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], Kim-Phuong Vu[2], Thomas Z. Strybel[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. Three conflict avoidance management concepts were studied: pilot, controller or automation primarily responsible. A modified Air Traffic Workload Input Technique (ATWIT) metric showed highest workload during the approach phase of flight and lowest during the en-route phase of flight (before deviating for weather). 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. The trend across multiple workload metrics revealed lowest workload when pilots had both conflict alerting and responsibility of the three concepts, while *all* objective and subjective measures showed highest workload when pilots had no conflict alerting or responsibility. This suggests that pilot workload was not tied primarily to responsibility for resolving conflicts, but to gaining and/or maintaining situation awareness when conflict alerting is unavailable.

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 (ATCo) performance; however, this impact can be reduced when ATCos are assisted by automated conflict resolution tools (Prevot et al., 2009). Another way to reduce ATCo workload is to adjust 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 ATCo to the flight deck or be automated to some extent. These strategies should alleviate a portion of ATCo workload, thereby allowing traffic loads to be increased while meeting or exceeding current safety and efficiency standards.

Currently, both air and ground-side conflict detection algorithms have been developed as proposed aids to human operators by identifying air traffic conflicts. These conflict detection algorithms feed data to programs such as an Auto-Resolver which can then provide conflict resolutions upon request from the user (Ezerberger, 2006). We refer to this use of the Auto-Resolver as the Auto-Resolver tool. Alternatively, the Auto-Resolver can be configured to automatically provide resolutions upon detection of a conflict and wait for confirmation from ATCo 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 send resolutions to the flight deck for execution without prior consent from ATCo. We refer to the use of the Auto-Resolver in this capacity as an "agent". 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 [available] and the amount required by the task" (Hart & Staveland, 1988). Many task factors can affect traffic management workload, with a major contributor being traffic density. According to Lee (2005), the relationship between workload and traffic count is non-linear: ATCo workload increases from low to high only when a certain traffic threshold is reached, indicating that workload cannot be predicted from traffic counts simply. 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.

Many strategies exist to automate conflict detection and resolution tasks, and determining the optimal 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 possibly loss of situation awareness (Parasuraman, Sheridan, & Wickens, 2000). Alternatively, high levels of workload forces the operator to focus their attention on the primary task, reducing the cognitive resources available for proactive acquisition of situation awareness relevant information in the environment (Parasuraman & Wickens, 2008). Automated tools that aid the human operator would support a performance benefit because workload could be reduced without causing the operator to lose 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 a human-in-the-loop system and though practice and continued use of the system (Ligda, Johnson, Latcher & Johnson, 2009). In the current investigation, pilot workload was measured under the following three functional allocation conditions (referred to as Concepts 1, 2 and 3, based upon where the responsibility for conflict avoidance primarily resided): (1) pilot, (2) controller, or (3) automation. This study was conducted in a larger context of trajectory oriented operations; the present paper will focus primarily on the impact of workload for flight operations conducted within each operational concept.

Measuring Workload

The Air Traffic Workload Input Technique (ATWIT; Stein, 1985) asks operators to rate their workload while managing traffic at certain intervals throughout a simulated trial. In its original form, an auditory alert is presented to the operator who is then instructed to rate his/her workload on a 1-7 response panel (1 = low workload and 7 = very high workload). We modified the procedure for administering the ATWIT to be consistent with the Situation Present Assessment Method (SPAM, Durso & Dattel, 2004), because one goal of the simulation was to measure situation awareness using 'real-time' queries. Most questions involved situation awareness queries, but a few simply asked pilots to rate their workload on a 5-point scale.

Every three minutes, a 'ready' alert was presented to the operator on a separate screen from the flight displays along with an auditory prompt. The operator pressed 'ready' as soon as s/he was ready to answer a query. Then either a workload or situation awareness question was presented on the screen. Three performance measures were obtained from this method: ready latency (the interval between appearance of the ready alert and when the participant responds ready); query latency (the interval between presentation of the query and the response to the question); and the response itself. The ready latency is assumed to quantify workload: 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 query latency is assumed to reflect situation awareness. This latency should not be related to workload per se, but the ease in which the operator can answer the question. The subjective rating measure should thus correlate with the ready latency but not with the query latency.

This study used these workload metrics to determine whether the three concepts will differentially affect pilot workload. We hypothesize that pilot workload will significantly increase with conflict avoidance responsibilities in concept 1. In addition, we examined how workload changes with the phase of flight to ascertain whether it is best to adjust to full or partial automation to maintain optimum levels of pilot workload. Each phase of flight was expected to have a differential impact on pilot workload considering engagement level systematically adjusts in each phase.

Scenarios

In the current study, pilots and controllers engaged in 'real time' simulations focused on en route, arrival and approach operations into Louisville International-Standiford Field Airport (SDF). Pilots were required to comply with spacing commands sent from the ground, hazardous weather avoidance, and in concept 1 traffic separation. Sector traffic density was higher than 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 concepts shifted responsibility for conflict detection and/or resolution between pilots, controllers, and automation:

Concept 1: Pilots were responsible for solving 75% of potential conflicts. Pilots initiated all traffic and hazardous weather maneuvers with no support from the controller, but had flight deck conflict detection and resolution capabilities. The controllers managed the remaining 25% of conflicts. The controllers were supported by an Auto-Resolver tool which generated conflict resolutions upon request.

Concept 2: Controllers were responsible for solving 75% of potential conflicts. 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 new routes to pilots without a human controller in the decision-making loop. The pilots were not responsible for any conflicts; however, pilots had onboard conflict detection and resolution tools, and were able to review all resolutions before executing.

Concept 3: The Auto-Resolver agent was responsible for 75% of potential conflicts. The remaining 25% was

allocated to the human controller. The pilots were not responsible for maintaining separation and had no capabilities to detect or predict conflicts, but could review all uplinked resolutions before executing. Concept 3 represents the highest contribution of automation in the study's ATM system and should produce the lowest 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. The ATPs were compensated $25/hr. In the rest of this report when the term pilot is used, it will be taken to mean participant pilot, and not an experimental confederate pseudo-pilot.

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

Pilots flew either a fixed based 777 simulator or a desktop simulator: Multi-Aircraft Control System, (MACS; Prevot et al, 2000) and the Cockpit Situation Display (CSD). 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. Pilots interacted with the CSD's Route Assessment Tool (RAT) to modify their flight plans for weather and traffic avoidance. The pilots either datalinked their modified flight plan to the ground automation and if needed the controller for review and approval, or executed it without review (depending on concept). A more detailed description of the RAT and an explanation of the Multi-Aircraft Control System (MACS) software can be found in Dao et al. (in preparation). 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 conflict avoidance delegated to the pilot, controller, or Auto-Resolver. Participants completed 4 trials per day over 3 days. On each day, one concept was tested. Two trials were repeated on day 4 due to software malfunctions. Each trial lasted about 80 minutes. Classroom training and practice trials were provided prior to the test days.

All pilots flew 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 Route Assesment Tool (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 agent before maneuvering for traffic. ATCo managed traffic based on the concept of operations, and provided re-sequencing instructions on request from the pilots. In addition, ATCo responded to datalinked requests for route modifications and other requests made over the radio.

Throughout the scenarios, pilots and controllers received query prompts to measure situation awareness or workload at three minute intervals. ATWIT 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 so that the participants could not predict when the workload queries would be administered. The participants were instructed that these prompts should not interrupt their primary pilot roles and responsibilities, and not to answer the prompt until after their primary duties were performed. If after one minute no response was made to either the ready prompt or query, the prompt/query 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 (or soon after) deviating for weather and spacing re-sequencing (27 min), at top of descent or beginning the descent phase of flight (45 min), and during the approach phase of flight into Louisville, Standiford Field (SDF) (69 min). The simulation did not pause while the participants were answering the queries.

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 (1 indicates low and 5 indicates 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 analysis included the four workload ratings (overall and peak for each of the pilot task) on each trial. See Table 1 for overall means and standard deviations.

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

n = 8	En-route Spacing	Weather Avoidance	Continuous Descent Approach	Arrival Spacing
Overall	1.87 (0.78)	1.74 (0.62)	1.87 (0.83)	2.09 (0.82)
Concept				
(1) Pilot Primary	1.69 (0.59)	1.54 (0.51)	1.74 (0.75)	1.85 (0.56)
(2) Controller Primary	1.78 (0.68)	1.75 (0.55)	1.61 (0.69)	2.11 (0.75)
(3) Auto-Rslvr Primary	2.14 (0.96)	1.92 (0.73)	2.26 (0.92)	2.29 (1.05)

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 working ratings in the CDA task ($F(2,10) = 5.91$, $p = .02$). Bonferroni post-hoc analyses revealed significant differences between concept 2 (controller responsible) and concept 3 (automation responsible). Interestingly, the difference between concept 1 (pilot responsible) and concept 2 (controller responsible) was not significant.

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 as an outlier 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

n = 8	Ready Latency	Query Latency	Workload Rating
Overall	5.36 s (3.51)	3.25 s (1.41)	1.76 (0.55)
Concept			
(1) Pilot Primary	4.47 s (2.29)	3.28 s (1.35)	1.70 (0.40)
(2) Controller Primary	5.42 s (4.89)	3.11 s (1.45)	1.78 (0.56)
(3) Auto-Resolver Primary	6.18 s (3.33)	3.36 s (1.44)	1.81 (0.68)

The number of timeouts signified a non-response 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 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. Most pilots reported 1 or 2 as their workload rating regardless of concept.

Figure 1: Frequencies of Workload Ratings
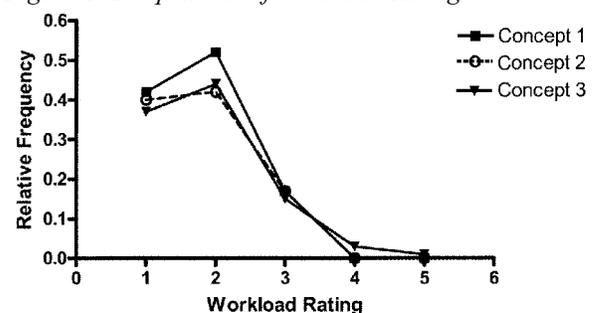
To examine assumptions of the modified ATWIT technique, Pearson's correlations between the pilots' ready latency, query latency and workload rating were computed, as shown in Table 3. Workload ratings were significantly correlated with ready latency, ($r(366)=.25$, $p<.001$) and not with query latency, ($r(366)=.05$, $p=.35$). The finding that ready latencies and subjective responses are correlated, but query latencies and subjective responses were not, suggests that the Ready Response Latency is a valid measure of workload. Workload ratings

were also correlated with 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, presented 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
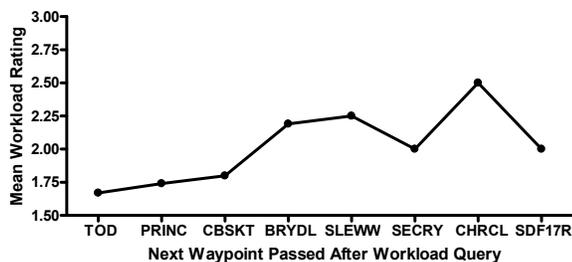
In Figure 2, it is evident that pilot workload increased throughout the descent phase, and reached a peak when queried before the CHRCL waypoint. The pilot's increased workload might be due to energy management performance during the arrival/approach phase of flight. Altitude and speed restrictions needed to be met at each waypoint while maintaining the specified spacing interval from their lead aircraft.

CONCLUSION

The prediction of pilots' workload increasing with an increase in roles and responsibilities was not supported. The trends found in multiple workload metrics suggest that workload was lowest when pilots had responsibility for avoiding traffic conflicts, and was significantly impacted by the energy management task during the constant descent arrival. Additionally, in *all* of the subjective measures, the trend for workload was highest in the automation responsible concept. This suggests that pilot workload is driven by gaining situation awareness and less driven by conflict avoidance responsibility. Furthermore, this study suggests that when conflict detection and resolution tools are available, workload did not increase, even with greater responsibility. These results also suggest that a higher traffic load did not have a significant effect upon the workload of pilots.

researchers. These researchers, located at CSU Long Beach, CSU Northridge, and Purdue, provided pseudopilots and ghost 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. (in preparation). The impact of air and ground based automated aircraft separation assistance on individual and shared situation awareness. Manuscript in preparation for publication

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.