pass behind*” if the conflict were left unresolved.

1. NASA-TLX

Workload data was analyzed to determine whether the passage orientation affected pilots’ perceived ratings of the amount of workload required during that scenario. There was a significant main effect of passage orientation on Temporal Demand, $F(1, 342) = 6.19, p = .01$ and Frustration, $F(1, 342) = 6.64, p = .01$. Specifically, pilots reported higher levels of workload when the IFR aircraft would pass in front than when it would pass behind (See Figure 9).

A potential explanation for this difference is that the “threat” of the approaching aircraft may have been perceived differently between the pass-in-front and pass-behind encounters. An intruding aircraft turning to pass behind the ownship, while still presenting a separation loss threat, may not be perceived as being a physical hazard, since they will be “behind the wings” by the time separation is lost. Conversely, an intruding aircraft turning in front of ownship may be seen as a greater collision threat and therefore command greater attention and workload.

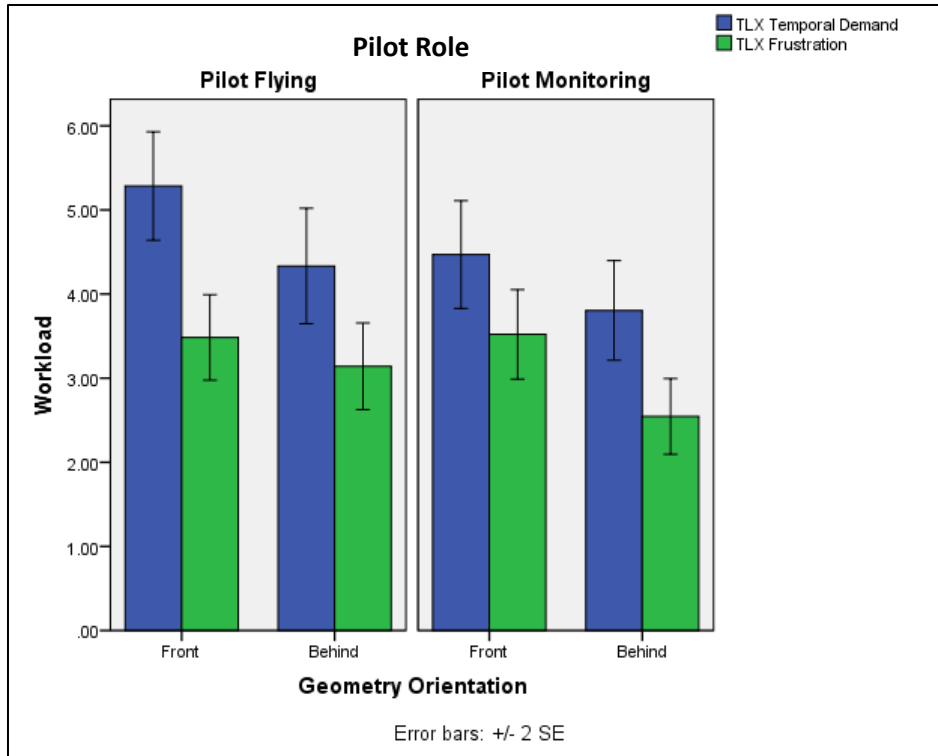


Figure 9. The effect of passage orientation on the NASA-TLX ratings of workload.

2. SART

Situation awareness data was analyzed to determine whether the conflict passage orientation affected pilots' perceived ratings of their level of situation awareness during that scenario. There was a significant main effect of passage orientation on the situation awareness Calculated score, $F(1, 342) = 5.06, p = .02$. Specifically, the pilots reported higher levels of situation awareness when the IFR aircraft would pass behind than when it would pass in front (See Figure 10). There were no significant differences in perceived situation awareness between the PF and the PM.

3. Discussion

Similar to encounter angle, pilots may have perceived the "threat" differently between pass-in-front and pass-behind encounters. An intruding aircraft turning to pass behind the ownship, while still presenting a separation loss threat, may not be perceived as being a physical hazard, since they will be "behind the wings" by the time separation is lost. Conversely, an intruding aircraft turning in front of ownship may be seen as a greater collision threat and therefore command greater attention and workload. Situation awareness may have also decreased given the reduced predictability of the future outcome, e.g., whether the intruder will turn again to increase the hazard. This effect could be mitigated with the inclusion of "broadcast intent" by traffic aircraft. The exploratory scenarios of this experiment directly address the value of broadcast intent¹⁵.

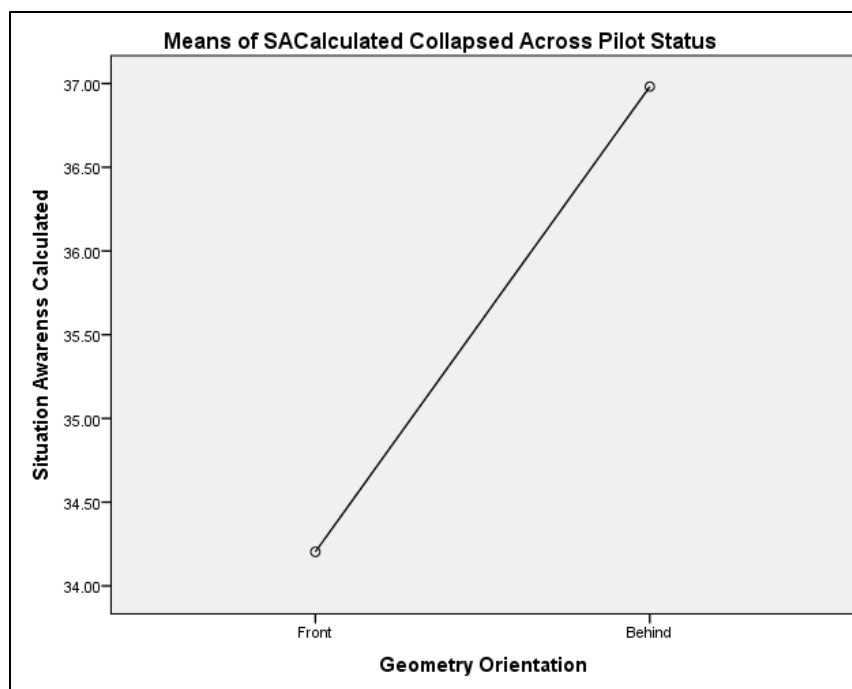


Figure 10. The effect of geometry orientation on Situation Awareness Calculated

VI. Conclusions

Generally across all conditions, pilots' perceived workload was low to medium, with the highest reported levels of workload occurring when the pilots experienced a loss of separation during the scenario. Pilot workload was perceived as being lower, and reported levels of situation awareness were higher, when pilots had at least one minute alerting time to resolve the conflict, during encounters with perceived short duration (acute conflict angle) and physical threat (when the IFR aircraft would pass behind), and when a loss of separation did not occur during the scenario.

In the current experiment, not having intent information became problematic when the alerting time to buffer loss was short and the pilots had very little time to react when detecting and resolving the conflict. Consistent with the results from the subjective assessments, comments from the pilots during the post-scenario questionnaires suggest that pilots would benefit from intent information. For example, "Having known the flight path of conflict would have made an easy strategic solution."

Both the observable data and the subjective data indicated that situation awareness was stable and acceptable when there was at least one minute alerting time to the conflict. Comments from the pilots were consistent with this finding. For example, "Could have been a little more time. An aircraft turned in front of us, at our altitude and was within 5 miles. We turned to pass behind but were still too close so AOP commanded a climb. Our climb wasn't aggressive enough so we had LOS. The AOP was effective but we were just too close."

Pilots did not experience any differences in workload or situation awareness between the lateral and vertical encounters, suggesting that maneuver dimension did not impact the pilots' experience. This result is interesting and positive given that previous research¹⁶ on airspace complexity for air traffic controllers has indicated vertical encounters as a significant contributor to controller workload. In the current experiment, pilot workload and situation awareness were not affected by maneuver dimension and may indicate that metrics of airspace complexity may be very different between AFR operations and conventional ATC-based separation.

It is interesting to note that pilots reported higher levels of workload and a lower level of situation awareness when there was a loss of separation, but *not* when they traversed into the buffer zone. This result indicates that providing the buffer zone was useful and allowed pilots the extra time needed to detect and resolve the conflict without increasing their perceived workload or reducing their situation awareness. Additionally, the subjective results confirm that adding an additional buffer zone is an acceptable way to handle unexpected lateral maneuvers of IFR aircraft.

Acknowledgments

The authors would like to thank Sally Johnson for her support and advice during the analysis and writing of this paper.

References

- ¹ Pearce, R.A. The NextGen JPDO Model of Interagency Planning. AIAA 2009-7010, 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO), Hilton Head, South Carolina, 2009.
- ² Wing, D., Prevot, T., Martin, L., Johnson, S., Cabrall, C., Commo, S., Sheth-Chandra, M., Mercer, J. and Morey, S. Pilot and Controller Evaluations of Separation Function Allocation in Air Traffic Management, Tenth USA/Europe Air Traffic Management R&D Seminar, Chicago, IL, June 2013.
- ³ Erzberger H. Transforming the NAS: The Next Generation Air Traffic Control System, 24th International Congress of the Aeronautical Sciences, Yokohama, Japan, 2005.
- ⁴ Prevot T., Homola J., Martin L., Mercer J. and Cabrall C. Toward Automated Air Traffic Control—Investigating a Fundamental Paradigm Shift in Human/Systems Interaction, *International Journal of Human-Computer Interaction*, 28:2, 77-98, 2012.
- ⁵ Karr, D.A., Vivona, R.A., Roscoe, D.A., DePascale, S.M., and Wing, D.J. Autonomous Operations Planner: A Flexible Platform for Research in Flight-Deck Support for Airborne Self-Separation. AIAA-2012-5417, Sept. 2012.
- ⁶ 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.
- ⁷ Kahneman, D. (1973). Attention and Effort. Englewood Cliffs, NJ: Prentice-Hall.
- ⁸ Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & D.R. Davies (Eds.). *Varieties of Attention*. (pp. 63-102). Orlando, FL: Academic.
- ⁹ Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). New York, NY: HarperCollins.
- ¹⁰ Dekker, S. & Lützhof, M. (2004). Correspondence, cognition, and sense making. A radical empirical view of situation awareness. In S. Bradbury and S. Tremblay (Eds) *A Cognitive Approach to Situational Awareness*. (pp. 22-41). Aldershot: Ashgate.
- ¹¹ Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32-64.
- ¹² Hart, S.G. and Staveland, L.E. (1988). Development of a NASA TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*, edited by P.S. Hancock and N. Meshkati, Amsterdam: North-Holland, 139-183.
- ¹³ Byers, J.C., Bittner, Jr., A.C., and Hill, S.G. (1989). Traditional and Raw Task Load Index (TLX) Correlations: Are Paired Comparisons Necessary? *Advances in Industrial and Ergonomic Safety I*, pp. 481-485.
- ¹⁴ Selcon, S. J. and Taylor, R. M. (1990, April). Evaluation of the situational awareness rating technique (SART) as a tool for aircrew systems design. In AGARD-CP-478, Situational Awareness in Aerospace Operations (pp. 5-1 to 5-8). Neuilly Sur Seine, France: Advisory Group Aerospace Research & Development. (AD-A223939).
- ¹⁵ Lewis, T. Analysis of traffic conflicts in a mixed-airspace evaluation of airborne separation assurance. To be presented at AIAA Aviation 2013, Los Angeles, California, August, 2013.
- ¹⁶ Kopardekar, P., Schwartz, A., Magyarits, S. and Rhodes, J. Airspace complexity measurement: An air traffic control simulation analysis. *International Journal of Industrial Engineering: Theory, Applications and Practice*, 16(1).