Human Systems
in Systems Design Engineering
Technical Report Series
HSSD-03-05

A Review of Experimentation Methodologies in the Area of Aviation Displays and Trust in Automation

Danny Ho
Catherine Burns

Technical Report
Advanced Interface Design Lab

Department of Systems Design Engineering
University of Waterloo
Waterloo, Ontario, Canada
N2L 3G1

Table of Contents

Abstract

This report highlights six publications in the area of aviation displays and trust in automation. Insight gained from their experimental methods was related to the TCAS 2 automation study by Ho and Burns (Ho, Burns, in press), and recommendations for its experimental design were made. Participants must show a suitable level of analytical performance. The independent factors shall be display formats and automation reliability. The dependent measures will be percentage of correct response, reaction time, and time spent in conflict. ANOVA will be performed to determine main factorial effects and their interactions. Finally, NASA-TLX will be administered to measure subjective workload and self-confidence levels, and questionnaires shall be used to determine levels of trust. This report established a suitable framework for the TCAS 2 study to proceed with experimental design.

Introduction

The evolution in the aviation industry towards free flight, automated flight planning and pilot support has fuelled research and development in the field of automated display systems. Alexander and Wickens approached the problem from a cognitive perspective to determine the effects of 3D and 2D coplanar display orientation on pilot maneuver performance (Alexander, Wickens, 2001). Krishnan et. al devised a 'cone analogy' perspective display and measured its performance compared to baseline displays (Krishnan, Kertesz, Wise, 2000). Still others challenged the automation itself, in determining how varying levels of information cueing and automation reliability can affect human decision-making in a sensor-to-shooter display (Rovira, McGarry, Parasuraman, 2002).

The current TCAS study conducted by Ho and Burns (Ho, Burns, in press) shares all three of the aforementioned approaches. It is the first study to apply Ecological Interface Design (EID) concepts to the existing Traffic Alerts and Collision Avoidance System (TCAS) where display enhancements will be tested with varying levels of information cued automation to measure the effect of EID-based displays on an operator's trust in automated systems.

To aid in the experimental design, six publications which share similar research directions to the TCAS study were chosen from the Human Factors and Ergonomics Society conference proceedings and their experimental methods were discussed. The result of this information gathering revealed the common approaches used for measuring human performance in their interaction with automated displays, and provided a suitable experimental framework for the TCAS study to build upon.

The publications will be summarized individually, with an emphasis on four major areas of the experimental method: participants, independent variables, dependent variables, and analysis methods. The direction of the TCAS research will then be described, and suitable experimental methods will be recommended based on a discussion of the publications as they relate to the TCAS study.

Publication Summaries

<u>Designing Situation Displays to Promote Conformance to Automatic Alerts</u>
Amy R. Pritchett and Balázs Vándor, (2001)

Pritchett et al. hypothesized that the interaction between information presented on situation displays and automatic alerts can encourage or discourage conformance to such alerts. Consonance (display provides alerting rationale) and dissonance (display does not provide rationale) studies were conducted.

45 participants took part in the study. They flew a simple side-stick controller while monitoring a parallel approach traffic display. They were told to evaluate the situation to decide when a potential collision hazard, in their judgement, requires an action. Instead of executing a traffic avoidance maneuver with the side-stick, they pressed a button. They were told that the automatic alerting system was not 100% reliable.

There were 3 independent factors in the study: three traffic displays, three alerting conditions, and four scenarios. Both the alerts and displays used one of three alerting algorithm conditions: NTZ (no transgression zone), MIT (miles-in-trail), and no alerting. The total nine (3 by 3) display and alert conditions tested alert effects only and display effects only. Consonance and dissonance conditions were set up with both NTZ and MIT algorithms by revealing more or less of the alerting logic. The four scenarios included low-convergence blunder (slow drift towards other approach path), high convergence blunder (fast drift towards other approach path), high-convergence missed approach intercept by the other aircraft (other aircraft crosses your path), and a placebo scenario (no conflict actually arises.

The primary dependent measure was the time taken by the participant to perceive a collision hazard and press the button to signal a corrective intent. Results analysis was divided into display effects, auto alert effects, and consonance and dissonance effects. In addition to comparing average reaction times based on varying display type and alerting algorithm, one-way ANOVA was performed and found significant differences among the consonance and dissonance conditions.

The study showed that consonance in displays improved the reaction time of pilots. Furthermore, it was found that the best reaction times were present in conditions where a high-performance MIT automation alert was given in consonance with the display of the MIT alert criteria. The findings therefore suggested that display consonance coupled with auto alerts may simplify the verification of traffic threat.

Although this study provided a focus on parallel approach threat displays, it can be extended to enroute traffic collision detection displays (such as TCAS) to determine the effect of display consonance and auto alerts on inflight traffic threat verification.

Cockpit Display of Traffic Information: The Effects of Traffic Load, Dimensionality, and Vertical Profile Orientation
Amy L. Alexander and Christopher D. Wickens, (2001)

Alexander and Wickens studied the effects of display dimensionality (3D vs. 2D coplanar views) and traffic density on conflict avoidance safety, maneuver choice, and efficiency. Pilots flew to predetermined waypoints while avoiding conflicts and minimizing deviations from target altitude, lateral position, and airspeed values.

18 flight instructors were used in the study. A within-subjects manipulation of display type was used. 3 display formats by 3 traffic loads were used to create 9 trials. Two replications were done, resulting in 18 trials flown by each flight instructor. The 3 display formats were 3D exocentric, 2D coplanar - top and side, and 2D coplanar - top and behind (see Figure 1), while traffic loads varied between 2, 6, or 10 aircraft in the radar vicinity.

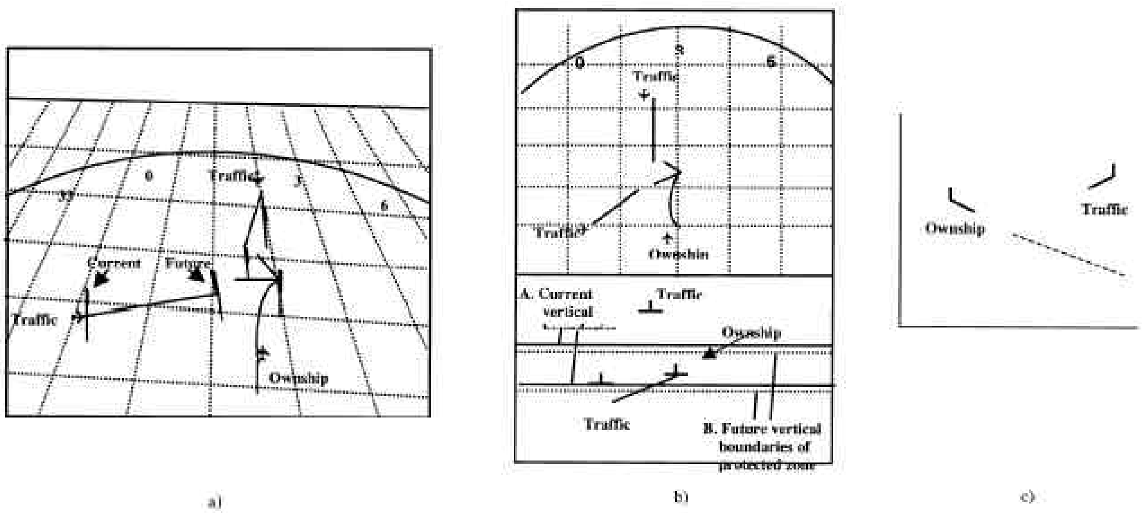

Figure 1. Three display versions of the CDTI: (a) 3D, (b) 2D coplanar rear-view, (c) 2D coplanar side-view. The rendering of the side view profile in (c) is not what participants actually saw, but is a schematic designed to highlight the different orientation from figure (b).

Figure 1 - Display Dimensionality (Alexander, Wickens, 2001)

The two primary measures were maneuver frequencies by display type, and time spent in predicted conflict being the measure of safety. Maneuver frequencies were categorized as lateral, descend, climb, airspeed change, and lateral/vertical direction change. Time spent in conflict was defined as a state of loss of separation within 45 seconds if no maneuver was taken.

A series of ANOVAs were conducted and determined that rear-view display supported marginally significant superior conflict avoidance performance over side-view display. Also, within climbing maneuvers, ANOVA results revealed a marginally significant effect within climbing maneuvers such that climbs were more accurately executed using 3D or rear-view than the side-view display.

Maneuver efficiency was measured by the deviation amount from the target altitude, airspeed, and heading. Mean absolute heading deviation was greatest with 3D, side-view, followed by rear-view. Mean absolute altitude deviation was greatest with 3D, rear-view, followed by side-view. Mean absolute airspeed deviation was highest with 3D and side-view, less with rear-view. In the three displays and four measures, a matrix was produced with + (best), 0 (second best), and – (worst) scores which were tallied to summarize the best performing display. The list of views ranked by accumulated participant scores was rear-view (3), side-view (0), and 3D view (-1).

This study helped to illustrate how display dimensionality and traffic load can affect pilot spatial awareness and maneuvering performance in potential conflict situations. With respect to the TCAS study, however, this study has a strong emphasis on the pilot task performance and less on the pilot's decision to perform a maneuver task. The TCAS study assumes that well trained pilots have sufficient piloting skills in conjunction with autopilot and other navigation systems, but it is the effects of display information on decision-making that is of larger interest.

Cockpit Traffic Displays Using Varying Levels of Intruder Intent Information
Richard Barhydt, R. John Hansman, (1999)

This paper describes a study in progress. By leveraging future datalink improvements to the flight deck, the team will assess the benefit of displaying varying levels of traffic intent for recognizing and resolving traffic conflicts. The four levels are:

Show intruder flight number and airline
Show predicted intruder location (based on heading, airseed, vertical speed)
Show predicted intruder location based on current state and autopilot state
Show predicted intruder location based on programmed FMS flight plan (waypoints, altitudes, tops and bottoms of climbs and descents)

The team is proposing an experimental design that will be conducted on a part task B747 flight simulator. Commercial airline pilots will be used as subjects, and will be presented with a series of traffic scenarios using each of the four displays. Pilots will be tested for their ability to identify traffic conflicts, assess their hazard level, and make an appropriate avoidance maneuver, if necessary. Similarities in the type of maneuver (lateral, vertical, or speed change) in addition to subjective pilot comments will be noted for each display.

This proposed study is similar to the TCAS study in that it will determine if increased exposure of system information will improve pilot situation awareness and decision making. However, the pilot response to the potential threats is open ended, even though Federal Aviation Association (FAA) guidelines have a clear policy for evasive maneuvers. For example, when TCAS alerts are given, the pilot may only maneuver in the vertical sense by climbing or descending away from the oncoming threat. The experimental design of the TCAS study will adhere to existing regulations and procedures, with the introduction of multiple EID-based interfaces.

Putting Four Dimensions in "Perspective" for the Pilot
K. Krishnan, Steve Kertesz Jr., John A. Wise, (2000)

Krishnan et al borrowed the 'space-time' paradigm outlined by Stephen Hawking to produce a 'space-time' based display that alerts the pilot to potential collision alerts in the near future. Threats that are within three minutes from occupying the protected volume of ownship are plotted on this display, and the imminent threats are drawn with directional threat vectors (see Figure 2). Two studies were conducted.

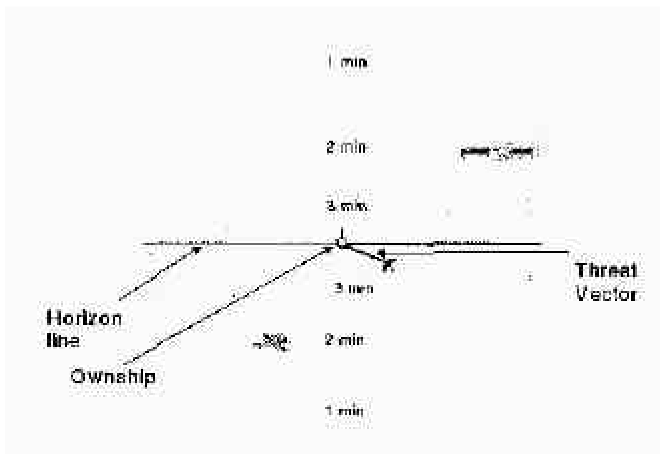

Figure 2 - Space-time display with threat vector (Krishnan et. al, 2000)

Study 1

The first study compared distance-based, time-based, and 'space-time'-based perspective displays. 18 participants were involved in a 3 by 6 two factor mixed design. Independent variables were display type and display sequence (order of presentation). Display type was varied within subjects, and sequence was varied between subjects. Dependent measures were percentage correct responses and the corresponding response times for the correct responses.

The task involved flying a Pathway-in-the-sky (PITS) display at level altitude while answering the following: Is a maneuver required? Which way should the initial maneuver be? Which aircraft is the most imminent collision threat?

Results were analyzed with the aid of Split-Plot ANOVA and determined a higher level of collision awareness and avoidance with the time-based displays. The main effect of display type on percent correct responses was significant (larger percentage with time-based displays), while main effect of display sequence and the interaction was insignificant. There was also a significant difference in the response time for correct responses for the main effect of display type (lower response time with time-based displays), but insignificant for display sequence and the interaction.

Study 2

Three displays based on ‘space-time’ were used. Pilot’s perspective, split-screen, and third person perspective (see Figure 3). 18 participants were used. The primary task was to maintain a certain altitude and heading on a predetermined flight route. The secondary task required the participant to answer two questions: Is a maneuver required and in which direction should the initial maneuver be? What direction would you look (outside the aircraft) to locate the aircraft that poses the most serious threat?

Results were analyzed using a one-factor within-subjects ANOVA for determining variance. The main effect of display type on mean reaction times was significant, with a lower reaction time for the pilot perspective display. For mean error rate for the maneuvering decision, the third person display showed higher error rate and differed significantly from the other two. Results from this second study suggested that mean reaction times and mean error rate for maneuvering decision were lowest for the pilot perspective display.

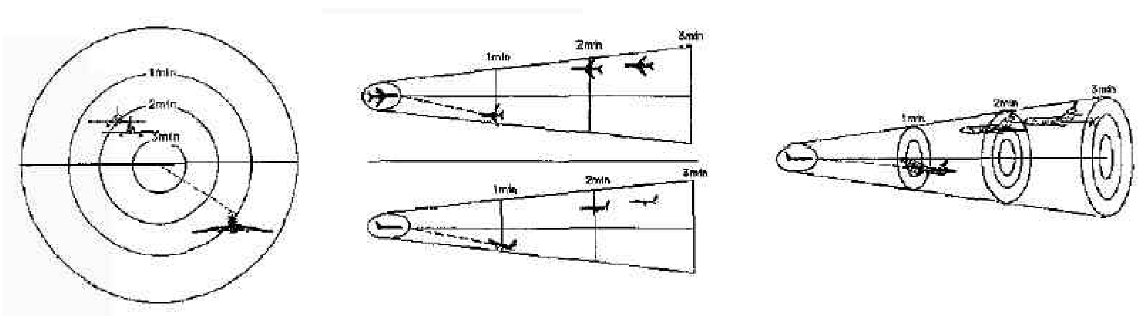

Figure 3 - Pilot's perspective, split-screen, and 3rd person displays (Krishnan et. al, 2000)

Scaling a display with the dimension of time is an interesting approach to conveying threat information. Since aircraft speeds are wide-ranging, a temporal measure of traffic proximity may be more representative of the ‘closer’ aircraft. However, other flight deck displays that show terrain, weather, and radar information are all distance based. Cognitively, there may be increased mental workload demand as the pilot switches between these frames of reference. Further investigation may be necessary to determine any performance effects caused by the switching between multiple frames of reference.

<u>Effects of Automated Cueing on Decision Implememtation in a Visual Search Task</u>
Scott M. Galster, Robert S. Bolia, Merry M. Roe, Raja Parasuraman, (2001)

A visual search task was used to examine the effects of status information automation cueing in a target detection task. Status information automation cueing is a method for aiding users in decision making. The visual search space consisted of a 'sky' portion and a 'ground' portion, where the ground was populated with the distractor set (¦ , - , ¦ , and +) and the target (-). Eight participants took part in the study, which was a within subjects 2 by 2 factorial design. The factors were automation condition (manual and information automation) and a distractor set (10 objects vs. 20 objects). Manual and information automation conditions were manipulated and participants responded to the absence or presence of the target within trial periods of 2.5 second duration.

Within the information automation condition, cue validity (a correct or incorrect target was highlighted) was varied to determine its effects. The manual condition required the participant to perform visual detection to determine the absence or presence of the target. The information automation condition was identical with the exception of a single highlighted element. In 67% of the trials, the correct target was highlighted to guide the participant's search task, and on the remaining trials a distractor element was highlighted instead, when a target was present elsewhere in the field.

Detection performance increased with the addition of information automation, despite varied levels of distractor sets (10 vs. 20). Repeated measures ANOVA was performed to measure the factor effects on correct responses, response times, and timeouts (where user did not provide an answer after 2.5 seconds). The qualitative summary of results is listed below:

- Participants made fewer correct responses as the distractor set size increased from 10 to 20, but the number of correct responses was higher when the automated status information aid was employed.
- Cue validity affected task performance only for sufficiently large distractor set sizes.
- Response times increased as the number of distractors increased, and was better in the automated condition than in the manual condition for the larger set size only.
- Reponse times increased with set size when invalid cues were employed. Response times were higher in the invalid cueing condition.
- The presence of automated cue reduced the number of timeouts.
- Timeouts accounted for the majority of the incorrect responses.
- Timeouts were not significantly affected by cue validity, but significantly affected by distractor set size. Timeouts increased with set size.

NASA-TLX workload and subjective confidence measures were taken after every 100 trials, and participants reported equal self-confidence in manual and automated status information conditions. The study showed that a performance benefit was achieved with the presence of status information cueing. This study reinforces the notion that reliable information automation cueing improves decision making under time pressure as the TCAS study will likely confirm through experimentation.

Effects of Unreliable Automation on Decision Making in Command and Control
Ericka Rovira, Kathleen McGarry, Raja Parasuraman, (2002)

This study examined the effects of different stages of automation and automation reliability on user performance in a simulation of a Sensor-to-Shooter system (STS). 18 undergrad students participated. A mock STS was prepared on a PC, and participants were required to identify the most dangerous enemy target and corresponding friendly unit to engage it. An action selection was required within 10 seconds.

A 4 (automation levels) by 2 (automation reliability) within-subjects study design was used.

The four levels of automation were:
- complete listing (all available permutation of actions, targets, shows distances)
- priority listing (like above, but ranked starting with 'best' selections)
- top choices (like above, but only shows top three choices for engagement)
- recommendation automation (shows single best engagement without distance)

Two levels of automation reliability were 60% and 80% correct automatic assessments.

20 manual trials without automation support was performed first. Both reliability conditions of each automation level were completed before progressing to the next automated condition. Each condition was followed by 20 trials of 100% reliability to recalibrate trust in the automated aid. In total, each participant completed 500 trials.

Dependent variables included accuracy and speed of engagement selections. NASA-TLX subject measures were taken to obtain mental workload of each participant..

Paired sample t-tests were performed to determine factorial significance in the accuracy of engagement decisions. Accuracy did not differ significantly between manual conditions and reliable automation aid conditions, but it did decrease significantly with unreliable automation aid.

Reaction times were reduced significantly from the manual condition when participants were given a reliable decision aid. There was speculation that the recommendation automation condition prompted a faster "okay" response, so it was excluded from t-test trials, and reduced reaction times still occurred due to reliable automation. Further, there was an inverse relationship between reaction times and decision aid reliability.

Paired samples t-test revealed no significant difference in ratings of mental workload between manual and automated conditions.

Candidate Methods From Publications Summary

The following list summarizes the experimental design features from the six papers.

<u>Participants</u>

- Flight instructors
- Flight school students
- Undergraduate college students

<u>Independent Variables</u>

- Traffic display types (Baseline, NTZ, MIT)
- Alerting conditions (detection algorithms: none, NTZ, MIT)
- Traffic scenarios (low-convergence, high convergence, missed approach)
- Display format (3D, 2D coplanar: front-side, 2D coplanar: front-back)
- Display sequence (order of presentation)
- Traffic load (2, 6, 10 aircraft)
- Traffic intent information
- Automation condition (manual vs. information cueing)
- Automation level (options listing, priority, top choices, recommendation)
- Automation reliability (variable)
- Distractor sets (10 vs. 20)

<u>Dependent Variables</u>

- Percentage correct (or accurate) response
- Reaction time (time limited trials may 'timeout', noted as incorrect user response)
- Maneuver frequency
- Time spent in conflict
- Path deviation
- NASA-TLX subjective workload, self-confidence, and trust rating

<u>Analysis Methods</u>

- One-way ANOVA
- Two-way ANOVA
- Split-part ANOVA
- Repeated measures ANOVA
- Paired samples t-tests

Highlighted Methods

Participants

There was a variety of participant profiles for the studies, from undergrad students to flight instructors. Although the institutional surrounding of a study may affect the type of participants chosen (A flight school has easy access flight students whereas a university has a large population of undergrad students), in all cases, participants were chosen as a capable group that possessed the mental capacity to compete the trials. Furthermore, some were trained or practiced, and had to meet a minimum performance level before experimental trials began. This is common in studies requiring tasks of skill such as simulator flying.

Independent Variables

The independent variables are the primary factors in question related to the area of study. These were the driving forces for the study, because they were manipulated within and between subjects, ordered, and varied in degree (such as degree of automation reliability). Alexander and Wickens (Alexander, Wickens, 2001) set out to determine the effects of display orientation, so one of their independent variables was orientation. They used three orientation views, and in post experiment analysis, ANOVA was used to distinguish the effects of each orientation on the whole task or parts of it. The majority of studies consisted of two factors. That is, two independent variables were manipulated to produce the trials. Where the first factor has 'x' possible values while the second has 'y' possible values, the experimental design was a described as having an 'x' by 'y' factorial design.

Dependent Variables

The participant always plays a responsive role in an experiment where actions can be measured, whether they react to stimuli or undertake actions as part of following procedure. At a minimum, they will respond by acknowledging the presence or lack of a signal. Although the independent measures can be as diverse as the research areas themselves, the range of dependent measures is remarkably fewer in number. This is partly due to the lack of facilities for measuring (such as eye-tracking equipment, pulse rate meters, etc), but it is also due to the inherent attributes of the domain being studied. Since the domain of aviation displays and automation focuses on the user's detection and response to visual and auditory stimuli, the common questions for measuring display performance are "Did they respond?", "How long did it take them?", and "What was their response?". A derivative of "how long it took" can be "time spent in conflict", for example, as measured by Alexander and Wickens (Alexander, Wickens, 2001).

Analysis Methods

Analysis of variance (ANOVA) is a very common and widely accepted data analysis tool for multi-factorial experiment designs (t-test is used for two groups of the same factor). It is an efficient way of determining main factorial effects all factors and interactions between them. In addition, NASA-TLX is often administered to determine workload and self-confidence measures to determine if these affected the experiment conditions.

Direction of TCAS 2 Study

The primary focus of the study is to measure the effect of EID-based displays on the pilot response to an automated alerting system such as TCAS 2. It is hypothesized that a display designed using EID principles will expose more useful system information to an expert operator (the pilot), thus improving his accuracy and performance in perceiving and reacting to an alert. In TCAS 2, the two classes of collision alerts are Traffic Advisories (TA) and Resolution Advisories (RA). They are analogous to yellow 'caution' and red 'emergency' stages of any warning system, with the addition that the RA also suggests a calculated and preventative course of action to all the involved aircraft.

A midair collision between a Russian airliner and DHL cargo jet over southern Germany in 2002 prompted concerns over human factors issues regarding TCAS (Ladkin, 2002). Although the larger culprit appears to centre around policy and training; Russian pilots obey air traffic control (ATC) over TCAS while North American pilots obey TCAS over ATC, the concern remained that the TCAS system may be inadequately conveying advisories to the pilot.

A simulated TCAS 2 gauge was developed for Microsoft Flight Simulator 2002 to mimic TCAS 2 functionality, including aural TA and RA announcements. Collision scenarios will be fabricated on Flight Simulator, and these will serve as trial conditions for participants in the normal TCAS 2 function display mode. From the work domain analysis conducted by Ho and Burns (Ho, Burns, in press), EID will be applied to create a set of similar displays with varying levels of automation information, and these will be tested against the control to determine the variabililty of pilot responses.

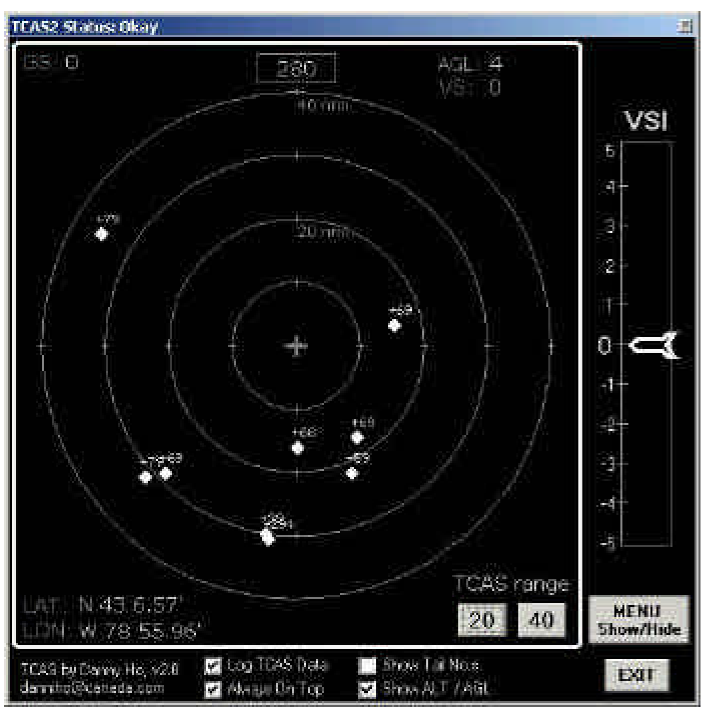

Figure 4 - TCAS 2 screenshot

Relating to the TCAS Study

Each of the six papers contributes to the discussion of the TCAS study and its experimental design, either directly or indirectly.

Pritchett et. al measured only the timing of a button press to indicate the pilot's desire to maneuver (Pritchett, Vándor, 2001). This is very relevant to the TCAS study since the focus is not on how the participant evades traffic, but on the timing of the decision itself. In addition, there is strong similarity between the TCAS study and Pritchett's consonance study, since consonance is described as "exposing automation rationale" (Pritchett, Vándor, 2001). This is exactly how EID may potentially improve the TCAS display.

Alexander et. al studied how display orientation affected maneuver performance, but that does not lend well to the intent of the TCAS study. However, they also measured reaction time of the participant by measuring "time spent in predicted conflict". If the predicted conflicted situation is clearly conveyed to the participant, then this measure becomes "reaction time to maneuver decision" as in the consonance study.

The publication by Barhydt et. al addressing aircraft intent information (such as autopilot settings, waypoints) is an interesting one, because it has potential overlap with the information requirements exposed by the work domain analysis model completed for the TCAS study. However, the datalink infrastructure for exposing such low-level data is not widely available, and as such, was beyond the scope of the TCAS study. The TCAS study focuses on information currently available to the TCAS system, as outlined by the FAA TCAS 2 specification (FAA, 2000).

The "space-time" study by Krishnan et. al showed an interesting approach to decision performance measure by asking the questions onscreen. This method seems more intrusive, where trial procedures to think their decision out loud may have sufficed. Alternately, they could have pressed a button to stop the simulation, then explain their maneuver choice.

In the visual search task study by Galster et. al, the variable automation condition is similar to that of the TCAS study. Instead of only two conditions (manual vs. information cueing), several EID displays will be used to determine if the effects are display specific or common to all EID generated displays. The distractor set is less relevant, since the TCAS system performs automatic highlighting of TA and RA alerts, but further investigation of TCAS traffic clutter is recommended.

The command and control study varied automation level and reliability and measured user accuracy and performance. This closely resembles the TCAS study, because although TCAS is theoretically flawless, there is still doubt cast on it's reliability. By varying the reliability of the TCAS study, the affect of reliability on user's perception and reaction to alerts can be explored.

Recommendations and Conclusions

This literature review has highlighted common research goals and experimental themes between the TCAS study and existing publications of a similar domain. By relating them with the current TCAS study, measuring techniques can be borrowed to achieve similar measurements of perception and reaction to automated TCAS alerts.

In the four major categories, the following recommendations are made in hopes of providing a suitable framework for the TCAS study to proceed with experimental design.

Participants

Participants must show a suitable level of analytical performance and must be trained to exhibit a minimum level of 'collision awareness' in order to accept or reject automated collision alerts. Since the TCAS study is being carried out in a university setting, undergrad students with mathematical proficiency may be suitable participants. It is recommended that test runs take place until the student can achieve a minimum level of conflict prediction performance based on parameters such airspeed and vertical speed, before experimental trials begin.

Independent Variables

The independent factors shall be display formats (non EID display, several EID displays) and a varying level of TCAS automation reliability (60%, 80%, 100%).

Dependent Variables

The dependent factors shall be percentage of correct response, reaction time, and time spent in conflict. NASA-TLX shall be administered to measure subjective workload and self-confidence levels. Questionnaires shall also be used to evaluate participant trust in each automation scenario.

Analysis Method

ANOVA shall be performed to determine main factorial effects and their interactions.

References

Alexander, A.L., Wickens, C.D., (2001). Cockpit display of traffic information: the effects of traffic load, dimensionality, and vertical profile orientation. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, 105-109.

Barhydt, R. and Hansman, R.J., (1999). Experimental studies of intent information on cockpit traffic displays. *Journal of Guidance, Control, and Dynamics, Vol. 22, No. 4, AIAA, July-Aug. 1999*, 520-527.

FAA (2000). Introduction to TCAS II Version 7. *U.S. Dept. of Transport. Federal Aviation Administration*. Nov. 2000.

Galster, S.M., Bolia, R.S. (2001). Effects of Automated cueing on decision Implementation in a Visual Search Task. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, 321-325.

Ho, D., Burns, C.M. (in press). Ecological interface design in aviation domains: work domain analysis of automated collision detection and avoidance.

Krishnan, K., Kertesz, S.Jr., Wise, J.A., (2000). Putting four dimensions in "perspective" for the pilot. *Proceedings of the IEA 2000/HFES 2000 Congress*, 3-81 - 3-84.

Ladkin, P.B. (2002). ACAS and the south german midair. (*Technical Note RVS-Occ-02-02*). Bielefeld, Germany: Universitat Bielefeld, Networks and Distributed Systems Research Group.

Pritchett, A.R., Vándor, B., (2001). Designing situation displays to promote conformance to automatic alerts. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*. 311 – 315.

Rovira, E., McGarry, K., Parasuraman, R. (2002). Effects of unreliable automation on decision making in command and control. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, 428-432.