Mode Monitoring and Call-Outs: An Eye-Tracking Study of Two-Crew Automated Flight Deck Operations

Caroline M. Björklund and Jens Alfredson
Swedish Defence Research Agency
Linköping, Sweden

Sidney W. A. Dekker
Center for Human Factors in Aviation, IKP
Linköping Institute of Technology

Mode awareness has been suggested as a critical factor in safe operations of automated aircraft. This study investigated mode awareness by measuring eye point of gaze of both pilots during simulated commercial flights, while recording call-outs and tracking aircraft performance. The results of this study show that the compliance to manufacturer or air carrier procedures regarding mode monitoring and call-outs was very low. However, this did not seem to have a negative effect on the flight path or safety during our observations. Crews exhibited a proliferation of strategies to keep track of status and behavior of the automation, often with little reliance on the flight mode annunciations of the primary flight display. The data confirm the limitations of current flight mode annunciator designs, and suggest that mode awareness is a more complex phenomenon than what can be captured by measuring eye point of gaze and communication alone.

Breakdowns in human–machine coordination have been a recurring problem in aviation, particularly in modern automated aircraft with glass cockpits (Degani, Shafto, & Kirlik, 1995; Dekker, 2000; Dornheim, 1995; Lyall, 1998; Nikolic & Sarter, 1999; Sarter, 1997; Vakil, Hansman, Midkiff, & Vaneck, 1995; Woods & Sarter, 2000). Glass cockpits contain considerable automation, the functions...
of which are governed by automation modes. A mode can be described as a condition in the machine that corresponds to a unique behavior or a manner of behaving as well as to a device state that controls or displays functions in a distinct way or has a distinct meaning (Lyall, 1998). Confusion about in which mode the automation operates can lead to automation surprises, where crew members think they have told the automation to do one thing, whereas it is actually doing another because it is in another mode. For example, the flight crew may have thought they commanded the automation to descend the aircraft at a particular flight path angle (expressed in degrees), but the same numeric command dialed into the automation gets interpreted by the system as a value in feet-per-minute because the automation is in vertical speed mode. The automation can transition from one mode to another either autonomously (because a predefined state is reached or exceeded, or an environmental parameter triggers the change) or manually by the flight crew.

Mode awareness is mentioned in the literature as a critical ingredient for avoiding automation-related problems (Funk, Lyall, & Niemczyk, 1997; Lyall, 1998; Sarter & Woods, 1995). It can be defined as “the ability of a supervisor to track and to anticipate the behavior of automated systems” (Sarter & Woods, 1995, p. 7). Lack of mode awareness, or loss of mode awareness, appears to be related to the nature of mode indications on typical flight decks (small alphanumeric state announcements in various hues) and pilots’ understanding of how the automation actually works—an understanding that can be rather limited and buggy (Woods & Sarter, 2000). The ability to track and anticipate the behavior of automated systems is thus compromised by features of the interface, properties of pilots’ mental models built up during training and operational experience, and more local challenges such as time pressure, unfamiliar situations, and plan changes (Huettig, Anders, & Tautz, 1999; Sklar & Sarter, 1999).

MONITORING OF MODE ANNUNCIATIONS

Monitoring mode annunciations on displays and calling out the mode transitions seen are both thought to be important for obtaining and keeping mode awareness on today’s flight decks. In the upper part of the primary flight display (PFD) a so-called flight mode annunciator (FMA) is dedicated to showing mode transitions. It displays the various two- to four-letter codes (e.g., V/S, FPA, LOC, LOC*) that serve to provide insight into the behavior of the automation. Experiments (Huettig et al., 1999; Mumaw, Sarter, & Wickens, 2001) discovered that pilots actually do not look at the FMA very often (typical fixation time is less than 5%), which could mean that the pilots do not value the information displayed on the FMA (Huettig et al., 1999) even though manufacturers stress that it is the only reliable source for current and expected automation mode informa-
tion (Huettig et al., 1999). Although pilots are supposed (and to some extent trained) to visually monitor the FMA, they are not given directives on how this should be done. During the first 10 sec after a mode transition, the annunciation on the FMA is surrounded by a box, highlighting the transition. Yet during these first 10 sec, a study by Mumaw et al. (2001) showed that pilots did not look at the FMA during 53% of the manually induced transitions, 45% of automation-induced transitions that were expected by the pilots, and 62% of automation-induced transitions that were not expected by them. Up to 10 sec after the box had disappeared (i.e., 20 sec after the mode transition), 32% (and 29% and 40%, respectively) of the mode transitions announced on the FMA had still not been looked at (Mumaw et al., 2001).

PROCEDURES FOR MODE MONITORING

In addition to monitoring the FMA, most manufacturers and air carriers have instituted procedures for double-checking mode transitions. One reason is to compensate for imperfect monitoring. Double-checking is supposed to occur by verbally announcing the mode transition (making a call-out) when it is visually noticed by the pilot, thereby directing the attention of the other pilot to the status and behavior of the automation. This in turn should lead to some type of common ground, whereby the entire crew is aware of what to expect of the automation next. This type of procedure is not always followed, as call-outs quickly get pushed aside by other more pressing tasks. Pilots report that this is especially true in higher workload situations, ironically the situations in which accurate mode awareness can be critical. Of course, call-outs are also not likely to occur when none of the crew members have discovered a mode transition in the first place. Finally, a number of call-outs are not made because the mode transition is as inevitable as it is normal and expected (e.g., transitioning into FLARE mode just before touchdown onto a runway).

The fact that the autoflight systems in commercial aviation are not standardized across manufacturers is not making the interaction easier for the pilots (Goteman, 1999). On the face of it, the features often seem to be the same, but different manufacturers have their own design philosophy for selection, activation, and annunciation of the different modes (Corwin, 1995). As system complexity increases, the process of standardizing procedures becomes more difficult and more costly, but also more important. It allows pilots who have not previously worked together to know what to expect of each other and thereby be able to perform their duties more efficiently without unnecessary small talk to reach a common ground (Degani & Wiener, 1994).

If these procedures are illogical or somewhat inconsistent, it may lead to deviations by the pilots. According to Degani and Wiener (1994) these deviations can
arise because of the fact that pilots are individuals with biases, prejudices, opinions, experiences, and self-concepts—the same reasons that often make the human a better problem solver than a procedure-following system. Other deviations from procedures may be because of decreasing vigilance, a tendency to be overtrustful of the system, attempts to break the monotony of the sometimes highly proceduralized situations, or simply frustration if the procedures somewhat hinder the pilots in other duties.

EMPIRICAL MEASUREMENTS OF MODE AWARENESS

Common ways of getting empirical access to mode awareness include ocular indication instruments such as eye point of gaze (EPOG) (e.g., Mumaw et al., 2001), verbal protocols (Sklar & Sarter, 1999), and pilot performance measures and flight progress data (Sarter & Woods, 1995). Measuring pilot EPOG to see if a mode transition is discovered can be a relatively objective method, at least when assuming that the eye–mind hypothesis is true (Fox, Merwin, Marsh, McConkie, & Kramer, 1996). This holds that information is extracted from the area in the visual field that is covered by the fovea. Although individual pilot EPOG studies have been performed, to date there are few or no studies that combine the EPOG of two pilots (the common situation in commercial aviation), together with verbal protocols and other indicators and determinants of performance (task load, flight parameters, pilot roles and responsibilities). The work described here aims to help fill that void. It aims to gain a better understanding of the creation and breakdown of crew mode awareness in a normal operating environment. Specifically, it empirically examines the links between the perception of mode display annunciations, verbal call-outs, pilot roles (flying or nonflying pilot) and pilot rank (captain, copilot). It also sheds light on how the procedures guiding these factors are handled by the crew, and the potential consequences these factors may have on flight progress and the aircraft flight path. As the interest here was in two pilots’ monitoring and call-out behavior during normal flight conditions, task load was not varied systematically except in the case of a go-around at the end of the flight.

METHOD

Participants

Twelve professional pilots participated voluntarily. The pilots came from four air carriers with minimal differences in their flight operational procedures, all of them prescribing visual verification and call-outs when mode transitions occur.
All pilots were male, aged 25 to 46 (M = 33.7 years), with an equal number of
captains and copilots. They had between 1,850 and 9,000 hr of flying time (M =
4,652 hr). The pilots were paired into crews based on their scheduling availability,
which is to say, as randomly as during real airline operations.

Apparatus

The flights were performed in a high-fidelity, motion-based flight simulator with
Boeing 737NG (New Generation) displays, where each pilot had a PFD located
next to the navigation display (Figure 1). The EPOG of both pilots was mea-
ured by a head- and eye-tracking system, with an EPOG recorder called
GazeTracker (Zon & Mooij, 1996). The GazeTracker is a helmet-mounted sys-
tem that consists of an eye-tracker and a head-tracker communicating with an
EPOG program. It allows pilots to perform free head movements during their
cockpit work. The eye-tracker receives input from a helmet-mounted
eye-camera optics module and a visor. The system works by detecting the center
of the pupil and the corneal reflection, when the eye is illuminated by infrared
light. The angle between the two detected points makes it possible to determine
the relative position of the eye. The head-tracker receives input from a hel-
met-mounted electromagnetic receiver. The data were used for both online mon-
itoring and offline analysis.

FIGURE 1  Eye point of gaze registration superimposed on a photograph of the primary flight
display (the display to the left) and of the navigation display. The FMA is located on top of the
primary flight display.
EPOG was registered across a defined space, and dwell times of the reflex vector from pupil to cornea were measured. The dwell time is specified as several single fixations within a visual angle of 1.5° during a minimum fixation time on 150 msec. EPOG was recorded through x-, y-, and z-coordinates for visual focus within the cockpit, and through x- and y-coordinates for focus on the simulator displays. The pilots, their respective visual fields, and the cockpit displays were recorded by a total of six video cameras. These also registered and kept the time of all flight deck sounds including pilot verbalizations. The simulator itself tracked a large number of parameters for aircraft performance and flight path.

Procedure

Before the flight, crews were informed about the procedure of the experiment and had a few minutes to acquaint themselves with the equipment. The pilots were not informed of the specific research interest in their automation mode monitoring or call-out behavior, but were asked to partake in the assessment of new EPOG and workload measurement instruments. After the simulation, the pilots gave permission for the analysis of their performance data.

The pilots were assigned according to their actual rank (captain, first officer) but, as per normal, decided for themselves who would be the pilot flying (PF) versus the pilot nonflying (PNF). Each crew was to fly a normal out-and-return flight from Amsterdam (AMS) to London (LHR). Flying time between Amsterdam and London is typically under 1 hr, and normally takes place in busy controlled airspace. The crews were to follow a normal climb-out from Amsterdam, stepwise climbing through flight levels 80, 100, 120, 140, and finally to cruising flight level. The route to be programmed into the flight management system, and to be followed, was a normal company route into the London Heathrow area, across its northeast cornerpost. At the end of the scenario an event was introduced to the pilots with the intention of increasing their workload. The higher workload event was a glide slope capture failure on ILS RWY 27L that forced a go-around. The glide slope signal was corrupted and the aircraft passed through the glide slope and stayed at 2,500 ft. The pilots were told that the cloud base was at 300 ft, so that a LOC/DME approach was not possible. The missed approach procedure required the pilots to descend to 2,000 ft. They turned left at 0.0 DME because they followed track 150. Finally, they landed on RYW 27R.

The crews’ mental effort was assessed by a rating scale mental effort (RSME)—a Likert scale indicator that pilots used to subjectively rate their effort on particular sets of tasks. The ratings of mental effort were taken after the flight to ensure that the pilots were not distracted from their tasks during the simulation. The results could be synchronized with other data traces. EPOG data, crew communication, and flight and system parameters were also synchronized.
Mode transitions derived from flight data files were used as triggers to study EPOG behavior and pilot communication around those times, employing the following coding categories:

Visual verification of mode transitions: Registration of EPOG on the FMA within 10 or 20 sec. The mode transition is highlighted for 10 sec on the FMA. If gazes fell on the FMA more than 20 sec after the mode transition, this was taken to be “no visual verification” of the mode transition.

Formal mode related call-out: Verbal announcement of manual mode selections and FMA entries made by crew members themselves, and of FMA annunciations. An example of a formal call-out could be “Altitude capture,” indicating that the autopilot is now leveling off at a preset altitude, using its altitude capture mode. The call-outs were categorized on the basis of occurrence before visual verification, after visual verification, or whether they did not occur at all.

Informal communication: Mode-related but informal communication is not a formal call-out as per the airline procedures. Informal communication surrounding altitude capture could, for example, be “Coming up to one-three-zero, capture” (referring to flight level 130). Like call-outs, informal communication was categorized on the basis of occurrence before visual verification, after visual verification, or of no occurrence at all.

The experiment was set up as a factorial between-subject design, using proportion tests as the statistical method for analyzing the data.

RESULTS

Mode transitions occurred 521 times during the 12 flights. All were available for verbal analysis, but due to technical and calibration issues, only 418 mode transitions were available for EPOG analysis. These consisted of 247 pilot-induced and 171 automation-induced transitions. The 418 useful mode transitions generated twice as many data points (for two pilots) and these 836 were taken as triggers for further investigation into the surrounding circumstances (e.g., task load), visual verifications, verbalizations, and aircraft behavior. The study generated the following results.

Visual Verification

About 40% of mode transitions were never visually verified. Within 10 sec, 47% were visually verified and 13% between 10 and 20 sec (Figure 2). There was no
significant difference in visual verification between manually induced or automation-induced mode transitions ($p > .05$). The captain was the PNF with the first officer as the PF in 64% of the studied mode transitions. There were differences in visual behavior between PF and PNF. PF visually verified mode transitions in 56% of the cases, compared with 65% for the PNF ($Z_{obs} = 2.52, p < .05$). PF also had a lower frequency of visual verifications within 10 sec (43% vs. 51% for PNF, $Z_{obs} = 2.24, p < .05$). There was no difference in verification rates between 10 and 20 sec ($p > .05$).

Differences existed between the captain and the first officer as well. The captain visually verified the transitions in 72% of the cases versus 47% for the first officer ($Z_{obs} = 7.21, p < .05$). The captain verified transitions within 10 sec in 57% of the cases versus 37% for the first officer ($Z_{obs} = 5.87, p < .05$). There was no significant difference between the captain’s and the first officer’s verifying rates for visual verification between 10 and 20 sec, which was 15% versus 11% ($p > .05$).

Call-Outs

Whereas there were 1,042 mode transition registrations (twice 512 because of two pilots), there were only 146 call-outs (Figure 3). Only 32 call-outs occurred after the pilot visually verified the mode transition on the FMA. Of those, 29 were formal call-outs as required as per the airline’s procedures, which means that they occurred in less than 3% of the mode transitions.

In addition to these formal call-outs, 44 mode transitions were accompanied by informal communication. Only 16% of these informal communications occurred after a visual verification of the mode transition on the FMA. The remaining 84% of informal mode communications, or references to changes in mode status, appeared to accompany a pilot’s reaching for a knob on the mode control panel. It could also have been triggered by some other visual cue, thus often preceding the actual mode transition. Informal mode communications accompanied mostly manual mode transitions, rather than automatic ones. Such informal references to mode status were often followed up by a visual check of the FMA.

There were no significant differences between the verbal behaviors of captain versus first officer ($p > .05$). PF and PNF differed in that PF more often verbalized
mode verifications than PNF (21% vs. 14%, $Z_{\text{obs}} = 2.71, p < .05$). PF also produced more informal communication than PNF (5.2% vs. 1.7%, $Z_{\text{obs}} = 3.06, p < .05$).

Crew mental effort was rated as moderate during most of the flight, with the go-around at London Heathrow as the only exception (Zon et al., 2003). During the go-around and missed approach, mental effort was consistently rated as higher. Although call-out frequencies were low throughout the flights, they were even lower during the go-around and the minutes following it. In the entire study, only one copilot (PNF in that case) made one mode-related call-out (“heading select”) while trying to navigate to the missed approach point. No other mode-related call-outs were recorded during the go-arounds and missed approaches of the other crews.

Strategies to Obtain Mode Awareness

Figure 4 shows the different alternatives of strategy that the pilots used as a response to mode transitions. The boxes represent the frequency of which different categories of the analysis occurs during 836 mode transition registrations. The arrows between the boxes show the direction of the separate actions that form the strategies. The size of the arrows roughly indicates the frequency of that particular strategy.

Further analysis of the data revealed a large number of strategies used by crews to keep track of mode transitions. During the flights, 18 different strategies were identified (12 when not considering the time for visual verification of the mode transition) and 9 were significantly separated from 0 ($p < .05$). The most widely used strategy was to visually verify only, without verbalization (409 times). No visual or verbal verification at all was the second most common strategy (271 times). Calling out without visual verification was also an often-used strategy (44 times before visual verification and 50 times with no visual verification at all). The strategy that airline procedures propose—to visually verify and then make a call-out—occurred 29 times in the data set.
During 12 flights, with 512 mode transitions, the official airline procedure was followed only 29 times. In other words, the experiment recorded 483 “procedure violations” with respect to the automation alone, and that within 12 hr of flying, or more than 40 automation-related procedure violations per hour. Rather than following official protocol, crews may look at the FMA and not say anything, or they may say something but not look at the FMA. Doing both, and in the right order, with a visual verification before a call-out, appeared rare in this experiment.

The FMA as it is designed today may not be a very useful basis for building mode awareness. Two out of five mode transitions on the FMA were never “seen” by the flight crews. In contrast to instrument monitoring in non-glass-cockpit aircraft, monitoring for mode transitions is likely to be based more on a pilot’s mental model of the automation that drives expectations of where and when to look. Such models are often incomplete and buggy (e.g., Sarter, 1995). Therefore, it may not be surprising that many mode transitions in this study are neither visually nor verbally verified by flight crews, and that the FMA triggered only 4% of call-outs in this study, of which one out of four was not the official call-out. The FMA did not get consulted for 40% of all mode transitions.

The flight crews did not seem to consistently check the FMA when changing modes manually. As can be expected, the mode control panel, where changes are actually made and selected settings (altitude, speed, etc.) can be seen, was used as a

**FIGURE 4** Pilot strategies when supervising mode transitions.

**DISCUSSION**

During 12 flights, with 512 mode transitions, the official airline procedure was followed only 29 times. In other words, the experiment recorded 483 “procedure violations” with respect to the automation alone, and that within 12 hr of flying, or more than 40 automation-related procedure violations per hour. Rather than following official protocol, crews may look at the FMA and not say anything, or they may say something but not look at the FMA. Doing both, and in the right order, with a visual verification before a call-out, appeared rare in this experiment.

The FMA as it is designed today may not be a very useful basis for building mode awareness. Two out of five mode transitions on the FMA were never “seen” by the flight crews. In contrast to instrument monitoring in non-glass-cockpit aircraft, monitoring for mode transitions is likely to be based more on a pilot’s mental model of the automation that drives expectations of where and when to look. Such models are often incomplete and buggy (e.g., Sarter, 1995). Therefore, it may not be surprising that many mode transitions in this study are neither visually nor verbally verified by flight crews, and that the FMA triggered only 4% of call-outs in this study, of which one out of four was not the official call-out. The FMA did not get consulted for 40% of all mode transitions.

The flight crews did not seem to consistently check the FMA when changing modes manually. As can be expected, the mode control panel, where changes are actually made and selected settings (altitude, speed, etc.) can be seen, was used as a
more dominant resource for knowledge about what the aircraft is going to do (despite manufacturer cautions). Nonetheless, in contrast to earlier findings, the results showed no significant difference between the visual verification of mode transitions induced manually or by the automation. This could in part be an experimental artifact, as the crew members may not have been as completely familiar with the layout and logic of the automated systems as in earlier studies where differences do appear.

PNF verified FMA-announced mode changes more frequently than the PF, and the captains verified them more frequently than first officers did. However, because the captain was the PNF in about two thirds of the studied mode transitions, this might reflect two viewpoints of the same reality. That the PNF visually verifies more often may be connected to the difference in visual workload between the PNF and the PF: the latter is likely occupied with flight parameters other than automation modes. The captain may verify visually more often because of his or her ultimate responsibility for the safety of the flight.

Although call-out rates were too low for statistical analysis, there is some support for the effect of workload on call-out frequency. Of all crews and flights studied, only one copilot made a mode call-out during the missed approach procedure. All others remained silent with respect to mode status, not even using informal communication to guide each other’s attention to current or future mode status. This would confirm that in higher workload situations, automation call-outs could be among the tasks to be put aside.

CONCLUSIONS AND PRACTICAL APPLICATIONS

Practical ways of dealing with mode automation awareness have traditionally included considerations of redesigning the FMA as well as tighter proceduralization of mode call-outs. When it comes to design, proposals for new regulations are taking shape around the same “old” display concepts. For example, Joint Advisory Circular ACJ 25.1329 (Joint Aviation Authorities, 2003), said:

The transition from an armed mode to an engaged mode should provide an additional attention-getting feature, such as boxing and flashing on an electronic display (per AMJ25-11) for a suitable, but brief, period (e.g., ten seconds) to assist in flight crew awareness. (p. 28)

The data from the experiment reported here show that flight mode annunciators may not really be attention-getting, whether there is boxing or flashing or not. Indeed, empirical data show (as it has before; see Mumaw et al., 2001) that the current FMA does not assist in flight crew awareness in a dominant way. In other words, regulations that affirm the design status quo may not produce better awareness.
What about proceduralizing for mode monitoring and call-outs? Is this an effective countermeasure to a possible lack of mode awareness with the current FMA? In this experiment, crews hardly followed the official procedures with respect to supervising and verbally coordinating mode transitions. There is likely a distance between official written guidance and the constraints and conditions governing actual practice. Fewer than 3% of all call-outs were formal expressions and occurred when they should occur (after visually checking the FMA). Crews actually did not communicate about the automation much, at least not in relation to mode changes. Only one out of five mode transitions were accompanied by verbal communication, and a substantial part of those were not formal call-outs. Especially manually induced mode changes are preceded by informal or implicit references to mode status.

There appears to be a diversity of locally and personally tailored “strategies” that pilots develop to try to stay ahead of what the automation is doing. In the 12 flights of this study, close to 20 such different strategies were documented. This either mirrors the lack of training and standardization, or is testimony to the difficulty of training and standardizing mode monitoring in modern cockpits.

Although automation surprises are generally rare and were not observed in this experiment, their lack or rarity may be testimony to a robustness of the pilot–pilot–machine constellation that does not wholly depend on the externally dictated logics of visually checking and verbally announcing the mode changes measured in this study. It may be due to an as yet underinvestigated resilience in crew strategies that draws on much more than an FMA and a call-out procedure. If this is true, then mode awareness is a very complex phenomenon that can be difficult to capture in experimental work alone, no matter how naturalistic.

ACKNOWLEDGMENTS

The work presented in this article was partly funded by the European Commission, and partly by the VINTHEC II project partners.

REFERENCES


Manuscript First Received: May 2004