Effects of Modes of Cockpit Automation on Pilot Performance and Workload in a Next Generation Flight Concept of Operation

Guk-Ho Gil,[1] David Kaber,[1] Karl Kaufmann,[1] and Sang-Hwan Kim[2]

1 Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, North Carolina
2 Department of Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn, Dearborn, Michigan

Abstract

The objective of this study was to compare the effects of various forms of advanced cockpit automation for flight planning on pilot performance and workload under a futuristic concept of operation. A lab experiment was conducted in which airline pilots flew simulated tailored arrivals to an airport using three modes of automation (MOAs), including a control-display unit (CDU) to the aircraft flight management system, an enhanced CDU (CDU+), and a continuous descent approach (CDA) tool. The arrival scenario required replanning to avoid convective activity and was constrained by a minimum fuel requirement at the initial approach fix. The CDU and CDU+ modes allowed for point-by-point path planning or selection among multiple standard arrivals, respectively. The CDA mode completely automated the route replanning for pilots. It was expected that the higher-level automation would significantly reduce pilot workload and improve overall flight performance. In general, results indicated that the MOAs influenced pilot performance and workload responses according to hypotheses. This study provides new knowledge about the relationship of cockpit automation and interface features with pilot performance and workload in a novel next generation–style flight concept of operation. © 2012 Wiley Periodicals, Inc.

1. INTRODUCTION

In a recent report on fatalities in aviation accidents, loss of control in flight was found to explain 65% of all accidents (Boeing, 2008). In a number of historical cases, loss of control has been attributed in part to pilot "out-of-the-loop" performance problems and loss of situation awareness due to high-level cockpit automation.

For example, regarding the American Airlines Flight 965 incident in Cali, Colombia, Endsley and Strauch (1997) reported that features of the automated aircraft flight management system (FMS) contributed to pilot error. Specifically, pilots were not aware that deletion of waypoints from a preprogrammed flight path using a multicontrol display unit interface caused vertical path controls to be disabled. This ultimately contributed to the crash.

In general, when a pilot must replan a route because of either non-nominal (e.g., runway closure or change, weather disturbance) or emergency (e.g., cargo fire, medical emergency, etc.) circumstances (Kalambi et al., 2007), there are required functions, including systems monitoring, generating decision alternatives, and selecting and implementing options. The functions

actually performed by the pilot depend on the level of aircraft automation. Levels of automation in human-in-the-loop systems define function allocation schemes for the human and machine (Endsley & Kaber, 1999). Pilots using low-level cockpit automation may need to perform route selection or implementation functions as compared to using high-level automation, where they may need to perform only route generation and monitoring functions. In general, lower level automation may increase pilot workload, but high-level automation may result in a loss of pilot awareness of the states of the aircraft or airspace, particularly during low workload phases of flight. A loss of pilot situation awareness in replanning can lead to errors and performance problems.

In a reroute scenario, the flight path of an aircraft is typically extended, and pilots must effectively monitor aircraft position changes to prevent flight into terrain, fuel consumption to ensure sufficient levels for flight before landing, and weather conditions to prevent damage to the vehicle. A loss of situation awareness in a reroute situation due to the occurrence of non-nominal or emergency conditions can also increase pilot workload as a result of a need for pilots to quickly reorient to the state of the automation and then to begin to process information (Sarter & Woods, 1995).

Previous research (Chen & Pritchett, 2001; Kaber et al., 2009; Kalambi et al., 2007; Wright, Kaber, & Endsley, 2003) has investigated the effect of various forms of cockpit automation on pilot performance and workload in replanning scenarios. The use of high-level automation for flight path control (e.g., the FMS) has been found to support superior piloting performance in advance of critical events, such as an approach revision (Chen & Pritchett, 2001; Wright et al., 2003). However, the use of low-level automation, engaging pilots to a greater extent in aircraft control loops, has been found to support superior pilot situation awareness and performance in dealing with an event, like an approach revision. Related to this, Chen and Pritchett (2001) investigated the effect of different levels of automation on generating and predicting a detailed trajectory to the runway threshold in a simulated emergency situation. Under the highest level of automation, pilots exhibited automation bias (i.e., tending to follow the automatically generated trajectory because of high workload and time pressure), even when the tool generated erroneous trajectories (Chen & Pritchett, 2001). Similarly, Wright et al. (2003) observed that pilots using high-level automation placed less emphasis on the

accuracy of flight planning in advance of critical events and greater reliance on data stored in the FMS and the automation capability. In a related study, Kaber et al. (2009) reported that when pilots used high-level cockpit automaton (the FMS) in advance of receiving a clearance for an alternate runway in a simulator-based landing scenario, they were more likely to forget critical procedures (e.g., backcourse for opposing instrument landing system runways) and to make errors in flight path control, as compared with when they used low-level automation. However, Kaber et al. (2009) also found that the use of high-level automation led to significantly lower objective and subjective pilot workload responses. Still other studies have demonstrated the same workload advantage of high-level (FMS-based) cockpit automation in simulated critical event (flight replanning) scenarios (Kalambi et al., 2007). It is important to note here that, in some investigations of the effects of levels of cockpit automation on pilot performance, higher levels of automation have also provided pilots with less complex and more usable interfaces as compared with modes of low-level automation, which may also contribute to reduced workload experiences.

Consequently, the design of contemporary advanced aircraft automation has been focused on the opportunity for reducing pilot workload but, at the same time, attempting to prevent the potential for loss of situation awareness in high-workload phases of flight, including airport approach and landing. Continuous descent approach (CDA) tools provide pilots with the capability to define an airport arrival at an increased distance from touchdown without the typical requirement of altitude "step-downs" and with the use of low continuous power throughout the descent (Coppenbarger, Mead, & Sweet, 2007). Coppenbarger et al. (2007) investigated the Tailored Arrival (TA) concept for enabling continuous descents under constrained airspace conditions with a datalink (DL)-enabled CDA. Continuous descent arrivals are "tailored" in the sense that the arrival is constructed for the pilot to take into account air traffic control (ATC) constraints, such as traffic patterns, airspace restrictions, and minimum altitudes, as well as the performance of the specific aircraft type. A clearance for a TA can be given well in advance of the point at which an aircraft begins its descent because routes are predetermined and have been coordinated across ATC facilities. From the pilot's perspective, the TA means far fewer flight tasks and actions at cockpit interfaces. It also allows for more time to deal with any off-nominal events that might develop. Beyond this,

TA clearances are generally given by DL (electronic commutation between a control center and the aircraft) to further reduce flight crew and ATC workload. (Although pilot performance effects of linkages of DL to the FMS of an aircraft are also important to understand in such scenarios, the focus of the present work was on assessment of the impact of various forms of cockpit automation for route replanning and not for conveying information from ATC or a CDA-like tool to the FMS.) In general, the TA scenario requires a high-level of aircraft and ATC automation for effective performance.

Very recently, Dao et al. (2010) investigated the effects of automated spacing tools during a CDA on pilot performance, including altitude, speed, and time-in-trail. However, no previous studies have empirically assessed the impact of CDA tool use on both pilot workload and performance (based on flight situation awareness) in comparison with conventional forms of cockpit automation, including the FMS. It is important to note that other prior investigations in the driving and military domains have used and evaluated measures of operator performance as a basis for non-intrusiveness capture and analysis of situation awareness (Kaber et al., 2005; Ward, 2000). The objective of this study was to compare the effects of various types and levels of cockpit automation for flight planning, including a CDA tool, on pilot workload, time-to-task completion (TTC), and decision making in a next-generation (NextGen)-style TA that followed the CDA model. NextGen is the Federal Aviation Administration (FAA) and National Aeronautics and Space Administration (NASA) acronym for referring to technologies and procedures to be developed as part of the Next Generation Air Transportation System with an implementation timeline extending approximately to the year 2020.

2. METHODS

2.1. Flight Simulator Setup

A high-end, PC-based flight simulator (see Figure 1), integrating X-Plane simulation software, was set up to present a B767-300ER cockpit with interfaces and functions of existing and futuristic forms of cockpit automation, including a control-display unit (CDU) for the FMS, an enhanced CDU (or CDU+) with predetermined route selection capability, and a CDA tool. Each type of automation was integrated with an ATC DL system for electronic relay of clearances be-
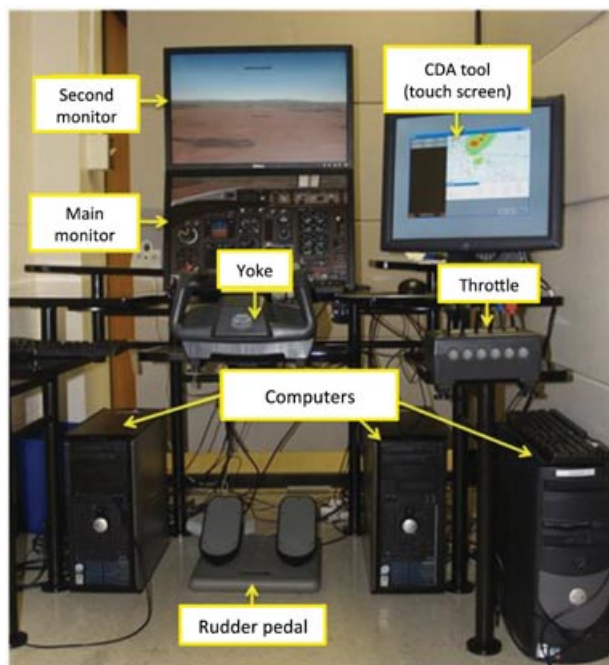

Figure 1 Flight simulator setup.

tween the simulated cockpit and control center. In addition to three MS Windows–based PCs, the simulator setup consisted of a cockpit workstation with flight deck controls, including a yoke, a throttle quadrant, and rudder pedals (see Figure 1). Two LCD monitors were arranged vertically with the lower display presenting the instrument panel for the B767-300ER, including the PFD (primary flight display), FCP (flight control panel or glare shield), and FMS with the CDU interface. The upper display presented an out-of-cockpit view. The X-Plane simulator software supported presentation of dynamic flight situations with high-fidelity graphics. A touch screen monitor was used to present the new CDA planning tool (see right side of Figure 1). Pilots used the screen to directly manipulate graphical buttons and select route waypoints or entire routes determined by the automation. The display content of the monitors was synchronized using a TCP/IP (Transmission Control Protocol/Internet Protocol) network supported by X-Plane.

2.2. CDA Tool Development

The CDA tool is a contemporary form of cockpit automation originally retrofitted in the UPS Boeing 757 and ABX Boeing 767 aircraft. Unfortunately, the current version of the X-Plane simulation software does not provide a CDA interface. For this reason, a custom
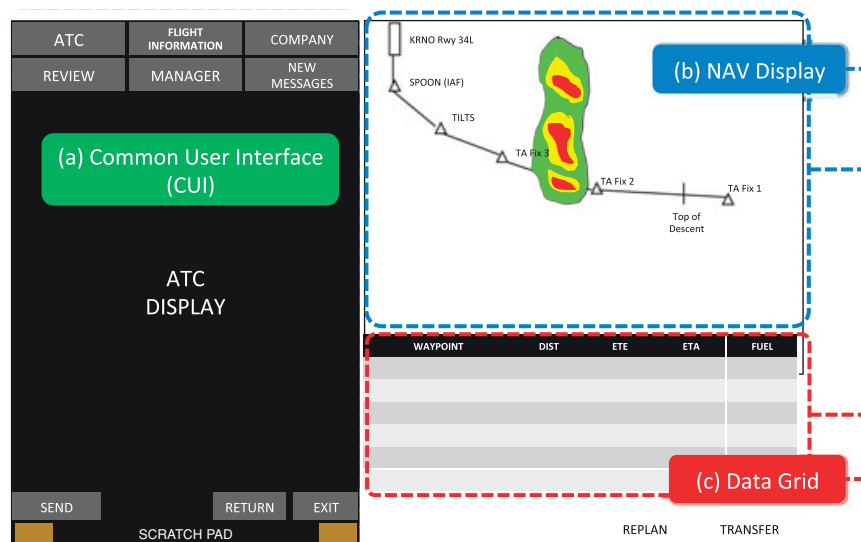
Figure 2   CDA tool interface design.

CDA tool was developed and integrated with the flight simulator. It includes a common user interface (CUI, see Figure 2a), a navigation display (ND, see Figure 2b), and a data grid (DG, see Figure 2c). In the experiment, pilots interacted with the CUI for receiving/requesting ATC clearances. The ND also provided weather information with color coding (e.g., red = severe weather conditions; blue = mild weather conditions). The CUI and ND are located in the middle of the cockpit display area. The DG is located below the CUI and ND. When a pilot uses the CDA to confirm a new route to avoid severe weather conditions, address fuel constraints, and so forth, the CDA sends route information to the X-Plane software. X-Plane then loads the new route into the CDU/FMS. In this way, the CDA tool provides pilots with the capability for a TA.

2.3.   Scenario Development

A realistic flight scenario was developed to support the objective of empirically assessing the impact of cockpit automation on pilot performance under high- and low-workload conditions. The Reno-Tahoe International Airport (KRNO) was selected as the termination of an approach with reroute followed by landing. The airport is located in a valley with terrain rising to approximately 6,000 feet above the field, making arrivals challenging for pilots even when using cockpit automation. The extreme terrain makes maintaining the published flight path more critical than when operating at most other airports and increases the utility of automation for pilots.

The scenario required a pilot to use all features of the CDA tool, including the ND to perceive and plot a route around a thunderstorm, the DG to determine fuel consumption for the reroute, and the CUI and DL to communicate with ATC for clearance of the new route. After confirming the new flight path with ATC, the pilot needed to perform different types of actions to change the flight path.

The scenario began with the simulated aircraft flying westbound toward a radio ground station approximately 150 nm (nautical miles) east of the airport [the Wilson Creek VOR (Very High Frequency/VHR Omni-directional Radio-beacon)]. The pilot received a DL clearance for the TA from ATC, with the route appearing automatically on the ATC uplink display. The pilot reviewed the route focusing on whether it was flyable under the constraints of weather and fuel consumption limits. (In all trials, the first clearance was to be rejected based on weather conflicts or fuel constraint violations. The pilot then had to replan the route and seek approval from ATC.) When an acceptable clearance was received, the pilot confirmed this via the DL using a button on the planning tool touchscreen. The TA route was then loaded into the CDU/FMS either by data connection between the DL system and FMS (for the CDA tool or CDU+ mode) or manually (for CDU mode).

2.4.   Independent Variables

There were three independent variables, including the mode of automation (MOA); the level of flight workload, manipulated in terms of the time available to

perform the reroute planning; and the phase of flight involving planning and decision making or flight plan implementation. The MOAs were CDU, CDU+, and CDA. The automation under all modes always presented feasible flight path waypoints or routes; however, other route recommendations might have violated specified flight constraints. Therefore, under each MOA, pilots needed to maintain their flight situation awareness and carefully review the automation recommendations for acceptability in terms of flight constraints. (We say more about the constraints later.)

The CDU was considered to be the lowest level of automation presenting pilots with all possible route waypoints for selection. Most combinations did not meet the flight scenario constraints, including the level of fuel consumption and weather to be avoided. After selecting waypoints, a pilot had to manually calculate fuel consumption to determine whether the aircraft would have the required amount of fuel when crossing the initial approach fix (IAF). Finally, the pilot had to manually program the new route into the CDU/FMS. That is, the automation assisted pilots only in terms of identifying the airspace and available waypoints and implementing the flight plan. The information-processing functions of generating and selecting a route were left to the pilot. The CUI of the CDA interface presented the confirmed waypoints on the selected route. Figure 3 shows the CDA and CDU/FMS interfaces for the CDU MOA.

The CDU+ MOA was considered to be an intermediate level of automation in which the CDA tool presented pilots with a choice of several alternative routes (instead of waypoints). Pilots selected one of the alternatives but were not required to calculate fuel consumption. The DG as part of the CDA interface presented the fuel consumption rates for route segments; however, pilots had to verify the level at the IAF when selecting a route. When the pilot decided on a route, he or she had to "load" the route into the CDU/FMS using the "RTE" (route) page of the CDU. That is, the automation supported pilots in perceiving the airspace and waypoints, generating alternative routes, and implementing the flight plan. The function of route selection was left to the pilot, as in the CDU mode. Figure 4 shows the CDA and CDU/FMS interfaces for the CDU+ MOA. In Figure 4a, the ND presents all the alternative route information. Figure 4b shows the "RTE" page of the CDU in which a pilot clicked the "SUGGEST RTE" option in the interface.

The CDA MOA was considered to be the highest level of automation and presented pilots with the best route to avoid severe weather and to meet the fuel constraint. This mode provided fully automated reroute capability as well as automatic loading of route waypoints into the CDU/FMS. When a pilot confirmed a suggested route, the CDA tool sent the route information to the FMS. The pilot was only required to click the "EXEC" (Execute) button on the CDU/FMS interface. That is, the automation supported pilots in monitoring the state of the airspace, generating flight path alternatives, selecting an optimal path alternative, and implementing the flight path. Pilots were only required to verify the actions of the automation. Figure 5 shows the CDA and CDU/FMS interfaces for the CDA MOA. In Figure 5a,

(a) CDA Tool   (b) CDU Interface

Figure 3   CDA and CDU/FMS interfaces under the CDU MOA.

Figure 4   CDA and CDU/FMS interfaces under the CDU+ MOA.

Figure 5   CDA and CDU/FMS interfaces under the CDA MOA.

the ND presents the suggested route, which meets all flight constraints. Figure 5b shows the "LEGS" page in the CDU, including the "EXEC" button used to invoke the route.

Flight workload was determined based on the distance between the starting position of the aircraft and the reroute decision point. Pilots were required to complete the reroute before reaching the Wilson Creek VOR, which is the first fix on the arrival. For the CDU MOA, the low- and high-workload conditions provided 80 and 60 nm of travel, respectively. For the CDU+ MOA, the low- and high-workload conditions provided 27 and 20 nm of travel, respectively. For the CDA MOA, the low- and high-workload conditions provided 13 and 10 nm of travel, respectively. Those distances and times that were provided were based on

the expert pilot input from preliminary experiment trials.

The decision phase of the flight involved pilot use of the CDA tool and ran from the start of the trial until acceptance of an arrival clearance. The implementation phase involved programming the FMS and ran from the start of pilot use of the CDU/FMS until the "EXEC" button was pressed to activate a route. Finally, total time to completion spanned from the beginning of a trial until the pilot pressed the "EXEC" button on the CDU/FMS.

Finally, there were three sequential blocks of trials (SET1, SET2, and SET3) in which all three modes of automation were presented. The order of presentation of the modes differed from block to block (e.g., SET1: CDU → CDU+ → CDA, SET2: CDA →

CDU+ → CDU) and was used to assess pilot learning effects.

2.5. Dependent Variables

TTC. The decision phase time, implementation phase time, and total time were recorded for each trial. This yielded three data sets for analysis.

Modified Cooper-Harper (MCH) Workload Rating. Previous research (Kaber et al., 2009) has used the NASA-TLX (Task Load Index; Hart & Staveland, 1988) as a subjective measure of pilot workload as influenced by level of cockpit automation. However, results revealed some marginal workload effects of MOA, and pilots were confused by ratings of six different workload demand components (i.e., mental demand, physical demand, temporal demand, performance, effort, and frustration) without direct reference to cockpit interfaces or aircraft controls. Although the NASA-TLX is considered to be a robust measure of subjective workload (Moroney, Biers, & Eggemeier, 1995), the MCH workload rating scale was used as it follows a decision-tree approach to extract pilot workload experiences and make direct reference to cockpit interfaces. The original "Cooper Harper assessment tool" was used to evaluate physical workload associated with use of aircraft controls (Harper & Cooper, 1986). The modified version was used in the present experiment to evaluate how well the displays supported pilot information processing and the general cognitive load experienced in the different flight segments (Cummings, Myers, & Scott, 2006).

Heart Rate (HR). To obtain an objective measure of pilot workload, pilot cardiovascular activity was recorded during each phase of flight. A Polar S810i HR monitoring system was used for this purpose. HR data were recorded every 2 seconds, and an average value was calculated for each phase.

After all experiment trials, pilots were instructed to relax without performing any tasks for approximately 10 minutes to measure their baseline HR. (Postexperiment baselines have been demonstrated to be more reliable than preexperiment baselines [Kaber, Perry, Segall, & Sheik-Nainar, 2007] because of subject anxiety in advance of testing.) During data analysis, the percent change in HR for test conditions, as compared to the baseline, was calculated using the following formula:

$$\frac{TestHr - BaselineHR}{BaselineHR} * 100.$$

Trial Success/Failure. At the end of each trial, the experimenters recorded whether the pilot completed route replanning by the first waypoint on the arrival.

2.6. Hypotheses

CDA trials were expected to produce the lowest workload response in terms of MCH and HR (Hypothesis (**H**)**1**). Whereas the CDU MOA required decisions on individual waypoints and manual programming of the FMS, the CDA MOA only required pilots to confirm a suggested route. With respect to TTC, the CDA MOA was expected to produce the best performance (shortest time) because it provided maximum automation (**H2**). When using the CDU MOA, pilots were expected to take longer to implement a new route because of manual calculations of fuel consumption. Last, the CDA MOA was expected to produce the highest performance in terms of task success rate (**H3**) because it provided the greatest assistance with information-processing functions (monitoring, generating, and selecting an optimal route). During CDU trials, pilots were required to perform many information-processing functions, including independently deciding on the most appropriate combination of waypoints to achieve safety and fuel constraints. Consequently, this mode was also expected to reduce the task success rate because of the number of functions in which pilots could make errors.

2.7. Design of Experiment, Participants, and Procedure

A split-plot design was followed in the experiment including: (1) the data collection block; (2) trial set; (3) MOA; (4) order of presentation of MOAs within a trial set; and (4) workload condition. The block represented a complete crossing of all the settings of the main effects (MOA and workload) for four subjects. The design included three blocks in total. Each subject performed three sets of trials (each set included three MOAs, CDU, CDU+, and CDA). The trial set variable represented the time period within a block.

Thirteen pilots were recruited for the experiment from several sources in the Raleigh, North Carolina, area including the local Airline Pilots Association chapter and the Experimental Pilots Association. An instrument rating and flight experience in aircraft equipped with an FMS were minimum requirements for participation. Eight subjects had FAR (Federal Aviation Regulations) Part 121 (scheduled airline) experience. Five subjects had FAR Part 135 (charter)
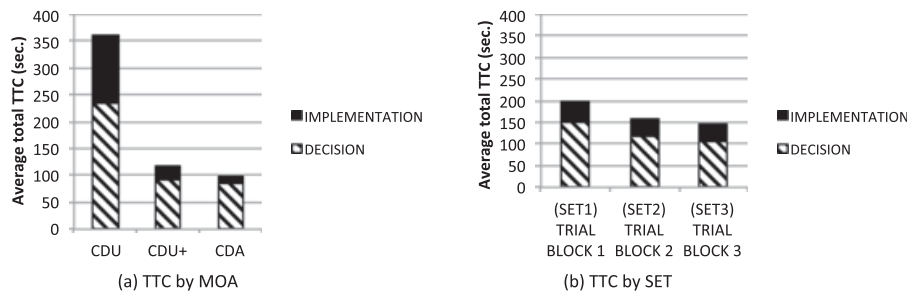
Figure 6   Average total TTC by MOA and SET.

or Part 91 (corporate) experience. Pilot age ranged from 25 to 72 years with an average of 51 years. Total instrument time ranged from 175 hours to 20,000 hours with an average of 3,706 hours. One participant's data were excluded from analysis, as he did not follow the MCH rating instructions.

After a brief survey session, pilots were familiarized with the simulator and the experimental setup. They trained on all three MOAs (CDA, CDU+, and CDU) following a TA into Raleigh-Durham International Airport (KRDU). The basic training regimen was defined by an expert Air Force transport pilot with ATP (airline transport pilot) certification, who had conducted check rides for 17 years. Pilots with "glass cockpit" experience (CRT-based displays and use of an FMS) were considered well qualified to participate in the experiment; however, all were provided with training under all MOAs until they felt comfortable with performing the TA. Subsequently, pilots were instructed on how to complete the MCH rating form, and they donned the HR monitor. During test trials, each participant followed the flight scenario under each of the three different MOAs in the three sequential trial blocks. At the close of each trial, the pilots were asked to complete the MCH scale. At the end of the experiment, and after the baseline HR recording, pilots were debriefed and they received a $100 honorarium.

3. RESULTS

Analyses of variance (ANOVAs) were conducted to identify effects of MOA, trial SET, and the flight workload (WL) condition on pilot TTC and HR within and across phases of flight. Because the age range of pilots in the study was substantial, an analysis of covariance (ANCOVA) was also conducted to assess age as a covariate with MOA and WL on TTC, HR, and MCH. Post-hoc tests were also conducted to compare the mean responses among the levels of MOA, SET, and

WL. Finally, contingency tables were used to examine the success rate (a binary response) and the frequency of MCH ratings under the three MOAs.

3.1. TTC

Figure 6 presents the mean TTC values for the various settings of the significant independent variables. ANOVA results revealed a significant effect of MOA ($F(2,64) = 454.15$, $p < .0001$) on total TTC. Post-hoc tests using Tukey's tests indicated the CDU MOA produced the longest TTC, whereas the CDA MOA required the shortest TTC. There was also a significant effect of SET ($F(2,64) = 21.27$, $p < .0001$) on pilot learning. Post-hoc tests indicated that the TTC for SET1 was significantly longer than the TTCs for SET2 and SET3. Simple effects analysis revealed total TTC to significantly vary among sets under the CDU MOA ($F(2,33) = 5.808$, $p = .007$). Tukey's tests showed that SET1 took longer than SET2 and SET3, whereas there was no difference between SET2 and SET3. Total TTC also significantly differed among sets under the CDU+ MOA ($F(2,33) = 4.124$, $p = .025$). Tukey's tests indicated that SET1 took longer than SET3, whereas there was no difference between SET1 and SET2 or SET2 and SET3. With respect to the CDA MOA, there was no significant interaction with SET. There was no significant main effect of the WL manipulation on total TTC (or decision or implementation phase TTC) across all trial sets.

The decision phase TTC also revealed a significant effect of MOA ($F(2,64) = 200.96$, $p < .0001$) on pilot route planning efficiency using the CDA tool. Tukey's tests indicated that the CDU MOA required more decision time than the CDU+ and CDA MOAs. There was also a significant effect of SET ($F(2,64) = 18.80$, $p < .0001$) on decision time. Tukey's tests indicated that there was pilot learning from SET1 to SET2 and SET3, but there was no difference in decision time between SET2 and SET3.

There was a highly significant effect of MOA ($F(2,64) = 1139.07$, $p < 0.0001$) on pilot time to implementation of a reroute. Tukey's tests indicated that the CDU MOA required the longest time, whereas the CDA MOA took the shortest time to implement. This particular finding was expected in that task time differences among the MOAs could, to some extent, be attributed to the nature of the automation interface with the CDU requiring more button presses and mouse clicks than the CDA mode. (The results on success rate [presented later in this article] reveal differences in pilot ability to achieve a reroute when considering enhanced automation capability and increased interface usability, as assessed by the expert pilot in advance of test trials.) There was also a significant effect of SET ($F(2,64) = 9.23$, $p = .0003$), with trials in SET1 having a longer implementation phase than those in SET2 and SET3, according to Tukey's test.

Regarding the results of the ANCOVA, there was no significant effect of age ($F(1,7) = 3.36$, $p = .1094$) and total instrument time ($F(1,7) = 0.07$, $p = .7993$) on pilot performance. The results of the ANCOVA were consistent with those of the ANOVA.

3.2. HR

ANOVA results revealed a significant effect of MOA ($F(2,63) = 6.33$, $p = .0031$) on pilot HR (see Figure 7a). Post-hoc tests using Duncan's method indicated that the CDU MOA produced a higher workload response (increase in HR over baseline) than the CDA MOA. There was also a significant effect of SET ($F(2,63) = 6.18$, $p < .0035$) on cardiovascular activity. Duncan's tests indicated that pilots experienced greater arousal during SET1 than SET2 and SET3 (see Figure 7b). Pilots likely experienced some anxiety in adapting to the experiment conditions. In general, there was no significant effect of the WL manipulations on the cardiac response for any time window across all trial sets. Related to this finding, under some experimental circumstances, a single process-oriented measure of workload, such as HR, may not be sensitive or accurate in revealing automation condition manipulations. Rye and Myung (2004) offered that a battery of objective and subjective measures is necessary to assess cognitive workload in multitask conditions where automation is applied. (The results of performance-based and subjective measures of the workload manipulation are reported in the subsequent sections.)

During the decision phase of test trials, there was a significant effect of MOA ($F(2,63) = 7.04$, $p = .0017$) on pilot HR. Duncan's tests indicated that the CDU MOA produced a greater increase in HR compared to the CDA MOA. There was also a significant effect of SET ($F(2,63) = 7.23$, $p = .0015$). Duncan's tests indicated that the decision phase HR for SET1 was different from that for SET2 and SET3. However, there were no significant effects of MOA or SET in the implementation phase of flight.

ANCOVA. There was no significant effect of age ($F(1,7) = 0.43$, $p = .5322$) and total instrument time ($F(1,7) = 0.16$, $p = .7048$) on pilot performance across MOAs. The results of the ANCOVA were consistent with those of the ANOVA.

3.3. Success/Failure of Trials

An additional contingency table analysis was conducted on the success/failure response. This measure was important because it essentially identified the effectiveness of the MOA for supporting pilot decision making in rerouting, while accounting for differences in usability among the interfaces as perceived by the expert pilot and the potential for pilots to achieve reduced planning and implementation times. Results revealed a significant effect of the MOA on success/failure rates ($\chi^2 = 21.70$, $p < .0001$) under the high task WL condition (see Figure 8b). This was not the case for the low WL condition. As can be seen in Figure 8b, there was no difference in success rates between the CDU and
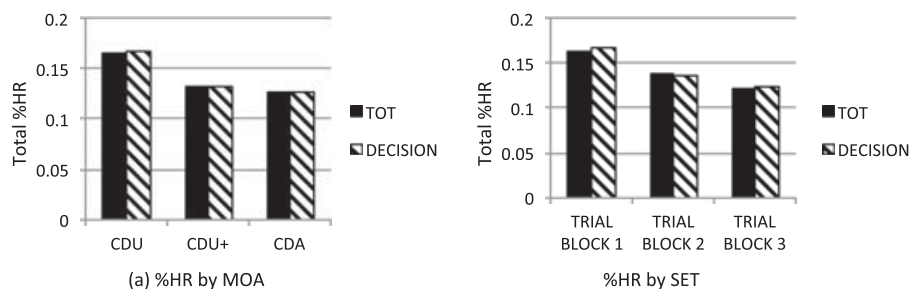

(a) %HR by MOA

%HR by SET

Figure 7 Total %HR by MOA and SET.
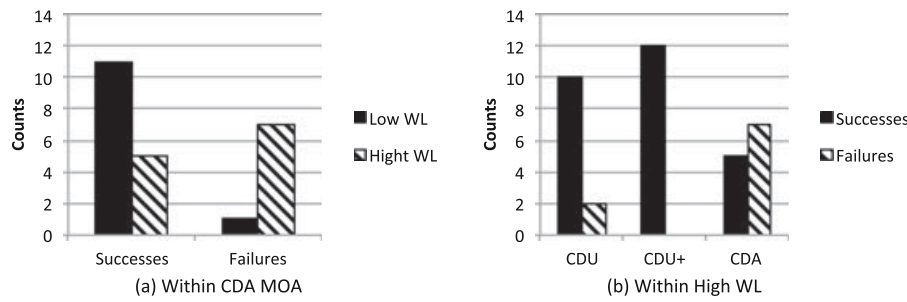
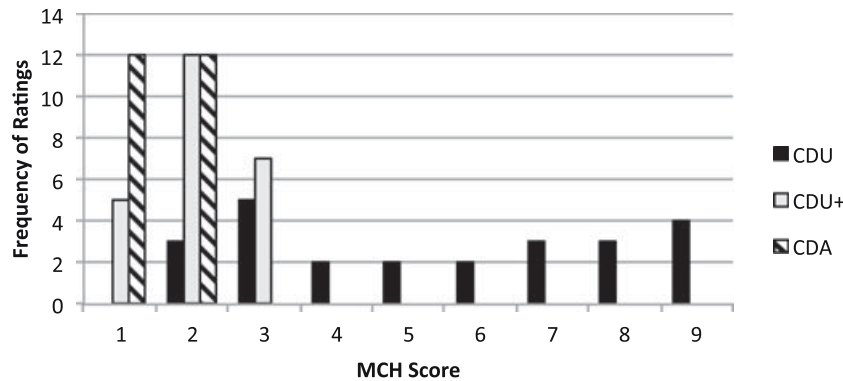**Figure 8**   Number of successes/failures within MOA and among WL settings.

Figure 9   MCH frequency of ratings.

CDU+ MOAs; however, both of these MOAs produced more successes than the CDA MOA. There was also a significant effect of the WL manipulation ($\chi^2 = 7.368$, $p = .007$) under the CDA MOA condition. As shown in Figure 8a, the low WL condition produced higher success rates than the high WL condition. Therefore, the outcome measure reflecting overall pilot performance was sensitive to the workload manipulation.

In general, when the pilots had more time available (lower temporal demand) for the decision under the CDU and CDU+ MOAs and for execution, the success rate was higher than under the CDA MOA, when less time was available (greater temporal demand). That said, reduced task times were allotted for the CDA mode because of the level of automation and the refined automation interface. These results reflect the importance of using a battery of measures to reveal workload effects that may be present in pilot–automation interaction.

3.4. MCH Ratings

The subjective workload rating data were analyzed using a contingency table by counting the frequency of ratings (0–9) under each MOA. Fisher's exact test was used to analyze the contingency table results because

of the relatively small sample size. Test results revealed mean pilot perceived workload to significantly differ among all MOAs (i.e., CDU vs. CDU+, $p < .001$; CDU vs. CDA, $p < .001$; and CDU+ vs. CDA, $p = .001$). These findings were in line with the findings on the HR response. Figure 9 presents the frequency of MCH ratings across MCH scores (1–9). Like the HR measure, the MCH ratings also appeared insensitive to the workload (task temporal demand) manipulation. It is possible that the MCH measure was less sensitive than the performance-based assessment of workload because of pilot recall bias regarding the timing of events in the simulation. This issue is common with subjective measures of workload. However, the lack of significance of the physiological and subjective assessments of the workload manipulation did not negate the effect revealed through pilot performance.

4. DISCUSSION AND CONCLUSIONS

With respect to the hypotheses on pilot workload and performance under the three different forms of cockpit automation, the use of low-level automation (the CDU) led to significantly higher pilot workload (HR and MCH) and longer TTCs when compared with the CDU+ and CDA modes. These results supported our

first and second hypotheses (**H1** and **H2**) and suggested that the CDU MOA supported lower situation awareness than the CDU+ and CDA modes with automated and optimal route generation, respectively.

Contrary to our third hypothesis (**H3**), the use of high-level automation (the CDA tool) led to the lowest success rate (correct route selections before the start of the arrival), whereas the use of intermediate-level automation (CDU+) led to the highest success rate. This result was likely due to an overly rigid success criterion defined for the CDA mode as compared with the CDU and CDU+ modes. The CDA mode trials allotted pilots a disproportionately short time (13 nm and 10 nm under low- and high-workload conditions, respectively) to accomplish the decision task, as compared with the CDU+ mode (27 nm and 20 nm under low- and high-workload conditions, respectively). Here, it is important to note that the criterion distances used to represent the specific workload settings under each MOA were set based on the judgment of an expert military transport pilot, who had some experience in using the CDA tool. The pilot's assessment of what was possible with the new MOA likely exceeded the capability of the line pilots who participated in the experiment and had limited training on the technology.

The CDU and CDU+ modes of automation revealed learning effects through the early test trials, in terms of TTC. The CDA mode, however, did not reveal a significant learning effect. This finding suggested that achieving SA and performance with the low and intermediate MOAs in a reroute scenario took a longer time than when using the new CDA MOA. Furthermore, the results suggested that pilots might be able to quickly adapt to the CDA form of automation for flight replanning tasks in NextGen operations.

There was no significant effect of the starting position (flight workload) manipulation on pilot TTC or workload responses (HR and MCH) across all modes of automation. In general, results indicated that the modes of cockpit automation influenced pilot performance and workload responses according to the hypotheses. Some caveats of the investigation include the use of a PC-based flight simulator. Use of a higher fidelity, motion-based flight simulator might promote the realism of the scenario for pilots, increase the sense of urgency in the reroute decision, and reduce pilot TTC. A motion-based simulator would also provide kinesthetic cues on aircraft states, an additional information source for pilot decision making, potentially reducing cognitive load. In general, there is a need

for experiments, such as the present, to achieve high-fidelity simulations relative to actual flight operations, including accurately simulating aircraft responses to control inputs and external forces, to promote the generalizability of results for system design purposes.

Another limitation of the study was that pilot SA was inferred based on performance with the various modes of automation. Other studies in the driving domain have used such an approach for analyzing driver SA (Ward, 2000). Although performance measures are easily integrated with simulator technologies, and provide objectivity and unobtrusiveness in assessing pilot SA, they are limited in terms of diagnosticity and sensitivity. Factors beyond pilot SA may impact performance. Pilots may also simply get lucky in performing flight tasks successfully without achieving the necessary level of SA. Future research should investigate the use of a direct measure of SA for assessing pilot perception, comprehension, and projection of aircraft states under various MOAs, such as real-time SA probes (Jones & Endsley, 2004) but ensure that any measurement technique is not disruptive to performance under time-constrained NextGen concepts of operation. This study provides new knowledge on the relationship of cockpit automation and interface features with pilot performance, cognitive workload, and SA in a novel NextGen-style flight concept of operation. The results of this research may be used as a basis for developing cockpit automation interface design guidelines and models of pilot cognitive behavior to predict performance and workload with other forms of futuristic automation, like the new CDA tool, as part of the aircraft systems design process.

ACKNOWLEDGMENTS

References

Boeing (2008). Statistical summary of commercial jet airplane accidents. Available at: http://www.boeing.com

Chen, T. L., & Pritchett, A. R. (2001). Development and evaluation of a cockpit decision-aid for emergency trajectory generation. Journal of Aircraft, 38(5), 935–943.

Coppenbarger, R., Mead, R., & Sweet, D. (2007). Field evaluation of the tailored arrivals concept for datalink-enabled continuous descent approach. In Proceedings of the 7th American Institute of Aeronautics and Astronautics Conference (ATIO). Belfast, Northern Ireland.

Cummings, M. L., Myers, K., & Scott, S. D. (2006). Modified Cooper Harper evaluation tool for unmanned vehicle displays. In Proceedings of the Unmanned Vehicle Systems Canada Conference. Montebello, Canada.

Dao, A., Lachter, J., Battiste, V., Brandt, S. L., Vu, K. P., Strybel, T. Z., et al. (2010). Automated spacing support tools for interval management operations during continuous descent approaches. In Proceedings of the 54th Human Factors and Ergonomics Society (HFES) (pp. 21–25). Santa Monica, CA: HFES.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics, 42, 462–492.

Endsley, M. R., & Strauch, B. (1997). Automation and situation awareness: The accident at Cali, Colombia. In R. S. Jensen & L. Rakovan (Eds.), Proceedings of the 9th International Symposium on Aviation Psychology (pp. 877–881). Columbus, OH: Ohio State University.

Harper, R. P., & Cooper, G. E. (1986). Handling qualities and pilot evaluation. Journal of Guidance, Control, and Dynamics, 9(5), 515–529.

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. Meshkati (Eds.), Human Mental Workload (pp. 5–39). New York: Elsevier.

Jones, D. G., & Endsley, M. R. (2004). Use of real-time probes for measuring situation awareness. International Journal of Aviation Psychology, 14(4), 343–367.

Kaber, D., Perry, C., Segall, N., & Sheik-Nainar, M. (2007). Workload state classification with automation during simulated air traffic control. International Journal of Aviation Psychology, 17(4), 371–390.

Kaber, D. B., Gil, G. H., Kaufmann, K., Kim, S. H., Veil, T., & Picciano, P. (2009). Effects of modes of cockpit automation on pilot performance and workload in a high workload approach and landing scenario. Poster presentation at the 15th International Symposium on Aviation Psychology (CD-ROM). Dayton, OH: Wright State University.

Kaber, D. B., Riley, J. M., Lampton, D., & Endsley, M. R. (2005). Measuring situation awareness in a virtual urban environment for dismounted infantry training. In Proceedings of the 11th International Conference on Human-Computer Interaction (Vol. 9, "Advances in Virtual Environments Technology: Musings on Design, Evaluation, & Applications"). Mahwah, NJ: Lawrence Erlbaum & Associates.

Kalambi, V. V., Pritchett, A. R., Bruneau, D. P. K., Endsley, M. R., & Kaber, D. B. (2007). In-flight planning and intelligent pilot aids for emergencies and non-nominal flight conditions using automatically generated flight plans. In Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting (CD-ROM). Santa Monica, CA: Human Factors & Ergonomics Society.

Moroney, W. F., Biers, D. W., & Eggemeier, F. T. (1995). Some measurement and methodological considerations in the application of subjective workload and measurement techniques. International Journal of Aviation Psychology, 5, 87–106.

Rye, K., & Myung, R. (2004). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. International Journal of Industrial Ergonomics, 35, 991–1009.

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. Human Factors, 37(1), 5–19.

Ward, N. J. (2000). Automation of task processes: An example of intelligent transportation systems. Human Factors and Ergonomics in Manufacturing, 10(4), 395–408.

Wright, M. C., Kaber, D. B., & Endsley, M. R. (2003). Performance and situation awareness effects of levels of automation in an advanced commercial aircraft flight simulation. In Proceedings of the 12th International Symposium on Aviation Psychology. Dayton, OH: Wright State University, 1277–1282.