Impact of Conflict Avoidance Responsibility Allocation on Pilot Workload in a Distributed Air Traffic Management System

Sarah V. Ligda[1], Arik-Quang V. Dao[1], Thomas Z. Strybel[2], Kim-Phuong Vu[2], Vernol Battiste[1], Walter W. Johnson[3]
[1]SJSURF/NASA Ames Research Center, MS 262-2, Moffett Field, CA 94035, USA
[2]California State University Long Beach, CA 90840, USA
[3]NASA Ames Research Center, CA 94035, USA

Pilot workload was examined during simulated flights requiring flight deck-based merging and spacing while avoiding weather. Pilots used flight deck tools to avoid convective weather and space behind a lead aircraft during an arrival into Louisville International airport. The study examined concepts placing conflict avoidance responsibility on different combinations of pilot, controller and automation. An ATWIT metric, modified to measure workload, showed highest workload during the approach, and lowest during the en-route phases of flight (before deviating for weather). The trend across multiple workload metrics showed workload to be lowest when pilots had both conflict alerting and responsibility for avoiding conflicts; while *all* objective and subjective measures showed workload was highest when pilots had no conflict alerting or responsibility for avoiding conflicts. The findings suggest workload is lowest when pilots have conflict alerting displays plus responsibility for conflict resolution; and highest when they had neither the displays nor responsibility. This suggests workload is primarily driven by an attempt to regain situation awareness when conflict alerting is unavailable. It also suggests that workload is not tied primarily to responsibility for resolving conflicts, but to maintaining situation awareness. In general, the modified ATWIT was shown to be a valid and reliable workload measure, providing more detailed information than post-run subjective workload metrics.

INTRODUCTION

It is predicted that demand for air travel will double within the next 15 years. To meet this demand, significant changes to the current air traffic management (ATM) system are being evaluated (Joint Planning and Development Office, 2007). It is known that increased traffic loads negatively affect air traffic controller (ATC) performance; however, this impact can be reduced when ATCs are assisted by automated conflict resolution tools (Prevot et al., 2009). One way to reduce ATC workload is to adjust the roles and responsibilities of operators in the ATM system. For example, a portion of the responsibility for maintaining safe separation distances between aircraft could be transferred from ATC to the flight deck or be automated to some extent. These strategies should alleviate a portion of ATC workload so that traffic loads can be increased while meeting or exceeding current safety and efficiency standards.

Currently, both air and ground-side conflict detection algorithms have been developed to aid the human operator with identifying air traffic conflicts. These conflict detection algorithms feed data to programs such as the Auto-Resolver which can then provide conflict resolutions upon request from the user (Ezerberger, 2006). Alternatively, the Auto-Resolver can be configured to automatically provide resolutions upon detection of a conflict and wait for confirmation from ATC before executing, shifting the controller to a more supervisory position than current day operations. In the most extreme case, the Auto-Resolver can be configured to automatically generate resolutions upon detection, as well as automatically execute resolutions without prior consent from ATC. Similarly, flight deck conflict detection algorithms contain logic to identify and highlight conflicts and provide automated resolutions on displays such as NASA Flight Deck Display Research Laboratory's (FDDRL) Cockpit Situation Display (CSD) (Granada, Dao, Wong, Johnson, & Battiste, 2005).

Impact of Workload

Subjective operator workload can be defined as "the perceived relationship between the amount of mental processing capability or resources and the amount required by the task" (Hart & Staveland, 1988). Many task factors can affect ATC workload, with a major contributor being air traffic density. According to Lee (2005), the relationship between workload and traffic count is non-linear. ATC workload increases from low to high only during which a certain traffic threshold is reached, meaning that workload cannot be predicted from traffic counts. Traffic management, which includes the detection and resolution of potential traffic conflicts, increases threefold with a linear increase in the number of aircraft (Wickens, 1992). This suggests that controllers reach their maximum workload capacity at a fixed traffic load, likely to be exceeded in the next 15 years.

There are many strategies for automating conflict detection and resolution tasks, and optimal selection of a strategy requires assessment of workload and situation awareness. In cases where automation is completely responsible for conflict detection and resolution, humans may be thrown out-of-the-loop leading to complacency and loss of situation awareness (Parasuraman, Sheridan, & Wickens, 2000). Moreover, when workload is high, the operator is forced to tunnel in on the primary task, reducing the cognitive resources required for proactive acquisition of situation-awareness-relevant information in the environment (Parasuraman & Wickens, 2008). On the other hand, automated tools that aid the human operator might support a performance benefit because workload could be reduced without losing situation awareness (Dao et al., 2009). However, operators must trust automation to be willing to use an automated system. Studies demonstrate that this is possible with some degree of a human-in-the-loop structure and with practice and continued use (Ligda, Johnson, Latcher & Johnson, 2009). Therefore, in the current investigation, pilot workload was measured under three function allocation strategies (Concepts) for conflict detection and resolution: pilot, controller, or automation primarily responsible for conflict avoidance. This study was conducted in a larger context of trajectory oriented operations, but the present paper will focus primarily on the impact of workload for pilots in each operational concept.

Measuring Workload

A modified Air Traffic Workload Input Technique (ATWIT; Stein, 1985) was used in the present study. The ATWIT is a 'real-time' workload measure that queries the operator for estimates of workload while managing traffic in human-in-the-loop scenarios. In its original form, an auditory alert is presented to the operator who is instructed to rate his/her workload on a 1-7 response panel (1 = low workload and 7 = very high workload). Because one goal of the larger simulation was to measure operator situation awareness using online queries, we modified the procedure for administering the ATWIT to be consistent with the Situation Present Assessment Method (SPAM, Durso & Dattel, 2004). Instead of receiving a situation awareness query, some queries simply asked pilots to rate their workload on a 5-point ATWIT scale.

With the SPAM, a 'ready' alert is presented first to the operator. When the operator responds that s/he is ready to answer a query, either a workload or situation awareness question is presented, and the operator answers the question. Three performance measures are obtained from this method: response latency, the interval between appearance of ready alert and when the participant responds ready; response latency, the interval between presentation of the query and the response to the question, and the response itself. The first latency (ready latency) is assumed to quantify workload because the time it takes the operator to indicate that a question can be answered should depend on his/her workload (i.e., lower workload = shorter ready latencies). The second latency (query latency) is assumed to reflect situation awareness. However, when workload rating queries are used, the workload rating itself (1 = low workload, 5 = high workload) should be related only to the ready latency and not the query latency. The query latency (the time to enter the workload rating), should not be related to workload per se, but the ease in which the operator can quantify his/her workload. The last measure, the subjective rating, should thus correlate with the response latency but not with the query latency.

The current study used these operator workload metrics to determine whether different concepts of operation will affect pilot workload. In addition, examination of phase of flight was also analyzed to ascertain when best to adjust to full or partial automation while maintaining optimum levels of pilot workload. Each phase of flight is expected to have a differential impact on pilot workload considering engagement level systematically adjusts in each phase. We predicted that these three concepts of operation will affect workload because of the shifts in responsibility for conflict resolution between controllers, pilots and ground/air automation.

Scenarios

In the current study, pilots and controllers engaged in 'real time' simulations focused on approach operations into Louisville International-Standiford Field Airport (SDF). Pilots were required to comply with spacing commands sent from the ground, and avoid hazardous weather and traffic. Sector traffic density was three-times that of current day in order to ensure that changes in workload would be detected. Controllers managed traffic and resolved conflicts using simulated radar software. Three concepts of operation were tested. These shifted responsibility for conflict resolution between pilots, controllers and automation:

Concept 1: Pilots were responsible for solving most traffic conflicts. Experimental pilots initiated all traffic and hazardous weather maneuvers with no support from the Auto-Resolver, but had conflict detection and resolution capabilities. The controllers in managed all conflicts that did not involve experimental pilots. The controllers were supported by the Auto-Resolver which generated conflict resolutions upon request.

Concept 2: Controllers were responsible for solving most traffic conflicts. Again, the controllers were provided with the Auto-Resolver tool for resolutions upon request. The Auto-Resolver agent solved the remaining conflicts by detecting and automatically datalinking the new routes to pilots. Here, the flight deck was not responsible for any conflicts; however, pilots had onboard conflict detect and resolution tools, and reviewed all resolutions before executing.

Concept 3: The Auto-Resolver managed most of the conflicts in the scenarios. The controllers managed the remaining portion. The flight deck was not responsible for maintaining separation and had no capabilities to predict conflicts, but reviewed all uplinked resolutions before executing. Concept 3 represents the highest contribution of automation in the study's ATM system and the lowest predicted workload for both controller and pilot participants.

METHOD

Participants

Eight air-transport rated (ATP) pilots with glass cockpit experience and two former Oakland-Center controllers participated in this study. They were compensated $25/hr for their participation.

Equipment

Participant pilots flew one of eight desktop simulators located at NASA Ames FDDRL. Additional air traffic was provided by pseudopilots. Pseudopilots and ATCs were located at FDDRL, Cal State University Long Beach, Cal State University Northridge and Purdue University in a distributed simulation network.

Pilots interacted with the CSD and a simulated 747 flight deck in the MultiAircraft Control System (MACS; Prevot et al, 2000). MACS and CSD provided an automated merging and spacing tool for each individual aircraft. The CSD, a PC-based 3D volumetric display, provided pilots with the location of surrounding aircraft and the ability to view the expected 4D trajectories of ownship and all traffic (Granada et al., 2005). Embedded within the CSD was logic that detected and highlighted conflicts. A more detailed description of the CSD's Route Assessment Tool (RAT) that pilots interacted with to modify their flight plans for weather and traffic avoidance, as well as an explanation of the Multi-Aircraft Control System (MACS) software, can be found in Dao et al. (submitted). A separate touch screen tablet computer was used to administer online queries.

Design and Procedure

The main independent variable discussed in this paper is concept of operations: primary responsibility for maintaining separation delegated to pilot, controller, or Auto-Resolver. Participants completed 4 trials per day over 3 days. On each day, one concept of operations was tested. Two trials were repeated on day 4 due to software malfunctions. Each trial lasted approximately 80 minutes. Classroom training and practice trials were provided prior to the test days.

All pilots flew within the same scenario in 'real time' and were assigned a spacing interval and lead aircraft by an automated ground station two minutes after the start of the trial. Pilots used the datalink panel on their display to load the information into the CSD and then executed the spacing command after manually selecting the lead aircraft on their display. In addition, pilots were trained to maneuver for convective weather using the RAT. Pilots adjusted their route relative to the weather based on their own safety criteria and constraints imposed by surrounding traffic. In Concept 1, pilots independently managed separation by maneuvering for traffic using the RAT. In Concept 2 and 3, pilots waited for commands issued by a human controller or the Auto-Resolver before maneuvering for traffic. ATC managed traffic based on the concept of operations, and provided re-sequencing instructions on request from the pilots. In addition, ATC responded to datalinked requests for route modifications and requests made over the radio.

Throughout the scenario, pilots and controllers received query prompts to measure situation awareness or workload every three minutes from the start of each trial. Workload ratings were obtained at 9, 27, 45, and 69 minutes from the start of the trial. The queries were not displayed until the participant responded to a ready prompt, meaning that the participants could not predict when the workload queries were administered. The participants were instructed that these prompts should not interrupt their primary pilot roles and responsibilities, and to answer the prompt after primary duties were performed. If no response was made to either the ready prompt or query after one minute, it was removed from the screen and scored as a time out.

The scenario times of the ATWIT query prompts roughly corresponded to significant events in the scenario: before deviating for weather (9 min), while deviating for weather (or immediately after deviating) and spacing re-sequencing (27 min), at top of descent or beginning the descent phase of flight (45 min), and the approach phase of flight into SDF (69 min). The simulation did not pause while the participants were answering the queries. Pilots completed a trial when they landed in SDF.

At the end of each run, pilots were given a post-trial questionnaire that contained eight workload questions. Pilots were instructed to rate overall workload and peak workload on a 5-point scale (with 1 indicating low and 5 indicating high workload) for 4 pilot tasks: en-route spacing, weather avoidance, continuous descent approach, and approach spacing. For example, questions regarding en-route spacing were: "Please rate your overall workload associated with maintaining spacing," and "Please rate your peak workload associated with maintaining spacing (If no peak event, circle N/A)."

RESULTS AND DISCUSSION

Pilot Subjective Workload: Post-Trial

The post-trial data analyzed included the four workload ratings (overall and peak for each) for each trial. See Table 1 for overall means and standard deviations.

Table 1: Overall Means & Standard Deviations of Post-Trial Workload (1= low workload; 5= high workload)

	Overall En-route Spacing	Overall Weather Avoidance	Overall CDA	Overall Arrival Spacing
Overall	1.87 (0.78)	1.74 (0.62)	1.87 (0.83)	2.09 (0.82)
Concept				
Pilot Primary	1.69 (0.59)	1.54 (0.51)	1.74 (0.75)	1.85 (0.56)
Controller Primary	1.78 (0.68)	1.75 (0.55)	1.61 (0.69)	2.11 (0.75)
Auto-Rslvr Primary	2.14 (0.96)	1.92 (0.73)	2.26 (0.92)	2.29 (1.05)

n = 8

For each flight task, workload ratings were submitted to a repeated measures analysis of variance (ANOVA), with Concept as a factor. A significant effect of concept was observed only for overall in CDA workload ratings ($F(2,10) = 5.91$, $p = .02$). Bonferroni post-hoc analyses revealed significant differences between Concept 2 (controller primarily responsible for conflicts) and Concept 3 (automation primarily responsible for conflicts). Interestingly, the difference between Concept 1 (pilot responsible) and Concept 2 (controller responsible) was not significant. This finding may be due to the absence of the conflict detection and resolution capabilities on the CSD in the Auto-Resolver Concept, even though the pilot was not responsible for conflicts in either Concept 2 or 3.

It also should be noted that the pilots selected "not applicable" on average of 67% of the time for the peak workload ratings (En-route Spacing: 64% [C1: 53%; C2: 78%; C3: 63%]; Weather Avoidance: 80% [C1: 81%; C2: 88%; C3: 63%]; CDA: 75% [C1: 75%; C2: 91%; C3: 63%]; Arrival Spacing: 56% [C1: 78%; C2: 47%; C3: 44%]). This suggests that there was no peak associated with the pilots' workload in the majority of the trials.

Modified ATWIT

Means and standard deviations for the ATWIT queries are presented in Table 2. Of the 366 ratings, pilots rated their workload as 5 only once. One possible reason why pilots avoided a response selection of 5 might have been due to a perception of 5 representing the inability to manage their aircraft. Nevertheless, the single rating of 5 was excluded from subsequent analyses.

All three ATWIT measures (response latency, query latency and workload rating) were submitted to separate repeated measures ANOVAs with concept as a factor. These analyses did not yield any significant main effects, suggesting that Concept did not affect pilot perceived workload. However, note that there are similar trends between the Post-Trial Subjective Workload ratings in Table 1, and all three measures of the ATWIT metric at each concept condition.

Table 2: Means & Standard Deviations of ATWIT Queries

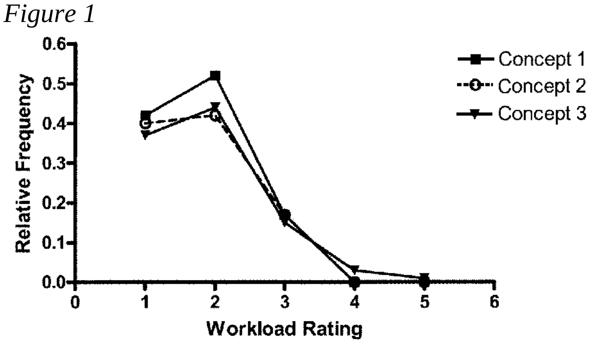
	Ready Latency	Query Latency	Response
Overall	5.36 s (3.51)	3.25 s (1.41)	1.76 (0.55)
Concept			
Pilot Primary	4.47 s (2.29)	3.28 s (1.35)	1.70 (0.40)
Controller Primary	5.42 s (4.89)	3.11 s (1.45)	1.78 (0.56)
Auto-Resolver Primary	6.18 s (3.33)	3.36 s (1.44)	1.81 (0.68)

n = 8

The number of timeouts represented when the participant did not respond to either the ready prompt or the ATWIT query after one minute. In these cases, it was likely that workload was too high to attend to the prompt. Overall, the percentage of timeouts was low (M = 4.7%) However, the differences between the percentage of timeouts when separated by Concept (M-concept 1 = 3.3%; M-concept 2 = 3.3%; M-concept 3 = 7.3%) were consistent with the Post-Trial Subjective Workload, and trends seen in all three measures of the ATWIT metric.

Lastly, when the relative frequencies of the subjective workload ratings are examined, they also suggest little effect of concept, as shown in Figure 1.

Figure 1



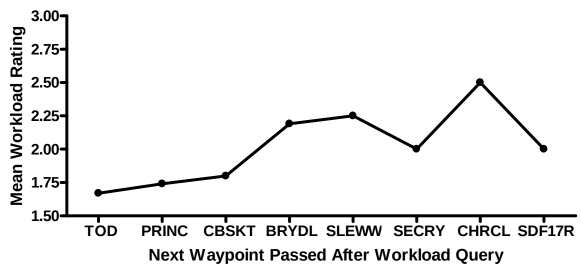
To examine assumptions of the SPAM technique, Pearson's correlations between the pilots' ready latency, query latency and workload rating were computed, as shown in Table 3. Workload ratings were significantly correlated with ready latency, (r(366)=.25, p<.001) and not with query latency, (r(366)=.05, p=.35). The finding that ready latencies and subjective responses are correlated, but query latencies and subjective responses were not, suggests that modified ATWIT is a valid and reliable measure of workload. Workload ratings were also significantly correlated with scenario time of query prompt (r(366)=.15, *p*=.004), also shown in Table 3. Because time of the query is related to the flight phase, the workload ratings were plotted in terms of the next waypoint passed after the workload question. These are shown in Figure 2.

Table 3: Pearson Correlation - Response Times

	Response	Ready	Query	Scen.Time	Next WP
Response	-	.25**	.05	.15**	.20**
Ready Latency		-	-.04	.09	.05
Query Latency			-	-.08	-.06

**= *p*< 0.01

Figure 2: Relationship between Response Rating and Next Waypoint After Workload Query



In Figure 2, it is evident that pilot workload increased throughout the descent phase, and reached a peak when queried before the CHRCL waypoint (directly before landing at SDF). Because pilots needed to perform energy management during the arrival/approach phase of flight, this task can account for their increased workload. Specifically, they needed to meet the altitude and speed restrictions at each waypoint while maintaining the specified spacing interval from their lead aircraft.

CONCLUSION

In this study, the prediction of pilots' workload increasing with more roles and responsibilities was not supported. The trend of the results in multiple workload metrics suggest that workload was lowest when pilots had the most responsibility for avoiding conflicts, and was significantly different for the subjective overall CDA workload. Additionally, in *all* of the objective and subjective measures, the trend for workload was highest in the auto-resolver primarily responsible concept. This suggests that pilots work hard to retain situation awareness when conflict detection and resolution tools are unavailable. Furthermore, this study suggests that when conflict detection and resolution tools are available, workload decreases, even with greater responsibility. These results also suggests that traffic load differentially affects pilots and controllers; a higher traffic load (3X) did not affect the workload of pilots to a large degree.

ACKNOWLDEGEMENT

This study was supported by the NASA Concepts and Technology Development Project, and in collaboration with NASA cooperative agreement (NNA06CN30A) researchers. These researchers, located at Cal State University Long Beach, Cal State University Northridge, and Purdue University, provided pseudopilots and controllers as part of a distributed simulation network. All participant pilots were tested at NASA Ames FDDRL.

REFERENCES

Dao, A. Q. V., Vu, K.-P., Battiste, V., Strybel, T. Z., & Johnson, W. W. (submitted). The impact of air and ground based automated aircraft separation assistance on individual and shared situation awareness. *54th Annual Meeting of the Human Factors and Ergonomics Society*, San Francisco, CA.

Dao, A. Q., Brandt, S. L., Battiste, V., Vu, K.-P. L., Strybel, T., & Johnson, W. W. (2009). The impact of automation assisted aircraft separation on situation awareness. *13th annual International Conference on Human-Computer Interaction*, San Diego, CA.

Durso, F.T, & Dattel, A. R. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application*.(pp. 137-153). Ashgate: Aldershot.

Erzberger, H. (2006). Automated Conflict Resolution for Air Traffic Control. *Proceedings of the 25th International Congress of the Aeronautical Sciences (ICAS)*, Germany.

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.

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.* (pp. 139-183). Amsterdam, Netherlands: North-Holland.

Joint Planning and Development Office. (2007). *Concept of operations for the Next Generation Air Transportation System, Version 2.0.* Washington, DC: Author.

Lee, P. U. (2005). A non-linear relationship between controller workload and traffic count. *Human Factors and Ergonomic Society Annual Meeting Proceedings, Human Performance Modeling, 5,* 1129-1133.

Ligda, S. V., Johnson, N., Lachter, J., & Johnson, W. W. (2009). Pilot confidence with ATC automation using cockpit situation display tools in a distributed traffic management environment. *13th Annual International Conference on Human-Computer Interaction,* San Diego, CA.

Prevot , T., Homola , J., Mercer, J., Mainini, M., & Cabrall, C. (2009). Initial evaluation of air/ground operations with ground-based automated separation assurance. *Proceedings of the 8th USA/Europe Air Traffic Management Research and Development Seminar,* Napa, CA.

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors, 3,* 511-520.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans, 3,* 286-297.

Stein, E. S. (1985). *Air traffic controller workload: An examination of workload probe.* Atlantic City, NJ: Federal Aviation Administration.

Wickens, C. D. (1992). *Engineering psychology and human performance (2nd ed.).* New York: Harper.