The ergonomics of flight management systems: fixing holes in the cockpit certification net

Gideon Singer\textsuperscript{a,b,*}, Sidney Dekker\textsuperscript{c}

\textsuperscript{a}Department of Aeronautics, Royal Institute of Technology, Stockholm, Sweden
\textsuperscript{b}Flight Operations, Saab Aircraft AB, SE-581 88 Linköping, Sweden
\textsuperscript{c}Division of Industrial Ergonomics, Department of Mechanical Engineering, Linköping Institute of Technology, SE-581 83 Linköping, Sweden

Received 16 August 1999; accepted 17 October 2000

Abstract

Recent air traffic control regulations mandate the installation of computer-based flight management systems in airliners across Europe. Integrating and certifying add-on cockpit systems is a long and costly process, which in its current form cannot meaningfully address ergonomics aspects. Two levels of problems occur: add-on systems carry many “classic” HCI failures, which could easily be addressed with modified certification requirements. Further, adding new technology changes practice, creates new skill and knowledge demands and produces new forms of error, which are more difficult to assess in advance. However, one innovative certification approach for add-on cockpit systems, based on the use of a representative population of user pilots, was found to be promising. This method minimizes the subjective bias of individual pilots in addition to defining pass/fail criteria in an operational environment. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Cockpit; Certification; Flight management system

1. Introduction

In most commercial modern aircraft today, navigation and flight path are controlled by a system called the flight management system (FMS). The interface with the crew in the cockpit is achieved by an active display that is called the control display unit (CDU). In addition, the FMS displays its planned/executed route on a navigation display (ND) and presents flight path guidance on the primary flight displays (PFD), both in front of the pilots. The FMS was originally introduced to large airliners during the eighties as an integral part of the new “glass cockpit”. It was evaluated and certified as a novel system by way of extensive simulations and crew evaluations. Support for the certification of such new systems on the basis of ergonomic criteria has slowly gained pace with researchers (Harris, 1997) as well as aviation regulators (FAA, 1996; Courteney, 1999).

In the late 1990s, ATC regulations were published that mandate the installation of FMS in all aircraft that use medium and upper airspace in Europe. This has fuelled a demand for cheap “add-on”, “Off-the-Shelf” systems for integration in older airliners of all sizes. Operators turn to these systems as a minimum-cost solution to comply with new requirements, but by doing so they buy a host of problems too. For aircraft that were designed without an FMS or any supporting hardware (data buses, large flat-panel displays, etc.), the addition of an FMS creates specific ergonomic problems at two levels, quite apart from obvious technical challenges:

- FMS’s were never designed for the after-market; instead they are manufacturer spin-offs from larger, integrated cockpit projects. The integration of such an FMS with older cockpit hardware often makes features common to other FMSs vanish, and produces new and exacerbates existing human–computer interaction (HCI) problems.
- The introduction of new technology in an older system creates unanticipated side effects, by demanding new knowledge and skills, upsetting common practices and creating new error opportunities.
The demand for add-on FMS creates a large ergonomics challenge. It is set against the backdrop of complex issues in human factors verification and validation (Wise et al., 1993). Currently, validation of how add-on FMS interfaces with the crew, displays information or reacts to erroneous crew inputs is ill defined, which will be shown later in this paper. Furthermore, each vendor is free to adapt their own design philosophy and method of showing compliance with the few underspecified certification criteria that do exist. The result is a confused market of subjectively approved systems, which pose both classic and novel ergonomic problems to pilots, possibly in circumstances where they can least afford them (e.g., Wiener, 1989; FAA, 1996; Singer, 1998). Verification and validation activities have waned and their development has basically stagnated (Hopkin, 1993). This while increased automation has actually not only amplified the need for them, but also shown repeatedly that existing validation criteria are inadequate or irrelevant with respect to highly computerized systems. One reason is that automation has increased the complexity and coupling of human–machine systems (Wise and Wise, 1993) to an extent that makes it hard to predict how and where performance problems will occur (Wise et al., 1991).

The role of the certification process was well documented in the book Human Factors Certification of Advanced Aviation Technologies (Wise et al., 1994) regarding modern cockpit designs and air traffic control (ATC) certification goals and methods. The need for testing each aircraft type integrated with each FMS make, when a significant difference is recognized, is highlighted (Gilson and Abbot, 1994) and the need for several layers of evaluation including simulation experiments is recommended (Small and Rouse, 1994). A more detailed review of existing certification requirements is presented and the adequacy of the present evaluation process questioned due to the limited experience of the evaluators with the new design (Paries, 1994). The kinds of errors and their relation to accident-risk is discussed and the conclusions made point towards the need for measuring the risk for accidents rather than taking an engineering approach to human reliability (Amalberti and Wilbaux, 1994). One could summarize this book by using the editors concluding remarks: There was remarkably little discussion during the workshop about research, and particularly about any supporting laboratory work for the application of human factors in certification (Hopkin, 1994). Our paper will try to show a method to achieve a more laboratory oriented but operationally valid way of addressing the concerns of certification that will capture risk for accidents using a wide population of users and pre-defined objective criteria.

This paper discusses the two levels of problems with add-on FMS and produces some guidance that is not only relevant to the aviation industry but to all kinds of domains that struggle with the integration of new technology in ongoing practice. For clarity, today’s aircraft certification process is briefly discussed first, followed by a presentation of applicable requirements and the problems associated with them.

2. Today’s certification process

Aircraft design, development and certification has evolved during the years into a well-structured and thorough process. When an aircraft manufacturer has installed a new system like the FMS, the following Airworthiness Approval process is usually being followed:

- An application is made to the certification authorities in the form of a detailed Test Description. This defines new system functions, explains which specific design requirements it will comply with and lays out how compliance will be shown.
- The system goes through a phase of Development Testing, which is an iterative process. Changes are made to optimize the system for its intended use.
- The Certification Test Plan is the document of the agreed upon process that must be presented to the aviation authorities. This testing is then performed by either the authority or is delegated to the manufacturer.
- Fit, Form and Function are evaluated by several test pilots engaged in the development phase. This includes the ergonomics of the interface units, interface with other controls in the cockpit and readability in different lighting conditions. This process is very subjective and is based on the experience of the reviewing pilots.
- A Certification Test Report is produced based on the previously agreed Test Plan. This report shows how each item has been tested successfully and forms the basis for aviation authority approval. Most system performance and reliability results must comply with objective guidelines and are tested and reported in such a manner.
- Crew Evaluation Certification Reports are used to cover all items that the engineers do not have test or analysis results for. These are all the qualitative requirements that are stated in the requirements and that relate to the way the pilot communicates with the system.
- When all the above-mentioned documents have been approved the system is declared Airworthy, meaning that the installation in this specific aircraft has been found safe for use.

In order to approve the system for use in the aircraft when carrying passengers (or cargo), each operator (airline) has to pass an Operational Approval that
includes the following:

- Documentation showing how the new system will be integrated into airline procedures is to be approved by the local operational authorities.
- The airline must submit a full training plan for pilots and maintenance personnel including detailed simulator training syllabus and aircraft system monitoring programs.
- In case of flight-critical systems, a track record for all aircraft and pilots in the fleet is required. This track record states the number of flights each aircraft/crew used the system and whether the process was successful.

The airworthiness and operational approval process is covered by detailed requirements and creates a clear trail of compliance documents. However, what are the contents of the rules that apply? Do they say anything meaningful about the ergonomic aspects of, in these cases, after-market cockpit automation systems? This question is addressed in the following section.

3. Present certification requirements

Any aircraft in service must meet rules and regulations set by aviation authorities. In the Western world today, most countries follow the rules set by the American Federal Aviation Administration (FAA), the Joint Aviation Authorities (JAA) for most European countries, or close derivatives of such rules. The Airworthiness rules see to it that the aircraft will be designed to certain standards, and meet system safety levels and minimum levels of control in all foreseeable conditions. In the case of large commercial aircraft (above 5.7t) the set of requirements are called Federal Aviation Regulations Part 25 (FAR 25) or the Joint Airworthiness Requirements (JAR 25).

The fact that an aircraft is built to the airworthiness standards does not clear it for operations. Separate requirements exist for each nation and type of operations that mandate the rules set for the airline regarding aircraft equipment; its use, crew training, duty time, maintenance etc. In the USA, operational rules are regulated by FAR 121 or FAR 135 and in Europe JAR-OPS. The mandatory FAR/JAR 25 paragraphs are usually very generic regarding new technologies. Nevertheless, when applying for approval of a new FMS installation, several FAR/JAR rules must be complied with.

The following paragraphs cover some of the mandatory design requirements that apply in this case:

\[
\text{FAR/JAR 25.671(a): Each control and control system must operate with the ease smoothness and positiveness appropriate to its functions.}
\]

The terms \textit{ease, smoothness and positiveness} are actually inherited from the “classic” cockpit with levers, switches and buttons. In the domain of electronic screens and keyboards this requirement is very vague since feedback from computers is rarely “felt”. How do these terms translate into software dialogue on a modern display?

\[
\text{FAR/JAR 25.771(a): Each pilot compartment and its equipment must allow the minimum flight crew to perform their duties without unreasonable concentration or fatigue.}
\]

Modern cockpits require almost no physical effort, but the means of measuring mental fatigue are not described and neither are the acceptable levels stated. How should one use this requirement to measure an acceptable level of concentration of fatigue when programming an obstinate FMS in difficult flying conditions?

\[
\text{FAR/JAR 25.777(a): Each cockpit control must be located to provide convenient operation and to prevent confusion and inadvertent operation.}
\]

This paragraph mainly addresses lever and switch design to prevent confusion and inadvertent operation. Implementing the same requirement for the FMS is impossible due to the infinite number of possibilities and probably impossible to achieve (since prevent is an absolute term). When using the CDU in turbulent flight conditions, even simple slips are unavoidable especially if the screen is of a Touch Screen-type. HCI terms (e.g., Fault tolerance, Undo) are missing and thus not required by the authorities.

\[
\text{FAR/JAR 25.1301(b): Each item of installed equipment must be labeled as to its identification, function or operating limitations, or any applicable combination of these factors. (d) \ldots function properly when installed.}
\]

This is a straightforward rule regarding normal cockpit controls. But how does one define labels on a display? When checking existing displays it becomes clear that there is no standard and that more guidance is needed. Even here, terms like effect of error should be used. The requirement to function properly when installed is easy to evaluate on mechanical levers, but how can this be achieved with any certainty on an FMS? In the PC environment, display architectures have become standardized to some extent (e.g., ISO 9241, 1997–1998). Such guidelines are missing for cockpit interface systems like the FMS.

\[
\text{FAR/JAR 25.1329(f): The system must be designed and adjusted so that, within the range of adjustment available to the human pilot it cannot produce hazardous loads on the aeroplane, or create hazardous deviations in the flight path, under any condition of flight appropriate to its use, either during normal operation or in the event of a malfunction, assuming that corrective action begins within a reasonable period of time.}
\]

This certainly is difficult to test in advance. For example, take the pilot who enters a plausible navigation beacon in the FMS, but which is interpreted differently...
by the computer (Aeronautica Civil, 1996). Such HCI
miscommunications are hard to anticipate.

FAR/JAR 25.1523(a): The minimum flight crew must be established so that it is sufficient for safe operation considering the workload on individual crewmembers (see Appendix D).

FAR 25-Appendix D: Criteria for determining minimum flight crew (a)(3) Basic workload functions ... Navigation. (b) Workload factors. The following workload factors are considered significant when analyzing and demonstrating workload for minimum crew determination: (1) The accessibility ease and simplicity of operation of all necessary flight, power, and equipment controls ... (8). The communications and navigation workload.

These rules were used for determining the ability of a crew of two pilots to handle the advanced cockpits without the aid of a flight engineer in the proof of concept in the early eighties. This does not address FMS aftermarket problems: How can one determine what is simplicity and how much training is needed for the crew to reach the level of ease stated in this requirement when existing practices, routines and vulnerabilities change as a result of the introduction of new technology?

4. What is missing in today’s requirements?

The discussion above illustrates how existing requirements are either vague or inapplicable, and do not aid a manufacturer or authority in evaluation or certification decisions. These requirements (that nonetheless need to be complied with) often result in lip service being paid. The reviewer will typically be a test pilot who is not representative of the end user population and whose employer may have a stake in expeditious certification. The evaluation is normally based on limited exposure to the system in very few operational conditions. Reviewer statements of a successful result are mostly as vague and generic as the requirements they attempt to meet: all they can basically do is mimic the requirement text to the letter. Statements typically say: “The new system was easy to use and did not result in excessive workload or fatigue in all flight conditions; this shows compliance with paragraphs x and y”. This is obviously not very meaningful and; in addition puts a very heavy burden on the evaluator, since his decision aiding tools are very limited.

If a reviewer finds problems with, for instance, workload or user errors beyond the certification requirements, it is difficult to make a case for change in the face of cost and schedule constraints. One problem is that terms such as workload, fatigue, ease and simplicity are meaningless without a frame of reference. Existing research on human interaction with cockpit automation (e.g., Wiener, 1989; Sarter et al., 1997) has amassed a corpus of data and produced some ergonomics design/evaluation guidelines (e.g., Wiener and Curry, 1980; FAA, 1996; Billings, 1996), but generally stops short of providing criteria that a manufacturer or authority could build a clear case on.

5. Deficiencies in current designs

Underspecified certification requirements that are generally inapplicable to aftermarket computer systems create ergonomics problems at two levels, as indicated in the introduction. At the first level, a catalogue of “classic” human–computer interaction failures can be found. Some typical ones are presented in this section, and point clearly to the holes that exist in the certification net with respect to cockpit computer systems today. These kinds of problems could reasonably be anticipated and countered by designers and authorities alike. The next section suggests a way of dealing with the more difficult problem of anticipating how the introduction of new technology creates new, unanticipated demands and vulnerabilities.

5.1. CDU Design

The control display unit (CDU) is the main interactive interface between the crew and the FMS. The standard interface includes a screen, line select keys, alphanumeric keys, specific function keys and warning lights. This design is typical to all manufacturers but, unlike the typewriter, there is no standard position for the letter keys. In addition, similarly named function keys have different uses in different designs. For example, the NAV key is used for navigating an active route in one design, while in another it is used for planning a route (e.g., before flight). The flight plan (FPLN) key is used in one design for planning while in the other it is used for navigating. These features increase the risk of negative transfer when transitioning between systems.

5.2. CDU Menu Layers

The CDU has a limited display area (8–10 rows of text) so several display layers are needed. This creates the “keyhole effect” (Woods, 1996), forcing serial user access to stacks of highly interrelated data. Since layers are unavoidable, one design aim should be to minimize the number of layers and, more importantly, to keep the operator oriented at all times by providing “visual momentum” (Woods, 1984). Users today get lost easily and the memory burden (where am I and how do I get back to where I came from?) can be very large. Most cockpits have dual FMSs installed (one for each pilot), which puts a premium on being able to transfer changes from one to another. On one design this cue is not available until the pilot has paged through the entire flight plan which may be spread across three to six pages.
5.3. Feedback on changes

For flight critical actions, such as amending a flight plan or selecting a more direct route, some designs allow the pilot to review the changes and then accept or cancel them before the change is implemented. Other designs are much less tolerant: once a change is initially prompted it is immediately implemented. With these systems, slips and erroneous inputs can easily propagate into greater trouble.

5.4. Representation on the navigation display

The FMS-generated flight track is normally depicted on a moving map display in front of the pilot (often called the navigation display, or ND). All FMS final flight plan changes are usually shown on the ND. But, when making a change such as DIRECT-TO or DELETE, only some designs depict both the active track and suggested change for pilots to review prior to activation. This represents another form of desirable error tolerance.

5.5. Database ambiguity

Today’s cockpit computer database of waypoints, navigation aids and arrival/departure routes is enormous. The same label used in the database may define several waypoints around the world. Usually, the risk for selecting the wrong waypoint is mitigated by the computer suggesting the one nearest the aircraft, but this does not always insulate against failure (e.g., Aeronautica Civil, 1996). With increased database complexity, it is essential to insert check functions for the reasonability of input to minimize the risk of selecting the wrong waypoints.

5.6. UNDO functions

Any user of PC software expects an UNDO function in order to recall inputs. This function is not standard in FMS software and each design has its own criteria for providing an UNDO feature — if indeed it has one. UNDO may be lacking even in the flight critical functions such as DIRECT-TO and DELETE. In some cases the selection of DELETE to the wrong line on the display cancels the entire flight plan without any Review or UNDO options. This can happen, for example, when a small change in a three to six-page flight plan is ordered by ATC prior to turning onto the runway, after which the aircraft is cleared for immediate take-off (with a deleted flight plan).

5.7. Color logic between CDU and ND

Interfaces between navigation display and FMS of different makes are common for “add-on” systems. This produces mismatches in standards to the point where different displays show different colors for the same waypoint. This, combined with aircraft manufacturers who have different color standards, produces displays that are confusing and misleading.

A host of HCI guidelines exist (e.g., Molich and Nielsen, 1990) that could counter the problems above and could be translated into certification requirements. Other features were discussed during flight test and certification projects including:

- It should always be visible to the operator which MENU is displayed, and what are the available options for direct transit.
- Active flight path changes (DIR, DELETE, HOLD, OFFSET, INSERT) may not be activated prior to a graphical review of the changes and a confirmation by the pilot.
- An UNDO function must always be available and a prompt visible for immediate activation.
- Not more than ONE page change should be needed for accessing functions for changing active flight path (ex: DIR, HOLD, DELETE, OFFSET, INSERT).
- If more than one CDU is available, the data transfer function must be available on each page, or a prompt must be visible without the need for page change.
- When inserting a waypoint name that has several possible locations, in addition to displaying all options in a logical order (closest first) the system should display a question requiring confirmation by the pilot. This question should include the full name of the selected waypoint.
- When the system does not accept a value or input it should display the reason for the denial in clear text. This is to expedite error analysis by the operator.
- When color-coding is used for waypoints on the interactive display (CDU) the same colors should be displayed on the graphical navigation displays.
- Since a CDU may be installed in different positions and angles relative to the pilot reference eye point, a means of aligning the display to the line select keys should be available.

6. Anticipating the effects of introducing new technology

Adding an FMS to an older cockpit represents an intervention in a field of ongoing practice (Winograd and Flores, 1987). According to Woods and Dekker (2001), introducing new technology is known to:

1) Re-define human–system relationships: Adding a machine changes the work required of humans and affects collaboration between humans. The idea that new technology can be introduced through simple substitution of computers for people,
without wider reverberations for the larger human-machine system, is a myth.

(2) Shape human performance: Technological change represents or demands new ways of doing things. It does not preserve the old ways with the simple substitution of one medium for another. As one pilot of an automated aircraft commented: “I can’t fly anymore, but I can type fifty words a minute now”. Technology changes practice, and through this it changes knowledge and skill requirements, as well as sources of vulnerability and failure.

(3) Change operational requirements and organizational expectations: Operators that invest in new technologies often unintentionally exploit the advances by asking operational personnel to do more, do it more quickly, do it in more complex ways, or do it with fewer other resources.

These effects have profound implications for the evaluation of new technology and how it will affect operational safety. One case history, however, lays out encouraging directions for assessing how add-on technology changes practice and affects reliability (see Singer, 1999; FAA, 1997).

The head-up display (HUD) is a recent system that underwent a thorough certification. HUDs are added to existing cockpits to improve an airline’s reliability in low visibility approaches. The HUD shows critical flight parameters from existing cockpit sources on a glass plate through which the outside world is visible. The pilot is put in the loop, both as a human servo and as a decision maker, in a critical system with new and novel means of guidance. In order to prove that the reliability of such a concept is equal or better than the Autoland systems already in use, new requirements were set, some of which addressed the crews’ ability to handle the system interface, display and guidance. Unlike existing human factors paragraphs, the HUD requirements represented more objective criteria than before.

The HUD certification process was innovative relative to existing methods described in this paper, particularly because:

- A large sample of typical users (line pilots, not test pilots) flew up to one thousand approaches with the HUD both in full flight simulators and actual aircraft.
- These approaches were flown in the full range of conditions (weather, airports, failures) the HUD is expected to perform in during its operational life.
- The certification base was not airworthiness (FAR/JAR25) but rather Operational Requirements for all-weather operation (JAR-AWO and FAR-AC-120.28C(D)). This allowed closer examination of how and whether HUD-transformed practice could live up to operational requirements in actual conditions.

- An objective Pass/Fail criterion for pilot performance (approach and landing parameters) and error rates resulting in failed approaches or other hazardous conditions was set as a precondition for approval. These values were also statistically tested to predict future system safety.

This method of certification combines compliance with airworthiness and operational requirements and is probably the only objective method of showing compliance with an acceptable human factors design. This method ensures that the typical operator, and not only the test pilot population, is represented. The point is not only to indicate how the behavior of human-machine systems should be measured, but also what is measured (Stager, 1993). It enables the evaluator to collect objective performance and error rate data and evaluate corrective procedures or design issues in order to rectify unacceptable results.

Such an evaluation is consistent with increasingly common ecological approaches to assessing human–computer interaction (see e.g., Rasmussen and Pejtersen, 1995). In these, the emphasis shifts from proving that the system works at a single point in time, to assessing where and how it works during an extended episode of operational exposure. Such an approach to assessment is more sensitive to how the addition of new technology transforms human roles and work, and to how humans in turn adapt the new device to make it work for them in their circumstances.

In the case of FMS installations, the approach and departure scenarios will be more appropriate. Each approach or departure can be divided into small segments that require the crew to use the FMS for making changes. Each segment can be evaluated and compared with an optimal flight path. Random changes in flight plan or clearance will be fed to the crew and their reaction checked from ATC and terrain perspectives. In addition to performance and system reliability, Human Factors aspects could be evaluated in the following way:

- Non-recoverable crew error leading to reduced safety (altitude busts, lateral separation violations on parallel runways, or reduced terrain clearance) will result in a failure to meet certification.
- Recoverable minor errors (affecting only crew response time) will be allowed up to 5% of the runs.
- Changes may be made to the system hardware, software or operating procedures of the crew and the test redone.
- Non-essential functions will be tested for risks of giving misleading information or conflicting with other display features in content or reference standards.

Such an evaluation should be done in a full mission simulator for the relevant type of aircraft with the new
system integrated to an acceptable level (minor deviations cannot be avoided). The simulator must show a high fidelity for the relevant flight phase and system integration must be correctly integrated. In the FMS, certification process simulator flight characteristics are not as critical as for a HUD system but system integration is critical for evaluating crew normal and abnormal procedures and situation awareness.

7. Conclusion

The installation of add-on flight management systems in transport aircraft is a good example of a process with significant shortcomings in ergonomic assessments. Existing certification rules and regulations cover technical issues in detail, while leaving the human-machine interface to vague subjective evaluations. As a result, designs have been introduced into cockpits that were actually not intended to accommodate them, and many opportunities for human error have been created. The problems identified in this paper could be addressed by (1) the application of more detailed and relevant ergonomics requirements, based on existing HCI standards, and (2) the evaluation of new designs with a more operationally oriented testing method. The proposed evaluation method measures objective performance and error criteria while exposing the representative user population in the full operational environment. In addition, it measures the final effects of crew error and reaction to technical failures, doing this while treating both critical and non-essential functions.

With the decommissioning of many land based navigation aids the FMS as a cockpit control interface is soon to become the primary tool for managing aircraft navigation and energy. At present, integration in older aircraft often incorporates traps that could lead to unsafe flight. Since these designs do meet the available requirements and advisory material, a more specific and performance-based method of certification is needed. Let us fill in the gaps in the human factors certification requirements to enhance future safety of flight.

Acknowledgements

The work for this article was supported in large part by Saab Aircraft AB and by a grant from the Flight Safety Department at the Swedish Civil Aviation Authority.

References


Federal Aviation Administration, 1996. On the interface between flightcrews and modern flight deck systems. FAA Human Factors Study Team Report, FAA, Washington, DC.


