

FLIGHTDECK AUTOMATION: EVIDENCE FOR EXISTING PROBLEMS

Beth Lyall
Mary Niemczyk
Rebekah Lyall
Research Integrations, Inc.
Tempe, Arizona

Ken Funk
Oregon State University
Corvallis, Oregon

ABSTRACT

Aviation human factors literature presents many problems and concerns with flightdeck automation. We are completing a study in which we documented these problems and concerns, and identified evidence associated with each of them. All evidence identified has been entered into a relational database along with the information on citations for each of the problems and concerns. The database is being made accessible through a world wide web site for use by the aviation community.

Our work has been conducted in two phases. Phase 1 included identifying the human factors problems and concerns with flightdeck automation. In Phase 2 we compiled empirical evidence supporting and refuting the 114 problems identified in Phase 1, and identified two additional problems for which we also identified citations and compiled evidence. Evidence was identified in existing publications and data made available to us, and generated through a survey of human factors experts with a broad knowledge of flightdeck automation research and/or broad flightdeck automation experience.

This paper focuses on our work identifying and documenting evidence, and presents general results of a survey of experts conducted to evaluate the problems/concerns and generate evidence. Specific examples of the evidence associated with a particular problem are presented. The database is described along with the preliminary design of the web site and how it will be made accessible. Our vision for the use of the database by the aviation community is also presented.

INTRODUCTION

Hundreds of articles and presentations have addressed problems and concerns with flight deck automation. Many have raised general concerns about the approaches taken to automation use philosophies and automation design. Others have addressed specific problems or concerns identified with particular designs or implementations of automation.

The first phase of our on-going research identified 114 human factors problems and concerns with flightdeck automation (Funk, Lyall, & Riley, 1995). It was necessary to evaluate each of these concerns to understand the level to which they were indeed problems and required

additional research to identify solutions. To accomplish this, we identified information and data that can be used as objective evidence related to each of the problems and concerns. This paper focuses on our methods for finding evidence and includes an example of how the evidence may be used to support development of solutions for the flightdeck automation problems.

EVIDENCE IDENTIFICATION

Five methods were used to identify and quantify evidence for the problems and concerns. We reviewed information and data available in published literature, in the analyses and conclusions from accident reports, and from a set of ASRS incident reports that we summarized for evidence. We also reviewed the information received from those who responded to our initial questionnaire to identify problems and concerns, and conducted another survey of experts to verify problems and their criticality. We have defined evidence as data or other results from experiments, surveys and incident studies, and findings or conclusions from official accident and incident reports. For each record of evidence a strength rating was assigned based on the type and value of the data. Our intent in developing these ratings was to provide a way of using the information from diverse sources in a consistent summarized form. The strength of any evidence record ranges from -5 to $+5$ where negative numbers represent evidence that refutes a problem and positive numbers represent evidence in support of a problem. In the following sections each of the first four methods are briefly described, and the expert survey methodology and results are presented in more detail.

Literature Review

To date, we have reviewed more than 200 documents to identify evidence and found evidence in 51 of those documents. The method used to assign a strength rating to an individual evidence record varied depending on the type of data presented by the authors. The strength ratings for evidence from survey data was based on the percentage of responses that the researcher obtained that supported or refuted the problem. For example, a strength of ± 5 was assigned to a record of evidence for survey data including 90 to 100% of the subjects, and ± 1 would be assigned for 1 to 24%. The strength ratings for evidence from experiments that tested a hypothesis related to one of the problems were based on the experience of the subjects and the fidelity of the performance situation. For example, a strength rating of ± 5 would be assigned to evidence from the results of an experiment that included qualified airline pilots as participants in actual airline operations, and a ± 1 would be assigned for results from using pilots without experience in the particular operations of interest (general aviation or student pilots) performing a low fidelity flight task.

Accident Report Review

To date we have reviewed more than 15 accident reports and found evidence in 14 of them. We extracted evidence from the conclusions and findings of the reports. Evidence was defined as a statement of findings or conclusion from the investigating board that was presented as a possible or probable contributor to the accident. The strength ratings were assigned to each evidence record based on how close the presented concern of the board was to our problem/concern (equivalent or similar to our statement) and its level of contribution to the

accident (possible or probable). A strength rating of +5 was given to a record of evidence that was equivalent to our statement and a probable contributor to the accident.

Incident Report Study

So far we have reviewed approximately 280 ASRS incident reports and are summarizing them for evidence. The reports were identified by searching the ASRS database for advanced technology aircraft incidents. We will define evidence from this study as the percentage of reports that include explicit statements by the pilot reporting that one of the problems contributed to the incident.

Expert Survey Study

One of our objectives for Phase 2 was to decide whether the concerns raised in Phase 1 could be considered problems on which resources should be spent to identify or develop solutions. As one step in meeting this objective we conducted a survey of individuals who have a broad experience or knowledge base related to flightdeck automation. In this survey we asked the respondents to what degree they agreed that the concern is a problem, how critical they believed the problem to be, and to identify the sources upon which they based their ratings.

Method. Forty-seven individuals were asked to participate in the survey based on their broad research or performance experience with flightdeck automation. Thirty-six agreed to complete the survey, and 30 completed surveys were received. The participants include pilots of several automated aircraft, university researchers, airline management pilots, industry designers and researchers, and government regulators and researchers.

Survey description. The survey requested general demographics information then presented 114 questions, one for each of the problems and concerns identified in Phase 1. For each statement of the problem/concern, the respondent was asked to rate their level of agreement that the statement described a real problem. The agreement rating was given on a scale from 1 (strongly disagree) to 5 (strongly agree) with 3 labeled as neutral. There was also a place to mark "cannot address" for each statement. In addition to the agreement rating each statement was rated for criticality to flight safety from 1 (not critical) to 3 (moderately critical) to 5 (extremely critical). Additionally, there was a place for the respondents to indicate the type of information upon which they based their ratings for each statement. The seven choices for type of information were personal experience, experience of others, personal research data, research data of others, aviation literature, personal opinion, and other (with a space to fill in). The respondents were asked to mark all the types of information that applied.

Results. The responses to each problem/concern statement were summarized. There were some statements upon which the ratings of the respondents were very similar. The statement that resulted in highest level of agreement between respondents was concerned with the requirements for monitoring the automation. The responses to this statement are shown in Figure 1. Five other statements resulted in the same pattern of responses that did not include any disagreement.

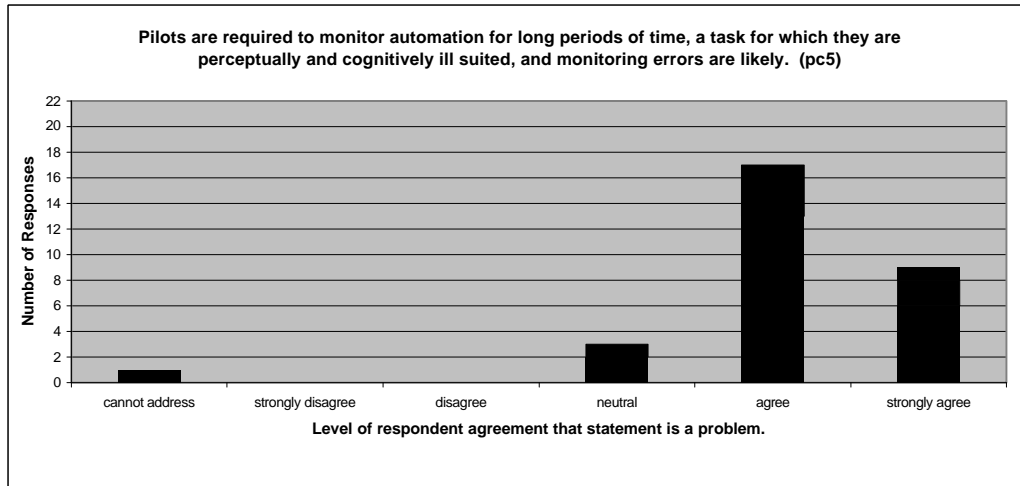


Figure 1. Survey respondents ratings that pc5 is a problem.

The five other problem/concern statements with similar distributions were:

- Cultural differences are not adequately considered in automation design, training, certification, and operations. Because they are not considered, they have resulting effects on performance and how automation is used. (pc 165)
- Transitioning back and forth between advanced technology aircraft and conventional aircraft increases pilot training requirements. (pc 129)
- Although automation may do what it is designed to do, design specifications may not take into account certain unlikely but very possible conditions, leading to unsafe automation behavior. (pc 125)
- When two pilots with little automation experience are assigned to an advanced technology aircraft, errors related to automation use are more likely. (pc 142)
- Side sticks are not coupled with each other or the autopilot, reducing awareness of the other pilot's or the autopilot's inputs, resulting in reduced situation awareness and/or improper control actions. (pc 30)

Only one statement generated consistent responses disagreeing that it represents a problem, "Automation induces fatigue which leads to poor performance (pc156)." The results for this statement are presented in Figure 2. The response distributions for the other 107 problem/concern statements were not as clear cut as for these seven. Distributions range from being somewhat skewed to one side or the other (most in support of the problem), to being bipolar, to being almost flat. Several of the bipolar and flat distributions seem to have resulted from some ambiguity in the problem/concern statement. We are using the feedback from the survey respondents to clarify the problem/concern statements during Phase 2. A complete description of the methods and results of this study is being prepared as an internet document.

Use of expert survey results as evidence. The responses for each problem/concern were summarized for evidence. Evidence was recorded for each problem dependent on the results of the agreement ratings. Evidence in support of a problem was designated with a strength rating associated with the percentage of responses of "agree" and "strongly agree" combined, and evidence refuting the problem was recorded based on the percentage of responses indicating

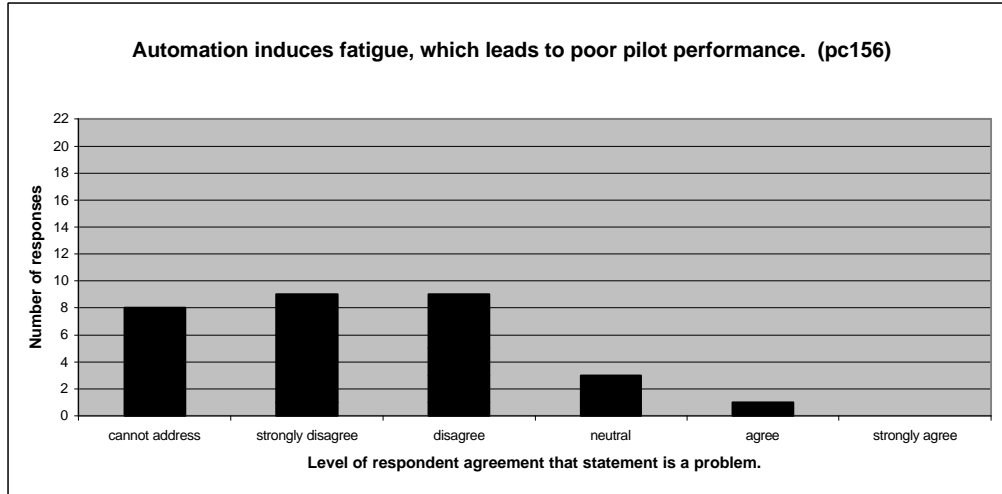


Figure 2. Survey respondents ratings of agreement that pc156 is a problem.

“disagree” or “strongly disagree.” The strength ratings ranged from -5 to +5. The survey results were used in the same fashion as other Likert scale results producing evidence records supporting and refuting each of the problems and concerns.

EVIDENCE APPLICATION

The database is being developed to allow access through the world wide web. We are still adding information to the database and developing the interface. The following example is given to show how the database can be used, and is based on the information in the database in its present form. A query of the current database asking for evidence associated with the problem “Training objectives, methods, materials, or equipment may be inadequate to properly train pilots for safe and effective automated aircraft operations (pc 133),” would identify 29 evidence records. Three of these records are from accident reports, 19 are based on data from surveys conducted by researchers in four studies, one is based on conclusions from an experiment, one is from an observational study, three are from incident studies, and 2 are from our expert survey. Twenty-two of the evidence records provide information about how training may be inadequate (support the problem) and seven provide information about how training is adequate (refute the problem). Each of the evidence records includes the strength of the evidence, the reference for the source, an excerpt from the document stating the information upon which the evidence was based, and the type of source it comes from (accident report, survey, etc.) A summary of this evidence query is presented in Table 1.

The evidence information can be used to summarize the perspectives and nuances that have been documented about this problem, as well as provide a starting point for identifying and evaluating potential solutions. It is our goal to keep this database current to allow researchers and practitioners to have a common base of information from which to work, and to provide a means for coordinating future efforts to address flightdeck automation problems.

Table 1 – Sample evidence records for: “Training objectives, methods, materials, or equipment may be inadequate to properly train pilots for safe and effective automated aircraft operation.”

Strength	Evidence Excerpt	Source Reference
-3	Question 6: "What is your opinion of the B767 training package?" ... "In general, the responses from the pilots was that the course is good. Thirty four [52%] pilots rated the course from 'okay' to 'excellent'."	Morters (1988, p. 75)
-1	"Training for the 757 at both airlines in this study was generally considered to be well planned and well conducted. A large number of pilots reported on their questionnaires that 757 school was the best training program they had ever been through."	Wiener (1989, p. 172)
+1	"Many pilots noted that the initial training program was very difficult and they felt unprepared for their first unsupervised flight. Insufficient information was provided concerning the actual aircraft and its systems, making it difficult to assess potential problems. They wished for more hands-on practice and more simulator time, to consolidate their knowledge."	Rudisill (1995, p. 7)
+3	Table 2 summarizes the "descriptive factors assigned to altitude-deviation reports" from traditional cockpits and glass cockpits. In the glass cockpit, 34 out of 50 (68%) reports suggested that training was a factor in the incident.	Palmer, Hutchins, Ritter, & VanCleemput (1993, p. 7)
+4	"It is recognized that the CAP and the F/O completed classroom, simulator and flight training based on the training syllabus prepared by China Airlines in accordance with Taiwanese civil aviation laws. However, it is recognized that this training was not necessarily sufficient to understand the sophisticated and complicated AFS system."	Aircraft Accident Investigation Commission - Ministry of Transport Japan (1996, p. 3-44)

REFERENCES

- Aircraft Accident Investigation Commission - Ministry of Transport Japan (1996). China Airlines A300B4-622R, B1816 Nagoya Airport April 26, 1994 (Report 96-5).
- Funk, K., Lyall, B, & Riley, V. (1995). Perceived Human Factors Problems of Flightdeck Automation (Phase 1 Final Report for Federal Aviation Administration Grant 93-G-039). Corvallis, OR: Dept. of Industrial and Manufacturing Engineering, Oregon State University. Internet address: <http://www.engr.orst.edu/~funkk/Auto/autoprob.html>
- Morters, K. (1988). B767 Flightdeck Automation Research. Unpublished research paper.
- Palmer, E.A., Hutchins, E.L., Ritter, R.D., & VanCleemput, I. (1993). Altitude Deviations: Breakdown of an Error-Tolerant System (NASA Technical Memorandum 108788). Moffet Field, CA: NASA Ames Research Center.
- Rudisill, M. (1995). Line Pilots' Attitudes About and Experience With Flight Deck Automation: Results of an International Survey and Proposed Guidelines. Proceedings of the 8th International Symposium on Aviation Psychology, pp. 288-293. Columbus, OH, April, 1995.
- Wiener, E.L. (1989). Human Factors of Advanced Technology ("Glass Cockpit") Transport (NASA Contractor Report 177528). Moffet Field, CA: NASA Ames Research Center.

ACKNOWLEDGEMENTS

The work presented was conducted under FAA Grant 93-G-039. Dr. Tom McCloy serves as our technical monitor and his guidance and assistance are much appreciated.