To be presented at the 12th International Symposium on Aviation Psychology, Dayton, OH, 2003.

MODERN FLIGHT DECK AUTOMATION: PILOTS' MENTAL MODEL AND MONITORING PATTERNS AND PERFORMANCE

N. Sarter[1], C. Wickens[2], R. Mumaw[3], S. Kimball[3], R. Marsh[2], M. Nikolic[1], and W. Xu[4]

[1] The Ohio State University, Department of Industrial and Systems Engineering and Institute for Ergonomics, Columbus, OH
[2] University of Illinois at Urbana-Champaign, Department of Psychology and Institute of Aviation - Aviation Research Lab, Savoy, IL
[3] Boeing Commercial Airplane Group, Aviation System Safety, Seattle, WA
[4] Intel, Sacramento, CA

Substantial empirical evidence from surveys and simulator studies indicates that "glass cockpit" pilots sometimes lose track of the status and behavior of automated flight deck systems and, as a result, experience "automation surprises." A number of related factors are assumed to contribute to these problems, including the nature of current automation feedback and gaps and misconceptions in pilots' understanding of the automation. To date, most research on pilot-automation interaction has focused on subjective accounts and on performance outcome measures. Little is known about underlying processes, including how pilots monitor the automation and at what stages their information processing tends to break down. To fill this gap, a simulator study was conducted where twenty 747-400 pilots flew a routine one-hour flight on a fixed-base 747-400 simulator. Several scenario events were introduced to assess pilots' monitoring behavior and their awareness of automation status and behavior. Throughout the scenario, behavioral and performance data as well as eye fixations were recorded. After the scenario was complete, pilots' mental model of the automation was probed based on a predefined set of questions. Overall, the findings from this research confirm that pilots experience considerable problems with monitoring the automation on modern glass cockpit aircraft. There is considerable diversity across pilots in terms of the frequency, duration, and pattern of scanning automation indications. Also, during the debriefing, pilots revealed significant gaps in their understanding of some of the automation features. The results from this study – both in terms of process and outcome measures – will be discussed in terms of their implications for improving training and design for effective pilot-automation collaboration.

Introduction

There is widespread agreement that the introduction of highly automated systems to a wide range of complex event-driven domains, such as aviation, has led to an increased precision and efficiency of operations. At the same time, it has created new training requirements, new forms of errors, a sometimes problematic redistribution of workload among crewmembers, and problems in the form of breakdowns in the communication and coordination between pilots and their modern flight deck systems (e.g., Abbott et al., 1996; Sarter and Woods, 1995; Wiener, 1989).

There is considerable empirical evidence for these problems, from pilot reports, numerous research efforts, and operational experience, including incidents and accidents (e.g., Abbott et al., 1996; Sarter and Woods, 1994, 1995, 1997; Sarter, Woods, and Billings, 1997; Wiener, 1989; Woods and Sarter, 2000). It is widely assumed that monitoring failures are responsible for breakdowns in pilot-automation coordination; however, only 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 the strategies are, and under what circumstances they may break down.

The goal of the present study was to fill this gap. It is the first study to collect and relate both performance outcome and eye-tracking process data from airline pilots who flew a one-hour scenario in a high-fidelity modern aircraft simulator. Additional data were gathered in a debriefing to assess pilots' knowledge and understanding of their autoflight systems. These mental model data can serve to explain the observed monitoring 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. Further details are provided in Mumaw et al (2000b). The focus will be on pilots' monitoring strategies and behavior and on their detection and interpretation of experimenter-induced mode transitions. These unexpected transitions were included in the scenario as they are known to be missed quite often in actual line operations. A limited overview of the findings from our mental model assessments will be provided also.

Methods

Subjects

Twenty B-747-400 line pilots (10 Captains, 10 First Officers) from two major U.S. airlines participated in this study. They had between 100 and 9000 hours of experience on the B-747-400 (mean=2600; SD=2100) and a minimum of 1000 hours of glass cockpit experience in general. Pilots were not paid for their voluntary participation.

Procedure

After signing a consent form and providing demographic information, pilots were briefed on the purpose of the study as well as the features and limitations of the simulator. They were asked to report any invalid or incorrect indications that they observed during the flight. This served to ensure that pilots would point out if they observed any of the experimenter-induced mode transitions, which resulted in mode settings that were inappropriate for the given flight context. Next, they were given as much time as needed to review their written clearance, the relevant navigational charts, and a set of dispatch papers and other flight-related information. Subsequently, pilots were taken to the simulator to be fitted and calibrated with the eye tracking equipment. They took their current crew position and were joined by a confederate pilot who 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 helping the subject notice or handle experimenter-induced scenario events. After a final check of the eye tracker calibration, the one-hour scenario was started.

Following the experiment, participants were given a 10-minute break. Next, the experimenters and the confederate pilot reviewed the scenario with the participating pilot to ensure that observer notes were complete and accurate. After the debriefing, participating pilots were asked a set of detailed questions about the functioning and operation of the autoflight systems to probe their knowledge and understanding of the automation. The first set of questions examined pilots' understanding of the three vertical navigation modes of the FMS: VNAV Path, VNAV Speed, and VNAV Alt. These modes and their interactions and transitions are highly complex and difficult to comprehend for many pilots. In addition, vertical navigation is usually not covered in sufficient detail during pilot training due to time constraints. Instead, pilots tend to learn about the intricacies of these modes during line operations and form their mental model of these modes based on those experiences, with few opportunities for uncovering and correcting erroneous assumptions. Other areas of interest in the mental model assessment included the mechanisms that can trigger a transition between the various VNAV modes, the availability of speed intervention in the context of various FMS modes, data propagation throughout the CDU page architecture, and different possible combinations of pitch and autothrottle modes.

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. Eyetracking data were collected using an ASL series 4000 head-mounted eye tracker (Applied Science Laboratory, Waltham, MA). An experimenter outside the simulator played the role of air traffic controller and provided clearances via a headset.

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 included twelve challenging autoflight-related events that required careful monitoring of the automation status and behavior and a thorough understanding of the Flight Management System (FMS) to be able to interpret the observed system indications. Most scenario events were introduced by means of an air traffic control (ATC) clearance that required the pilot to invoke the automation in order to comply with the controller’s instructions. Four other events involved an experimenter-induced flight mode transition that led to a mode indication that was inappropriate given the current flight context. Note that only the mode annunciation, but not the aircraft behavior, changed. This manipulation helped examine the claim that pilots tend to notice changes in automation status based on feedback about resulting aircraft behavior rather than associated visual feedback (see Sarter and Woods, 1995). Experimenter-induced mode transitions were introduced to simulate 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 transitions are sometimes missed in actual line operations – a situation that results in a loss of mode awareness and involves the potential for undesirable, or even unsafe, aircraft behavior.

The following paragraphs briefly describe the most critical scenario events in domain-specific terms:

Event 1. Runway change
ATC requested a runway change during initial taxi. This change lead to several consequences: a) the take-off speeds were deleted in the FMC and had to be reselected on the Control Display Unit, b) a route discontinuity was created right after the Standard Instrument Departure (SID), c) an altitude constraint that was not part of the SID but had been entered by the pilot was lost and had to be re-entered.

Event 2. Request to expedite climb
ATC requested an expedited climb to cross waypoint D6 at 4000ft.. This request in a high-workload situation is likely to make the pilot leave the active VNAV mode, revert to a lower-level automation mode, and return to VNAV at a later point in time to recapture the preprogrammed flight path. Also, this clearance leads to resetting the altitude on the Mode Control Panel (MCP) and thus removing the reminder of the altitude restriction.

Event 3. Inappropriate pitch mode
During climb, the experimenter triggered a pitch mode transition from VNAV SPD to VNAV PTH. On the actual airplane, the latter mode is never active during climb.

Event 4. Revise CRZ altitude
When the airplane reached FL315 on its climb to the preprogrammed cruise altitude of FL350, ATC requested a level-off at FL330. Shortly thereafter, ATC indicated that FL330 would be the new final cruise altitude. Reprogramming the automation for this new altitude leads to VNAV ALT (instead of the usual VNAV PTH) becoming the active pitch mode upon altitude capture. This implies that the airplane will no longer automatically begin its descent upon reaching the top-of-descent point.

Event 5. Inappropriate pitch mode
After the airplane was established on the VNAV descent path, the experimenter triggered a mode change from VNAV PTH to VNAV SPD. In actual operations, this mode would not be active in this circumstance.

Event 6. Inappropriate autothrottle mode
Some time after the experimenter-induced mode transition to VNAV SPD, the autothrottle mode annunciation was changed to THR – again, this mode would not normally be active in this situation.

Event 7. Loss of glideslope
The ground signal for the glideslope was failed. As a result, the glideslope diamond on the Primary Flight Display never filled in nor centered itself. This failure was introduced to observe monitoring performance in the case of cueing-by-absence, which tends to be problematic as it fails to capture and guide pilots' attention.

Results

The complete set of results from this study is beyond the scope of this paper. It was reported completely in an earlier report (see Mumaw et al., 2000b; see also Mumaw et al., 2001, where some of these same results were reported). Here, we focus on pilots' overall scanning patterns and strategies, on their awareness of both the naturally occurring and the experimenter-induced flight mode transitions, and on a brief overview of the mental model data.

Overall Scanning Patterns

The analysis of the eye-fixation data focused initially on the distribution of fixations across the following seven areas of interest (AOIs): the Primary Flight Display, the navigation display, outside the window, the Control Display Unit, the upper and lower EICAS displays, and the Mode Control Panel. At a more fine-grain level, fixations within the Primary Flight Display were analyzed for 14 pilots. The following AOIs were considered: the attitude indicator, the altitude and airspeed tapes, the heading indicator, and each of the three flight mode annunciations (autothrottle, pitch, and roll modes) at the top of the PFD.

Tables 1 and 2 show the percentage of dwell time for each AOI for each of five major flight phases. Note that Table 2 indicates dwell times for each AOI as a percentage of the overall PFD dwell time during that phase. Note also that in these two tables the AOI percentages do not add up to 100% because the Table does not include those times in which pilots were fixated on regions outside the designated AOIs.

Table 1. Percent Dwell Time for each major AOI by flight phase.

	TO	CLB	CRZ	VNAV DES	DES/ Land
PFD	14	38	22	32	40
Nav Display	2	26	22	33	23
CDU	0	6	13	6	2
MCP	1	4	2	3	4
Up EICAS	3	2	2	1	1
Low EICAS	0	2	4	1	0
Out Window	70	4	1	1	12

Table 2. Percent Dwell Time for each PFD AOI by flight phase.

	TO	CLB	CRZ	VNAV DES	DES/ Land
PFD airspeed	26	13	16	22	22
PFD attitude	29	36	35	28	34
PFD altitude	1	28	16	24	18
PFD heading	0	2	7	3	3
PFD FMAs	14	4	7	5	5

Table 1 reveals that the PFD and navigation displays are monitored the most, except in the Takeoff phase when pilots spend most of the time looking out the window. Within the PFD (see Table 2), the airspeed, attitude, and altitude indications receive the most attention. The dwell time on the attitude indication remains relatively high and unchanged across flight phases. Monitoring of the altimeter and airspeed tape increases during climb and descent and during takeoff and descent, respectively. The flight mode annunciations are monitored more during take-off, than during other phases.

Awareness of Flight Mode Transitions

For the purpose of this analysis, flight mode transitions were divided into the following three categories:

- *Manual (M)* – the pilot manually selects a new pitch or roll mode using a switch on the MCP.
- *Automatic-Expected (AE)* – a mode change that is expected by the pilot is initiated by the automation
- *Automatic-Unexpected (AU)* – a mode change is initiated by the automation but is not expected by the pilot nor is it associated with any airplane performance changes.

Overall, pilots failed to fixate the FMAs in 53% of the manually induced mode transitions, 45% of the automatic-expected mode changes, and 62% of the automatic-unexpected mode changes within the first 10 seconds of the transition. During that time, the appearance/presence of a green outline box around the affected mode is supposed to summon their attention. Even within the first 20 seconds after the appearance of the green box, the failure rates for monitoring the FMAs are still high: 32% for manual transitions, 29% for automatic-expected transitions, and 40% for automatic-unexpected mode changes.

In case of the three experimenter-induced mode transitions (events 3, 5, and 6), the following monitoring behavior and detection performance was observed (see Table 3).

Table 3. Monitoring behavior and detection performance for the three experimenter-induced mode transitions

	Fixated	Detected/ Reported
Event 3	12/18 pilots	0/18 pilots
Event 5	10/19 pilots	1/19 pilots
Event 6	10/19 pilots	0/19 pilots

Pilots' Mental Model of the FMS

After the simulator session, pilots were asked to explain and comment on specific features and functionalities of the FMS to assess the completeness and accurateness of their mental model of the system.

Overall, most pilots were able to describe correctly how frequently used components of the system work in common situations. Interestingly, not all pilots applied this knowledge effectively during the simulator session. Also, most pilots were much less knowledgeable about rarely used and more intricate aspects of the automation. In particular, they showed considerable gaps in their understanding of VNAV SPD and VNAV ALT. For one example of the findings in this area see Table 4, which shows how many of the 20 pilots made a correct or incorrect statement or did not comment at all.

Table 4. Pilot responses to questions about the VNAV PTH mode

In VNAV Path:	Correct answer	Incorrect answer	No comment
the FMC is flying to an FMC-calculated fixed path	12		8
the FMC flies path on elevator and airspeed on autothrottle	1		19
FMC altitude and airspeed constraints will be met during descent	2		18
when in cruise at FMC cruise altitude	15	1	4
when on the FMC-created descent path	15	1	4
during climb when there is an FMC altitude constraint tied to a waypoint	3	1	16

Discussion

Overall, the findings from this study show that pilots use a wide range of sometimes ineffective monitoring strategies and thus experience considerable problems with tracking the status and behavior of the automation on modern glass cockpit aircraft (e.g., Abbott et al., 1996; Sarter and Woods, 1994, 1997; Wiener, 1989). The observed monitoring patterns in this study are consistent with the smaller glass-cockpit data set obtained by Huettig et al. (1999) and with data gathered for general aviation pilots, independent of whether or not those pilots were presented with an outside view (with outside view:Helleberg Goh and Wickens, 2003; Wickens Helleberg & Xu, 2002; Helleberg and Wickens 2000; without outside view: Bellenkes et al., 1997). For the PFD in particular, the observed average percent dwell time of 35% is consistent with findings by Huettig et al. (1999) but it is substantially smaller than the 50-60% values reported by Wickens et al. (2002). In addition to a difference in pilots' skill level, the presence of a navigation display in the current study may account for this difference. Indeed, if we combine the dwell times for the navigation display and the PFD, the total percentage (60%) agrees closely with the data from the earlier GA studies.

A second parallel with the glass cockpit data of Huettig et al. (1999) is the important finding that pilots spent very little time attending to the FMAs. Additionally, even if they fixated the FMAs, pilots in this study did not process the information at sufficient depth to notice problematic automation states, as shown by the data for the experimenter-induced transitions and for the inappropriate VNAV ALT indication at cruise altitude, which was triggered by pilot actions.

One contributor to these problems is the design of current automation feedback. In particular, the eyetracking data show that pilots fixated the FMAs only about half of the time after a mode transition had occurred. This suggests that the green outline box which appears in case of a mode transition to capture and guide pilots' attention is not effective. Its 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). One possible solution to the problem is the distribution of information across different modalities. For example, the sense of touch which is currently underutilized on the flight deck, could serve to alert pilots to unexpected changes without disrupting their performance on concurrent visual tasks (see Sarter, 2000; Sklar and Sarter, 1999).

The debriefing data from this study confirm that another contributor to breakdowns in monitoring are gaps and misconceptions in pilots' knowledge and understanding of the aircraft automation (e.g., Sarter and Woods, 1997, 2000). This problem could be addressed by adopting 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.

Also, 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. It turns the difficult cognitive task of integrating and interpreting FMAs and projecting their implications for future aircraft behavior into a less effortful and less error-prone perceptual task.

Acknowledgments

This work was funded partially by NASA Ames Research Center through Dr. Key Dismukes (Contract #NAS2-99074). The authors would also like to thank the collaborating airlines and the participants in this study as well as Xidong Xu, Bobbie Seppelt, Jack Bard, Bob Reid, Dean Holt, and Bill Shontz for their contributions to this research.

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., and Woods, D. (1996). *The Interface Between Flightcrews and Modern Flight Deck Systems*. Federal Aviation Administration Human Factors Team Report. Washington D.C.: FAA.

Bellenkes, A.H., Wickens, C.D., and Kramer, A.F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine*, 68, 569-579.

Helleberg, J. and Wickens, C.D. (2000). *Datalink modality in general aviation* (Technical Report ARL-00-7/FAA-00-4). Savoy, IL: University of Illinois, Aviation Res. Lab.

Helleberg, J., Wickens, C.D., and Goh, J (2003). Traffic and data link displays: auditory, visual or redundant? A visual scanning analysis. In R. Jensen (Ed). *Proceedings 12th International Symposium on Aviation Psychology*.

Huettig, G., Anders, G., and Tautz, A. (1999). *Mode awareness in a modern glass cockpit attention allocation to mode information*. Proceedings of the 1999 Ohio State University Aviation Psychology Conference. Dayton, OH: Ohio State University.

Mumaw, R.J., Boorman, D., Griffin, J., Moodi, M., and Xu, W. (2000a). *Training and design approaches for enhancing automation awareness*. (Boeing Document D6-82577). Seattle, WA: Boeing.

Mumaw, R.J., Sarter, N., Wickens, C., Kimball, S., Nikolic, M., Marsh, R., Xu, W., and Xu, X. (2000b). *Analysis of pilots' monitoring and performance on highly automated flight decks.* (Final Project Report: NASA Ames Contract NAS2-99074). Seattle, WA: Boeing Commercial Aviation.

Mumaw, R.J., Sarter, N.B., & Wickens, C.D. (2001). *Analysis of pilots' monitoring and performance on an automated flight deck.* Proceedings of the 11th biennial meeting of the International Symposium on Aviation Psychology, Columbus, OH.

Nikolic, M., Orr, J., and Sarter, N.B. (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. (2000). The need for multisensory feedback in support of effective attention allocation in highly dynamic event-driven environments: The case of cockpit automation. *International Journal of Aviation Psychology*, 10(3), 231-245.

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

Sklar, A.E. and Sarter, N.B. (1999). "Good Vibrations": The Use of Tactile Feedback In Support of Mode Awareness on Advanced Technology Aircraft. *Human Factors, 41(4)*, 543-552.

Wickens, C.D. (2000). *Pilot expectancy and attentional effects for hazard awareness: Implications for free flight and datalink.* (Final Technical Report ARL-00-6/FAA-00-3). Savoy, IL: University of Illinois, Aviation Res. Lab.

Wickens, C.D., Helleberg, J., and Xu, X (2002). Pilot maneuver choice and workload in free flight. *Human Factors, 44 (2)*. Pp171-188.

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. and 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.