Are the Intrusive Effects of SPAM Probes Present When Operators Differ by Skill Level and Training?

Hector I. Silva[1], Jason Ziccardi[1], Tristan Grigoleit[1], Vernol Battiste[2], Thomas Z. Strybel[1], and Kim-Phuong L. Vu[1]

[1] Center for Human Factors in Advanced Aeronautics Technologies, Department of Psychology, California State University Long Beach, 1250 N Bellflower Blvd, Long Beach CA 90840

[2] San Jose State University Foundation and NASA Ames Research Center

hector_silva91@yahoo.com, jbziccardi@gmail.com, t_grub@msn.com, Vernol.Battiste@nasa.gov, {Thomas.Strybel,Kim.Vu}@csulb.edu

Abstract. The Next Generation Air Transportation System (NextGen) plans to implement a series of automated tools into the National Airspace System to aid air traffic controllers (ATCos) in managing a two to three times increase in air traffic density. However, introducing automated technologies into a system like air traffic management (ATM) changes the responsibilities of the ATCo from an active controller to a passive monitor, which can result in lower levels of situation awareness (SA). To measure SA objectively in such a dynamic task as ATM, the Situation Present Awareness Method (SPAM) is often used. SPAM provides the operator with SA probes while the operator is performing the task. Some studies have shown that the use of SPAM to measure SA is intrusive because it provides the operator with a secondary task. The present study examines whether these intrusive effects of SPAM are present when the operator has achieved a high skill level at the time of test, and whether training operators to rely more or less on NextGen automated tools influence their performance when SPAM queries are presented as a secondary task.

Keywords: situation awareness, online probe methodology, situation present assessment method.

1 Introduction

The Next Generation Air Transportation System (NextGen) plans to implement a series of automated tools into the National Airspace System in order to aid air traffic controllers (ATCos) in managing a two to three times increase in air traffic [1]. NextGen seeks to accommodate this increase in traffic efficiently and safely, and at the same time reduce its impact on the environment. To do this, NextGen will implement a series of automated tools into air traffic management (ATM) systems that will aid ATCos in managing the increased number of aircraft (AC). Some of the tools that NextGen seeks to implement will aid controllers with conflict detection and resolution [1].

S. Yamamoto (Ed.): HIMI/HCII 2013, Part I, LNCS 8016, pp. 269–275, 2013.

However, the lack of attention on the part of the operator when a system is highly automated is likely to cause detrimental effects. For example, introducing automated technologies into a system like ATM changes the responsibilities of the ATC from an active controller to a passive monitor [2], which can result in lower levels of situation awareness (SA), and out-of-the-loop syndrome. With highly automated tasks, the operator is less vigilant, and, as a result, will have a harder time regaining awareness should something go wrong [3-4].

The most common real-time probe method used to measure SA in a dynamic task such as ATM is the Situation Present Awareness Method (SPAM). With SPAM, operators are presented with a "ready" prompt to indicate that a probe question is ready to be administered. If the operator's workload is low enough for him/her to be able to take a question, the operator will indicate that s/he is ready (typically by pressing a designated button). The time between the presentation of the ready prompt and the operator's response is a measure of workload. The time needed to answer the SA probe question, and the accuracy of the answer, is taken to be a measure of SA. Participants who show shorter SA probe latencies and higher SA probe accuracy are considered to have higher SA. Using SPAM to measure SA has many benefits over other methods by providing a more sensitive measure of SA (i.e., providing latency data as well as accuracy data), being a method that can be applied in the field while the operator is performing his or her task, and by being consistent with theoretical notions of SA, such as the distributed approach where the user off-loads information into the environment or displays [5-9].

Besides the benefits of using SPAM, it is important to note the drawbacks of using this method. Despite the use of the ready prompt, SPAM does not eliminate all effects of workload, which makes the method somewhat intrusive. Answering probe questions is a secondary task which could cause negative effects on the primary task [9]. Pierce [10] had participants perform an air traffic management task in conjunction with a secondary task that consisted of answering SPAM queries, or one that involves a high or low cognitive load. Pierce reported that use of SPAM queries resulted in greater disruption in performance compared to other secondary tasks. In addition to this, asking the operator questions may change the information that the user is paying attention to and therefore can affect the user's attention post-query [9, 11]. On the other hand, Bacon and Strybel [12], in a simulation designed to manipulate the operators' attention to off nominal events, found no evidence that SPAM SA probes affected the operator attention.

The present study examines whether these intrusive effects of SPAM are present when the test participants are given extensive practice with the task, and whether the effects of SPAM differs depending on the operator's skill level at the time of test. In addition, we wanted to determine if there were differences in the effects of administering SA probe questions using SPAM when operators were given different types of training. The data reported in the present study comes from a larger study examining the effects of training participants to rely more or less on NextGen automated tools. It could be that being trained to rely more on automated NextGen tools would lower workload enough and thus allow the operator the cognitive resources to be able to complete the secondary task of answering probe questions without disturbance to the

primary task. Conversely, it could also mean that training to rely on the harder-to-learn manual skills provides a strong foundation, which could lead to the development of multi-tasking skills that can lead to successful completion of both primary and secondary tasks. In addition, highly skilled ATCo students may have gained multi-tasking skills and be efficient enough to allow them to complete a secondary task without much interference to the primary task.

2 Method

2.1 Participants

Fourteen students enrolled in Mount San Antonio College, a Federal Aviation Administration Collegiate Teaching Initiative school, were recruited to participate in this study. The students were seeking an Aviation Science degree and took part in a 16 week internship at the Center for Human Factors in Advanced Aeronautics Technologies at California State University, Long Beach.

2.2 Measures

To measure mental workload, the NASA Task Load Index (TLX) was administered after each scenario. The TLX is a questionnaire that measures workload on six dimensions: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. The scores on these dimensions can be added and multiplied by 1.11 to yield a combined score on a 100 point scale. The combined score of these dimensions were used for data analysis.

Measures of air traffic controller task performance were also recorded. Performance as a measure of safety was recorded as the number of losses of separation (LOS) that occurred. A LOS occurs when two aircraft come within 1,000 feet vertically or 5 nautical miles laterally of each other. The lower the number of LOS, the safer the controllers were in managing the sector. Along with safety, two measures of efficiency were recorded. The average time and distance that aircraft took to fly through a sector show how efficient the student controllers were in maneuvering the aircraft through the sector.

2.3 Simulation Environment

The simulation environment was presented using the Multiple Aircraft Control System (MACS), a medium fidelity simulation software [13]. The program simulated Indianapolis Center, Sector ZID-91, where traffic consists of en-route flights and arrivals/departures to Louisville airport. Highly trained “pseudopilots” maneuvered all the aircraft in the sector and communicated with the ATCo students. Verbal radio communications were achieved with push-to-talk headsets and a voice IP server between the ATCo students and pseudopilots.

The NextGen equipped aircraft had Integrated Controller-Pilot Data Communication (Data Comm) to allow ATCo to issue clearances and request information digitally, Conflict Alerting, which alerts the ATCo of conflicts between NextGen equipped aircraft 6 minutes before a loss of separation (LOS) occurs, and a Trial Planner with Conflict Probe, which allows ATCos to modify aircraft routes by clicking and dragging them on the screen, while simultaneously probing for potential conflicts with other aircraft.

2.4 Training Procedure

The students that participated in the internship were separated into two groups. Both groups were trained to manage traffic in a mixed equipage airspace where some aircraft were equipped with NextGen technologies and others were not. One group trained with 25% equipped and 75% non-equipped aircraft, and thus trained to rely mostly on their manual skills to manage aircraft for the first 8 weeks. Because this group only had a few equipped aircraft, the students were much less likely to benefit from NextGen technologies.

The second group was trained in 75% equipped aircraft and 25% non-equipped aircraft for 8 weeks, which had the students rely heavily on NextGen technologies. Seventy five percent of the aircraft in the sector are equipped because prior research has shown that at least 50% of the aircraft must be equipped in order for the benefits of NextGen to be realized [14]. By having mostly equipped aircraft in their sector, the students were able to benefit from NextGen technologies as well as have much more opportunity to use them.

Both groups were trained to identify and resolve conflicts between aircraft using manual and NextGen tools before a LOS occurs. To resolve and/or avoid these conflicts, the students were trained in four separation techniques. One skill is vectoring aircraft, where the ATCo changes the flight plan or heading of one or more aircraft. The ATCo must assess the current heading of the aircraft and be able to calculate the safest heading change. The ATCo may change the heading angle in which the aircraft is traveling (determined by the degrees on a compass), or by having the aircraft move a number of degrees to the left or right of its present course. Another skill is altitude separation, where the altitude of the aircraft is manipulated. If two or more aircraft are at the same flight altitude and their paths will converge, the ATCo can issue clearances to one or both of the aircrafts to climb or descend from their current altitude to maintain separation between the aircraft. The students can also use speed separation, where the speed of the co-altitude aircraft is increased or reduced to maintain lateral separation. Lastly, the students were trained to use structured traffic flows, where the controller can create a "highway" for aircraft at certain altitudes, headings, and speeds. This helps the ATCo by grouping aircrafts together in a structure that will maintain separation. The students had to complete a full scenario using only one of these separation methods to successfully move to the next technique. Upon successful completion of all four techniques, the student was awarded "Journeyman" Status by the instructor. Skill level was determined by whether or not the students had earned Journeyman Status by the time testing occurred after the 8th week. Seven out of the 14 participants earned Journeyman Status by the time of the midterm testing.

2.5 Testing Procedure

As noted earlier, the data for this report is a small part of the larger study. The present study included data obtained after the first 8 weeks of the internship when the students completed a midterm examination. The scenarios used in the midterm test varied in the percentage of equipped aircraft, with participants completing scenarios with 0%, 100%, and two 50% equipped aircraft scenarios. The order of the scenarios was counterbalanced between all participants using a partial Latin square. The Situation Present Awareness Method was employed during only one of the two 50% scenarios. To establish a baseline measure, no SPAM queries were presented during the other 50% scenario. This allowed us the opportunity to directly compare the effects of SPAM query presentation.

When administered, SPAM queries operators with questions about their task as the task is ongoing and all displays remain visible and active. Therefore, SA is measured in response time to answering correct questions (with lower response times indicating greater SA) and with accuracy to answering the probe questions (with higher accuracy indicating greater SA). Before a query is presented, the operator must first accept it. Although more realistic in terms of real-world performance than many other freeze-probe methods, answering probes is still a secondary task that may disrupt the primary task of air traffic management. In the present experiment, a query was presented on a touch screen computer approximately every 3 minutes for the complete duration of the scenario. The ATCo students responded to these questions by touching one of the multiple choice responses on the touch screen monitor.

3 Results

Because we are mainly interested in whether the effects of using the SPAM probes are intrusive, we report data from the two 50% scenarios. The performance measures of interest were analyzed using separate 2 (training emphasis: manual/NextGen) × 2 (Journeyman Status: yes/no) × 2 (probes presented/not presented) ANOVAs. NextGen training emphasis and earning Journeyman Status were between subjects factors, and use of SPAM probes (presented or not presented) a within-subjects factor.

3.1 Workload

For workload as measured by the NASA TLX, a main effect of Journeyman Status was present, $F(1, 9) = 13.783$, $p = .005$, $\eta^2 = .605$. Air traffic control ATCo students who achieved Journeyman Status experienced significantly lower workload ($M = 45.18$, $SE = 3.14$) than ATCo students who had not achieved Journeyman Status ($M = 62.79$, $SE = 3.56$). We did not find any differences in workload as a function of whether the SPAM queries were presented, and the use of SPAM probes did not interact with any other variables.

3.2 Performance

Only the main effect of Journeyman Status was found to be significant for time and distance through the sector, Fs(1, 9) = 11.00 and 12.66, ps < .010, η^2s = .550 and .585, respectively. Students who achieved Journeyman Status managed their aircraft in less time (M = 652.78 s, SE = 3.52 versus M = 670.45 s, SE = 4.00), and had their aircraft travel a shorter distance (M = 78.65 nm, SE = .42 versus M = 80.89 nm, SE = .47) through their sector, than students that were not awarded Journeyman Status.

The analysis on the number of losses of separation revealed a marginal main effect of Journeyman Status, $F(1, 9) = 3.57$, $p = .079$, $\eta^2 = .303$, where students awarded Journeyman Status tended to commit fewer LOS (M = .60, SE = .26) than those who were not (M = 1.375, SE = .29). In addition, a main effect of training emphasis was also found, $F(1, 9) = 5.280$, $p = .047$, $\eta^2 = .370$. Students trained to rely on manual skills committed fewer LOS (M = .542, SE = .26) than students trained to rely on NextGen tools (M = 1.438, SE = .29).

We did not find any differences in performance as a function of whether the SPAM queries were presented, and the use of SPAM probes did not interact with any other variables.

4 Discussion

The results showed that the main factor determining whether the student ATCos experienced less workload and better performance was their individual skill level—that is, whether they were awarded Journeyman status by the time they were tested. There was no evidence in the present analysis that SPAM probes were intrusive to performance or affected operators differently as a function of their individual skill level or training method. However, the lack of significant effects involving the use of SPAM queries may be a result of the low observed power observed in our analyses. Upon closer look of the analyses executed, we see modest values of observed power for the significant effects reported above (0.91 for Journeyman effect of TLX scores; 0.84 and 0.89 for Journeyman effect on time and distance through sector, respectively; 0.54 for training effect on LOS, and 0.42 for Journeyman effect on LOS). However, we see relatively small observed power values for all other effects that were expected (all < 0.35). Thus, although our sample of 14 participants allowed us to detect some effects, it may not have been sufficient to capture the intrusive effects of SPAM probes, if any were evident.

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

References

1. Joint Planning and Development Office. Concept of operations for the next generation air transportation system version (2010), http://www.jpdo.gov/library/NextGenConOps.pdf (retrieved January 10, 2012)
2. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 30(3), 286–297 (2000)
3. Wickens, C.D., Dixon, S.R.: The benefits of imperfect diagnostic automation: a synthesis of the literature. Theoretical Issues in Ergonomics Science 8(3), 201–212 (2007)
4. Kaber, D.B., Onal, E., Endsley, M.R.: Design automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. Human Factors and Ergonomics in Manufacturing 10(4), 409–430 (2000)
5. Chiappe, D., Strybel, T.Z., Vu, K.-P.L.: Mechanisms for the acquisition of Situation Awareness in situated agents. Theoretical Issues in Ergonomic Sciences 13, 625–647 (2012)
6. Durso, F.T., Dattel, A.: SPAM: The real-time assessment of SA. In: Banbury, S., Trembley, S. (eds.) A Cognitive Approach to Situation Awareness: Theory, Measures and Application, pp. 137–154. Aldershot, New York (2004)
7. Durso, F.T., Hackworth, C.A., Truitt, T.R., Crutchfield, J., Nikolic, D., Manning, C.A.: Situation awareness as a predictor of performance for en route air traffic controllers. Air Traffic Control Quarterly 6, 1–20 (1998)
8. Jeannot, E., Kelly, C., Thompson, D.: The development of situation awareness measures in ATM systems. EATMP Report, HRS/HSP-005-REP-01 (2003)
9. Salmon, P.M., Stanton, N.A., Walker, G.H., Green, D.: Situation awareness measurement: A review of applicability for C4i environments. Applied Ergonomics 37, 225–238 (2006)
10. Pierce, R.S.: The effect of SPAM administration during a dynamic situation. Human Factors 54, 838–848 (2012)
11. Langan-Fox, J., Sankey, M., Canty, J.: Human factors measurement for future air traffic control systems. Human Factors 51, 595–637 (2009)
12. Bacon, L.P., Strybel, T.Z.: Assessment of the validity and intrusiveness of online-probe questions for situation awareness in a simulated air-traffic-management task with student air-traffic controllers. Safety Science (in press)
13. Prevot, T.: Exploring the many perspectives of distributed air traffic management: The Multi-Aircraft Control System MACS. In: Chatty, S., Hansman, J., Boy, G. (eds.) HCI-Aero 2002, pp. 149–154. AIAA Press, Menlo Park (2002)
14. Hah, S., Willems, B., Schulz, K.: The evaluation of data communication for the future air traffic control system (NextGen). In: Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society, Santa Monica, CA, pp. 99–103 (2010)