A SIMULATOR STUDY OF PILOTS' MONITORING STRATEGIES AND PERFORMANCE ON MODERN GLASS COCKPIT AIRCRAFT

Randall J. Mumaw[1], Mark I. Nikolic[2], Nadine B. Sarter[2], Christopher D. Wickens[3]

[1] Boeing Commercial Airplane Group, Seattle, WA
[2] The Ohio State University, Institute for Ergonomics, Columbus, OH
[3] University of Illinois, Aviation Research Lab, Savoy, IL

Monitoring failures are widely assumed to be responsible for breakdowns in pilot-automation coordination; however, only limited subjective and anecdotal data are available on the monitoring behavior of flight crews on glass cockpit aircraft. This study is the first to collect and relate both performance and eye-tracking data from pilots who flew a challenging scenario in a high-fidelity modern aircraft simulator. Twenty B-747-400 line pilots were recruited from two major U.S. airlines and were given a one-hour scenario that was designed specifically to assess pilots' use and monitoring of the automation. The findings from this study confirm that pilots experience considerable difficulties with tracking the status and behavior of the automation on modern glass cockpit aircraft. They appear to monitor flight mode annunciations to a much lesser extent and at a more superficial level than intended and expected by designers and training departments. Possible ways of supporting pilots more effectively through improved automation feedback and training are discussed.

INTRODUCTION

The introduction of highly automated systems to a variety of complex dynamic domains such as aviation has been a mixed blessing. It has led to an increased precision and efficiency of operations but it also involves the risk of breakdowns in the communication and coordination between pilots and their modern flight deck systems.

There is considerable empirical evidence for these problems, both from pilot reports, research, and operational experience, including incidents and accidents (e.g., Abbott et al., 1996; Sarter & Woods, 1992, 1994, 1995, 1997; Sarter, Woods, & Billings, 1997; Wiener, 1989; Woods & Sarter, 2000). Monitoring failures are widely assumed to be responsible for breakdowns in pilot-automation coordination; however, only limited subjective and anecdotal data are available on the monitoring behavior of flight crews on glass cockpit aircraft. It is not known what monitoring strategies pilots adopt to track automation behavior, how effective these strategies are, and under what circumstances they tend to break down.

The present study was aimed at filling this gap. It is the first study to collect and relate both performance and eye-tracking data from pilots who flew a challenging scenario in a high-fidelity modern aircraft simulator (B-747-400). Additional data were gathered in a debriefing to assess pilots' knowledge and understanding of the autoflight system components to explain their observed performance and scanning strategies during the scenario. The scope of this paper allows us to report only a small subset of the large amount of data collected in this study. The focus will be on pilots' detection and interpretation of unexpected mode transitions, which were included in the scenario to learn more about the reasons why pilots sometimes miss these events in actual line operations.

METHODS

Subjects

Twenty B-747-400 line pilots (10 Captains, 10 First Officers) were recruited from two major U.S. airlines. Pilots had between 100 and 9000 hours of experience on the B-747-400 (mean=2600;

SD=2100), and they had a minimum of 1000 hours total of glass cockpit experience. Pilots were not paid for their voluntary participation.

Procedure

First, each participant signed a consent form and provided demographic information. Pilots were then briefed on the purpose of the study as well as the limitations of the simulator. They were asked to report any invalid or incorrect indications that they observed during the flight. Next, they were given a written clearance, the relevant navigational charts, and a set of dispatch papers and other flight-related information. They had as much time as they needed to review the documents. When pilots completed the review, they were taken to the simulator to be fitted and calibrated with eye tracking equipment. The participating pilots took their current crew position and were joined by a confederate pilot who reviewed the clearance, flight plan, and flight deck settings that were entered while the subject was being briefed. The confederate pilot helped ensure that the scenario evolved as designed. He performed his regular pilot not-flying duties without creating any problems for the subject but also without being proactive to help the subject notice or handle experimenter-induced scenario events. When the subject was comfortable with the simulator and the planned flight, there was a final check of the eye tracker calibration, and the 1-hour scenario was started. Upon completion of the scenario, the subject was given a 10-minute break. Subsequently, the experimenters and confederate pilot reviewed the scenario with the participating pilot, primarily to make sure that observer notes were complete and accurate. After the debriefing, a set of detailed questions was asked about the functioning and operation of the autoflight systems in order to probe pilots' 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 & 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 about one hour from take-off to landing. The scenario was designed specifically to assess pilots' use and monitoring of the automation. It included twelve challenging autoflight-related events that required a thorough understanding of the Flight Management System (FMS). Most of these events consisted of an air traffic control (ATC) clearance that required the pilot to invoke the automation in order to comply with the controller's instructions. Another four events involved experimenter-induced changes to the flight mode annunciations (FMAs) on the Primary Flight Display (PFD) (see Figure 1).

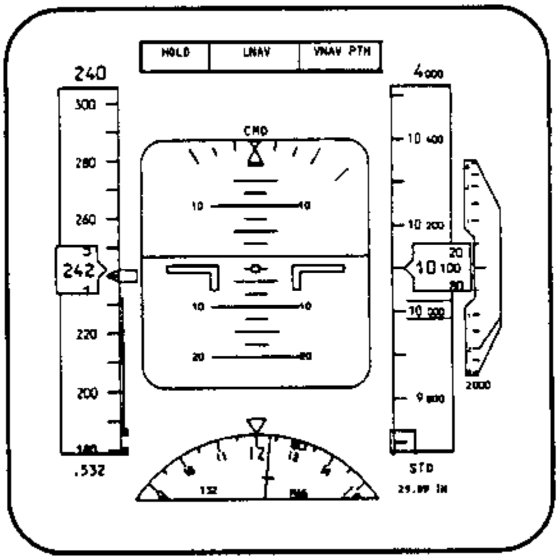

Figure 1. **The primary flight display (PFD) of a B-747-400. The flight mode annunciations (FMAs) are shown at the top in three columns (in boxes).**

These changes were introduced to simulate the unexpected mode transitions that can occur on modern flight decks due to system coupling, as a result of a delayed system responses to pilot instructions, or as a consequence of input by another flight crew member. These unexpected mode changes are sometimes missed by pilots in actual line operations – a situation that can lead to a loss of mode awareness and that involves the potential for undesirable, or even unsafe, system behavior. This paper will focus on only these four events, which are directly relevant to maintaining awareness of the present and future status and behavior of the automation.

Event 1: Inappropriate pitch mode #1. During climb, the vertical pitch mode annunciation was artificially altered on the primary flight display (PFD). With VNAV engaged on climb, the airplane pitches to a speed target (not to a path target) to climb to the cruise altitude, and VNAV SPD is normally annunciated on the PFD. In this scenario, however, just after the plane started climbing, the pitch mode was changed to display VNAV PTH, a mode that exists but would never be active during this phase of flight. The mode change was made without using the green outline box that normally appears around the changing mode indication in order to assess the extent to which effective monitoring depends on top-down and bottom-up processes, i.e., on pilots' expectations about specific modes and on the attention capture power of the green box, respectively. The event was introduced to determine whether pilots notice the transition and whether they understand that the displayed automation state is inappropriate under the given circumstances.

Event 2: Revised cruise altitude. When the airplane reached 31,500 ft. during its climb to cruise altitude, the pilot received an ATC request to level off at 33,000 ft. Shortly thereafter, the pilot was told that 33,000 ft. would be his final cruise altitude. This clearance required that the pilot re-program the flight management computer, which had been set initially to an expected cruise altitude of 35,000 ft. The event was set up in such a way that the reprogramming of the automation resulted in the activation of an inappropriate pitch mode (VNAV ALT) in which the aircraft would not automatically commence the descent phase upon reaching the top-of-descent point. Again, the question of interest was whether pilots would notice and understand the implications of the inappropriate mode setting.

Event 3: Inappropriate pitch mode #2. After the airplane was established on the VNAV descent path, it showed the correct indication (VNAV PTH) as the active pitch mode. At this point, similar to the first event, the annunciation was changed by the experimenter to VNAV SPD, which would never appear under these conditions in real line operations.

Event 4: Inappropriate autothrottle mode. After the change to VNAV SPD, while the airplane was still on the VNAV descent path, the autothrottle mode annunciation was changed to THR - yet another and the last inappropriate mode indication given the actual flight context in this scenario.

RESULTS

The following sections report, for each of the four described scenario events, whether pilots fixated the relevant flight mode annunciation, noticed that it was inappropriate, and/or understood its implications for aircraft behavior. Due to unexpected variations in the scenario, not every pilot was exposed to each scenario manipulation (see Table 1).

Event 1: Inappropriate pitch mode #1

Eighteen subjects experienced the experimenter-induced pitch mode change. None of those pilots reported that they had noticed the inappropriate pitch mode nor did they take any action to correct the situation even though the eye tracking data show that twelve of them fixated the pitch mode annunciation at least once during a three-minute period following the transition.

Event 2: Revised cruise altitude

Sixteen subjects found themselves in the VNAV ALT mode after the cruise altitude change. Recall that this would prevent the aircraft from starting its descent automatically upon reaching the top-of-descent (TOD) point. Of the nine pilots who fixated the pitch mode (eye tracking data were missing for five pilots), only two commented on it. Twelve pilots failed to intervene and take the necessary steps to ensure that the aircraft would descend at TOD. Two pilots attempted to correct the situation but used an inappropriate technique for doing so.

Event 3: Inappropriate pitch mode #2

Of the nineteen subjects exposed to this manipulation, only one indicated that he had noticed the unusual pitch mode annunciation. The eye tracking data show, however, that a total of ten of the nineteen subjects fixated the pitch mode at least once during the three-minute period following the change.

Event 4: Inappropriate autothrottle mode

Of the nineteen subjects who experienced this scenario event, none of them indicated that he had observed the unusual autothrottle mode annunciation. Again, the eye tracking data indicate that ten of the nineteen subjects fixated the inappropriate pitch mode indication at least once following the transition.

Event	N	Fixated	Reported	Intervened
1 - Inappropriate pitch mode #1	18	12	0	0
2 - Revised cruise altitude	16	9	2	4
3 - Inappropriate pitch mode #2	19	10	1	0
4 - Inappropriate autothrottle mode	19	10	0	0

***Table 1.* Summary of findings for the four scenario events. Shown are the numbers of subjects that were exposed to each manipulation (N), as well as how many of those subjects fixated the relevant indicator, reported an anomaly, and intervened to correct the anomalous event.**

DISCUSSION

The above findings confirm that pilots experience considerable problems with monitoring the automation on modern glass cockpit aircraft (e.g., Abbott et al., 1996; Sarter and Woods, 1994, 1997; Wiener, 1989). Overall, pilots appear to monitor flight mode annunciations to a much lesser extent and at a more superficial level than intended and expected by designers and training departments. The debriefing data from this study confirm that this problem is related, in part, to gaps and misconceptions in pilots' knowledge and understanding of the aircraft automation (e.g., Sarter and Woods, 1997, 2000). These gaps make it difficult for pilots to form proper expectations of system status and behavior, which, in turn, represent the basis for knowledge-driven, timely and effective attention allocation. Even if pilots looked at the flight mode annunciations, as indicated by the eye tracking data, they did not realize when the indication was not appropriate for the given flight context. This was true both for the experimenter-induced mode transitions and for the VNAV ALT indication at cruise altitude, which was triggered by pilot actions.

Our findings suggest a need for both improved training and design. Pilots' mental model of the automation can be improved through a more exploratory approach to training (e.g., Woods and Sarter, 2000), which helps avoid the creation of inert knowledge, i.e., knowledge that is not activated in the proper context. More complete and accurate context-conditioned knowledge about the automation can be expected to contribute to more effective attention allocation and an increased depth of processing of the observed indications.

At the same time, pilots' monitoring performance could be enhanced through a more adequate system image and the design of more effective automation feedback. In particular, changes in the automation status and behavior need to be highlighted in more effective ways to capture pilots' attention more reliably. In this context, it is interesting to note that pilots did not only miss most of the experimenter-induced mode transitions without the green outline box. The eye-tracking data show that they also fixated only about half of the other mode changes that *were* accompanied by the onset of a green outline box. This failure to capture attention can be explained, to a large extent, by the fact that current flight mode annunciations (and, with them, the green box) are embedded in the context of the highly dynamic and data-rich Primary Flight Display (e.g., Nikolic, Orr, and Sarter, 2001).

Finally, effective visualizations of current and future aircraft and automation behavior, especially with respect to vertical navigation and thrust management, are needed. The map display on current flight decks illustrates how such an approach can aid pilots in keeping track of the intentions and actions of the automation.

ACKNOWLEDGMENTS

This work was funded partially by NASA Ames Research Center through Dr. Key Dismukes (Contract #NAS2-99074). Other contributors to this research effort were Steve Kimball, Roger Marsh, Wei Xu, Xidong Xu, Bobbie Seppelt, Jack Bard, Bob Reid, Dean Holt, and Bill Shontz.

REFERENCES

Abbott, K., Slotte, S., Stimson, D., Bollin, E., Hecht, S., Imrich, T., Lalley, R., Lyddane, G., Thiel, G., Amalberti, R., Fabre, F., Newman, T., Pearson, R., Tigchelaar, H., Sarter, N., Helmreich, R., & Woods, D. (1996). *The Interface Between Flightcrews and Modern Flight Deck Systems.* Federal Aviation Administration Human Factors Team Report. Washington D.C.: FAA.

Nikolic, M., Orr, J., & Sarter (2001). Why onsets don't always capture attention: The importance of context in display design. *Proceedings of the 11th International Symposium of Aviation Psychology,* Columbus, Ohio.

Sarter, N.B., & Woods, D.D. (1992). Pilot interaction with cockpit automation: Operational experiences with the flight management system. *International Journal of Aviation Psychology, 2,* 303-321.

Sarter, N.B., & Woods, D.D. (1994). Pilot interaction with cockpit automation: II. An experimental study of pilots' model and awareness of the flight management system. *International Journal of Aviation Psychology, 4,* 1-28.

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,* 5-19.

Sarter, N.B., & Woods, D.D. (1997). Team play with a powerful and independent agent: operational experiences and automation surprises on the Airbus A-320. *Human Factors, 39,* 553-569.

Sarter, N.B., Woods, D.D., & Billings, C.E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd edition, pp. 1926-1943). New York: Wiley.

Wiener, E.L. (1989). Human factors and advanced technology (glass cockpit) transport aircraft. (TR 177528). Moffet Field, CA: NASA Ames Research Center.

Woods, D.D. & Sarter, N.B. (2000). Learning from automation surprises and "going sour" accidents. In N.B. Sarter and R. Amalberti (Eds.), *Cognitive Engineering in the Aviation Domain* (pp. 327-353). Hillsdale, NJ: LEA.