Presented at the 11th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University. 2001.

ANALYSIS OF PILOTS' MONITORING AND PERFORMANCE ON AN AUTOMATED FLIGHT DECK

Dr. Randall J. Mumaw
Boeing Commercial Airplanes
Seattle, Washington

Dr. Nadine B. Sarter
Ohio State University
Columbus, Ohio

Dr. Christopher D. Wickens
University of Illinois
Champaign-Urbana, Illinois

ABSTRACT

In order to understand the role of pilot monitoring in the loss of mode awareness on automated flight decks, we studied 20 Boeing 747-400 line pilots in a simulated flight. We developed a set of scenario events that created challenges to monitoring. We measured automation use, eye fixations, and pilot mental models. The results showed that, at an aggregate level, pilot monitoring patterns were consistent with those found in the few previous studies. However, mode awareness was affected by both failures to verify mode selections and an inability to understand the implications of autoflight mode on airplane performance.

INTRODUCTION

A key issue for enhancing flight operations safety is to support more effectively the interaction between flight crews and flight deck automation (specifically, autopilot, autothrottle, and the Flight Management Computer) (Abbott et al., 1996). The introduction of automation to the "glass cockpit" has provided numerous benefits, such as increased precision and efficiency. However, these benefits occur primarily in situations where the automation performs tasks that don't require pilot involvement. In circumstances that require cooperation and coordination between pilots and automated systems, unexpected problems are being encountered (Sarter, 2000).

Numerous recent studies (e.g., Sarter & Woods, 1994, 1995, 1997) have demonstrated that pilots can become confused about the state and/or behavior of flight deck automation. One consequence of breakdowns in pilot-automation coordination is the pilot's loss of mode awareness. Mode awareness refers to the knowledge and understanding of the current and future state and behavior of the automation. This loss of mode awareness can lead to mode errors and automation surprises. Mode errors, generally speaking, occur when a pilot performs an action appropriate for the assumed system state but not for the actual state. Or, a mode error can refer to the omission of a required action or intervention with automation actions. Mode errors lead to automation surprises when the pilot notices that the automation is engaged in activities that were not commanded (or, fails to engage in activities that were thought to be commanded). Both mode errors and automation surprises have played a role in recent incidents and accidents and can lead to poor or slow compliance with ATC clearances (e.g., deviations from assigned altitudes).

Several factors can contribute to a loss of mode awareness:

- the pilot can have an incomplete and/or inaccurate mental model of the flight deck automation.
- the automation feedback can be inadequate because it fails to support pilots in predicting, assessing, and understanding current system state and behaviors.
- the highly complex logic underlying flight deck automation behavior that differs from pilots' reasoning about their flying tasks, and differs considerably across manufacturers, aircraft types, and in some cases, across individual planes within type (due to software upgrades).

One avenue for removing these problems is to modify the flight deck interface to increase the salience of changes that can occur without explicit pilot commands. In addition, pilots need better support for interpreting the indicated automation state in terms of its implications for current and future aircraft behavior. New flight deck interface designs are being developed with these requirements in mind. However, design changes take considerable time to find their way into the fleet, and solutions are needed for the existing fleet. Therefore, new approaches to automation training also need to be developed and implemented. In fact, efforts are currently underway to enhance pilot training and improve pilot mental models of the flight deck automation (e.g., Mumaw et al., 2000a).

A second approach to addressing mode confusion relates to the fact that pilots are not well supported in learning how to monitor automation-related indications effectively. The "accepted wisdom" on scanning cockpit indications for years was based on the "T" pattern of primary indications (airspeed, attitude, altitude, and heading). However, with the advent of integrated flight deck displays and highly complex automated systems on glass cockpit aircraft, the pilot needs to monitor a larger, more diverse, and more distributed set of indications. Although the primary flight display (PFD) and the navigation (Nav) display integrate most of the primary indications, the pilot also needs to monitor the mode control panel (MCP) and the flight management computer (FMC), which is accessed through the control data unit (CDU). There are no documented strategies for effectively monitoring this diverse set of indications, and, as a result, pilots often develop their own—not necessarily effective—approaches to the task.

To better understand monitoring, we sought to answer several questions:

1. How do pilots monitor indications on automated flight decks?
2. What information do they access? When, in what sequence, and for what purpose do they access it?
3. How does the interface design support or hinder effective monitoring?

The present study serves to address these questions. Pilots were asked to fly a scenario involving events that are known to challenge monitoring. Our goal was to identify pilots' monitoring skills and strategies and relate those to performance outcomes, their understanding of the automation, and the feedback provided by the system. Particular scrutiny was focused on the pilots' scanning of the Flight Mode Annunciations (FMAs), which are designed to alert pilots to mode changes. Failure to monitor these when changes occur could be tied to of a loss of mode awareness. In addition to assessing mission performance, we recorded eye fixations and assessed each pilot's understanding of key automation concepts.

GENERAL METHOD

We recruited twenty 747-400 line pilots (10 Captains and 10 First Officers; all male) from two U.S. airlines. Ages ranged from 45-59, with a mean age of 53.3. Pilots had between 100 and 9000 hours on the 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 participation.

Each data collection session proceeded as follows:

1. complete informed consent and demographics
2. review flight plan and clearance, charts, and dispatch papers
3. eye-tracker calibration
4. simulator familiarization
5. scenario/data collection
6. 10-minute break
7. debrief on pilot actions
8. conduct mental model test

The study was carried out in a 747-400 fixed-base simulator. Each pilot's front window view covered 45° horizontally and 34° vertically, with a 2° look-down angle. Pilots were encouraged to use flight deck automation (i.e., to not fly manually) until they descended to about 5000 ft, and then a visual approach.

SCENARIO

In collaboration with one of the participating airlines we developed a scenario that allowed us to evaluate monitoring in situations where the flight deck interface can create impediments to a full assessment of current and future automation behavior. We developed the following set of scenario events:

<u>Event 1.</u> Runway change - ATC requested a runway change during initial taxiing. This change affects the FMC: the take-off speeds (v-speeds) are deleted and need to be reselected, a route discontinuity is created after the SID, and a hard restriction at a waypoint is lost since it wasn't part of the standard SID.

<u>Event 2.</u> Expedite to cross D6 (waypoint) at 4000 ft - ATC requested an expedited climb to cross D6 at 4000 ft., which can compel the pilot to leave VNAV (into a lower-level mode) in a high-workload situation and then recover it later. Also, this clearance leads to resetting the MCP altitude and removing the reminder for a waypoint altitude restriction.

<u>Event 3.</u> Inappropriate pitch mode - After a waypoint altitude restriction was passed and the airplane began climbing, we altered the pitch mode annunciation from

VNAV SPD to VNAV PTH, which cannot occur in this situation.

Event 4. Loss of LNAV/VNAV and visual airplane target - When the airplane reached FL200, ATC requested a level-off and new heading for traffic. Re-intercepting the course is tricky and requires monitoring which waypoint is active.

Event 5. Revise CRZ altitude - When the airplane was at FL315, ATC requested to level at FL330 (the FMC cruise altitude was FL350). Then, at some later point ATC indicated that FL330 would be the final CRZ altitude. Going through this sequence, with this FMC, results in a VNAV ALT pitch mode (instead of the typical transition to VNAV PTH for cruise).

Event 6. Reduction in airspeed to 260 kts - Late in cruise, ATC requested a speed reduction to 260 kts. This requires careful monitoring of airspeed when the transition from CRZ to DES phase occurs.

Event 7. Inappropriate pitch mode - After the airplane was established on the VNAV descent path, we changed the pitch mode annunciation from VNAV PTH to VNAV SPD, even though the vertical path indicator on the Nav display showed it was on path.

Event 8. Inappropriate autothrottle mode - After the change to VNAV SPD, while the airplane was still actually on the VNAV descent path, we altered the autothrottle annunciation to THR, which is not a mode one would see in this situation.

Event 9. Loss of glideslope diamond and glideslope - We failed the ground signal for the glideslope. As a result, the glideslope diamond on the PFD never filled in and centered itself. This failure was introduced because it relies on the disappearance on an indication, which is more difficult to notice than a failure that is associated with a positive indication, such as an aural warning.

SELECTED RESULTS

The NASA report (Mumaw et al., 2000b) contains the analysis of the full data set. In this short paper, we focus on just two issues: overall scanning patterns and awareness of FMA indications.

Overall Scanning Patterns

The initial analysis of the eye-fixation data focused on how fixations were distributed to each area of interest (AOI). A single data set could be analyzed at two levels, depending on the precision of the data. At the coarsest level, data could be analyzed into the following set of seven AOIs: PFD, Nav display, out the window, CDU, upper EICAS display, lower EICAS display, and the MCP. We were able to reliably analyze the data from 17 pilots with this set of AOIs. At a more fine-grained level, fixations within the PFD could be analyzed into the following smaller set of seven AOIs: attitude, altitude, airspeed, heading, and each of the three FMAs. We were able to reliably analyze the data from 14 pilots with this set of AOIs.

Tables 1 and 2 show the percentage of dwell time that these pilots spent in each AOI for each of five major flight phases. Note that the Table 2 values are percentages of PFD fixations. For example, during take-off, PFD airspeed was fixated 26% of the time that the PFD was fixated. Thus, PFD airspeed fixations were 26% of 14%, or approximately 3.6% of all fixations. 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—e.g., looking down at a clipboard. We can use these data to answer two questions:

1. How do the percentages relate to those found in other studies of glass cockpit (Heuttig et al., 1999), and non-automated cockpit (Wickens, 2000; Wickens et al., 2000, Helleberg & Wickens, 2000; Bellenkes et al., 1997) visual scanning?
2. How do the scanning parameters change across the five major phases of flight: take-off, climb, cruise, VNAV/FMS descent and ATC vectors descent and landing?

Table 1 reveals the dominance of the PFD, followed closely by the Nav display. The PFD value of around 35% (averaged across the last four flight phases) is consistent with the glass cockpit study of Heuttig et al. (1999), who reported a value of around 40%. Correspondingly, the Nav display value of around 25% is similar to the 20% value reported by Heuttig et al. The PFD value of 35% observed in the current data is substantially less than the 50-60% PFD values reported by Wickens et al. (2000) in the general aviation (GA) studies involving CDTI and data link technology. In addition to possible differences in pilot skill level, a major reason for the difference between these studies can be attributed to the presence of the Nav display in the glass cockpit airplanes of the current study. This display contains much of the information that would otherwise be presented in the heading display of the GA aircraft. Indeed, when we sum pilot attention allocated to the Nav display and the PFD, the total of 60% agrees closely with data from the GA studies.

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

A second parallel with the glass cockpit data of Heuttig et al. is the very small percentage of time spent attending to the FMAs (Table 2). These values were consistently slightly less than 2% of the total fixations. Also consistent with the data of Heuttig et al. is the relatively small amount of time fixated out the window (Table 1). Heuttig et al. found this to be approximately 10%. In the current study, this value was also about 10% during the vector final approach (and less during earlier phases in the air). Interestingly, the relatively low out-of cockpit scanning value is consistent with the conclusions drawn by Wickens (2000) in the two GA technology studies; those pilots also "undersampled" the outside world, relative to the desired value (and that specified in pilot training guidelines).

The second way of examining the data in Tables 1 and 2 is to consider the shifts in scanning behavior (and hence perceived AOI importance) across phases. There are a number of predictable, and expected, effects. The most obvious of these is the large drop in out the window scanning after the plane leaves the ground (Table 1) and its subsequent but smaller increase after the plane approaches the ground in the final segment. Also of interest, although small in magnitude but still statistically significant, is the increase in the MCP attention during climb and descent, relative to cruise. In Table 1, the Nav display receives its greatest attention during the FMS portion of the descent, whereas the CDU receives greatest interest during the cruise phase. Note that across the four in-flight phases, there is a nearly reciprocal relation between attention to the CDU (automation concerns) and attention to the PFD (flying concerns).

Within the PFD (Table 2), attention to the dominant ADI remains relatively unchanged across flight phases. The altimeter shows an appropriate increase of interest during vertically changing flight (climb and descent) with slightly greater interest on climb than descent. Airspeed, in contrast, shows the reversed relationship, with greater attention received during the descent than the climb. Finally, as we have noted before, interest in (attention to) the FMAs is greatest during take-off.

In summary, the general pattern of monitoring is consistent with the smaller glass-cockpit data set obtained by Heuttig et al. (1999), and also shows a pattern consistent with the data collected in GA flight, whether pilots are given an outside scene (Wickens, 2000; Wickens et al., 2000; Helleberg & Wickens 2000) or not (Bellenkes et al., 1997).

Awareness of FMA Indications

One concern with the current glass cockpit interface is that pilots are required to monitor the FMA area of the PFD to know autoflight configuration. Pilots are told during training that they need to monitor there, but because the MCP is where they make mode selections, FMA monitoring may take on a lower priority.

We determined how frequently pilots actually scanned the FMA across a set of representative cases. We classified mode changes into three groups:

- Manual (M) – when the pilot manually selects a new pitch or roll mode (e.g., FLCH, HDG SEL) by engaging a switch on the MCP. The pilot should monitor to observe that the selected mode engages.

- Automatic-Expected (AE) – when a mode change is initiated by the automation, but it is a change that the pilot expected. E.g., when the pilot is climbing in fight level change (FLCH) mode to an altitude he set in the MCP altitude window, the pitch and autothrottle modes will change as he begins to capture and level off. The pilot should monitor to verify that the level-off actually occurs and new modes engage.
- Automatic-Unexpected (AU) – when a mode change is initiated by the automation and there is no perceptible airplane performance difference associated with the change. For example, as the airplane descends from a level altitude, the auto-throttle mode will start in IDLE and eventually will transition to HOLD mode, meaning that the pilot can reposition the throttles if he desires. Our mental model data suggest that pilots are not as familiar with these transitions and their timing.

The data showed that pilots often "failed" to fixate the FMA within the first 10 seconds (during which time a green box appears to highlight it). The percentages of cases in which pilots failed to fixate in that first 10 seconds were as follows: M = 53%, AE = 45%, AU = 62%. In other cases, pilots fixated near the time or near the FMA. However, if we examine (more liberally) failures to fixate any FMAs in the first 20 seconds after the green box appears, the failure rates are still high: M = 32%, AE = 29%, AU = 40%. Thus, for a considerable percent of cases pilots do not verify the FMA change, and further, Unexpected mode changes are verified less frequently than the Manual or Automatic-Expected mode changes. This failure suggests that the attention-attracting properties of the green box that accompanies every mode change may be an insufficient cue (Nikolic, Orr, & Sarter, 2001).

In three separate Events (3, 7, 8), we changed the FMA artificially (and unknown to the pilot; in fact, for these cases the standard green box did not accompany the mode change). Table 3 shows the three cases where a change occurred. The second column indicates how many pilots fixated the relevant FMA while that change was in effect (note that data were not available for all pilots at this level of precision). The last column shows that in only 1 of the 32 total cases did a pilot notice that the FMA was inappropriate. Thus, even when scanning included the FMA, pilots failed to understand the implications of the FMA.

This finding was also true for Event 5, in which most pilots were in VNAV ALT (instead of VNAV PTH) during cruise. Twelve of the 16 pilots who were in VNAV ALT failed to take action to descend at or prior to the top of descent point. Fixating the pitch mode did not lead pilots to generate correct implications for airplane performance.

Other Data Collected

We collected detailed performance data on automation use during the scenario and found several areas in which pilots were not proficient. One notable finding is that there was considerable confusion about the altitude intervention button (which provides very poor feedback). No pilot used this button correctly in all cases.

At the conclusion of the simulator session we interviewed each pilot to assess his understanding of various automation concepts. Through these interviews, we found that pilots typically could state correct expectations about common mode behavior. For example, 16 of 20 pilots indicated that they expected to see VNAV PTH as the pitch mode during cruise. However, few pilots applied this knowledge effectively during the simulator session, where they flew cruise in VNAV ALT. Also, although pilots were generally correct in the information they offered, they provided little information on subtler automation features. Detailed knowledge of VNAV SPD and VNAV ALT, in particular, was not offered. Note, however, that we recorded only what pilots volunteered, and it is possible that pilots knew these details but chose not to offer them in this setting.

DISCUSSION

The larger set of results of this study (see Mumaw et al, 2000b)

1. reconfirmed that pilots have difficulty using certain elements of flight deck automation correctly.

Table 3. Number of pilots detecting artificial mode changes.

	# Fixating	**# Detecting Change**
VNAV PTH on CLB	12	0
VNAV SPD on DES	10	1
THR on DES	10	0

2. demonstrated that pilots “fail” to verify mode changes on the FMA in 30-60% of cases, calling into question the effectiveness of the green-box cue.
3. revealed that even when pilots do fixate the mode annunciation, they often fail to understand the implications of that mode for airplane performance.
4. bolstered the idea that pilots have shallow knowledge of automation concepts, especially regarding VNAV SPD and VNAV ALT.

Given the complexity of the flight deck interface, we believe that effective knowledge-driven monitoring is critical for effective fight operations; even more important than data-driven monitoring. These data suggest that pilots have insufficient knowledge of automation behavior to anticipate important automation state changes. Training to improve pilot mental models and to improve monitoring strategies needs to be developed until flight deck interface improvements can be established in the fleet.

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, Mark Nikolic, Roger Marsh, Wei Xu, Xidong Xu, 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., and Woods, D. (1996). *The Interfaces Between Flightcrews and Modern Flight Deck Systems.* Federal Aviation Administration Human Factors Team Report. Washington, D.C.: Federal Aviation Administration.

Bellenkes, A.H., Wickens, C.D., & 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., & 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.

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

Mumaw, R.J., Boorman, D., Griffin, J., Moodi, M., & 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., & 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.

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

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., & Woods, D.D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots’ model and awareness of the flight management system. *The 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.

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., Xu, X., Helleberg, J.R., Carbonari, R., & Marsh, R. (2000). *The allocation of visual attention for aircraft traffic monitoring and avoidance: Baseline measures and implications for free flight.* (Technical Report ARL-00-2/FAA-00-2). Savoy, IL: University of Illinois, Aviation Res. Lab.