DIAGNOSTICITY OF AN ONLINE QUERY TECHNIQUE FOR MEASURING PILOT SITUATION AWARENESS IN NEXTGEN

Thomas Z. Strybel, Kim-Phuong L. Vu, L. Paige Bacon, Joshua Kraut, Center for Human Factors in Advanced Aeronautics Technologies, California State University, Long Beach, Long Beach CA

Vernol Battiste, San Jose State Foundation and NASA Ames Research Center, Moffett Field, CA

Walter Johnson, NASA Ames Research Center, Moffett Field, CA

Abstract

An online probe technique for measuring situation awareness and workload was evaluated for its ability to detect changes in awareness and workload caused by changes in roles and responsibilities for traffic separation. Three plausible NextGen concepts of operation were evaluated: Pilot primary (pilots responsible for traffic separation), ATC primary (ATC responsible for most traffic separation) and Automation Primary (automated conflict detection and resolution agent responsible for most traffic separation). Pilots were queried about task relevant information throughout a ninety-minute scenario. Queries were categorized into conflict, command and communications, and status information. Situation awareness was measured in terms of response latency and accuracy to the queries. Response latency to conflict queries changed with concept of operation, suggesting that online queries for specific task-relevant information can determine changes in situation awareness for task-specific information.

Introduction

The Next Generation Air Transportation System (NextGen) will be a comprehensive overhaul of the National Airspace System (NAS). In NextGen, airspace operators will assume new roles and responsibilities, utilize new air-traffic-management technologies and work with new concepts of operation in order to increase the capacity of NAS, improve its safety, and increase the reliability of air transportation. According to the FAA NextGen Implementation Plan, human factors' considerations will play a large role in system development. For example, NextGen concepts of operation are being developed to determine the effect of changes in roles and responsibilities between air and ground operators in order to decrease workload and increase operator-reliance on automation. Validating these concepts will require the development of quantitative metrics to define and validate human performance [1].

Adequate validation and testing of alternative concepts of operation (CONOPS) requires measurement methods and quantitative metrics of operator and system performance that are reliable, valid, diagnostic, sensitive and usable [2.3]. These qualities are, of course, essential for any rigorous performance measure. Reliability refers to the consistency of the measure. Validity is the extent to which a measure actually measures what it is supposed to measure. Sensitivity is the ability to detect changes in the amount of the measured construct, and diagnosticity is the extent to which a measure can identify the causes of changes in levels of performance. Usability refers to how easily the measurement method can be applied, which influences user acceptance of the method. This paper reports on the development of metrics for situation awareness and workload, two constructs that are important determinants of the effectiveness of any human-machine system, and known to be limiting factors in current day NAS. Specifically we investigated the validity and diagnosticity of situation awareness and workload probes for pilots operating under plausible NextGen CONOPS, as part of a NASA NRA, "Metrics for Situation Awareness, Workload and Performance in Automated Separation Assurance Systems, NNA06CN30A ").

Situation Awareness and Workload

Both researchers and NAS operators consider situation awareness and workload as critical factors influencing performance in current air traffic management (ATM) systems. Air traffic controller (ATC) workload, for example, is a major determinant of current airspace capacity. Workload is the amount of effort, either physical or mental, required to perform a task [4]. Several workload measures have been validated for capturing pilot and controller

978-1-4244-6618-4/10/$26.00 ©2010 IEEE

performance [5] but most have not been validated for NextGen operations. Measures of situation awareness are not as fully developed, despite years of research on the topic and the fact that it is a critical component of current airspace operations. Poor situation awareness has been implicated as a major factor in carrier aircraft accidents, ATC-reported incidents [6,7] and the severity of ATC operational errors [8,9].

One reason for the lack of good measures of situation awareness is that the concept itself is widely debated in human factors. Endsley's definition of situation awareness is the most widely cited [6]. Endsley defines situation awareness in terms of its information-processing components. In this view, situation awareness consists of three levels of knowledge: Perception, detecting relevant information in the task environment, Comprehension, understanding and integrating the information detected, and Projection, predicting future states based on information acquired in stages 1 and 2. Endsley's model is consistent with basic cognitive/information processing models of task performance, but this similarity makes separating situation awareness from existing cognitive constructs difficult [10]. Alternatively, situation awareness can be viewed as being distributed between operators and the task environment through interactions with external information sources. This "distributed" view of situation awareness stems from recent shifts in models of perception and cognition and challenge the view that human perception requires a complete internal representation of the external world. Currently, perceptual researchers state that perception is accomplished through partial, incomplete representations that are not updated unless they are attended to. Human factors' researchers [11,12] have applied this model of perception to situation awareness. According to Durso [11] situation awareness is not "all in the head," but distributed between the operator and his/her task environment. In other words, operators do not construct a complete situation awareness picture in the head but instead represent some information internally, and know the location of critical task information for quick access as the need arises. Note that offloading information to the environment alleviates some operator workload because developing and maintaining a complete representation of all relevant information is resource intensive [10].

Measuring Situation Awareness

The information- and distributed-processing approaches to understanding situation awareness use similar methods for measuring situation awareness. These techniques have been called "probe" techniques because the operator is queried about his/her awareness while the task is being performed. Endsley's Situation Awareness Global Assessment Technique (SAGAT; [13]) presents a series of questions to the operator during pauses in a simulation in which all displays are blanked. The number of correctly answered queries, in this view, corresponds to the amount of situation awareness in the operator's head, because the operator must rely solely on memory for accurate responses. Durso's online method, known as the Situation Present Awareness Method (SPAM; [9]), also queries operators for situation awareness information, but quieries are presented while the simulation is ongoing and all displays are active. Because information relevant to the online probe queries is always available, situation awareness is measured as both the number of correct responses and response latency, the time to respond to the query. According to Durso, queries that can be answered with information in the head should be responded to more quickly than queries in which information must be extracted from task-relevant information displays. Moreover, if the information is not in memory, the speed of response will be determined by the operator's knowledge of where the information is displayed. Therefore, the speed of the response is a measure of the operators' awareness of task-relevant situation awareness information.

Probe techniques have been evaluated for validity. Endsley [14,15] reports criterion validity for the measure, but most of these data are indicators of sensitivity, the ability of the measure to detect subtle changes in situation awareness. Durso [16] showed that SPAM reaction times predicted novice ATC performance after variance due to individual differences in cognitive skills was removed. SPAM reaction times have been shown to be related to measures of ATC and pilot performance. These data suggest that SPAM and SAGAT have some validity as measures of situation awareness. However, data on the diagnosticity of these probe techniques are less prevalent. Diagnosticity should be based on the information contained in the queries. Endsley recommends a Goal-Based Task Analysis for creating situation awareness queries in order to obtain measures of situation awareness that reflect the three information processing stages of perception, comprehension, and projection, according to her 3-level situation awareness model. Durso's method of developing queries is less formal, although he does recommend that queries be prepared in consultation with subject-matter and operations' experts. Another difference in these methods is the types of questions that are asked. SAGAT queries are based on memory for absolute information (e.g., "Is an aircraft at a prescribed altitude?"), yet controllers normally store relative traffic information (e.g., "Is aircraft A higher than aircraft B?"; [9]), which tend to be the focus of SPAM queries.

To improve the diagnosticity of situation awareness probe techniques, we have been assessing measures of situation awareness by systematically querying the operator for different categories of information. We have developed queries based on categories that are consistent with Endsley's levels of processing model, and categories based on relevant task components identified by subject matter experts for both ATCs and pilots. Strybel [17] investigated different processing categories of online probe questions for experienced and student ATCs and found that certain categories of questions predicted changes in performance within operators over multiple scenarios varying in traffic density. Questions based on task components such as conflict detection and resolution were significantly related to the number of LOS, average vertical distance between aircraft, and the number of traffic advisories issued. The proportion of speed changes made by ATCs were related to questions based on subjective assessments future events. No significant relationships were obtained between questions based on levels of processing (perception vs. comprehension) and performance, however.

In two airside studies of situation awareness, response latencies to online probes based on subjective assessments of threat of encroachment were good predictors of a subjective measure of situation awareness known as Situation Awareness Rating Technique (SART) and performance measures [18,19]. Strybel et al. [19] compared probe queries presented during scenario freezes to the same probe queries presented while the scenario was active in a human-in-the-loop simulation of pilots flying approaches into Dallas-Fort Worth (DFW) airport. Both online and offline queries were categorized in terms of level of processing (recall, comprehension, subjective assessment) and time frame (past, present or future events). Pilots were instructed to vary the speed of their aircraft in order to meet spacing restrictions at the final approach fix. Overall, performance on both SPAM and SAGAT were correlated with a measure of spacing performance, indicated airspeed (IAS) variability. However, only SPAM latencies were related to the number of missed traffic clearances. Moreover, SPAM latencies to subjective assessment queries were correlated with IAS variability. For SAGAT no category was predictive, suggesting that SAGAT may be limited to assessing an overall level of awareness without pinpointing whether operators are more aware of some information than others.

Dao et al. [20] assessed probe accuracy and latency of pilots performing a conflict resolution task. In this experiment the queries were categorized in terms of levels of processing (recall vs. comprehension) and time frame (past, present or future). These queries were administered after a single conflict resolution task (similar to SAGAT) but displays were not blanked. Both accuracy and response latency were measured (as in SPAM). On each trial, pilots were presented with a traffic conflict on a 3-D Cockpit Situational Display (CSD). Three conflict resolution conditions were tested. In the automated condition a suggested resolution was displayed, and pilots were instructed to accept it. In the interactive condition a suggested resolution was displayed, but pilots could either accept or modify it. In the manual condition, pilots resolved the conflict themselves using conflict probe and resolution tools. Dao et al. showed that probe latency was highest in the automated resolution condition compared with both interactive and manual resolution conditions. Response latencies for past and present questions were significantly faster than future questions in all conflict-resolution conditions but this difference was largest for queries of future events. Lastly, an interaction between level of processing and time frame was obtained. Probe latencies were lower for queries that required recall compared with comprehension for past and present queries, but this trend was reversed for queries based on future events. The results of Dao et al. suggest that situation awareness is lowest (probe latencies highest) when pilots were not directly involved in resolving conflicts, either manually or by modifying suggested resolutions, and that processing levels affect response latencies.

Although differences in latencies between queries based on comprehension and event-time frame were obtained, these information-processing categories may have limited diagnostic value for evaluations of NextGen CONOPS and automation technologies. To be diagnostic of changes in roles and responsibilities, the situation awareness probes should be sensitive to changes in awareness of the necessary information for performing the new task only. For example, if responsibility for traffic separation is shifted to the flight deck, a diagnostic measure should detect changes in pilot awareness for traffic conflicts and not other flight tasks. Of course, awareness for unrelated tasks may change if the additional role significantly increases operator workload. Therefore, a good diagnostic measure of situation awareness should also include an assessment of workload.

In the present investigation we examined online probes that were either categorized on the information required to complete specific flight tasks or assessed workload to determine if queries would be diagnostic of changes in pilot roles and responsibilities. Some of these probes measured pilot workload. Three plausible NextGen CONOPS were tested. These changed the responsibility for conflict detection and resolution between pilots, ATCs, and an automated, ground-based conflict detection and resolution agent. Pilots flew desktop simulators in ninety-minute scenarios while probe queries were administered at regular intervals. The scenarios in the simulation required pilots to deviate from their planned trajectory route to avoid weather, perform merging and spacing operations, and execute a continuous descent approach (CDA) under one of three CONOPS for separating traffic. We hypothesized that pilot performance on online queries of traffic conflicts would reflect changes in their responsibility for traffic separation.

Method

Participants

Eight experimental pilots were tested in the second week of a two-week simulation. All were air transport pilots. Five pilots were captains and three were first officers. All pilots had glass cockpit experience, but none had any prior experience with merging and spacing operations. Three pilots had experience flying CDAs.

Simulation Configuration

Participants “flew” the aircraft using desktop PCs with a standard keyboard and mouse. Two pieces of software composed the pilot’s simulated flight deck, the Multi-aircraft Control System (MACS) and the 3D Cockpit Situational Display of Traffic Information (CSD). The MACS system provided pilots with an interface that allowed flying their aircraft with tools normally found in current day Boeing 777 aircraft [21]. MACS includes a datalink tool that displayed spacing instructions clearances from ATCs. The CSD displayed traffic and weather within a range of 160 nm. Weather was displayed in either current day radar format (NextRad) in a 3-dimensional NextRad (3D CSD NextRad) view. The CSD also had a Route Assessment Tool (RAT) that allowed pilots to modify their current flight plan through a graphical user interface to avoid weather and traffic [22]. In some conditions, automated conflict alerting algorithms provided visual and auditory alerts when traffic conflicts were detected.

The CSD was also equipped with an automated spacing tool. When prompted by spacing clearances, pilots used the spacing function located in the tool bar at the bottom of the CSD to open spacing tool options. Pilots selected a spacing interval in the submenu, then visually located and selected the assigned lead aircraft from the traffic display before engaging the spacing tool. The spacing tool modified speed to achieve the desired spacing interval at the final approach fix, based on Eurocontrol CoSpace logic [23]. When engaged, the merging and spacing tool calculated if the aircraft would achieve its assigned spacing by the runway. When coupled with the auto throttles, the spacing tool gradually modified the aircraft speed to achieve the assigned spacing interval. Aircraft data tags which provided aircraft callsign, altitude, and speed information could be displayed on the CSD at any time at the pilot's discretion.

A distributed simulation network consisting of four research labs provided the real-time simulation environment. Participant pilots were located at NASA Ames' Flight Deck Display Research Laboratory (FDDRL). Participant ATCs, "ghost" ATCs, and pseudopilots were located at the Center for Human Factors in Advanced Aeronautics Technologies (CHAAT) at California State University Long Beach. Additional pseudopilot stations were located at the Systems Engineering Research Laboratory (SERL) at California State University Northridge, and the Human Integrated Systems Engineering Laboratory (HISEL) at Purdue University.

Scenarios

As shown in Figure 1, Pilots were instructed to fly an arrival route into Louisville Standiford International Airport (SDF). All pilots flew within the same scenario in real time and were assigned a spacing interval and lead aircraft two minutes after the start of the trial by an automated ground station. Pilots used the data link panel on their display to load the information into the CSD and then they executed the spacing command after manually selecting the lead aircraft on their display. Additionally, pilots were responsible for weather avoidance in all conditions and were trained to maneuver using the RAT. Pilots adjusted their route relative to the weather based on their own safety criteria and constraints imposed by surrounding traffic. Onboard conflict alerting was provided in some concepts. These alerted the pilot to conflicts up to 8 minutes in advance of loss of separation.

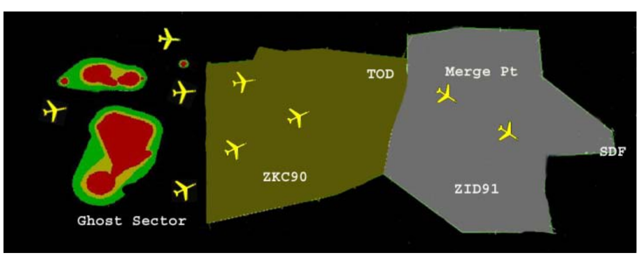

Figure 1. Simulated Airspace

ATCs managed a proportion of the traffic in each scenario. Their task was to manage the air traffic and, if requested, re-sequence arrival aircraft for merging and spacing after the aircraft deviated for weather. In addition, controllers responded to data linked requests for route modifications, as well as requests made over the radio. Aircraft rerouting was accomplished manually using the MACS Trial Planner. Conflicts were also resolved from the ground by requesting automatically generated resolutions from the Auto-Resolver tool. Controllers also responded to situation awareness questions in similar fashion to pilots, but controller data are not presented here (for additional information regarding ATC performance see [24,25].)

Probe Questions

During each scenario run, probe questions were administered at regular intervals throughout the scenario. The probe questions were constructed by experimenters in consultation with subject matter experts to fit in four task information categories and administered so that an equal number of questions were presented in each category. These categories are as follows.

Conflicts

These questions asked pilots about nearby traffic. Some, asked about conflicts directly (e.g., "Are you currently in conflict with AAL123?"; "What is the current clock position of the aircraft most likely to be in conflict with you in the next 10 minutes?"), others asked about information related to conflicts ("What is the heading of the aircraft closest to you?"; "Will any aircraft cross your path at the same altitude in the next 10 minutes?").

Command and Communications (C2)

These queries asked about information related to communications with ATC, and commands either just executed or about to be executed (e.g., "How many minutes before you change heading?"; "Will you change frequency within the next 5 minutes?"). Some questions in this category queried the pilots on status of merging and spacing operations (e.g., "Will the status of the spacing box relative to the spacing goal be late five minutes from now?"; "What is your commanded speed (kts)?"). During the later stages of the scenario, probe questions in this category requested information regarding the status of the CDA (e.g., "Will you be below your assigned altitude at CBSKT?").

Status. These questions queried pilots on the status of the aircraft and the airspace surrounding the aircraft (e.g., "What area relative to ownship has the most aircraft?"; "What is the difference in altitude between ownship and FLG309?"). In the early stage of the scenario this category included questions about weather in the vicinity of the aircraft (e.g., "What is the distance between the two largest weather cells?"; "How far will you deviate laterally (nm) for weather?").

Workload

These queries asked pilots to rate their current workload on a scale of 1 (very low) to 5 (very high).

Probe questions were presented on a small touch screen located adjacent to the pilots' flight displays. To ensure that response latency was related to situation awareness and not workload, each query began with a "Ready" prompt and audio alert. Pilots were instructed to respond "yes" to the ready prompt only when they had sufficient time to take a probe question. If the pilot responded affirmatively to the ready response, the probe question was immediately presented and the pilot responded by selecting the answer via touch input. Probe queries were presented every three minutes beginning four minutes into the scenario. If the ready response was not accepted after one minute, it "timed-out" and was removed. In this case, the probe question was skipped. The sequence of queries based on task information was counterbalanced for each pilot and scenario.

Design

Twelve scenarios were run, based on three plausible NextGen concepts, as shown in Table 1. These CONOPS allocated responsibility for conflict detection and resolution between pilots, ATCs and an automated ground-based conflict resolution agent. Note that these concepts were selected only to test shifts in operator roles and responsibilities. In Concept 1, pilots independently separated traffic using the RAT and a conflict probe, and executed modified flight plans without prior approval from ground. In Concepts 2 and 3, pilots were required to data link a request for route modification either to ATC in Concept 2 or an automated agent in Concept 3. As shown in Table 2, the major difference in pilot roles and responsibilities was in Concept 1, in which they were responsible for resolving traffic conflicts with Ownship. In concepts 2 and 3, responsibilities were shifted between ATCs and automation, in Concept 2, pilots could use the CSD to request routes changes with the conflict resolution tool, but in Concept 3, the tool was not available.

Two additional variables were manipulated, weather complexity (high vs. low) and type of weather display (NextRad vs. 3D NextRad). High-complexity weather contained more cells, and covered more airspace. Low-complexity weather consisted of fewer weather cells. A total of twelve scenarios were run, four at each CONOP. Within each CONOP, two scenarios each were run at each complexity condition, one each with the different weather displays. All scenarios for a particular CONOP were run before going on to the next condition. Because all pilots flew in the same scenario, the order of CONOP was fixed; Concept 3 on day one, Concept 2 on day two, and Concept 1 on day three.

Table 1. Plausible NextGen CONOPS Manipulated in the Simulation

Concept 1: Pilot Primary -ATC Secondary	Concept 2: ATC Primary - Autoresolver Agent Secondary	Concept 3: Autoresolver Agent Primary - ATC Secondary
• Pilots responsible for resolving 75% of traffic conflicts with ownship • ATCs responsible for resolving remaining conflicts • Autoresolver agent not responsible	• Pilots not responsible for resolving traffic conflicts • ATCs responsible for resolving 75% of conflicts • Autoresolver agent responsible for resolving remaining conflicts	• Pilots not responsible for resolving traffic conflicts • ATCs responsible for resolving remaining conflicts • Autoresolver agent responsible for resolving 75% of conflicts

Table 2. Pilot Roles, Responsibilities and Major Tasks in Each Simulated CONOPS

Concept 1: Pilot Primary -ATC Secondary	Concept 2: ATC Primary - Autoresolver Agent Secondary	Concept 3: Autoresolver Agent Primary - ATC Secondary
• Avoid weather • Maintain spacing with assigned lead AC, based on uplinked clearances • Fly CDA to SDF • *Resolve all traffic conflicts with own ship*	• Avoid weather • Maintain spacing with assigned lead AC, based on uplinked clearances • Fly CDA to SDF • *Suggest/request conflict resolutions, but no responsibility*	• Avoid weather • Maintain spacing with assigned lead AC, based on uplinked clearances • Fly CDA to SDF • *No conflict resolution responsibility*

Procedure

Within the first 10 minutes of the scenario, pilots were assigned a lead aircraft and given spacing instructions (in trail of lead by 105s at the final approach fix). Pilots were instructed to fly the Sea Biscuit One arrival into (SDF) while maintaining separation from other traffic (in Concept 1 only), avoiding convective weather, maintaining the assigned spacing interval relative to a lead aircraft at the final approach fix, and complying with Sea Biscuit One's altitude and speed restrictions. At top of descent (which was determined separately for each individual aircraft based on its speed and altitude), participant aircraft began the CDA to the 17-R runway. Participant pilots were to notify ATC whenever they discontinued spacing; at that time the aircraft was under the control of the air traffic controller.

Results and Discussion

Accuracy of probe responses presented was high, ranging from 71% to 89% (M=82%, SD=31%). Response latencies to correctly-answered probe questions were determined for each probe category. Separate three-way ANOVAs were run on latencies in each probe category with the factors CONOPS, Weather Complexity, and Weather Display. Workload probe performance has been reported elsewhere [26] and will not be discussed here.

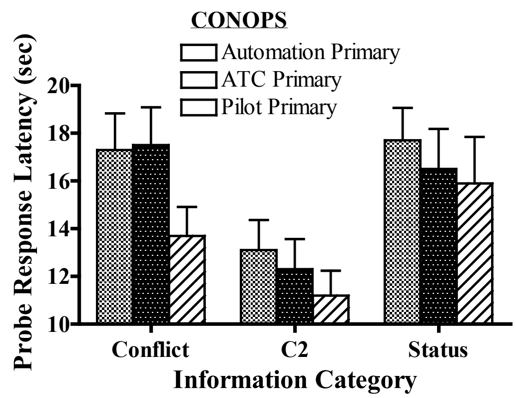

Figure 2. Probe Response Latency by Information Category for each CONOPS

However, it is important to note that pilot response latencies to the ready prompt, which is assumed to be an indicator of workload, was indeed significantly correlated with workload ratings. That is, higher latencies to the ready prompt were associated with higher workload ratings. Furthermore, pilot latencies to the probe questions being analyzed here were unrelated to the latencies of the ready prompt. Most important, pilot workload was not affected by to the probe questions being analyzed here were unrelated to the latencies of the ready prompt. Most important, pilot workload was not affected by changes in operation concepts, indicating that changes in awareness or performance could not be attributed to increased workload.

Separate three-way ANOVAs were run on latencies in each probe category with the factors CONOPS, Weather Complexity, and Weather Display. Workload probe performance has been reported elsewhere [26] and will not be discussed here. However, it is important to note that pilot response latencies to the ready prompt, which is assumed to be an indicator of workload, was indeed significantly correlated with workload ratings. That is, higher latencies to the ready prompt were associated with higher workload ratings. Furthermore, pilot latencies to the probe questions being analyzed here were unrelated to the latencies of the ready prompt. Most important, pilot workload was not affected by changes in operation concepts, indicating that changes in awareness or performance could not be attributed to increased workload.

A significant main effect of operating concept was obtained for the conflict questions only (conflict: $F(2,6)=4.90$; $p=.04$; Commands and communication (C2): $F(2,6)=1.80$; $p=.20$; status: $F(2,6)=2.20$; $p=.15$). As shown in Figure 2, response latencies for C2 queries overall were much lower than for conflict and status queries. Because commands and communications are issued recently by the operators, the shorter latencies may reflect the fact that this information is still active in working memory or still "held in the head." However, only conflict questions were affected by CONOPS, which differed with respect to responsibility for separation, and the effect was in the predicted direction: response latencies for conflict probes were on average four seconds faster in the pilot primary condition compared with the remaining CONOPS (ATC primary: M=17.3 s SEM=1.5 s; Automation primary: M=17.5 s SEM=1.6 s; pilot primary: M=13.7 s SEM=1.2 s). That is, pilots were more aware when they were actively engaged in the conflict resolution task.

Main effects of weather complexity were obtained for status and C2 queries, but not conflict queries (conflict: $F(1,6)=1.10$; $p=.33$; C2: $F(1,6)=21.77$; $p<.001$; status: $F(2,6)=69.78$; $p<.001$). As shown in Figure 3, response latencies for status queries indicated that pilots were more aware during scenarios with highly-complex weather (high: M=13.1s, SEM=2.8s; low: M=18.3s, SEM=3.1s). On the other hand response latencies for C2 queries were reversed: greater awareness of commands is shown for the less complex weather scenarios (high: M=14.7s, SEM=2.5s; low: M=9.3s, SEM=1.6s).

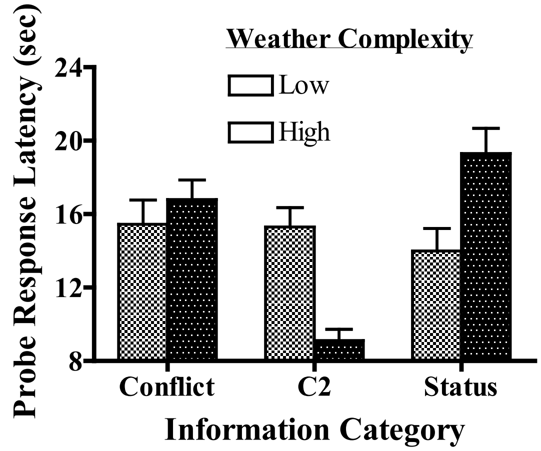

Figure 3. Probe Response Latency by Information Category as a Function of Weather Complexity

Therefore, pilots were more aware of status information (which included weather) when flying through more complex weather, and because their attention was focused on the weather, it could have resulted in less aware of command and communication information. When the weather pattern was not as complex, pilots were more aware of C2 information.

Table 3. Correlations Between Probe Performance and Conflict Resolution Performance

	Number Conflicts	Conflict Resolution Time	LOS Time at Resolution
Conflict Latency	-.16	.26 *	-.16
Conflict Accuracy	-.11	-.39 **	.39 **
C2 Latency	-.09	.24 *	-.03
C2 Accuracy	-.02	.11	-.19
Status Latency	-.22	-.08	-.02
Status Accuracy	.08	-.22	.07

*p<.09, ** p<.05

Probe vs. Conflict Resolution Performance

To determine if probe categories were related to conflict resolution performance, probe latencies in each category were correlated with measures of pilot conflict resolution performance: Number of conflicts, conflict resolution time, and time to LOS at resolution. On average, pilots resolved 1.7 (SEM=.16) conflicts per run, and took 154 s (SEM=14 s) to resolve a conflict, with the average time to LOS at resolution being 311 s (SEM=12.2 s). As shown in Table 3, queries on conflict information latency and accuracy were significantly related to conflict resolution time. Longer response latencies were associated with longer conflict resolution times. Higher accuracy on conflict probes was associated with lower conflict resolution times. These patterns are what one would expect if the probes are measuring situation awareness. Higher accuracy was also related to LOS time at resolution: higher accuracy on probe questions was associated with longer times to LOS at resolution. C2 probe latencies were marginally related to conflict resolution time; status probes were unrelated.

Probe vs. Merging and Spacing Performance

The relationship between performance on probe questions and merging and spacing performance was also analyzed. Merging and spacing accuracy was assessed by computing the absolute value of the average deviation from the target spacing value (105 s), and the root mean square (RMS) spacing error for each pilot over the course of a scenario. These values were analyzed with a three way, CONOPS – weather complexity-weather display repeated measures ANOVA. CONOPS significantly affected RMS spacing error ($F(2,,6)= 8.17$; $p<.001$). RMS errors were largest in the automation primary condition (M=23.4 s; SEM=1.2 s) compared with ATC primary and pilot primary conditions (M = 16.9 and 15.8 s; SEM = 1.3 and 1.0 s for ATC and pilot primary, respectively), suggesting that human intervention improves spacing performance. However, it is difficult to determine, the extent to which spacing errors are determined by pilot performance because the merging and spacing tool, once activated, managed the speed of the aircraft without pilot intervention. Therefore, to create a measure of performance based on pilot performance, we calculated for each pilot the proportion of time that the spacing tool was activate (i.e., managing aircraft speed). This value was significantly correlated with both aircraft measures (r's = -.12 and -.32) suggesting that longer spacing engagement was related to higher accuracy.

Table 4 shows the correlations between probe performance and measures of spacing accuracy. Probe accuracy for conflict questions was significantly and negatively correlated with RMS spacing error. Accuracy for C2 and status questions were significantly and negatively correlated with percentage of time engaged, suggesting that greater accuracy on these questions was related to lower percentage of spacing engagement. C2 latencies were positively correlated with percentage of time that spacing was engaged, meaning that longer probe latencies were associated with longer times spacing was engaged. Lastly, accuracy of responses to status questions were associated with greater percentages of times in which spacing was engaged.

Table 4. Correlations between Probe and Spacing Performance

	Average Spacing Error	RMS Spacing Error	Percent Time Spacing Engaged
Conflict Latency	-.07	-.03	.05
Conflict Accuracy	-.14	-22**	.05
C2 Latency	.11	.11	.27 **
C2 Accuracy	-.25 **	-.01	.19 *
Status Latency	-.01	-.08	-.05
Status Accuracy	-.21*	-.05	.19 *

*p<.09, ** p<.05

Summary and Conclusion

The results from the simulation show that an online probe method for assessing pilot situation awareness was effective in detecting changes in pilot roles and responsibilities for traffic separation. In effect, the manipulation of roles and responsibilities for traffic separation was detected by the probe latencies for questions related to traffic conflicts. Note that these changes in awareness cannot be attributed to workload because workload probes showed no effect of CONOPS [26]. The significant difference in latencies to conflict-related queries is consistent with Dao et al [20] who showed that awareness was highest when pilots were engaged in the conflict resolution task. Also noteworthy is the finding that probe questions based on commands/communication and sector status was not affected by changes in roles and responsibilities for traffic separation. Response latencies for these categories were affected, though, by the complexity of the weather cells in the scenario. Based on response latencies to status questions (which included specific queries on the weather), pilots were more aware of surrounding traffic and weather when presented with highly complex weather, as indicated by lower response latencies. There may have been a cost, however, to attending to sector information because latencies for queries on C-2 information such as merging and spacing or communications with ATCs were higher in these scenarios. One possible explanation for this result is cognitive tunneling , in which awareness is focused on a salient event at the cost of knowledge for peripheral events.

Latencies for probe questions addressing traffic separation were significantly correlated with conflict resolution time, further suggesting that they are valid measures of situation awareness for information related to specific task responsibilities. Response latencies in this category were positively correlated with conflict resolution time, meaning that longer response latencies were associated with longer conflict resolution times. Pilots who were less aware of potential conflicts would take longer to resolve those that became actual conflicts.

The relationship between probe queries addressing C-2 information and performance is less clear. Latencies to probe questions addressing merging and spacing were not related to spacing error. However, the spacing tool used by pilots in this study automatically managed the aircraft's speed in order to meet the spacing target at final approach fix. Therefore it is difficult to determine the performance of the pilot on the task. Strybel et al. [19] did show a relationship between probe performance and spacing accuracy, but pilots managed speed manually without any spacing tool. Therefore, although there is some suggestion that the percentage of time the tool is engaged is related to pilot awareness, the results are not straightforward. This brings up a potential problem in validating NextGen CONOPS and evaluating new technologies: it is necessary for new measures of operator performance be developed that can assess the contribution of the operator to the automation.

Nevertheless, not only have we provided additional validation of conflict probes but also we determined that probe questions are sensitive to changes in responsibility for traffic separation. These results suggest that online probing may be an effective method of assessing changes in situation awareness due to changes in roles and responsibilities being considered in NextGen.

References

[1] Federal Aviation Administration, NextGen Integration and Implementation Office, 2010, NextGen Implementation Plan, Federal Aviation Administration

[2] Boff, K. R., & Lincoln, J. L., 1988, Engineering Data Compendium: Human Perception and Performance. Wright-Patterson AFB, OH: Harry G. Armstrong Medical Research Laboratory.

[3] Salmon P., Stanton, N., Walker, G., & Green, D., 2006, Situation awareness measurement: A review of applicability for C4i environments, Applied Ergonomics 37, 225–238.

[4] Proctor, R. W., & Van Zandt, T., 2008, Human Factors in Simple and Complex Systems, Boston, MA: Allyn and Bacon.

[5] 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. Meshkahi (Eds.), Human Mental Workload, Amsterdam, Netherlands: North-Holland, pp. 139-183.

[6] Endsley, M.R., 1995, Measurement of situation awareness in dynamic systems, Human Factors, 37(1), 65-84.

[7] Endsley, M. R., & Jones, D. G., 1995, Situation awareness requirements analysis for TRACON air traffic control (TTU-IE-95-01), Lubbock, TX: Texas Tech University.

[8] Rodgers, M.D., Mogford, R.H., & Mogford, L.S., 1997, The relationship of sector characteristics to operational errors, Air Traffic Control Quarterly, Vol. 5, No. 4, 241-263.

[9] Durso, F.T., Truitt, T.R., Hackworth, C.A., Crutchfield, J.M. & Manning, C.A. 1997, En route

operational errors and situation awareness, The International Journal of Aviation Psychology, 8 (2), 177-194.

[10] Chiappe, D., Strybel, T.Z., & Vu, K-P.L., 2010, Review of situation awareness from the perspective of cognitive science. Manuscript in preparation.

[11] Durso, F., Rawson, K., & Girotto, S., 2007, Comprehension and situation awareness. In F. Durso, R. Nickerson, S. Dumais, S. Lewandowsky & T. Perfect (eds) Hanbook of Applied Cognition (2nd Edition), Hoboken, NJ, Wiley.

[12] Salmon, P., Stanton, N. Walker, G., Baber, C., Jenkins,D., McMaster, R., & Young, M., 2008, What is really going on? Review of situation awareness models for individuals and teams, Theoretical Issues in Ergonomic Science, 9(4), 297-323.

[13] Endsley, M. R., Bolte, B. & Jones, D. G., 2003, Designing for situation awareness: An approach to human-centered design. London: Taylor & Francis.

[14] Endsley, M. R., 2000, Direct measures of situation awareness: validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds). *Situation Awareness, Analysis and Measurement.* New Jersey: Lawrence Erlbaum.

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

[16] Durso, F. T., Bleckley, M. K., Dattel, A. R., 2006, Does situation awareness add to the validity of cognitive tests? Human Factors, 721-733.

[17] Strybel, T. Z., Minakata, K., Nguyen, J., Pierce, R. & Vu, K-P., L., 2009, Optimizing online situation awareness probes in air traffic management tasks, In M. J. Smith and G. Salvendy (Eds.): Human Interface, Part II, HCII 2009, Lecture Notes in Computer Science, 5618, 845-854.

[18] Dwyer, J. P., Strybel, T. Z., Vu, K-P. L., 2006, Simulation of Multiple Uninhabited Aerial Vehicles Operating in an Airport Terminal Area. Human Factors of Uninhabited Military Vehicles as Force Multipliers Meeting Proceedings RTO-MP-HFM-135. (pp 24-1 – 24-16), Neuilly-sur-Seine, France: RTO. CD-ROM.

[19] Strybel, T. Z., Vu, K.-P.L, Kraft, J., 2008, Assessing the Situation Awareness of Pilots Engaged in Self Spacing, Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society, NY:HFES

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

[21] Prevot, T., 2002,. Exploring the many perspectives of distributed air traffic management: The multi aircraft control system MACS. International Conference on Human-Computer Interaction in Aeronautics, HCI-Aero 2002, MIT: Cambridge, MA.

[22] Granada, S., Dao, A. Q., Wong, D., Johnson, W. W., & Battiste, V., 2005, Development and integration of a human-centered volumetric cockpit display for distributed air-ground operations, Proceedings of the 12th International Symposium on Aviation Psychology, Oklahoma City, OK.

[23] Eurocontrol Experimental Center, 2006, Flight deck requirements for airborne spacing (sequencing and merging) retrieved from http://eurocontrol.int/eec/gallery/content/public/documents/EEC_SSP_documents/CoSpace/2006/Sequencing_and_merging_FD_URD_2-3.pdf

[24] Vu, K-P., L., Strybel, T. Z., Battiste, V., & Johnson, W. , 2010, Factors influencing the decisions and actions of pilots and controllers in three plausible NextGen environments, In press.

[25] Vu, K-P, L., Strybel, T. Z., Kraut, J., Bacon. L.P., Minakata, K., Nguyen, J., Rotterman, A., Battiste, V., & Johnson, W., 2010, Pilot and controller workload and situation awareness with three traffic Management concepts. 29th Digital Avionics Systems Conference, this volume.

[26] Ligda, S., Dao, A-Q., V., Vu, K-P., L., Strybel, T.Z., Battiste, V., & Johnson, W., 2010, Impact of conflict avoidance responsibility allocation on pilot workload in a distributed air traffic management system. Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society, CA:HFES.

Acknowledgements

The simulation described in the paper was supported in part by NASA cooperative agreement NNA06CN30A, *Metrics for Situation Awareness, Workload, and Performance in Separation Assurance Systems*. Preparation of this paper was supported by NASA cooperative agreement NNX09AU66A, *Group 5 University Research Center: Center for the Human Factors in Advanced Aeronautics Technologies*.

Disclaimer

Any opinions or conclusions are those of the authors alone, and do not necessarily reflect the official position or policy of NASA, SJSU Foundation, CSULB, or the CSULB Foundation.

Email Address

Dr. Thomas Strybel: tstrybel@csulb.edu

29th Digital Avionics Systems Conference
October 3-7, 2010