Pilots' Monitoring Strategies and Performance on Automated Flight Decks: An Empirical Study Combining Behavioral and Eye-Tracking Data

Nadine B. Sarter, Randall J. Mumaw and Christopher D. Wickens

DOI: 10.1518/001872007X196685

The online version of this article can be found at: http://hfs.sagepub.com/content/49/3/347

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
Human Factors and Ergonomics Society

Additional services and information for Human Factors: The Journal of the Human Factors and Ergonomics Society can be found at:

Email Alerts: http://hfs.sagepub.com/cgi/alerts
Subscriptions: http://hfs.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav
Citations: http://hfs.sagepub.com/content/49/3/347.refs.html

>> Version of Record - Jun 1, 2007
INTRODUCTION

Breakdowns in human-automation interaction have long been a concern in a number of complex event-driven domains, such as aviation, for which considerable empirical evidence of these problems exists in the form of pilot reports, research findings, and operational experience (e.g., Abbott et al., 1996; Eldredge, Dodd, & Mangold, 1991; Funk, Lyall, & Riley, 1996; Sarter & Woods, 1994, 1997, 2000; Sarter, Woods, & Billings, 1997; Wiener, 1989). One consequence of breakdowns in pilot-automation coordination is a loss of mode awareness (a lack of knowledge and understanding about the current and future state and behavior of the automation; see Sarter & Woods, 1994, 1995). A loss of mode awareness can lead to mode errors, which occur when a pilot performs an action appropriate for the assumed, but not the actual, system state. Mode errors, in turn, can lead to automation surprises when the pilot notices that the automation is engaged in activities that were not commanded or that were commanded unintentionally (Sarter et al., 1997). Mode errors and automation surprises can lead to accident precursors and have played a role in a number of actual aviation incidents and accidents (e.g., Dornheim, 1995; Sparaco, 1994). It is therefore critical to understand why they occur and how they can be avoided.
In the literature, several factors have been proposed to contribute to breakdowns in pilot-automation coordination on modern flight decks: (a) low observability, attributable to poor automation feedback that fails to support pilots in managing their attentional resources and in assessing and interpreting system states and behaviors; (b) gaps and misconceptions in pilots’ mental model of highly complex flight deck automation, attributable to limited training, inappropriate manuals, and a problematic system image; (c) a high degree of system autonomy (the ability of the automation to initiate actions without immediately preceding pilot input); and (d) a high degree of system coupling (a high level of interdependence between various components of the automation), which can lead to unexpected side effects of pilot input.

These factors have the potential to contribute to mode errors and automation surprises by affecting both data-driven and knowledge-driven monitoring. Inaccurate or incomplete mental models of the automation, in combination with high levels of system autonomy, coupling, and complexity, create a challenge for knowledge-driven monitoring. They make it difficult for pilots to form valid expectations of system behavior and thus to monitor automation indications in a timely and effective manner. Low observability and a high degree of dynamism can interfere with data-driven attention capture and guidance, which is critical in case of unexpected changes in automation status and behavior. The possible role of these factors in monitoring breakdowns has been discussed for some time, yet only limited and mostly anecdotal data are available on pilots’ actual visual monitoring behavior and performance on glass cockpit aircraft.

To our knowledge, the only empirical studies to examine these issues in the context of a glass cockpit aircraft were conducted by Huettig, Anders, and Tautz (1999) and by Diez et al. (2001). In the Huettig et al. (1999) experiment, three flight crews were asked to fly a line-oriented scenario and another crew flew a total of six instrument landing system (ILS) approaches (four under normal conditions; two involving nonnormal events) on a full-flight Airbus A-340 simulator. Three important findings from this study were that (a) pilots tended to monitor indications of aircraft behavior (such as airspeed, altitude, and attitude) rather than flight mode annunciations (FMAs), which indicate automation states and modes on the three axes of flight (lateral, vertical, speed); (b) the pilots monitored FMAs on average only 4.7% of the time; and (c) pilots did not employ a standardized, context-independent scanning pattern but, rather, monitored the automation based on expectations associated with specific flight contexts.

Diez et al. (2001) asked five B-777 pilots to fly two scenarios (50 and 30 min in duration, respectively) on a B-747-400 desktop simulator. Eye-tracking data were collected, and pilots were interrupted six times throughout the scenario and asked to recall as many details as possible about the flying situation and values from specific instruments. Participants were quite accurate at remembering basic flight parameters, such as altitude, airspeed, engine power, and aircraft position. In contrast, they performed significantly worse with respect to recalling automation-related indications – in particular, the throttle and pitch FMAs. They were able to report whether or not some mode was engaged but failed to remember the specific mode annunciations. This was true especially for the submodes of vertical navigation (VNAV), the most complex of the automation systems for the three axes of flight (Sherry & Polson, 1999). A trend toward longer fixation times (on the order of an additional 200–300 ms) was observed for those indications that pilots were able to recall correctly. It is important to note that the conclusions drawn from those studies, although plausible, were of limited generalizability because small sample sizes prevented conventional statistical analysis and inferences from being drawn.

The present research replicates some of these earlier findings on automation monitoring and mode awareness for a different type of glass cockpit. It involves a larger sample size and a higher fidelity flight simulator, which allows greater generalizability of the results. It also expands on those studies by collecting and relating continuous behavioral and eye-tracking data in an effort to not only assess but also explain pilots’ monitoring strategies and performance and their management of the automation across a wide range of circumstances. Pilots in this study witnessed various types of expected and unexpected changes in automation modes, some experimentally imposed and others following the natural progression of an automated flight. Our goal was to establish the extent to which annunciations associated with these events are attended (as assessed via scanning...
data), the extent to which variations in monitoring behavior are coupled to differences in behavioral responses, and the extent to which pilot’s “mental model” or understanding of the automation might mediate this coupling. Ultimately, the findings from this study can inform improvements of automation training and design and thus help reduce problems with pilot-automation coordination.

**Autoflight and the Flight Management System**

A brief, somewhat simplified introduction to the autoflight and flight management system (FMS) is warranted to help the reader understand the reported research (see Figure 1). The FMS supports a variety of functions on modern flight decks, including automatic flight path control. Pilots can use two interfaces to enter data into the flight management computer (FMC): the mode control panel (MCP) and the control display units (CDUs; one for each pilot). The MCP is a tactical interface that is used to enter individual airspeed, vertical speed, altitude, and heading targets and to activate autoflight modes related to thrust (e.g., the speed mode [SPD]), vertical navigation (e.g., VNAV), and lateral navigation (e.g., LNAV). The CDU is a more strategic interface that allows pilots to enter an entire flight plan (e.g., way points with associated altitude and airspeed constraints).

After the FMS has been instructed via either of these two interfaces, the pilot can activate either the autopilot, which will then execute the programmed flight path, or the flight director, which will provide guidance to the pilot who is manually flying the airplane. Information on the current and future status, targets, and behavior of the automation is distributed across four displays: the CDU data display, the MCP target windows, the primary flight display (PFD, which also shows basic flight parameters, e.g., airspeed and altitude), and the map display, which depicts a plan view of the own aircraft and its future flight path. Importantly, at the top of the PFD, FMAs indicate active automation modes, any armed modes – those that will be triggered by future conditions, such as capturing an altitude level or navigation signal (e.g.,

![Figure 1](image-url)

**Figure 1.** Flight deck controls and displays related to pilot-FMS interaction. FMS = flight management system; PFD = primary flight display; FMAs = flight mode annunciations; CDU = control display unit; FMC = flight management computer.
ILS) – and the status of the autopilots and flight directors. On the airplane type studied here, changes in the FMAs are always highlighted by the appearance of a green outline box around the changing mode for 10 s. This feature is designed to capture the pilot’s attention in a data-driven fashion.

METHODS

Participants

Twenty male B-747-400 line pilots (10 captains, 10 first officers) were recruited from two U.S. airlines. Pilots had between 100 and 9000 hr of experience on the B-747-400 (mean = 2600, SD = 2100), and they had a minimum of 1000 hr total of glass cockpit experience. Pilots were not paid for their voluntary participation.

Procedure

Each participant signed a consent form and provided demographic information. Pilots were briefed on the study and provided with all relevant flight-related paperwork for their review. Next, pilots were taken to the simulator to be fitted and calibrated with the eye-tracking equipment. The participant took his current crew position and was joined by a confederate pilot, who helped ensure that the scenario evolved as designed. The confederate performed his regular pilot-not-flying duties without creating problems but also without being proactive to help the participating pilot notice or handle experimenter-induced scenario events. When the volunteer pilot was comfortable with the simulator and the planned flight, the 1-hr scenario was initiated. Upon completion of the scenario, the experimenters and confederate pilot reviewed the scenario with the participant, and, for the purpose of mental model elicitation, the pilot was asked a set of questions concerning the functioning and operation of the autoflight systems in order to probe his knowledge and understanding of the automation.

Apparatus

The study was conducted in a B-747-400 fixed-base simulator with outside view, which was created by an Evans and Sutherland ESIG 3350 image generation system. Visual monitoring measures were made using an ASL Series 4000 head-mounted eye tracker (Applied Science Laboratory, Waltham, MA). An experimenter outside of the simulator provided live air traffic control clearances to the pilots through headsets.

Scenario

In collaboration with one of the participating airlines, a scenario was developed that lasted approximately 1 hr from takeoff to landing. It included 12 challenging autoflight-related events that required a thorough understanding of the FMS to be able to manage and monitor the system effectively. Because the scope of this manuscript does not allow us to report all findings from this study (for a complete account, see Mumaw, Sarter, et al., 2000), only those scenario events that focused on pilots’ monitoring (rather than management) of the automation are described in the following section.

Scenario Event 1: Experimenter-induced mode transitions. Three times during the scenario (once during climb, once when established on the descent path, and once later during the descent), an experimenter-induced mode transition occurred. These transitions led to the display of a pitch or autothrottle mode on the FMA that would not normally appear in the given flight context. They did not lead to any changes in airplane behavior and, because of simulator limitations, they also did not involve the appearance of a green outline box around the changing FMA. Note that this box, which accompanies all mode transitions during regular training and flight operations to capture pilots’ attention, was shown for all other mode transitions during the scenario. Its absence in the context of experimenter-induced transitions was considered acceptable because these probes were introduced specifically to determine whether pilots processed the available information in sufficient depth to notice the inappropriateness of the active mode for the current flight context.

Scenario Event 2: Revision of cruise altitude. When the airplane reached 31,500 feet during climb to the expected and programmed cruise altitude of 35,000 feet, the pilot received an air traffic control request to level off at 33,000 feet. Shortly thereafter, air traffic control indicated that 33,000 feet would be the final cruise altitude. When the pilot entered this change in the FMC via the CDU interface, the automation transitioned to the VNAV V altitude (VNAV ALT) pitch mode. In this “tactical” mode, the airplane will not automatically start its descent at the top-of-descent point unless the pilot takes an extra action to
activate the VNAV path (VNAV PTH) mode, which is usually active at cruise altitude and which leads to the desired automatic start of descent. Pilots who fail to do so will overfly the top of descent point and descend late.

Scenario Event 3: Loss of glide slope diamond and glide slope. The ground signal for the glide slope was failed during the ILS approach. As a result, the glide slope diamond on the PFD never filled in and centered itself. This event was introduced to probe the effectiveness of a “cuing by absence” or “cuing by disappearance” approach to provide a warning.

Dependent Measures

Behavioral data. Data sheets were developed that laid out the various scenario events and the range of associated possible pilot responses to allow a trained observer to collect data on actual responses during the experimental runs. These data were reviewed and edited during a debriefing session with the participating pilot.

Eye-tracking data. Eye-tracking data were collected using an ASL™ Series 4000 head-mounted eye tracker (Applied Science Laboratory, Waltham, MA), which is designed to measure a pilot’s eye line of gaze with respect to the head. When combined with an optional head-tracking device and eye/head integration software, the eye tracker can also measure a pilot’s eye line of gaze with respect to stationary surfaces in the environment.

Mental model assessment. Following the flight, an extensive debriefing period was administered in which pilots were asked a series of questions to probe their “mental model” of the automation (Sarter & Woods, 1992). In particular, they were asked about VNAV and LNAV mode activation, targets, logic, and transitions. Most of the results of these cognitive assessments are described elsewhere (Mumaw, Sarter, et al., 2000). The focus of the current mental model analysis is pilots’ knowledge of three different VNAV modes (VNAV PTH, VNAV ALT, and VNAV speed) that play an important role in the context of Scenario Event 2. The accuracy of pilots’ answers was assessed relative to an expert model, and pilots were categorized into three groups based on their number of correct answers.

RESULTS

The following sections first describe pilots’ general monitoring strategies and behavior. Next, the eyetracking and performance data for the three scenario events are presented and related (for further details, see Mumaw, Sarter, et al., 2000).

Pilots’ Monitoring Strategies and Performance for Automation-Related Indications

In our initial analysis, six areas of interest were defined: the outside world, MCP, PFD, map display, CDU, and Engine Indication and Crew Alerting System. Importantly, this analysis revealed an expected dominance of PFD scanning (attended 31% of the time, averaged over flight phases), compared with the next most dominant area of interest, the map display (25% of the time). The outside world was monitored only 3% of the time until the final approach phase, when this percentage jumped to 12%.

In order to examine statistically the scanning of the three FMAs, we performed a more detailed analysis on six areas of interest within the PFD. These six areas were defined by two orthogonal factors: (a) axis of flight control (lateral, vertical, and speed [or longitudinal]), and (b) type of monitored data (raw data vs. FMAs). The raw data instruments were the heading indicator, the altimeter and vertical speed indicator (both relating to altitude), and the airspeed indicator. The corresponding automation indicators were the three FMAs relating to roll (lateral), pitch (vertical), and autothrottle (longitudinal/speed) mode. The percentage dwell times for these indicators are shown in Figure 2, plotted by three phases of flight: climb, cruise, and descent (not including final approach). An ANOVA on these data revealed a pronounced preference for looking at the raw data over the FMAs, $F(1, 13) = 425.5, p < .01$, with the latter being scanned only 2.5% of the time on average, an effect that can be readily accounted for by the higher bandwidth of these instruments (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003). Noteworthy in the vertical plane is the markedly greater interest in altitude raw data during the climb and descent segments, an interest not shared by the vertical FMAs and a pattern unique to the vertical axis; three-way interaction, $F(4, 52) = 8.2, p < .01$.

Mean dwell duration (the time spent per glance at an instrument) was also examined. Most notable here is the finding that glances at the FMA were considerably shorter (mean = 0.40 s) than they
Figure 2. Percentage dwell times (PDTs) in six primary flight display areas of interest (raw data vs. flight mode announcements for speed, heading, and vertical plane) by phase of flight ($n = 14$) (continued next page).
were to the other instruments within the PFD (mean = 0.60 s).

Next, pilots’ monitoring of the FMAs was analyzed independent of specific scenario events for three different types of mode transitions: manual, automatic-expected, and automatic-unexpected. These three transitions are defined as follows:

- Manual: The pilot manually selects a new pitch or roll mode (e.g., flight level change and heading select) by engaging a switch on the MCP.
- Automatic-expected: A mode change is initiated by the automation in the absence of immediately preceding pilot instructions, but the change is likely to be expected by the pilot (e.g., because it has been preprogrammed earlier).
- Automatic-unexpected: A mode change is initiated by the automation, and the pilot is unlikely to expect the mode change (e.g., transitions that have little meaning to airplane performance, such as a transition between autothrottle modes).

The first two cases are likely to involve knowledge- or expectation-driven monitoring, whereas the last case — unexpected automatic transitions — requires automation feedback that is capable of capturing pilots’ attention in a data-driven fashion.

Two aspects of our analyses of these fixations are of note (see Table 1). First, there were no statistically significant differences in fixation frequency across the three classes, suggesting a relatively muted role of expectancies in driving FMA fixations. Second, across the three classes, fixation rates were surprisingly low. Over pilots and occasions, only 48% of the FMAs were fixated in the first 10 s following the change (while the green box remained on), and only 17% were fixated in the following 10-s interval.

**Experimenter-Induced Mode Transitions**

Pilots’ monitoring behavior was analyzed for the three inappropriate mode annunciations that were triggered by the experimenter. Recall that these mode transitions did not involve the appearance of a green outline box around the changing FMA. Table 2 shows, for each case, (a) how many pilots indicated during the debriefing what the appropriate mode annunciation would be for the

---

**Figure 2 (continued).** Percentage dwell times (PDTs) in for six primary flight display areas of interest (raw data vs. flight mode annunciations for speed, heading, and vertical plane) by phase of flight (n = 14).
particular phase of flight; (b) how many pilots fixated the indication at any point in time while the inappropriate mode was displayed (inappropriate FMAs appeared for 3–7 min, depending on how the scenario played out); (c) the conditional probability that a pilot fixated the FMA, given that he had mentioned the appropriate mode indication during the interview; and (d) how many pilots processed the indication at sufficient depth to notice that it was inappropriate. Depth was operationally defined here by the duration of the dwell, with durations greater than 300 ms indicating true information extraction, rather than just confirmation of a preexisting value (Harris & Christhilf, 1980). Note that for those pilots who did not mention the appropriate mode annunciation during the debriefing, it is not clear whether they lacked this knowledge or simply forgot to include this information in their statement. For that reason, no conditional probabilities were calculated for this group.

A considerable number of pilots (between 10 and 12 pilots in each case) fixated the FMAs; however, with the exception of 1 pilot during the early descent phase, they failed to notice the inappropriateness of the mode annunciations. The single pilot who noticed the problem during early descent showed the largest number of fixations (10, compared with 0–6 fixations for the remaining pilots) of the pitch mode annunciation, and the dwell duration of his first pitch FMA fixation was substantially longer (1.2 s, compared with less than 600 ms for all other pilots).

**Revision of Cruise Altitude**

In this analysis, we closely examined the linkages among bottom-up attention allocation (inferred from fixations on the appropriate FMA during cruise); pilots’ knowledge of the automation (captured by their answers to debriefing questions and scored as low, medium, or high according to their accuracy); and two important behavioral indices: (a) whether pilots noticed the inappropriateness of the VNAV ALT mode and therefore reset it to VNAV PTH, and (b) whether or not they initiated descent after the top of descent point (see Table 3).

The distribution of pilots in Table 3 reveals the following: First, the two best indices of “appropriate” behavior are well correlated. All 4 pilots who noticed and responded to the inappropriate mode descended early, and half (4 of 8) who descended early had noticed the inappropriate mode. Second, scanning the FMA appears to be

<table>
<thead>
<tr>
<th>TABLE 1: Responses to FMA Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMA Change</td>
</tr>
<tr>
<td>Manual</td>
</tr>
<tr>
<td>Automatic-expected</td>
</tr>
<tr>
<td>Automatic-unexpected</td>
</tr>
</tbody>
</table>

Note. These percentages are calculated on the basis of the total number of opportunities to detect, pooled over the number of pilots, and the number of changes experienced by each pilot.

<table>
<thead>
<tr>
<th>TABLE 2: Responses to Experimenter-Induced FMA Changes (n = 16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase of Flight</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Pilots who mentioned appropriate mode annunciation</td>
</tr>
<tr>
<td>Pilots who fixated annunciation</td>
</tr>
<tr>
<td>Conditional probability of fixation, given knowledge of appropriate annunciation</td>
</tr>
<tr>
<td>Pilots who detected problem</td>
</tr>
</tbody>
</table>

Note. The second row includes pilots who did not mention the appropriate mode annunciation during the debriefing.
necessary (but not sufficient) for noticing the inappropriate mode change. Second, scanning the FMA appears to be necessary (but not sufficient) for noticing the inappropriate mode change (8 pilots scanned but did not notice). There is also some association between scanning and the other performance indicator – early descent – although 1 pilot was successful in the early descent without scanning the FMA. Here again, scanning was not sufficient, as pilots scanned but descended late nevertheless. Third, we examined how knowledge influenced appropriate behavior and/or modified the influence of scanning on that behavior. The 3 of the 4 pilots with the lowest knowledge of VNAV modes neither noticed the mode change nor descended early (even though 2 did scan). In contrast, 4 of the 5 pilots with the most accurate VNAV knowledge descended early, and 3 of these 5 also noticed the inappropriate FMA. An intermediate level of knowledge – even in the presence of scanning (7 pilots) – was insufficient to guarantee appropriate behavior. Chi-squared analyses confirmed that a successful early descent was significantly more likely to be performed by those with high knowledge (4/5) than those without low or medium knowledge (3/15). $\chi^2 = 5.43, p < .05$, and that detection of the inappropriate mode was also more likely for those with high knowledge (3/5) than those without it (1/15). $\chi^2 = 6.19, p < .05$.

**Loss of Glide Slope Diamond and Glide Slope**

Eight pilots noticed that the glide slope signal was missing prior to intercepting the final leg, whereas 6 pilots detected the problem after intercepting the final leg. The remaining 6 pilots noticed the problem only after it was pointed out by the confederate pilot or air traffic control.

**DISCUSSION**

The converging data from this study provide an informative picture of automation use and monitoring by highly skilled pilots in the context of a high-fidelity simulation. Its results provide specific and confirming evidence that monitoring failures constitute a major contributor to breakdowns in pilot-automation interaction. More specifically, we found that pilots do not always respond appropriately to unanticipated changes in automation settings that can happen because of the high level of complexity and coupling of modern flight deck technologies. Pilots’ inappropriate responses reflect a lack of mode awareness, which, in this study, was most clearly indicated by the failure to descend at the top of descent point by the large number of pilots who found themselves in the wrong mode. Although considerable prior research has confirmed the prevalence of these mode confusions (Sarter & Woods, 1994, 1995, 1997, 2000), we can assess here, for the first time, the extent to which they may be directly attributable to monitoring failures.

A first, strong hypothesis could be that if pilots had looked at the inappropriate mode status (VNAV ALT during cruise), they would have noticed it and changed modes appropriately. We can clearly reject this hypothesis because a large
number of pilots did fixate the pitch mode during cruise and yet failed to change the mode and/or to descend on time. We found that such failures were attributed, at least in part, to inappropriate or incomplete knowledge (a “buggy mental model”) about vertical navigation modes, as assessed during the debriefing (for detailed findings, see Mumaw, Sarter, et al., 2000). Although scanning is not sufficient, the data in Table 2 reveal that it did appear, in our population, to be necessary; hence, we can attribute automation use breakdowns observed here, as well as elsewhere, in part to inappropriate monitoring.

Other aspects of our data spoke to nonoptimal aspects of pilot visual sampling of automation as well. For example, the steady-state rate of glances to automation was quite low (Figure 2), replicating other studies that had lower statistical power (Huettig et al., 1999). On the one hand, it may be argued that this low rate of scanning is approaching optimal scanning behavior given the low frequency of changes in such indices (Moray, 1986; Wickens et al., 2003). On the other hand, stronger evidence that automation sampling was still suboptimal is provided by two sources of our data. First, the experimenter-induced changes were rarely noticed at all. One may argue that these went unnoticed because they were not accompanied by the onset of the green box and thus failed to capture attention in a data-driven fashion. However, the analysis of scanning contingent upon aircraft-triggered mode changes (and resulting green box onsets) indicated that even in cases when an FMA change might be highly expected, of value to confirm, and coupled with a discrete color onset, it was still fixated around only 50% of the time, hence seriously departing from what might be described as optimal data- and knowledge-driven scanning strategies (Wickens et al., 2003).

Thus, the following overall picture emerges from our data. Four factors might lead a pilot to sample an FMA: (a) a change in its value (actually having very little influence); (b) the onset of the green box surrounding this change (which increased the likelihood to between 50% and 67%); (c) some contextual event preceding the change; and (d) accurate knowledge of the automation contingencies that might drive the change. We found that the first three factors can, but do not necessarily, trigger a fixation in a data-driven fashion. Even if a fixation occurs, it is not sufficient to lead to an appropriate response – that requires the fourth factor: a complete and accurate model of the automation. This knowledge-driven influence on monitoring is necessary not only to anticipate such a change, but also to interpret its meaning and thus understand its implications for airplane behavior.

In summary, the present study confirms and expands on earlier findings on automation-monitoring strategies and performance breakdowns, which mirror operational experiences with pilot-automation interaction (Bellenkes, Wickens, & Kramer, 1997; Helleberg & Wickens, 2003). Most importantly, it is the first study to be able to explain these performance breakdowns based on converging behavioral, eye-tracking, and mental model data. Our findings call for more conceptual and exploratory approaches to training (e.g., Casner, 2003; Chappell, Crowther, Mitchell, & Govindaraj, 1997; Mumaw, Boorman, & Griffin, 2001; Mumaw, Boorman, Griffin, Moodi, & Xu, 2000; Sarter & Woods, 2000) and improved feedback design, which could take the form of multimodal interfaces (see Nikolic, Orr, & Sarter, 2004; Sarter, 2000; Sklar & Sarter, 1999; D. Javaux, personal communication, 2004) and visualizations of automation intent and aircraft behavior (see Boorman & Mumaw, 2004), to avoid future mode errors and automation surprises.

ACKNOWLEDGMENTS

This work was funded, in part, by NASA Ames Research Center (Technical Monitor: Dr. Key Dismukes; Contract #NAS2-99074 from the Aviation Safety Program). We would like to thank the participating airlines, the pilots, and our colleagues who contributed to this research: Jack Bard, Dean Holt, Steve Kimball, Roger Marsh, Mark Nikolic, Ashley Nunes, Bob Reid, Bill Shontz, Wei Xu, and Xidong Xu.

REFERENCES


Nadine B. Sarter is an associate professor of industrial and operations engineering at the University of Michigan. She received her Ph.D. in industrial and systems engineering from Ohio State University in 1994.

Randall J. Mumaw is an associate technical fellow in human factors at Boeing Commercial Airplane. He received his Ph.D. in cognitive psychology from the University of Pittsburgh in 1985.

Christopher D. Wickens is a senior scientist at Alion Science Corporation, Micro Analysis & Design Operations, Boulder, Colorado, and professor emeritus at the University of Illinois at Urbana-Champaign. He received his Ph.D. in psychology from the University of Michigan in 1974.

Date received: December 1, 2003
Date accepted: April 27, 2006