The Impact of Automation Assisted Aircraft Separation on Situation Awareness

Arik-Quang V. Dao[1], Summer L. Brandt[1], Vernol Battiste[1], Kim-Phuong L. Vu[2], Thomas Strybel[2], and Walter W. Johnson[1]

[1] NASA Ames Research Center
Moffett Field, CA 94035, United States of America
{quang.v.dao,summer.l.brandt,vernol.battiste-1,
walter.w.johnson}@nasa.gov
[2] California State University Long Beach, Dept of Psychology
1250 N Bellflower Blvd, Long Beach, CA 90840, USA
{kvu8,tstrybel}@csulb.edu

Abstract. This study compared situation awareness across three flight deck decision aiding modes. Pilots resolved air traffic conflicts using a click and drag software tool. In the automated aiding condition, pilots executed all resolutions generated by the automation. In the interactive condition, automation suggested a maneuver, but pilots had the choice of accepting or modifying the provided resolution. In the manual condition pilots generated resolutions independently. A technique that combines both Situation Global Assessment Technique and Situation Present Awareness Method was used to assess situation awareness. Results showed that situation awareness was better in the Manual and Interactive conditions when compared to the Automated condition. The finding suggests that pilots are able to maintain greater situation awareness when they are actively engaged in the conflict resolution process.

Keywords: automation, conflict resolution, situation awareness, cockpit display of traffic information, CDTI, cockpit situation display, CSD.

1 Introduction

Without additional tools and automation, increases in air traffic densities during the next few decades will exceed the capabilities of current-day controllers to manage them [1]. However, adding tools and automation to the air traffic management system will require a new distribution of roles and responsibilities across ground control, the flight deck, and automation. In response, the Joint Planning and Development Office (JPDO), in its vision of the Next Generation Air Transportation System (NextGen), has suggested multiple concepts of operation for distributing roles and responsibilities, including the delegation of aircraft separation responsibilities to flight crews, aided by automation, or to automation alone.

Automated systems have been successfully used in many areas of transportation to increase operator workload capacity and system safety. However, automation has

M.J. Smith and G. Salvendy (Eds.): Human Interface, Part II, HCII 2009, LNCS 5618, pp. 738–747, 2009.

been shown to impact operators' information acquisition, information analysis, decision making, and action [2]. Thus, a major issue in implementing automation in a system where humans remain "in-the-loop" is its impact on operator situation awareness (SA). As a first step towards addressing this issue, researchers from the NASA Ames Flight Deck Display Research Laboratory (FDDRL) and California State University Long Beach conducted a study comparing SA of commercial pilots performing a conflict resolution task across three levels of automation aiding.

1.1 Automated Separation Assurance on the Ground and the Flight Deck

One responsibility of an air traffic controller is to detect and resolve upcoming conflicts (losses of legal separation) between aircraft. Since the number of conflicts increases with traffic density, this can lead to an unacceptable increase in controller workload. One solution is to provide conflict detection and/or resolution automation that take some of the work off the controller. In addition, researchers have been investigating the possibility of delegating separation responsibility to appropriately equipped flight decks [3]. On such flight decks, the pilot would be provided with technology that supports conflict detection and resolution.

As one of the primary proponents of such automation, Erzberger [4], at the NASA Ames Research Center, has been developing an "auto-resolver". While his efforts have been aimed mainly at a ground-based implementation, several concepts of operation could be envisioned which would utilize this automation. For example, Homola [5] examined controller acceptance of concepts where this automation would be used 1) to autonomously generate resolutions without involving pilots or controllers (fully automated mode), and 2) to allow air traffic controllers to request conflict resolutions on demand, and then modify them as they wished. In turn then, the general goal of the present study was to examine pilot acceptance of this same automation, and for this paper, to look at its impact on pilot SA.

1.2 The Impact of Automation on Situation Awareness

Situation awareness refers to the operator's understanding of the state of the relevant environment and his or her ability to anticipate future changes and developments in that environment [6]. A widely cited definition of SA specifies three levels of the construct in terms of 1) perception, 2) comprehension, and 3) projection [7].

Factors that impact SA in the context of automation include automation complacency, automation mistrust, workload, and automation transparency. Automation complacency occurs when the human falls out-of-the-loop due to over trust in the system [8]. In the extreme case, the human no longer actively processes information to maintain an awareness of the system state, diminishing his or her ability to recover from automation failure. Automation mistrust occurs when the human perceives the automation to be unreliable and devotes excessive attention to monitoring the automation. SA can also be diminished when workload is very high, resulting in attentional tunneling [2] where all attentional resources are drawn to the primary task, reducing the amount of resources available for perceiving and processing other information in the environment. SA is also impacted by the requirement to evaluate choice alternatives when interacting with decision support automation [9]. In these, operators must often integrate the information provided by the automation with their assessment of the situation. This additional workload could also reduce the resources available for maintaining SA. A system is "transparent" when the underlying reason and information behind the behaviors of automation are understood by the operator [10]. In a fully transparent system an operator may be led to attend to too much system information, with resulting high workload and diminished SA [11]. In the other extreme, a system that does not display information or provide adequate feedback regarding system behavior may reduce workload at the cost of transparency. Thus, the lack of transparency can also lead to diminished SA. A system that supports good SA provides transparency at a manageable workload level.

1.3 SA Probing Techniques

SA measures are classified as probing techniques, rating scales, or performance-correlated measures. Probe techniques have been identified as the most promising of the measures because they are sensitive to the operator and task environment [12] and can provide diagnostic information regarding the cause(s) of poor SA. The Situation Global Assessment Technique (SAGAT) and Situation Present Awareness Method (SPAM) are two commonly used SA probe techniques [7][13]. SAGAT questions are administered by stopping the scenario at random intervals so that SA probes do not interfere with the current tasks. Because the operator does not have access to the display during the simulation pauses, this technique is highly dependent on his or her memory for information in the simulated environment.

Unlike SAGAT, SPAM does not index SA solely on the basis of the information the operator can remember. Instead, SPAM probes are administered during the course of the scenario while operators have access to information on their displays. This technique is based on the notion that information can be stored in memory or looked up on the display. If SA information, or the location of the information, is in the operator's awareness, then response times are expected to be faster than when the information had to be located. One drawback of SPAM is that the questions may interfere with the operator's primary task [14].

The present study attempted to avoid the problems associated with SPAM and SAGAT probes by using short 3-minute scenarios, where a conflict was presented and resolved prior to answering probe questions. The probe questions were administered after the conflict resolution and while pilots had full access to their displays, so much of the information in the environment did not have to be held in memory. And, because the scenario stopped, there was no additional workload imposed on the operator's primary tasks while answering the questions.

1.4 The Current Study

Commercial pilots viewed scenarios requiring conflict resolutions, and were presented three aiding conditions: automated, interactive, and manual. In the automated condition, pilots were required to evaluate resolutions generated by the automation. They were not allowed to modify this resolution. In the interactive condition, pilots were given the automated resolution, but were free to modify it with a Route Assessment Tool (RAT) prior to executing it. The RAT allowed the pilot to graphically stretch and bend the flight plan to generate a preferred conflict free path but did not itself provide a recommended resolution. In the manual condition, pilots resolved the conflict using the RAT alone (no automation).

Using the above SA probe technique, the current study assessed pilot SA across the aforementioned three automation levels (manual, automated, and interactive). We predicted that workload in the manual condition would impose a greater demand on the pilot's attentional resources, reducing SA compared to conditions with automation aiding. We expected that SA in the fully automated condition would be lower than in the interactive condition due to complacency. We reasoned that the SA obtained by actively examining alternatives to the suggested resolutions in the interactive condition would more than offset the loss due to the increase in workload.

2 Method

2.1 Participants

Seventeen commercial airline pilots with glass cockpit experience participated in this study and were compensated $25/hr. Data from five of the pilots were excluded due to difficulty in learning to use the display and conflict resolution tools.

2.2 Apparatus

All scenarios were presented on a 30" monitor using a Cockpit Situation Display (CSD; see Figure 1), developed by NASA FDDRL.

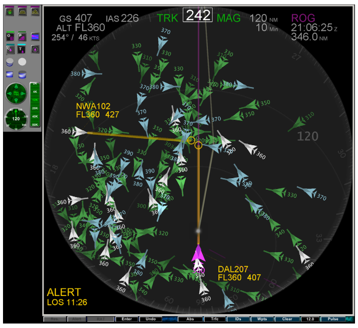

Fig. 1. CSD with conflict alerting and route proposal

The CSD, a PC-based 3D volumetric display, provided pilots with the location of surrounding aircraft, plus the ability to view the expected 4D trajectories of ownship and all traffic [15]. Embedded within the CSD was logic that detected and highlighted conflicts. Pilots were told that conflict detection was 100% reliable. In addition, the CSD had pulse predictors that emitted synchronous bullets of light that traveled along the displayed flight plans at a speed proportional to the speeds of the associated aircraft. In this way, a prediction of up to 20 minutes into the future was added to the display and provided a graphical way of confirming how close ownship was expected to come to any other aircraft.

A version of Erzberger's auto-resolver was implemented to aid pilots with the conflict resolution task in the automated and interactive conditions [4]. In addition, in the interactive and manual conditions participants were asked to resolve conflicts by modifying their flight plan using the RAT [16]. This tool was linked to conflict detection software allowing the pilots to find conflict-free paths. Proposed resolutions created by the automation or the operator were color coded in gray to distinguish them from the current route (amber if in conflict; magenta in nominal conditions).

A pool of candidate questions was generated to capture two important dimensions of SA, time frame and processing category. A question's time frame refers to whether it queried awareness involving past, present, or future events. A question's processing category was divided into recall and comprehension. Recall refers to queries where the pilot had to recall the information, or where the information was located on the display. Comprehension refers to queries where the pilot had to process information before a response was made. Together, these dimensions generated six distinct probe types (see Appendix A for sample questions). All questions were checked by a commercial airline pilot for relevancy to the task. The six types of probes were counterbalanced so that each aiding condition received four probes from each category. Presentation order was randomized for each trial.

2.3 Design and Procedure

The SA data were collected within a larger simulation that examined pilots' acceptance of, and preferences for, automated conflict resolutions, and the impact of these automated resolutions on SA [17]. This paper analyzes the SA probes as a function of 3 aiding conditions, 3 question time frames, and 2 question processing categories.

Participants completed 3 blocks of 16 trials. Each block presented only one level of aiding. For the automated and interactive conditions, half of the proposed resolutions were vertical and half horizontal. All participants received the automated condition in the first block due to constraints imposed by goals of the larger study [17]. Following the first block, participants were trained on how to use the RAT to create or modify the conflict resolutions. Half of the participants then received the manual condition followed by the interactive condition, while the other half received the opposite order. Each trial lasted about 3 minutes with the entire experiment lasting approximately 4 hours. Participants were provided breaks after each block of trials.

During the initial 15 seconds of each trial participants could manipulate the display in any manner, but could not modify the route. The automation proposed resolution was then displayed in the automated and interactive conditions. Pilots could modify the route in the interactive condition. No automation resolution was shown in the manual condition, but the pilots used the RAT to create a resolution. For all aiding conditions, pilots had a 90 second window in which to examine/modify/create a resolution. Pilots terminated this window by executing the resolution.

Pilots were encouraged to change their display between 2D and 3D views when analyzing the conflict to fully visualize the possible resolutions, especially those generated by the automation. Transitioning between 2D and 3D views prevented the pilots from making perceptual errors in differentiating vertical from horizontal resolutions. The pilots were also trained to use a number of decluttering features and could zoom in or out of the situation by varying the horizontal and vertical range.

At the end of each trial, pilots were given three SA questions. It was emphasized that the questions were being used to assess SA. Therefore, participants were NOT to deliberately search for potential information being queried but to focus on the primary task, conflict resolution, during the active trial. The display remained accessible while the probe questions were presented in order to allow pilots to look up information.

3 Results

Responses to the SA questions were reviewed for anomalies and missing values, and three comprehension questions were thrown out as unscorable because definitive answers could not be determined from the information available on the display at the end of the trial. The review also uncovered missing data points (4%) due to computer error and incomplete trials. These data were replaced with unweighted means calculated from the responses for each individual participant for the specific condition. No participant was missing all responses for any condition.

3.1 Analysis of Correct Responses to SA Probes

The percent of correct responses to the SA questions was analyzed in a 3 (aiding condition: automated, manual, interactive) x 3 (time frame: past, present, future) x 2 (process: recall, comprehension) within-subjects analysis of variance (ANOVA). Means and standard deviations are presented in Table 1.

The main effect of aiding condition was not significant, $F(2, 22) = 1.37$, $p = .27$. Participants answered a similar number of the questions correctly in the automated (73%), manual (72%) and interactive (69%) conditions.

Table 1. Means and standard deviations for percent correct responses and reaction times in seconds to SA questions

	% Correct	RT
Aiding Condition		
Automated	73 (8)	27.17 (9.01)$_{ab}$
Manual	72 (8)	18.20 (5.60)$_{a}$
Interactive	69 (10)	19.62 (6.33)$_{b}$
Time Frame		
Past	78 (7)$_{a}$	18.12 (4.40)$_{a}$
Present	78 (9)$_{b}$	17.38 (4.52)$_{b}$
Future	58 (8)$_{ab}$	29.49 (8.33)$_{ab}$
Process		
Recall	80 (7)$_{a}$	19.07 (5.79)$_{a}$
Comprehension	63 (10)$_{a}$	24.26 (5.50)$_{a}$

Note. Means in the same column within a condition that share a subscript differ at p<.05

There was a significant main effect of time frame on percent correct, $F(2, 22) = 71.29$, $p < .001$. Post-hoc comparisons performed using the Bonferroni adjustment for multiple comparisons showed significantly fewer questions were answered correctly for future (58%) than past (78%) or present (78%) queries, $p < .001$. The low accuracy for future questions may be a result of pilots not needing to know the “big” picture (as compared to air traffic controllers). In addition, the RAT automatically provided information about potential conflicts along proposed routes. If pilots solely focused their attention on the immediate task of resolving a conflict, processing of information about potential future traffic events may have been prevented.

A main effect of process on percent correct was also found, $F(1,11) = 31.18$, $p < .001$. Accuracy was lower for comprehension (63%) than recall (80%) questions. This is consistent with the idea that the information required by recall queries was readily available from memory or the display compared to the comprehension queries, which required more in depth understanding of the situation.

There was a significant interaction between time frame and process, $F(2, 22) = 28.05$, $p < .001$. More recall queries were correctly answered than comprehension queries in the past and present time frames, but there was no difference in recall and comprehension accuracy for the future queries (see Figure 2). The low accuracy for future recall questions is surprising because the information needed to answer the query could have been found directly on the display.

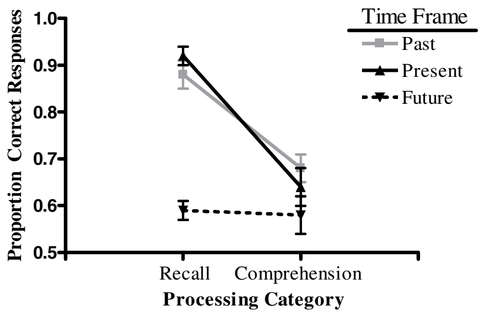

Fig. 2. Mean percent correct responses for SA categories

3.2 Analysis of Response Latencies to SA Probes

A 3 x 3 x 2 ANOVA was performed on the response time (RT, in seconds) to SA queries (see Table 1). The p-values were adjusted using Greenhouse-Geisser for violations of sphericity, where appropriate. There was a main effect of aiding, $F(2, 22) = 8.90$, $p < .001$, where RT was longer in the automated (27.17s) condition than in the manual (18.20s) and interactive (19.62s) conditions, *p*s< .01, which, in turn, did not differ, $p = 1.00$. High RTs in the automated condition support the notion that factors such as automation complacency or system transparency have a negative impact on SA. SA appears higher in the interactive and manual conditions, where pilots were engaged in manually modifying or creating new resolutions.

There was also a main effect of time frame on RT, $F(2, 22) = 57.17$, $p < .001$. Consistent with the accuracy data, pilots had the most difficulty with future queries. RT for past (18.12s) and present (17.38s) questions did not differ, $p > .99$, but were faster than future (29.49s) questions, $p < .001$. Similarly, there was a main effect for process on RT, $F(1,11) = 41.43$, $p < .001$, with faster response to recall (19.07s) queries than to comprehension (24.26s) queries.

The aiding by time frame interaction was significant, $F(4, 44) = 4.37$, $p < .01$ (Figure 3, left panel). Response times for past and present questions were significantly faster than future in all aiding conditions (p <.01), but more so for the future than past or present time frame, $p < .001$. The time frame by process interaction was also significant, $F(2, 22) = 42.40$, $p < .001$. Again, recall queries were responded to quicker than comprehension queries for past and present queries, $p <.001$, but the reverse was true for future queries, $p < .01$ (see Figure 3, right panel).

The SA response latencies for correct and incorrect responses to the SA questions were then reviewed. Although not statistically analyzed, trends in the data show that response times were shorter when a participant answered the question correctly compared to when questions were answered incorrectly. This implies that response latencies do reflect SA and are not due to a speed accuracy trade-off.

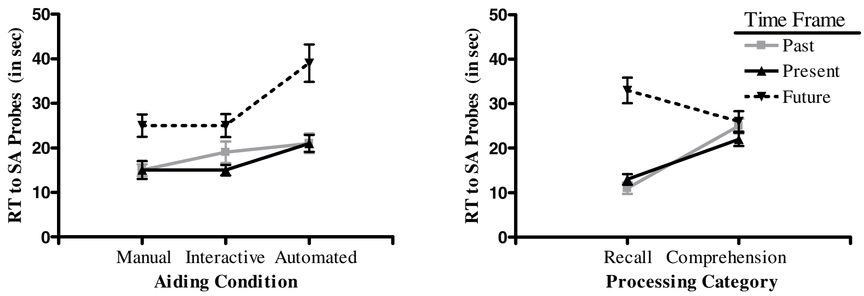

Fig. 3. Reaction time to SA probe questions as a function of aiding and time frame (left panel) and processing category and time frame (right panel)

Given that the automated condition was always presented first, poorer performance in that automated condition could have been due to a general learning effect. To see if learning was occurring, performance on the first and second halves of the automated trials were compared. There were no significant difference in percent correct between the first half (*M* = 71%, *SD* = 12%) and the second half (*M* = 77%, *SD* = 10%) of the automated aiding trials ($F(1,11) = 1.81$, $p = .21$); or in SA probe latencies (*M* = 26.05s, *SD* = 9.86s for the first half and *M* = 26.80s, *SD* = 10.56s for the second half, $F(1,11) = .06$, $p = .82$). Thus, it is likely that the aiding effects reported are due to the automation level and not to learning.

4 Discussion

The goal of this study was to examine SA across two levels of automation and a fully manual condition. SA queries were designed to probe pilot recall of information, as well as comprehension for information about past, present, and future events in the scenarios. There were no overall differences for aiding levels for SA probe accuracy, but there was a difference for SA response latencies. This finding implies that response time may be a more sensitive index of SA than the accuracy of responses. Response times were faster in the interactive and manual conditions when compared to the automated condition, suggesting that fully automated conflict resolution does not equally support SA. The effect of diminished SA in the automated condition may have been due to complacency, perhaps exacerbated by the lack of transparency in the flight deck implementation of the auto-resolver. Relatively better SA in the interactive and manual conditions implies that conflict resolution systems may profit from keeping the human operator actively engaged in the task. However, the lack of a difference between the manual and interactive condition shows that conflict resolution automation can be implemented without a cost to SA.

Unlike aiding, probe accuracy varied as function of time frame. Pilots scored lower on probes for information about the future status of an event than on probes for present and past information. It is possible that the 90 second time pressure of the conflict resolution task required pilots to focus more on the immediate traffic situation. We speculate that pilots have more egocentric awareness and do not require the “big picture” of expert controllers. In addition, the increased frequency of response errors to future questions may have been due to limited access to information that would otherwise provide more clues about future traffic events. Although pilots had access to information on the display while responding to the SA probes, the traffic in the scenarios was frozen. Therefore, pilots were not able to access additional information that could have been attained if a traffic event was continued.

Consistent with results found for the frequency of correct responses, analysis of response times for time frame revealed slower response times for future questions than for past and present questions. Patterns in the data reveal that latencies were greater when accuracy scores were lower, thus the data does not reflect a speed accuracy trade-off, but imply that the response latencies reflect SA [13].

In general, an operator can experience diminished SA when control of system state changes is delegated to either automation or to other human operators [8]. Future studies will assess team SA in scenarios where air traffic management responsibilities are coordinated across automation, pilots, and controllers.

Acknowledgement. This simulation was partially supported by NASA cooperative agreement NNA06CN30A.

References

1. Joint Planning and Development Office: Next Generation Transportation System: Concept of Operation V 2.0. Washington D.C: Government Printing Office (2007)
2. Parasuraman, R., Wickens, C.D.: Humans: Still vital after all these years of automation. Human Factors 50, 511–520 (2008)
3. Lee, P.U., Mercer, J.S., Martin, L., Prevot, T., Shelden, S., Verma, S., Smith, N., Battiste, V., Johnson, W., Mogford, R., Palmer, E.: Free Maneuvering, Trajectory Negotiation, and Self-Spacing Concept In Distributed Air-Ground Traffic Management. In: USA/Europe Air Traffic Management R&D Seminar, Budapest, Hungary (2003)
4. Erzberger, H.: Automated Conflict Resolution For Air Traffic Control. In: Proceedings of the 25th International Congress of the Aeronautical Sciences, Hamburg, Germany (2006)
5. Homola, J.: Analysis of Human and Automated Separation Assurance at Varying Traffic Levels. Master’s Thesis, San Jose State University, San Jose, CA (2008)
6. European Air Traffic Management Programme: The Development of Situation Awareness Measures in ATM Systems. HRS/HSP-005-REP-01 (2003)

7. Endsley, M.R.: Measurement Of Situation Awareness In Dynamic Systems. Human Factors 37(1), 65–84 (1995)
8. 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), 286–296 (2000)
9. Endsley, M.R., Bolte, B., Jones, D.G.: Designing for Situation Awareness: An Approach to User-Centered Design. Taylor & Francis, New York (2003)
10. Mark, G., Kobsa, A.: The Effects Of Collaboration And System Transparency. Presence 14(1) (2005)
11. Duggan, G.B., Banburry, S., Howes, A., Patrick, J., Waldron, S.M.: Too Much, Too Little Or Just Right: Designing Data Fusion For Situation Awareness. In: Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting, pp. 528–532 (2004)
12. Strybel, T.Z., Vu, K.-P.L., Kraft, J., Minakata, K.: Assessing The Situation Awareness Of Pilots Engaged In Self Spacing. In: Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting, pp. 11–15 (2008)
13. Durso, F.T., Bleckley, M.K., Dattle, A.R.: Does Situation Awareness Add To The Validity Of Cognitive Tests? Human Factors, 721–733 (Winter 2006)
14. Pierce, R.S., Vu, K.-P.L., Nguyen, J., Strybel, T.: The Relationship Between SPAM, Workload, and Task Performance on Simulated ATC Task. In: Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting, pp. 34–38 (2008)
15. Granada, S., Dao, A.Q., Wong, D., Johnson, W.W., Battiste, V.: Development And Integration of a Human-Centered Volumetric Cockpit Display for Distributed Air-Ground operations. In: Proc.12th International Symposium on Aviation Psychology, Oklahoma City, OK (2005)
16. Canton, R., Refai, M., Johnson, W., Battiste, V.: Development And Integration Of Human-Centered Conflict Detection And Resolution Tools For Airborne Autonomous Operations. In: Proc.12th International Symposium on Aviation Psychology, Oklahoma City, OK (2005)
17. Battiste, V., Johnson, W.W., Dao, A.-Q., Brandt, S., Johnson, N., Granada, S.: Assessment of Flight Crew Acceptance of Automated Resolution Suggestions and Manual Resolution Tools. In: 26th International Congress of the Aeronautical Sciences, Anchorage, AK (2008)

Appendix A: Sample Situation Awareness Questions

Past Recall	What was your heading at the start of the trial?
Past Comprehension	What was the difference in altitude of Ownship and the Intruder prior to the resolution?
Present Recall	What is your current heading?
Present Comprehension	Currently, what is the difference in altitude between Ownship and the nearest proximal traffic?
Future Recall	What will be your groundspeed at the push point?
Future Comprehension	How much additional time will it take you to reach the next original waypoint, given the executed resolution?