

From Webster's Universal Dictionary:

cat's paw - one used by another as a dupe or a tool.

See also the official emblem of the great CLEMSON TIGERS!

See 04
92180

Controlling Air Traffic (Simulated) in the Presence of Automation (CATS PAu) 1995



**A study of measurement techniques
for Situation Awareness
in Air Traffic Control**

Jennifer R. French
NASA Langley Aerospace
Research Summer Scholar
Clemson University

Debbie Bartolome, Ed Bogart, DanBurdette,
Ray Comstock, and Alan Pope
Human Engineering Methods Team
Crew/Vehicle Integration Branch
Flight Dynamics and Control Division
Research and Technology Group
NASA Langley Research Center

Abstract

As automated systems proliferate in aviation systems, human operators are taking on less and less of an active role in the jobs they once performed, often reducing what should be important jobs to tasks barely more complex than monitoring machines. When operators are forced into these roles, they risk slipping into hazardous states of awareness, which can lead to reduced skills, lack of vigilance, and the inability to react quickly and competently when there is a machine failure. Using Air Traffic Control (ATC) as a model, the present study developed tools for conducting tests focusing on levels of automation as they relate to situation awareness. Subjects participated in a two-and-a-half hour experiment that consisted of a training period followed by a simulation of air traffic control similar to the system presently used by the FAA, then an additional simulation employing automated assistance. Through an iterative design process utilizing numerous revisions and three experimental sessions, several measures for situational awareness in a simulated Air Traffic Control System were developed and are prepared for use in future experiments.

Introduction and Background

Just as in the field of aviation, in which the technological advances that make aircraft safer and more reliable are the same ones that may have negative psychological effects on the flight crew (Burt, in press), the FAA's current efforts to upgrade and automate many of the tasks involved in Air Traffic Control may have detrimental psychological effects on controllers. As aviation situations become more automated, the amount of involvement required of operators tapers off, which can lead to dangerous states of awareness. In "monitoring tasks," (tasks requiring less active participation by the operator when the automated system performs most activities and requires a human operator only to monitor the system) mental engagement may drop to a level that precludes satisfactory performance (Pope et. al., 1994). Among the effects of decreased involvement are declined level of control or loss of skills (Endsley and Kiris, 1994); however, more dangerous is loss of vigilance and other symptoms of "boredom" that are not associated with fatigue, especially a decrease in situation awareness (Pope and Bogart, 1992). When this occurs, operators of automated systems become slower responding to errors, or may fail to notice system errors entirely. Additionally, as the decision-making process becomes increasingly facilitated by automated systems, the operator may slip from an active mode of information processing to a passive one. This, also, can lead to a dangerous decline in situation awareness and may have a drastic effect on performance (Endsley and Kiris, 1994).

In order to facilitate a study of these declines in performance using automated systems, Endsley and Kiris (1994) defined five specific levels of automation. The first level, incorporating no automation, leaves all decisions and actions to the operator. The second level, dubbed "decision support," calls upon the operator to make decisions and actions, while the automated system makes suggestions. In the third, or "consensual" level, the automated system makes the decisions and actions, but requires concurrence on the part of the operator. The fourth, "monitored" level, sees all decisions and actions made by the system, while the operator has only veto power. The fifth, fully automated level omits the human from the process entirely.

In Air Traffic Control today, most tasks reside in the first level. Currently, automation is rarely used for anything beyond transmitting information. Most often, even this process relies on outdated technology that Scientific American had dubbed "winking, blinking, aged hardware" that is "often less powerful than the personal computers used by agency secretaries for word processing" (1994).

In response to the rapid growth of air traffic that is quickly becoming too large to be serviced by existing ATC technology, the FAA has undertaken a wide-sweeping plan to update and upgrade, called Automated En Route Air Traffic Control (AERA). In addition to communicating the location of aircraft, the AERA computers will notify controllers of future conflicts, check for deviations, advise alternate routes, devise the most time- and fuel-efficient routes, and communicate directly with the airplane. In one textbook of air traffic control, the author tells us "as system capacity increases, and

confidence is gained in the computer's capability, the AERA system may be permitted to formulate alternative clearances, choose the most practical clearance, and transmit it directly to the aircraft without controller intervention. The air traffic controller will only be required to monitor system performance and to intercede in unusual conditions." (Nolan, 1994) Software has been designed at NASA Ames Research Center to help steer aircraft through traffic jams, advise the best sequencing for landing, and suggest landing maneuvers for individual airplanes (Scientific American, 1994).

However, in light of findings that warn of declining performance when decision-making processes become automated, extreme caution must be taken. Although numerous tests have been conducted on components of the new automated systems by NASA, and the FAA (Credeur et. al.), and international agencies (Beniot et. al.), none have been performed with an eye to the levels of automation as pertains to situation awareness.

So the question remains: What is the maximum level of automation that can be utilized to improve Air Traffic Control situations without exceeding the point at which controllers cease to be sufficiently involved and mentally engaged?

To answer this, an informal research project was conducted by this researcher in the summer of 1994. From that study, it was observed that a more in-depth and comprehensive means of data collection would need to be developed before the question posed above could be answered with any certainty. Since there is no definitive means for measurement of situation awareness in air traffic control, studies focusing on measurement of situation awareness and measurement of air traffic controllers were examined.

As early as 1980, David Hopkin discussed measurements of air traffic controllers at length in a special issue of *Human Factors* dedicated entirely to ATC. He sited several empirically proven techniques of testing air traffic controllers, from performance, errors, delays and omissions to physiological indices to interviews, discussions, questionnaires and case histories. One technique he favored was task performance as it pertains to workload and involvement, stating that, "All (ATC activities) do not have equal importance, and some, thought desirable, may often be postponed for awhile or omitted altogether," and that "measures of the least important activities of the controller may provide the most sensitive indices of the effects of high task loading." More recently, Hopkin (1994) has stated that, "Measures of errors, omissions, the time scale of decision, options considered and discarded, and tasks that are desirable rather than essential may all be more sensitive indices of the benefits of automation in air traffic control and of its other consequences than direct measures of core task performance," and that, "measures of performance that relate directly to these core tasks may therefore be insensitive to the effects of automation and computer assistance, whereas more peripheral activities may be changed greatly."

As recently as March 1995, Mica Endsley discussed measurement of situation awareness at length in a special issue of *Human Factors*. She began by establishing that the criteria for a measurement technique for situation awareness must measure the construct it claims to measure and not other processes, will provide the required insight in the form of sensitivity and diagnosticity, and will not interfere with the process being

tested. Beyond that, the technique should be able to predict performance and be sensitive to changes in workload and/or attention. Endsley then proceeds to analyze in detail physiological measures, performance measures, subjective techniques, and questionnaires. But the best measure of situation awareness, she concludes, is to freeze the simulation briefly to quiz the operator on his/her awareness of many different facets of the simulation and the information with which he/she should be familiar at all times.

To satisfy both sets of specifications as well as numerous others, a synergy of data collection techniques was developed and/or adopted for the CATSPAu experiment. Building on many of Hopkin's and Endsley's techniques, they are designed to test a subject's awareness of the air traffic control situation (See "Data Collection and Analysis" below, and Appendices for details).

The CATSPAu experimental task itself was also altered to reflect more realistically air traffic control situations that are in use/planned. A four-post system, which many air traffic controllers use to simplify their task by filtering all incoming aircraft through four main points on the radarscope (Erzberger and Nedell, 1993), was applied. Also, the timing of the automated system was altered to more closely emulate the Direct Course Error timer recommended as a part of the Final-Approach Spacing Aids, designed by LaRC researchers in 1993 (Creuder, et. al.)

Beyond that, conclusions from last summer's project were incorporated into CATSPAu. Based on the number of subjects from those experimental sessions who quickly lost patience with the automated system, additional instructions encouraged them to adhere to its recommendations. Further, the extended training session detailed below is reflective of last summer's conclusions as well.

Approach and Equipment

Equipment and Facilities: The Air Traffic Control simulation software TRACON, produced by Wesson Software, was run on an IBM PC with graphics capabilities. An additional IBM PC was used to run a program that simulated "automated assistance" written in quick basic by Dr. Ray Comstock. Additionally, a headset and a second monitor were used to aide a concealed confederate researcher to simulate higher levels of automation. Data was collected by pen-and-paper means. All facets of the experiment were conducted in the Human Engineering Methods offices and laboratory (Bldgs 1168 and 1268) at Langley Research Center.

Subjects: Three volunteer subjects, recruited from the pool of LARSS students and the researchers' personal contacts, were utilized. All three were male and ranged in age from 17 to 28. Subjects were screened to insure they had normal vision, had not been diagnosed with Attention Deficit Disorder (ADD) or Attention Deficit-Hyperactivity Disorder (ADHD), and had no prior experience with air traffic control.

Experimental Design: Subjects performed a task similar to that of Air Traffic Control by engaging in variations of TRACON, which realistically simulates the ATC radar scope and contains a computerized version of the paper strips used in Air Traffic Control (Wesson and Young, 1988). The experience of communicating with aircraft was simulated by having subjects speak into a headset, and their verbal commands were translated to TRACON keyboard commands by the concealed confederate.

From Endsley and Kiris' five levels of automation, the first two were selected and applied to ATC through TRACON:

Level 1: No automated decision-making aides. Subjects engaged in TRACON's ATC tasks with no automated assistance, much as in status quo ATC.

Level 2: Suggestions from automated system. Subjects engaged in TRACON while an automated assistant provided suggestions on the safest and most efficient commands, much like the proposed improvements to future ATC environments.

Subjects completed a 45 minute training session on the use of the simulation software and the philosophy behind ATC. Subjects then engaged in extended training sessions that were similar in duration and demand to data collection. The data collection sessions consisted of seventeen and a half minutes of TRACON at Level 1, and seventeen and a half minutes at Level 2.

Data collection and analysis: Several types of measurement techniques were designed and arrived at through an iterative process of testing and revision. A Freeze Technique Questionnaire, quizzing subjects on the location of aircraft, and their destinations and status (see Appendix A), was developed to administer to subjects at various intervals during the simulation in which the program was paused and the screen was covered. The Task Load Index (see Appendix B) developed and empirically proven by Hart and Staveland (1988), was also administered before the simulation resumed. Errors, in the form of missed approaches (aircraft that are not successfully prepared for landing at the airport), were counted during the simulation. Omissions less vital to the overall success of the task, in the form of hand-offs (aircraft that are not successfully passed on to the next controller), were counted as well. Finally, subjects were verbally de-briefed at the end of the simulation regarding their comfort and confidence with respect to the automated assistance they received.

Results and Discussion

Because only three subjects were run and conditions were altered for each subject through the process of iterative design, no cross-subject results can be derived. However, Table 1 illustrates a sampling of the types of data that would be available using the final form of the data collection techniques developed for CATSPAu. Data is organized by freeze number for each condition by subject. Comparisons among the number of aircraft of which each subject was aware and the number that were actually present can be made, as can the destinations of current aircraft, number of aircraft about to enter the sector, and

status of aircraft next to landing. Additionally, errors, omissions, and TLX ratings (composite) can be compared by subject and condition.

In the verbal de-briefing, two subjects reported that the automated assistant made them less comfortable and confident and more pressured, while one reported the opposite. These answers, however, seem to be directly related to the extent to which each subject trusted the automation. When asked if they felt the automated assistant had helped or hindered their performance, the subject who claimed it helped him relied almost entirely on the automated assistant, the subject who claimed it hindered him frequently strayed from the recommendations, and the subject who said it neither helped nor hindered him later said that he used it as a self-check.

Unfortunately, due to the nature of the simulation, there are limitations that may contribute to subject's lack of trust in the automated assistant. The script that feeds commands to the automated assistant receives no genuine feedback and is not dynamic, and it cannot respond in any way to changes or deviations from the pre-planned flight paths. In other words, once subjects begin to second-guess the automation, they cannot surrender that control until they have cleared the individual aircraft from their sector. Also, if the subject makes an error or fails to issue a command, the automated assistant is unforgiving and cannot incorporate those mistakes back into the flight plan, thus causing the subject to lose confidence in the system. This ability to allow for controllers' errors is vital to the success of any automation, and has been incorporated into currently proposed automated systems for Air Traffic Control (Erzberger, 1992).

Despite these limitations, the techniques discussed above should prove to be an adequate means of investigating the situation awareness of air traffic controllers in the presence of automation. A future study that would utilize these measurement techniques to test situation awareness in different levels of automation would ameliorate or knowledge of human awareness in the presence of automation.

Works Cited

- Burt, Jennifer L., Bartolome, Debbie S., Burdette, Daniel W., Comstock, J. Raymond Jr. "A psychophysiological evaluation of the perceived urgency of auditory warning signals." *Ergonomics*, in press.
- Benoit, Andre, and Swierstra, Sip, et. al. "The Air Traffic Controller Facing Automation: Conflict or Co-operation," *Aircraft Trajectories Computation-Prediction-Control*, Volume 2. AGARDograph No. 301, 1990.
- Credeur, Leonard, and Capron, William, et.al. *Final-Approach Spacing Aids (FASA) Evaluation for Terminal-Area, Time-Based Air Traffic Control*. NASA Technical Paper 3399, December 1993.
- Endsley, Mica R., and Kiris, Esin O. "The Out-of-the-Loop Performance Problem: Impact of Automation and Situation Awareness." *Human Trends in Automated Systems: Current Research and Trends*, 1994.
- Endsley, Mica R. "Measurement of Situation Awareness in Dynamic Systems." *Human Factors Special Issue: Situation Awareness*, Vol. 31, No. 1, March 1995.
- Endsley, Mica R. "Toward a Theory of Situation Awareness in Dynamic Systems." *Human Factors Special Issue: Situation Awareness*, Vol. 31, No. 1, March 1995.
- Erzberger, Heinz. *CTAS: Computer Intelligence for Air Traffic Control in the Terminal Area*. NASA Technical Memorandum 103959, July 1992.
- Erzberger, Heinz, and Nedell, William. *Design of Automated System for Management of Arrival Traffic*. NASA Technical Memorandum 102201, June 1989.
- Hart, S. G. and Staveland, L. E. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Human Mental Workload*. North Holland: Elsevier Science Publishers B.V, 1988.
- Hopkin, David. "Human Performance Implications of Air Traffic Control Automation." *Human Trends in Automated Systems: Current Research and Trends*, 1994.
- Hopkin, V. David. "The Measurement of the Air Traffic Controller." *Human Factors Special Issue: Air Traffic Control I*. Vol. 22, No. 5, 1980.
- Nolan, Michael S. *Fundamentals of Air Traffic Control*. Belmont, CA: Wadsworth Publishing Co., 1994.
- Pope, Alan T., and Bogart, Edward H. "Identification of Hazardous Awareness States in Monitoring Environments." *SAE 1992 Transactions Journal of Aerospace*, Section 1-Volume 101, 1992.
- Pope, Alan T., Bartolome, Debbie S., Bogart, Edward H., Burdette, Daniel W., and Comstock, J. Raymond Jr. "Biocybernetic System Validate Index of Operator Engagement in Automated Task." *Human Performance in Automated Systems: Current Research and Trends*, 1994.
- Stix, Gary, "Trends in Transportation: Aging Airways." *Scientific American*, May 1994.
- Wesson, Robert B., and Young, Dale. *TRACON Air Traffic Control Simulator*. Austin, TX: Wesson International, 1988.

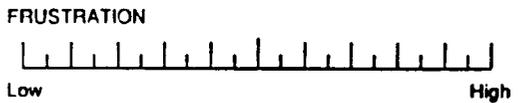
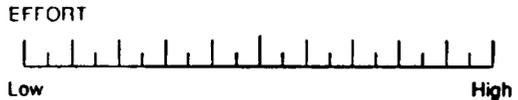
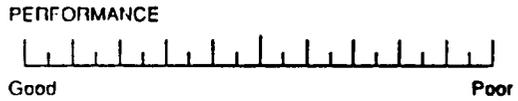
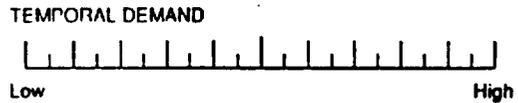
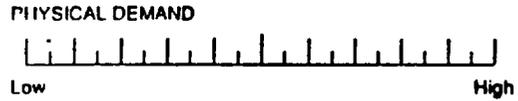
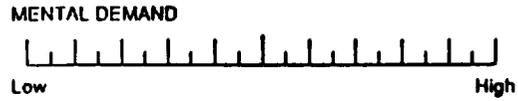
Appendix B

Subject ID: _____ Study ID: _____

Rating Scale Definitions

Title	Descriptions
MENTAL DEMAND	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Place a mark at the desired point on each scale:



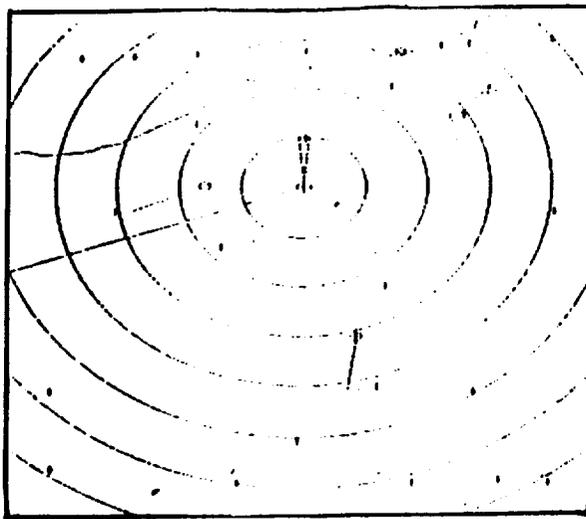
			A/c on screen	A/c reported	Non-landing a/c	Non-landing a/c reported	Pending a/c	Pending a/c reported	Next-to-land a/c correct (y/n)	Missed approaches	Handoff errors	TLX
Subject 1	Manual	Freeze 1	5	5	2	2	0	n/a	no	2	0	22
		Freeze 2	2	2	0	0	0	1	yes			4
	Automated	Freeze 1	6	5	2	2	0	2	yes	1	0	26
		Freeze 2	2	2	0	0	0	0	yes			5
Subject 2	Manual	Freeze 1	5	5	2	2	2	2	yes	0	0	59
		Freeze 2	5	7	2	2	4	2	yes			72
	Automated	Freeze 1	6	5	1	1	3	7	no	1	1	84
		Freeze 2	9	7	2	1	2	2	no			87
Subject 3	Manual	Freeze 1	9	9	4	0	2	8	yes	0	1	28
		Freeze 2	10	10	4	2	4	10	yes			24
	Automated	Freeze 1	5	4	1	1	3	4	no	3	2	9
		Freeze 2	7	7	2	2	2	7	yes			17

Appendix A

Subject number _____ L of A _____

Freeze Technique Questionnaire

1. Please indicate the general positions of all aircraft you are currently handling on the image of your sector below. Use an arrow pointed in the direction your aircraft is traveling to represent each airplane and its heading.



2. In the diagram above, place a box around any aircraft that you are currently handling which are not landing.
3. How many aircraft are currently shown under the heading of "Pending" on your paper strips?
4. What are the altitude and speed of the aircraft that is closest to landing?
5. What are the point of origin and destination of the aircraft with which you most recently established radar contact?

Appendix C

Subject Data and Performance

Subject ID _____
 Level of Automation in second simulation: suggest / concur / veto

Timing:
 Tr1 _____ start _____ end _____ XTr1 _____ start _____ end _____ DC1 _____ start _____ end _____
 Tr2 _____ start _____ end _____ XTr2 _____ start _____ end _____ DC2 _____ start _____ end _____

Data
 Number of unsuccessful handoffs in NON-automated mode: _____
 Number of missed approaches in NON-automated mode: _____

Number of unsuccessful handoffs in automated mode: _____

Number of missed approaches in automated mode: _____

Deviations from Automation: _____

Questions prior to Verbal Debriefing
 Were you more comfortable with or without automated assistance?
 Did having automated assistance make you more or less confident?
 Did having the automated assistance make you feel more or less pressured?
 Do you think having automated assistance helped or hindered your performance, or neither?

Observations: _____