2005

Measurement of Situation Awareness Effects of Adaptive Automation of Air Traffic Control Information Processing Functions

Carlene M. Perry

Noa Segall

David B. Kaber

MEASUREMENT OF SITUATION AWARENESS EFFECTS OF ADAPTIVE AUTOMATION OF AIR TRAFFIC CONTROL INFORMATION PROCESSING FUNCTIONS

Carlene M. Perry
Noa Segall
David B. Kaber

Department of Industrial Engineering
North Carolina State University
Raleigh, North Carolina

The goal of this research was to define a measure of situation awareness (SA) in an air traffic control (ATC) task and to investigate the effect of adaptive automation (AA) of various information processing functions on SA. An ATC simulation was used that was capable of presenting four different modes of control, including information acquisition, information analysis, decision making and action implementation automation, and a manual mode. Eight subjects completed two trials under each mode of control. Operator workload, assessed using a secondary task, was used to trigger automation of the primary ATC task. The SA measure was an adaptation of the Situation Awareness Global Assessment Technique (SAGAT), involving cueing of aircraft positions as well as objective weighting of the relevance of aircraft to controllers for queries. The SA response measure revealed a significant effect of AA on subject perception and overall SA, with superior SA under the information acquisition mode of automation. ATC performance was significantly superior ($p<0.05$) when automation was applied to lower-order sensory processing functions, including information acquisition and action implementation, as compared to higher-order functions, specifically information analysis. During manual control periods as part of AA trials, ATC performance was significantly superior when following automation of information acquisition and information analysis functions. Secondary task performance was significantly worse under information analysis and decision making automation.

Introduction

Air traffic control (ATC) requires high levels of cognitive processing, and one approach for alleviating stress and workload of controllers is to use automation to perform some controller activities (National Research Council (NRC), 1998; Parasuraman & Riley, 1997). Automation has many potential advantages for controllers, including reduced task load (Laois & Giannacourou, 1995) and increased system reliability (NRC, 1998). However, automation in ATC may also present disadvantages (Dillingham, 1998), including a loss of controller situation awareness (SA) (Endsley & Jones, 1995). As machines perform more and more ATC functions, controllers have less interaction with the traffic management system, impairing their ability to detect when a problem has occurred, determine the current state of the system, understand what has happened and what courses of action are needed, and react to the situation (Endsley, 1996). Thus, maintaining SA in ATC is critical for accurate decision making and performance (Endsley, 1996), and this issue needs to be addressed through automation design.

Currently, advanced forms of adaptive automation (AA) are being considered for ATC to mitigate out-of-the-loop (OOTL) performance problems associated with conventional automation, and to preserve operator SA. AA refers to complex systems in which the level of automation or the number of system functions being automated can be modified in real time (Scerbo, 1996). Some research has explored the use of dynamic function allocations (DFAs) in the context of ATC simulations. Results provide evidence that AA may improve ATC performance over completely manual control and static automation (e.g., Hilburn, Jorna, Byrne & Parasuraman, 1997). They also indicate that the effectiveness of AA in the context of ATC may be dependent upon the type of automation presented to an operator. For example, Clamann, Wright, and Kaber (2002) found that, in the context of a low-fidelity ATC simulation, humans are better able to adapt to AA (from a performance perspective) when applied to lower-order sensory and psychomotor functions, such as information acquisition and action implementation, as compared to AA applied to cognitive (planning and decision making) tasks.

Measures of SA and AA

Many measures of SA have been developed over the past 10-15 years, including direct, objective measures such as the Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1995). SAGAT involves comparing an operator's perceptions of a task environment to some "ground truth" reality. This is accomplished by freezing a simulation exercise at random points in time and hiding task information sources (e.g., blanking visual displays) while subjects quickly answer questions about their current knowledge of the simulation. Subject responses are then graded based on actual data on the real situation, thus providing an objective measure of SA.

AA research has demonstrated SAGAT to be sensitive to dynamic changes in system states (Kaber & Endsley, 2004), as well as changes in adaptive interface content over time (Kaber & Wright, 2003). Kaber and Endsley (2004) also observed that operators achieved better SA with DFAs of levels of automation that applied computer assistance to decision-making aspects of the dynamic control task, as compared to levels applying automation to monitoring and implementation roles. This research also suggests that the impact of AA on SA may be dependent upon the human-machine system information processing (IP) functions to which AA is applied, but that SA may be affected in a different way than performance.

Recent research examining the use of SAGAT in an ATC simulation (McClernon, 2003) found its sensitivity for identifying differences among manual and automated conditions to be limited. Nunes (2003), who applied SAGAT to evaluate aided and unaided display conditions in an ATC task, also found that the technique did not reveal differences between conditions. Another study by Hauss and Eyferth (2003) suggests that SAGAT may not be a sensitive measure of SA in the ATC environment due to different aircraft having different relevance to controllers at different times. They argued that aircraft which had recently been contacted by a controller, or required control actions, demanded more attentional resources than other displayed aircraft. Consequently, controllers may focus on certain aircraft to the exclusion of others at various times and may recall their flight parameters in responding to SAGAT queries more accurately.

Hauss and Eyferth's (2003) concerns with the SAGAT for assessing controller SA led them to develop a new measure of SA which assigned weights to aircraft based on their relevance to the current control scenario. In addition, rather than having subjects recall aircraft positions on a blank radarscope, as an initial query, they employed cued recall in which participants were given the positions for the aircraft they were to be queried on. Hauss and Eyferth (2003) compared the new SA measure with SAGAT using a high-fidelity air traffic management simulation. Their results confirmed controllers used event-based mental representations, since significantly more relevant parameters than irrelevant parameters were reproduced using the new measure.

In the current study, we developed a modified approach to implementation of the SAGAT measure in order to assess the impact of various forms of AA of ATC IP functions on SA. Cued recall of aircraft positions in a simulated ATC task was implemented and aircraft relevance was objectively weighted as a basis for SAGAT queries. Different from Hauss and Eyferth's (2003) measure, this approach involved real-time identification of aircraft in conflict, as well as those that had recently been issued clearances (e.g., hold, reduce speed, etc.), as predictors of aircraft relevance to controllers. It was expected that these modifications would lead to a more sensitive measure of the impact of AA on controller SA, as compared to the SAGAT measures implemented by McClernon (2003) and Nunes (2003).

Method

We evaluated the SA, performance, and workload effects of AA of four different stages of IP in ATC. The forms of automation included information acquisition, information analysis, decision making and action implementation (see Parasuraman, Sheridan, & Wickens, 2000).

Tasks

The Multitask© Simulation. Multitask© is a lab simulation of ATC developed for studies of workload-matched AA of ATC IP functions (see, for example, Clamann et al., 2002). The task interface (see Figure 1) includes a radarscope, control and status boxes and a menu bar. Near the center of the radarscope are two airports. Each airport has two runways. Eight equally spaced holding fixes are also represented on the display by small circles (approximately 30 nm from the airports).

Simulated aircraft are represented on the display by triangle icons and data tags presenting their call signs. The aircraft icons represent one of three possible aircraft types: commercial, private, or military. The type of aircraft also dictates the possible range of speeds for the vehicle. During simulation run time, aircraft first appear toward the perimeter of the display on one of eight approach trajectories and move toward one of the two airports, destined for one of the two runways at an airport.

The control box includes eight buttons. Five control buttons facilitate clearance change commands, including reduce speed, hold, resume, change airport, and change runway. Two action commands are used to submit and cancel these clearances. Finally, the query command is used to initiate communication with an aircraft and obtain its flight parameters.

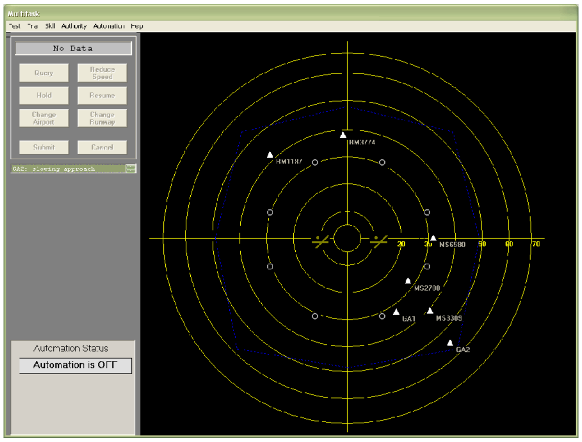

Figure 1. *Multitask© display in manual control.*

The simulation is capable of operating under one of the following five modes of automation:

(1) Manual control – No assistance is provided.

(2) Information acquisition – A scan line rotates around the radar display, and as it passes over an aircraft icon, a Trajectory Projection Aid (TPA) for that aircraft is presented for 2 sec. The TPA shows the aircraft destination and route, as well as its speed and destination airport and runway identifiers. This form of automation assists operators with acquisition of data on aircraft.

(3) Information analysis – Information on each aircraft on the radarscope is displayed in a table, including the aircraft's call sign, destination airport, destination runway, speed, and distance from the airport. A final column denotes the call sign of aircraft that are in conflict with each other. This form of automation assists operators with the integration of aircraft information.

(4) Decision making – In addition to conflict alerting, recommendations for conflict resolution are provided. Information on conflicting aircraft, the recommended clearance change, and which aircraft to advise of the change, are all displayed. This form of automation assists operators with IP requirements associated with decision and response selection aspects of the task.

(5) Action implementation – This form of automation simulates the "hand-off" of aircraft control from approach control to local-tower control, and the tower automatically maintains full control responsibility for aircraft within 20 nm of the center of the radarscope. This type of automation prevents any conflicts after "hand-off" to tower control. Action implementation automation assists the operator with the requirement of response execution as part of the ATC simulation.

Under all modes of automation, the objectives of the controller are to contact aircraft appearing on the radar display and make any necessary changes to pre-existing aircraft clearances (based on their potential to cause a conflict) while maintaining landing efficiency. Multitask© performance is measured in terms of the number of aircraft cleared, the number of trajectory conflicts, and actual collisions. This data is recorded during simulation trials. Aircraft arriving safely at an airport are considered cleared. Aircraft traveling within 3 nm of other aircraft, or two aircraft that are within 20 nm of the center of the radarscope and destined for the same runway at the same airport, are considered to be in conflict. Aircraft that simultaneously arrive at the same airport destined for the same runway, or aircraft that come in contact with each other, constitute actual collisions. During experimental trials, the various modes of automated assistance can be switched "on" or "off", based on operator workload states; however, only one mode can be used per trial.

Secondary Gauge-Monitoring Task. The experiment used a dual-task scenario involving simultaneous subject performance of the Multitask© simulation and a gauge monitoring task to objectively assess operator workload. The gauge task included a fixed scale, moving pointer display with a central "acceptable" region bordered on either side by two "unacceptable" regions. The user's goal was to detect and correct pointer deviations into either unacceptable region by using a keyboard. Gauge task performance was recorded as a hit-to-signal ratio.

Experimental Design and Procedures

Approach to AA. The gauge-monitoring task provided an index of operator workload in the Multitask© simulation. A low score in the gauge task implied a high level of workload in the ATC simulation and vice versa. The gauge task served as a basis for triggering DFAs in Multitask©. When secondary-task performance was poor, suggesting an increase in operator workload, the ATC simulation shifted from manual control to automated control. If operator secondary-task performance was good, the simulation returned to manual control.

Experiment Design. The experiment followed a within-subjects design with blocking on the subject. Eight subjects completed two trials under each of the five modes of Multitask© control. Each trial lasted approximately 50 minutes, including 30 minutes of simulation time and approximately 20 minutes to answer SA questions.

Situation Awareness Measure. The modified SAGAT measure (described above) was developed based on a goal-directed task analysis (GDTA) of ATC operations and application of the GDTA methodology to the Multitask© simulation. Following Endsley's (1995) methodology, three simulation freezes were conducted at random points in time during experimental trials to deliver SA queries. Subjects were posed with 9 questions during each freeze, including three targeting each level of SA (1 – perception; 2 – comprehension; 3 – projection), as defined by Endsley (1995). When a freeze occurred, the simulation displays were temporarily blanked and subjects were asked to move to a secondary computer workstation and respond to queries. At the same time, an experimenter collected information from the Multitask© software by accessing an automated aid which provided information on aircraft in conflict with each other and recommended clearances. Based on this information, the experimenter identified the three aircraft with the highest priority, or greatest "relevance", at that point in time in the simulation. Aircraft priority was determined based on a hierarchy of simulation events, e.g. aircraft in conflict were considered to have the highest priority, followed by aircraft issued a "hold" clearance, etc. The experimenter then sketched the locations of the "high priority" aircraft on a blank graphic of the Multitask© radarscope. The subjects were given the graphic and asked to respond to each of the 9 SA queries for each "high priority" aircraft. Composite scores for Level 1, 2, and 3 SA were computed based on the accuracy of subject responses to the sets of questions across freezes.

Hypotheses

(H1) We expected the modified SAGAT measure to be sensitive to changes in controller SA as a result of the AA manipulations. Counter to Kaber and Endsley's (2004) findings, because of the complexity of the version of Multitask© used in this study, subjects were expected to do better at responding to SA queries under lower levels of automation (information acquisition) and manual control as compared to high-level automation (information analysis and decision making), as a result of the potential for OOTL performance problems. We also speculated that under high levels of automation, such as decision making or information analysis, operators would exploit the additional capabilities of the automation, including conflict warnings and recommendations, pay less attention to the actual radarscope, and spend less time on low-level control functions which may be important to achieving SA.

(H2) On the basis of Clamann et al. (2002) findings, we expected Multitask© performance to be superior during trials in which AA was applied to lower-order sensory/ response functions, such as information acquisition and action implementation.

(H3) Based on Hilburn et al. (1997) results, AA of Multitask© was expected to affect performance on the secondary gauge-monitoring task, or operator workload. It was expected that higher levels of automation, including information analysis and decision making, presenting complex displays for operator interpretation, might demand high levels of visual attention and increase workload.

Results and Discussion

Situation Awareness

An ANOVA on the SA response measures revealed a significant effect of the specific forms of AA on Level 1 SA queries ($F(4,227)=3.78$, $p=0.005$) and the total SA score ($F(4,227)=2.7$, $p=0.032$). These findings support our expectation (H1) that the modified version of the SAGAT-based measure was sensitive to AA manipulations. Figure 2 shows the average Level 1 SA scores under each mode of automation. The pattern of results on Total SA was similar.

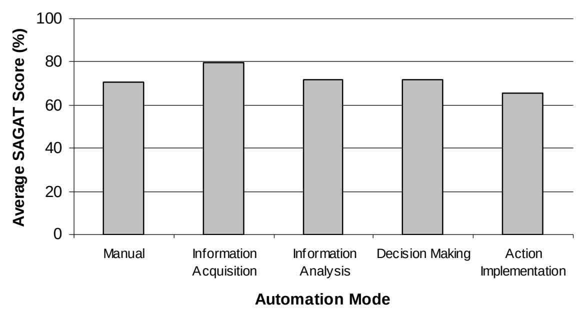

Figure 2. *Mean Level 1 SAGAT scores for the different modes of automation.*

Duncan's test showed Level 1 SA to be significantly superior under information acquisition automation, compared to information analysis, decision making, action implementation, and manual trials ($p<0.05$). However, manual control was not found to increase Level 1 SA. With respect to total SA, Duncan's test also revealed information acquisition to be superior to action implementation automation ($p<0.05$), which is consistent with our hypothesis (H1). However, SA during information analysis and decision making trials was not inferior to SA during other automation trials. In general, these results suggest that perception of system states may be most critically affected by demanding automation displays.

Figure 3 summarizes the mean Level 1, Level 2, Level 3 and total SA scores for automated and manual control periods, as part of AA trials (only). A marginally significant effect of the mode of automation was found for Level 2 SA queries (F(1,227)=3.51, p=0.062), indicating that subject comprehension was, on average, higher during manual control periods compared to automated control periods. This finding supports the notion that introducing some forms of automation in ATC may remove the controller from the loop (Endsley, 1996) and lead to decrements in higher levels of SA (H1).

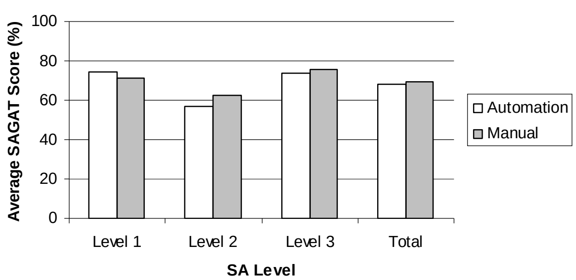

Figure 3. *Mean SAGAT scores during manual and automation control periods.*

Primary Task Performance

Results of ANOVAs on data collected during the automated control periods as part of AA revealed a significant effect of mode of automation on the number of cleared aircraft (F(3,41)=3.62, p=0.021) and the number of aircraft conflicts (F(3,41)=3.97 p=0.014), but not on the number of collisions (Figure 4). Duncan's MR test indicated that the number of cleared aircraft was higher for the information acquisition, decision making, and action implementation modes of automation, as compared to information analysis (p<0.05), in support of our hypothesis (H2). The high number of cleared aircraft during decision making may be attributable to the longer automated control periods under this mode of automation, as compared with the other modes. Duncan's test also revealed decision making to be significantly worse than information analysis for preventing aircraft conflicts (p<0.05). This finding was not surprising given that the decision aid made recommendations to subjects for dealing with conflicts. It is possible that subjects developed a strategy of waiting for the automation to warn them of a conflict and then to think about how to appropriately clear aircraft.

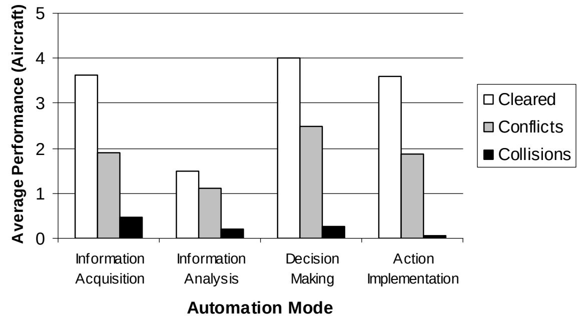

Figure 4. *Primary task performance during automated control periods.*

ANOVA results on manual control periods as part of the AA conditions revealed a significant effect of mode of automation on only the number of cleared aircraft (F(4,68)=7.58, p<0.0001). Safe landings were significantly higher for the information acquisition and information analysis modes of automation than for decision making and action implementation (Duncan's test, p<0.05). The results on the decision making condition are in agreement with our hypothesis (H2). It is possible that participants needed more time to shift from using a complex mental model for interaction with the decision aid back to their manual control mental model after the decision aid disappeared from the display and they had to identify conflicts themselves.

Workload (Secondary Task Performance)

An ANOVA on the workload data revealed a significant mode of automation effect when analyzing the automated control periods as part of AA trials (F(3,41)=4.01, p=0.014). Duncan's MR tests showed that action implementation, a lower-order sensory/response function, yielded higher average secondary-task performance than information analysis and decision-making automation (p<0.05). These findings are in line with our hypothesis (H3).

An ANOVA on workload data comparing the manual control condition with the manual control periods as part of AA also revealed a significant effect of the control mode (F(4,68)=2.66, p=0.04). The pattern of results under the manual control periods was almost exactly opposite to that observed during automated control periods. Duncan's test indicated that average workload was significantly lower (p<0.05) under decision-making automation, as compared to workload during manual control periods in AA of the information acquisition and action implementation functions, as well as the completely manual control condition. It is possible that when decision-making AA was applied and the recommendations for conflict avoidance were followed, the result was a lower workload when the simulation returned to manual control.

Conclusions

We designed a modified SAGAT approach to measuring SA in the context of an ATC task, which proved to be effective in terms of assessing the impact of specific forms of AA on controller perception, comprehension and projection. Using queued recall of aircraft, and establishing relevance weights for various aircraft at the time of SAGAT freezes, caused the SA response measures to be sensitive to the AA of information acquisition, information analysis, decision making, and action implementation functions. In general, our findings support a dependence of SA, performance, and workload effects of AA in ATC on the specific controller IP functions to which automation is applied. With a more complete understanding of the effects of AA on SA, additional research is needed to develop methods for real-time assessment of SA in ATC. Such a method could be used as a basis for triggering DFAs in complex, adaptive systems control on the basis of SA.

Acknowledgments

This research was supported by a grant from NASA Langley Research Center, Human Measures & Performance Element of the Airspace Systems Program, Hampton, VA (Grant No. NAG-1-03022). Lance Prinzel was the technical monitor. The views and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of NASA.

References

Clamann, M.P., Wright, M.C., & Kaber, D.B. (2002). Comparison of performance effects of adaptive automation applied to various stages of human-machine system information processing. In *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 342-346). Santa Monica, CA: Human Factors and Ergonomics Society.

Dillingham, G.L. (1998). *Air traffic control: Evolution and status of FAA's automation program.* (Tech. Report GAO/T-RCED/AIMD-98-85). Washington, DC: United States General Accounting Office.

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

Endsley, M.R. (1996). Automation and situation awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 163-181). Mahwah, NJ: Lawrence Erlbaum.

Endsley, M.R., & Jones, D.G. (1995). *Situation awareness requirements analysis for TRACON air traffic control* (Tech. Report TTU-IE-95-01). Atlantic City, NJ: Technology Center, Federal Aviation Administration.

Hauss, Y. & Eyferth, K. (2003). Securing future ATM-concepts' safety by measuring situation awareness in ATC. *Aerospace Science and Technology, 7,* 417-427.

Hilburn, B.G., Jorna, P., Byrne, E.A., & Parasuraman, R. (1997). The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload. In M. Mouloua & J. M. Koonce (Eds.), *Human-automation interaction: Research and practice* (pp. 84-91). Mahwah, NJ: Lawrence Erlbaum Assoc., Inc.

Kaber, D.B. & Endsley, M.R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5*(2), 113-153.

Kaber, D.B. & Wright, M.C. (2003). Automation-state changes and sensory cueing in telerobot control. In *Proceedings of the XVth Triennial Congress of the International Ergonomics Association* (CD-ROM). Seoul, Korea, August 24-29, 2003.

Laois, L. & Giannacourou, M. (1995). Perceived effects of advanced ATC functions on human activities: Results of a survey on controllers and experts. In *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 392-397). Columbus, OH: The Ohio State University.

McClernon, C.K. (2003). *Situation awareness effects of adaptive automation of various air traffic control information processing functions.* Unpublished masters thesis, North Carolina State University, Raleigh, NC, USA.

National Research Council (1998). *The future of air traffic control: Human operators and automation.* Washington, DC: National Academy Press.

Nunes, A. (2003). The impact of automation use on the mental model: Findings from the air traffic control domain. In *Proceedings of the Human Factors & Ergonomics Society 47th Annual Meeting* (pp. 66-70). Santa Monica, CA: Human Factors and Ergonomics Society.

Parasuraman, R., & Riley, V.A. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230-253.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model of types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics, 30,* 286-297.

Scerbo, M.W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 37-63). Mahwah, NJ: Lawrence Erlbaum.