
**Mode monitoring and call-outs:
An eye-tracking study of 2-crew automated flight deck operations**

Caroline. M. Björklund¹, Jens. Alfredson^{1*}, Sidney W. A. Dekker²

(1) Man-System-Interaction
Swedish Defence Research Agency, SE-581 11, Linköping, SWEDEN
E-mail: caroline.bjorklund@foi.se
E-mail: jenalf@foi.se *(corresponding author)
(2) Center for Human Factors in Aviation, IKP
Linköping Institute of Technology, SE-581 83, Linköping, SWEDEN
E-mail: sidde@ikp.liu.se

Abstract

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. Hardly any crew follows manufacturer- or air carrier procedures on mode monitoring and call-outs. However, this does not seem to have a negative effect on flight path or safety. Crews exhibit a proliferation of strategies to keep track of automation status and behavior, with little reliance on the flight mode annunciations of the primary flight display. The data confirms 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 EPOG and communication alone.

Keywords: Aviation, Communication, Flight Mode Annunciator, Mode Awareness, Procedures, Eye-point-of-gaze (EPOG).

Introduction

Breakdowns in human-machine coordination have been a recurring problem in aviation, particularly in modern automated aircraft with “glass cockpits” (Dornheim, 1995; Lyall, 1998; Degani, Shafto & Kirlik, 1999; Woods and Sarter, 2000). Glass cockpits contain considerable automation, whose functions are governed by automation modes. A mode can be described as a condition in the machine that corresponds to unique behavior or a manner of behaving, as well as to a device state which controls or displays functions in a distinct way or has a distinct meaning (Lyall, 1998). Confusion about which mode the automation is in can lead to automation surprises—where crewmembers think they have

told the automation to do one thing, while it is actually doing another (since 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 as feet-per-minute because the automation is in vertical speed mode. The automation can transition from one mode to another either autonomously (because a pre-defined state is reached or exceeded, or an environmental parameter triggers the change) or by the flight crew.

“Mode awareness” is mentioned in the literature as a critical ingredient for avoiding automation related problems (Sarter and Woods, 1995; Funk, Lyall & Niemczyk, 1997; Lyall 1998). It can be defined as “the ability of a supervisor to track and to anticipate the behaviour of automated systems” (Sarter and 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 flightdecks (small alphanumeric state annunciations 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 pilot mental models built up during training and operational experience, as well as more local challenges such as time pressure, unfamiliar situations and plan changes (Sklar and Sarter, 1999; Huettig, Anders and Tautz, 1999).

Monitoring mode annunciations on displays and calling out the mode changes seen are both thought to be important for obtaining and keeping mode awareness on today’s flightdecks. In the upper part of the Primary Flight Display (PFD) a so-called Flight Mode Annunciator (FMA) is dedicated to showing mode changes. 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. Mumaw *et al.* (2001) discovered that pilots actually do not look at the FMA much (typical fixation time is less than 5%, nor do they value what the FMA has to tell them even though manufacturers stress that it is the only reliable source for current and expected automation mode information (Huettig *et al.*, 1999). While 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 seconds after a mode transition, the annunciation on the FMA is surrounded by a box, putatively highlighting the transition. Yet during these first 10 seconds, one study showed that pilots did not look at the FMA during 53% of the manually induced transitions, 45% of automation-induced transitions expected by them, and 62% of automation-induced transitions not expected by them. Up to 10 seconds after the box had disappeared (that is, 20 seconds after the mode change), 32% (and 29%, and 40% respectively) of the mode transitions announced on the FMA had still not been looked at (Mumaw *et al.* (2001).

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 change as it is discovered (making a “call-out”), 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 either as call-outs quickly get pushed aside by other more pressing tasks, especially in higher workload situations (ironically the situations where accurate mode awareness can be critical). Pilots do not seem to suffer a “cattle-auction cockpit” gladly, to use Earl Wiener’s colorful characterization. And, of course, no call-outs are likely to occur when none of the crewmembers have discovered a mode transition in the first place (Goteman, 1999). 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 touch-down onto a runway).

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 and Sarter, 1999) and pilot performance and flight progress data (Sarter & Woods, 1995). Measuring pilot EPOG to see if a mode change 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. While individual pilot EPOG studies have been done, to date there are few or no studies that combine the EPOG of *two* pilots, 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 lacuna. It aims to gain a better understanding of the creation and breakdown of crew mode awareness in naturalistic settings. Specifically, it empirically examines the links between the perception of mode display annunciations, verbal call-outs, pilot roles (flying or non-flying pilot) and pilot rank (captain, co-pilot). It also sheds light on the potential consequences these factors may have on flight progress and the aircraft flight path. As our interest was in two pilots’ monitoring and call-out behavior during normal flight conditions, taskload was not varied systematically except for a go-around at the end of the flight.

Method

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 calling out of mode changes. All pilots were men, aged 25 to 46 (mean = 33.7), with an equal number of captains and co-pilots. They had between 1850 to 9000 hours of flying time (mean = 4652 hours). Pilots were paired into crews based on their scheduling availability, which is to say, as randomly as during real airline operations.

The flights were performed in a high fidelity, motion-base flight simulator with Boeing 737NG (New Generation) displays, where each pilot has a Primary Flight Display located next to the Navigation Display. Eye Point of Gaze of both pilots was measured by a head- and eye tracking system, with an EPOG-recorder called GazeTracker (Mooij & Associates, 1996). The GazeTracker is a helmet mounted system 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 was 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 helmet-mounted electro-magnetic receiver. The data were used for both on-line monitoring and off-line analysis.

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 degrees during a minimum fixation time on 150 milliseconds. EPOG was recorded through x-, y- and z-coordinates for foci within the cockpit, and through x- and y-coordinates for foci 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 aircraft performance and flight path parameters.

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, pilots were asked for their permission to analyze their performance data even for the purposes of learning about mode monitoring and call-out behavior.

The pilots were assigned after their actual rank (captain, first officer) but, as is normal, decided for themselves who would be pilot flying versus pilot-not flying. Each crew was to fly a normal out-and-return flight from Amsterdam (AMS) to London Heathrow (LHR). Flying time between Amsterdam to London is typically under an hour, 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 2500 ft. The pilots were told that the cloud base was at 300, so that a LOC/DME approach was not possible. The missed approach procedure required the

pilots to descend to 2000 ft. They turned left at 0.0 DME since they followed track 150. Finally, they landed on RYW 27R.

Crew 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 behaviour and pilot communication around those times, employing the following coding categories:

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

Mode related call-out: Verbal announcement of mode selections made by crewmembers themselves, FMS-entries made by crewmembers, and FMA annunciations. The call-outs were categorized on the basis of occurrence before visual verification, after visual verification or whether did not occur at all.

Implicit communication: Mode related, but implicit communication that is not a call-out as per the airline procedures. Like call-outs, implicit 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. The statistical method used for analyzing the data was a proportion test.

Results

521 mode changes occurred during the 12 flights. All were available for verbal analysis, but due to technical and calibration issues, 418 were available for EPOG analysis. These consisted of 247 pilot-induced and 171 automation-induced transitions. The results were not corrected for pair-wise comparison. The 418 useful mode changes generated twice as many data points (for two pilots). These 836 were taken as triggers for further investigation into the surrounding circumstances (e.g. task load), visual verifications, verbalizations and aircraft behavior. This generated the following results.

Visual verification

40 % of mode changes were never visually verified. 47 % were visually verified within 10 seconds and 13 % between 10-20 seconds. There was no significant difference in

visual verification between manually induced or automation induced mode transitions ($p < 0.05$).

There were differences in visual behaviour between Pilot Flying (PF) and Pilot Not Flying (PNF). PF visually verified mode transitions in 56 % of the cases, compared with 65 % for the PNF ($Z_{\text{obs}}=2.52$, $p < 0.05$). PF also had a lower frequency of visual verifications within 10 seconds (43 % versus 51 % for PNF, $Z_{\text{obs}}=2.24$, $p < 0.05$). There was no difference in verification rates between 10 to 20 seconds ($p < 0.05$).

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 ($Z_{\text{obs}}=7.21$, $p < 0.05$). The captain verified transitions within 10 seconds in 57 % of the cases versus 37 % for the first officer ($Z_{\text{obs}}=5.87$, $p < 0.05$). There was no significant difference between the captains' verifying rate for visual verification between 10 to 20 seconds, that was 15 % versus 11 % for the first officer ($p < 0.05$).

Call-outs

While there were 1042 mode transition registrations (twice 512), there were only 146 call-outs. Only 32 call-outs (or 3%) occurred after the pilot visually verified the mode change on the FMA. 29 of those were the formal call-out as required per the airline's procedures. An example of a formal call-out could be "Altitude capture", indicating that the autopilot is now leveling off at a pre-set altitude, using its altitude capture mode.

In addition to these formal call-outs, 44 mode transitions were accompanied by implicit communication. Implicit communication surrounding altitude capture could for example be "Coming up to one-three-zero, capture" (referring to flight level 130). Only 7 (16%) of these implicit communications occurred after a visual verification of the mode change on the FMA. The remaining 84% of implicit mode communications, or references to (changes in) mode status, appeared to accompany a pilot's reaching for a knob on the mode control panel or be triggered by some other visual cue, thus often preceding the actual mode change. As can be expected, informal or implicit mode communications accompanied mostly manual mode changes, rather than automatic ones. Such informal references to mode status were neither followed up by a visual check of the FMA, nor by a formal call-out.

There were no significant differences between the verbal behaviours of captain versus first officer ($p < 0.05$). PF and PNF differed in that PF more often verbalized mode verifications than PNF (21 % versus 14 %, $Z_{\text{obs}}=2.71$, $p > 0.05$). PF also produced more implicit communication than PNF (5.2 % versus 1.7 %, $Z_{\text{obs}}=3.06$, $p < 0.05$).

Crew mental effort was rated as moderate during most of the flight, with the go-around at London Heathrow as only exception (Zon et al., 2003). During the go-around and missed approach, mental effort was consistently rated as higher. While call-out frequencies were low throughout the flights, they are even lower during the go-around

and the minutes following it. In the entire study, only one co-pilot (PNF in that case) made one mode call-out (“heading select”) while trying to navigate to the missed approach point. No other mode call-outs were recorded during the go-arounds and missed approaches of the seven other crews.

Different ways of dealing with mode changes

Further analysis of the data revealed a large number of strategies used by crews to keep track of mode transitions. 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 ($p < 0.05$). 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 occurring “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.

Discussion and conclusions

During twelve flights, with 512 mode transitions, the official airline procedure was followed only 29 times. In other words, the experiment recorded 438 “procedure violations” with respect to the automation alone, and that within 12 hours of flying: 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 (FMA, then call-out) is extremely rare. The simplistic reflex would be to conclude that these are either bad crews, bad procedures, bad interfaces (or a bad experiment).

First to the “bad interface”. Indeed, the FMA as it is designed today may be close to useless as a basis for building mode awareness. Two out of five mode transitions on the FMA are never even “seen” by entire flight crews. In contrast to instrument monitoring in non-glass cockpit aircraft, monitoring for mode transitions is based more on a pilot’s mental model of the automation (which drives expectations of where and when to look) and an understanding of what the current situation calls for. Such models are often incomplete and buggy (e.g. Sarter, 1995) and it is not surprising, then, that many mode transitions in this study are neither visually nor verbally verified by flight crews. When flight crews changes mode manually, they do not consistently check the FMA either. As can be expected, the mode control panel, where changes are actually made and selected settings (altitude, speed, etc.) can be seen, is used as a more dominant resource for knowledge about what the aircraft is going to do (despite manufacturer warnings that this is not the case).

The FMA, presented in aircraft flight decks as the main source of knowledge about automation status, is actually nothing but. It triggers a mere 4% of call-outs (of which 1 out of 4 is not even the official call-out), and gets ignored altogether for a whole 40% of all mode transitions. Proposals for new regulations are unfortunately taking shape around the same “old” display concepts. For example Joint Advisory Circular ACJ 25.1329 (JAA, 2003, p. 28), says that

“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.”

The data from the research reported here show that flight mode annunciators are not at all “attention-getting”, whether there is boxing or flashing or not. Indeed, empirical data shows (as it has before, see Mumaw et al., 2001) that the FMA does not “assist in flight crew awareness” in any dominant or relevant way. If design really is to capture crew’s attention about automation status and behavior, it will have to do radically better than annunciating abstruse codes in various hues and boxing or flashing times.

Now to the “bad procedures”. Crews hardly follow the official procedures with respect to supervising and verbally coordinating mode transitions. There is a large distance between assumptions of official written guidance and the constraints and conditions governing actual practice. Fewer than 3% of all call-outs are the formal expression and occur when they should occur (after visually checking the FMA). But crews did not communicate about the automation much at all, 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 that was not the official call-outs. Especially manually induced mode changes are preceded by informal or implicit references to mode status. Indeed, there appears to be a large number of locally and personally tailored strategies that pilots develop in order to try to stay ahead of what the automation is doing. In the twelve flights of this study alone, close to twenty different strategies were documented. This either mirrors the lack of training and standardization, or is testimony to the sheer impossibility to train and standardize mode monitoring in modern cockpits. While call-out frequencies were altogether too low to make statistical comparisons, there is some support for the effect of workload on call-out frequency. Of all crews and flights studied, one co-pilot made one mode call-out during the missed approach procedure. All others remained silent with respect to mode status, not even using implicit communication to guide each other’s attention to current or future mode status. This would confirm that in higher workload situations, automation call-outs are among the first tasks to be put aside.

To “bad crews”. There is no evidence that the 12 pilots of this study would constitute an unusually bad (i.e. procedure-violating) sample. Their unfamiliarity with each other is no argument either, since airline pilots often do not know each other. In fact, there are indications that these crews, probably like any operational crew, are highly adaptive in

their strategies to stay ahead of the automation and the aircraft. The type of strategy differs if the pilot is the PF or PNF or the captain or first officer—particularly when it comes to visual verification of the mode transitions. Captains and PNFs make more visual verifications. 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. Interestingly, and 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 artefact, as the crewmembers may not have been as intimately familiar with the layout and logic of automated systems as in earlier studies where differences appear.

Finally to the “bad experiment”. Even though actual monitoring and verification performance appeared to be far from what official airline procedures would typically prescribe, crew communications on the 12 flights revealed no automation surprises that could be traced to a lack of mode awareness. Even the high workload of an unexpected missed approach (while accompanied by essentially no mode changes at all) produced no automation surprises or noticable difficulties in managing the autoflight system. One conclusion can be that mode awareness perhaps does not matter that much for flight safety after all. But there is an interesting experimental side-effect here: if mode awareness is measured mainly by visual verification and verbal call-outs, and crews neither look nor talk, then are they unaware of modes, or are the researchers unaware of pilots’ awareness? This poses an epistemological puzzle: crews who neither talk nor look can still be aware of the mode their automation is in, but how, in that case, is the researcher to know? Scenario design is one answer: by simply looking at where the aircraft is going, it can capture whether pilots “missed” something or not (see Sarter, 1995). In the research reported here, however, pilots “missed” nothing: there were no unexpected aircraft behaviors from their perspective. They were either not aware of the modes and it did not matter, or they were aware but the research did not capture it. The latter may be testimony to a robustness of the pilot-pilot-machine constellation that does not depend on the externally dictated logics of visually checking and verbally announcing (the things measured here). 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 phenomenon much more complex than what can be captured in experimental work alone—no matter how naturalistic.

Acknowledgement

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

References

Dekker, S. W. A. (2000): Crew Situation Awareness in High-Tech Settings: Tactics for Research into an Ill-Defined Phenomenon. *Transportation Human Factors*, 2 (1), 49-62.

Dornheim, M. A. (1995): Dramatic Incidents highlight Mode Problems in Cockpits, In: *Aviation Week and Space Technology*. January 30, 1995. 57.

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

Joint Aviation Authorities (2003). *Advisory Circular Joint ACJ 25.1329: Flight guidance system, Attachment 1 to NPA (Notice of Proposed Amendment) 25F-344*. Hoofddorp, NL: JAA.

Lyall, B. (1998): Autoflight Mode Awareness Issues: An Overview. Paper. *Presented at the FAA Mode Awareness Workshop*. Seattle, WA.

Mooij & Associates (Zon, G. D. R. and Mooij, H. A.) (1996). *GazeTracker for Real-Time, Omni-Directional and Digital Point-Of-Gaze Determination – Installation and Operation Guide*. EPOG Version 4.3, MA-95-15, December 1996.

Mumaw, R.J., Sarter, N.B., and Wickens, C.D. (2001): Analysis of Pilots' Monitoring and Performance on an Automated Flight Deck. *11th International Symposium in Aviation Psychology*. Columbus: Ohio State University.

Sarter, N. B. (1995): Knowing when to look where: Attention Allocation on Advanced Automated Flightdecks. *International Symposium on Aviation Psychology, 8th, Columbus*. Columbus, OH, Ohio State University. 239-242

Sarter, N. 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(1), 5-19.

Woods, D. D. and Sarter N. B. (2000): Learning from Automation Surprises and Going Sour Accidents. In Sarter N. B. and Amalberti R. (Eds): *Cognitive Engineering in the Aviation Domain*. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Zon G.D.R et al. (2003). *Full-scale simulator experiment*. (Tech. Rep. No. VINTHEC II-WP7-TR01). Amsterdam: NLR

Figure X. 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 occur 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.
