Image Cover Sheet

CLASSIFICATION	SYSTEM NUMBER 509688
UNCLASSIFIED	

TITLE

EXPERIMENTAL QUESTIONS AND PRELIMINARY EXPERIMENT DESIGN TO EVALUATE ELBOWS DISPLAY ALTERNATIVES

System Number:

Patron Number:

Requester:

Notes:

DSIS Use only:

Deliver to:

DCIEM NO. 98-CR-50

September 1998

PWGSC Contract Number W7711-7-7431/A

Experimental Questions and Preliminary Experiment Design to Evaluate ELBOWS Display Alternatives

Laura Lynne Churchill

Daniel Edward Computing Solutions Inc.
50 Qualicum Street
Nepean, ON K2H 7H4
e-mail: Laura.Lynne.Churchill@DECSI.com

Table of Contents

Introduction

The critical question that must be answered before the design of the ELBOWS interface is finalized is "What warning style most effectively presents the required information to the aircrew?" However, because of the differing characteristics of the proposed warnings, answering this question is not as straightforward as simply rating each proposed warning style on a scale and selecting the one with the highest score. In order to compare the potential warnings they must be quantitatively evaluated through experimental studies with real operators. Possible warning styles identified include: voice directions; alerting tone; visual alert; alerting tone and voice directions; alerting tone and visual directions; visual alert and visual directions; heads-up display(HUD); or helmet mounted display (HMD). Due to financial and technical limitations, HUDs and HMDs are not currently considered feasible alternatives. In an attitude survey administered to CF helicopter pilots and flight engineers in January and February 1998, operators were asked to report their preferred warning style for an obstacle avoidance system. The three most commonly reported preferences were: voice directions; audio alert and visual directions; and visual alert and visual directions. Because of the small difference in the number of responses for each of the three warnings, the relative effectiveness of each should be compared. Several supplementary research questions must be answered before a comparison of the three candidate warnings can be conducted. Many of these questions were identified in the project plan for ELBOWS (Mack, Arrabito, and McFadden, 1997, p.2) and the remainder arose from the attitude survey and a task analysis performed in January 1998 (Churchill, 1998). The questions to be addressed are as follows:

1. What style of warning – command or information – is most appropriate for ELBOWS?
2. Which warning style of the three most commonly reported in the attitude survey results most effectively presents the required information to the pilot?
3. What are the characteristics of an optimal audio alert?
4. What are the characteristics of an optimal voice warning?
5. What are the characteristics of an optimal visual display?
6. Do 3-dimensional audio warnings have an advantage over 2-dimensional audio warnings under realistic workload conditions?
7. Can audio warnings be "tuned out" during periods of high auditory workload?
8. What is an adequate warning time for ELBOWS based on human and aircraft performance data?
9. Is there a false alarm rate at which operators stop trusting warnings systems? If so, what is that rate for ELBOWS?

Experimental Questions

Command vs. Information

A critical question that must be addressed is the general style of warning – command or information – that is most appropriate for ELBOWS. A common measure of the effectiveness of warning systems is the reaction time of the operator. However, this is not always an appropriate measure, especially if considered independently of the context in which the system is used. In the case of an obstacle warning system, the nature of the command must be considered with as much, if not more, importance than the reaction time. An example of a simple, yet effective, command-style warning is the Ground Proximity Warning System (GPWS) installed in commercial aircraft. This warning is straightforward, instructing the pilot to "Pull up" when the aircraft comes too close to the ground or descends too rapidly. Although the number of accidents that have been avoided because of the warning cannot be documented accurately, since the introduction of the system into the cockpit analysis of accident rates has identified a decrease (Wickens, 1992, p. 535). However, the scenario in which GPWS is used is a simple one, with only one possible course of action to avoid the collision. In contrast, there are several alternative actions available to a pilot attempting to avoid an obstacle such as wires. The impact and probability of inappropriate commands must be considered in addition to the reaction time to a command-style warning. Indeed, to effectively warn the pilot a command-style warning display may need to evaluate all of the possible evasive manoeuvres (e.g. climb, descend, turn left) in the context of other flight data, such as enemy aircraft. Because of the number of variations in evasive manoeuvres pilots may use to avoid obstacles, an informational warning may be more appropriate. It is believed that the answer to this question may be developed through literature review and operator interviews.

Optimal Warning Style

In an attitude survey administered in January and February 1998, pilots and flight engineers were asked to report their preferences for display of a wire warning system. Several suggestions were presented to the participants, and they were encouraged to report any ideas not included on the survey. The three most commonly reported preferences were: Voice directions; Alerting tone and Visual Display; and Visual alert and display (Churchill, 1998). Because of the small number of operators surveyed, there was not a significant difference between the number of responses for the three most commonly reported preferences. Consequently, it was decided that the efficacy of the three most commonly reported displays must be compared quantitatively through experimental study.

However, it must be remembered that the preferred warning style is not necessarily the optimal one. During the proposed experimental work evaluating the alternative displays, quantitative analysis of the data may contradict preferences reported by participants. Also to be considered when selecting the optimal warning is the trade-off between reaction time and accuracy of localization. For an obstacle warning system such as ELBOWS, simply minimizing operators' reaction times is insufficient. Because of the myriad of evasive maneuvers available to the pilot, providing cues that aid in the accurate localization of the obstacle may be of even greater importance than initiating an automatic response on the part of the pilot.

Characteristics of Audio Alerts

A proposed warning style for ELBOWS is an alerting tone coupled with a visual display. This style of warning would use the tone to direct the operator's attention to the visual display. In addition to directing operators' attention to the visual display, it is hoped supporting information can be incorporated into the audio alert. As an example, a three-dimensional audio alert can provide cues to aid in the localisation of the obstacle. Alternatively, the urgency of the alert can be encoded in the audio alert by varying the onset speed or repetition rate (Hellier, Edworth, and Dennis, 1993).

An audio alert is simple conceptually, yet many characteristics must be considered for the alert to be effective. While there are many attributes of an audio alert that must be assessed, the two primary factors are fundamental frequency and volume. Perhaps the most crucial factor in selecting a candidate tone for an audio alert is its distinctiveness from other audio warnings in the aircraft. Because of the applicability of ELBOWS to both Search and Rescue (SAR) and tactical missions, it is likely that the system will be installed on multiple platforms. Ideally, the same alerting tone will be used on both aircraft to minimize training and allow greater movement of operators between aircraft. This implies that the candidate warning must be distinct from all audio alerts on both aircraft. A set of distinct attensons has been developed and evaluated by the Defence Research Evaluation Agency (DERA) (1997). It is anticipated that this attenson set will be used on the new SAR platform, and for consistency the alerting tone should be selected from this set. It is hoped that at least one of the identified attensons is unique from the existing audio warnings in the tactical aircraft (CH 146 Griffon).

Characteristics of Voice Warnings

The primary question that must be answered is what text is used to provide warning information to the operators. The attitude survey administered in January and February 1998 asked participants who preferred a voice warning to suggest the content of such a warning. The responses were: Obstacle; Obstacle, clock heading; Obstacle, clock heading, distance to obstacle (ft); and Obstacle, clock heading, distance to object (ft), elevation of object (ft) (Churchill, 1998). An alternative not suggested in the attitude survey but that bears investigation is the presentation of distance information in time units. Thus, instead of presenting "Wires, 2 o'clock, 500 ft," the system would give the warning "Wires, 2 o'clock, 4 seconds."

Theoretically, the more information that is provided to the operator, the better able he or she will be to respond to the emergency. The difficulty that exists for ELBOWS, and many other systems, is that if all of the information is presented to the operator, he or she may not have sufficient time to prevent or minimize the severity of an accident. Consequently, a trade-off must be made between the completeness of the information and the time available to the operators to respond. This trade-off is what must be explored when during the development of the voice warnings.

As an alternative to selecting a fixed length voice warning from one of the four reported, it was suggested that the length of the warning be varied with the time to striking the obstacle. An example would be to have the system present "Obstacle" if the object is less than 250 ft. from the aircraft, "Obstacle, clock heading" if the object is between 250 and 500 ft., etc. However, variable length warnings may influence operator performance differently than fixed length warnings. It is possible that operators may come to depend on the longest version of the text warning, and in situations where only some information is presented, their performance may be worse than if they consistently heard a shorter warning.

Characteristics of Visual Displays

Developing a visual display often asks many questions of the designer. How should a three-dimensional system be represented on a two-dimensional display? What role should colour play on the display? Where should the display be located to minimize scanning time or time away from the primary task? How can the information be presented on the visual display to minimize the time the pilot requires to process the information? How can an additional visual display be incorporated into an environment that already has a heavy visual workload? (Wickens, 1992, pp. 82-83; 101-102; 130 and Newman and Lamming, 1995, pp.392-395) Operator feedback obtained during the attitude survey suggested pilots might prefer a display similar to the threat-warning system for the CH 146 Griffon. This display uses icons on a clock-like circle to represent the location of threats. Two essential issues must be addressed if this style of visual display is selected for ELBOWS. The first is the use of color on such a display, as it is anticipated ELBOWS will be used with and without night vision goggles (NVG). Also to be considered is the position of such a display on the instrument panel. The display must be centrally located, so pilots can quickly observe the position of the obstacle, yet it must not interfere with their scan of the primary flight instruments. Of the two questions, the position of the display on the instrument panel is likely to be the more difficult to resolve.

Two-Dimensional vs. Three-Dimensional Audio under Realistic Workload

The use of three-dimensional sound has been shown to improve accuracy and reduce the time required to localize a target in laboratory experiments and flight simulator studies (McKinley, Erickson, and D'Angelo, 1994; Bronkhorst, Veltman, and vanBreda, 1996). However, the majority of studies investigating the use of three-dimensional sound as a localization aid place participants in a contrived situation. Often, the workload or background noise is not representative of the conditions under which the three-dimensional audio system is to be used. Because operators using ELBOWS will require information to localize the obstacle in situations where fractions of seconds can be critical, it is hoped that the use three-dimensional audio in an alerting tone or voice warning will improve operator performance. However, three-dimensional audio has not been studied under the high workload and noise conditions associated with flying a helicopter. It is vital to determine if the advantage of three-dimensional sound over two-dimensional sound observed in laboratory studies is preserved under realistic conditions. Because of the costs associated with three-dimensional audio, if a reduction in reaction time or improvement in localization is not observed, there would be no benefit realized by installing such a system.

False Alarm Rate

A system's false alarm rate is an important design consideration. This is especially true for critical applications such as nuclear power or avionics. The result of an excessively high false alarm rate is the "cry wolf" effect: operators stop trusting the system, believing each alert to be just one more false alarm. However, developing a system with a zero false alarm rate is not practical, if even possible. This is true of ELBOWS: a trade-off must be made between the number of false alarms presented and the detection range of the system. Determining quantitatively the false alarm rate at which pilots' behavior is affected is an important part of the ELBOWS design. This is not an easy task. Determining the optimal false alarm rate is in essence a Signal Detection Theory (SDT) experiment, which requires many presentations of warnings and false alarms. It has been shown that operators match their performance to the true behavior of the system. That is, over time, operators identify true warnings and false alarms at the same ratio as they are presented (Bliss and Gilson, 1998). To establish the ratio of false alarms to true warnings at which pilots stop trusting ELBOWS, each participant must be presented with sufficient stimuli at a variety of levels. A between subject experimental design is inadequate in this case, because any result obtained could well be the result of individual response characteristics, rather than true differences.

Performance Characteristics of Aircraft

The response characteristics of the target platform, combined with those of the operator, limit the minimum range of the system. Of particular importance is the time required for the aircraft to change direction or attitude after a control input is initiated by the pilot. If, for example, a pilot requires 2.5 seconds to initiate a control input in response to an alert and the aircraft then takes 3 seconds to begin responding to that input, there would be little point in presenting alerts for obstacles closer than 5 seconds to the aircraft. Fortunately, the characteristics for one of the target platforms, the CH 146 Griffon, can be readily obtained. One component of the training simulator is a computer workstation that performs collects data on aircraft position, speed, attitude, etc. By initiating specific control maneuvers at precisely documented times during the flight, the delay between the initiation of that maneuver and the aircraft response can be calculated. This, in combination with human reaction time data, can be used to determine the minimum range required for ELBOWS.

Can Audio Warnings be Ignored?

This research question arose from concerns reported by operators in the attitude survey conducted in January and February 1998. When asked to explain their preference for most effective warning style, several operators reported that visual warnings were more effective because, unlike audio warnings, they could not be "tuned out" during periods of high auditory workload. If this is true, the addition of another audio alert will exacerbate the problem. However, the veracity of the survey participants' claim has not been established. This concern can be investigated in while the efficacy of audio alerts and voice directions is measured in a realistic environment.

Experimental Design

Distinctiveness of Audio Alerts

Two approaches may be adopted to measure the uniqueness of each candidate tone. The first, used by Chillery and Collister (1988a) teaches participants ten named tones at their own pace. During the learning phase of the experiment, new tones are introduced one at a time in a Latin Square block randomized order. After a new tone is presented, an identification test supplements the learning phase. Feedback is provided to the participants, and the next tone is not introduced until the participant can correctly identify all of the tones. A separate test phase follows the learning phase. After the participant learns all of the tones, each tone is presented once, in random order, and participants are asked to identify each tone (without feedback). While appropriate for Chillery and Collister's study, this method may not be so for the current project. The proposed experiment is not attempting to measure the differences between the candidate tones; instead, the variable of interest is the difference between any candidate tone and the current warning set in the helicopter. If Chillery and Collister's approach were to be used, it would only be possible to have each participant evaluate one candidate tone because of the probable learning effects. After the participant learned one candidate tone and the current warning set, he or she would be better able to identify the candidate tone as being different from the existing tones.

An alternative approach to that of Chillery and Collister exists. Instead of teaching the participants to identify all of the warnings, it is only necessary to teach participants to identify the candidate tone from a set of distractors. To assess the distinctiveness of the candidate warning(s) from the existing alerts, the distractor tones should be based on the current audio warning set. Given the current number of audio tones in the Griffon helicopter, a signal (candidate warning) to noise (current warnings) ratio of 1:4 is believed to be appropriate. For a parametric statistical analysis to be valid, a minimum of twelve to fourteen data points per variable are required (Drew, Hardman, and Hart, 1996, p. 373). Given a signal to noise ratio of 1:4, this requires the random presentation of a minimum of sixty alerting tones (twelve candidate warning and forty-eight distractors) per candidate. To prevent "saturation" of the participants, the presentation of each candidate could be broken down into two groups of thirty.

As mentioned previously, it is hoped the candidate warnings can be selected from the attensons identified by DERA. Existing audio alerts in the CH 146 Griffon were recorded for analysis during flight trials conducted at CFB Borden on June 5, 1998. Based on the analysis of the current warnings and the specifications for the DERA attensons, it is unlikely that more than four attensons will be identified as unique. The number of trials required to evaluate the candidate tone has been calculated using this maximum value. Clearly, if fewer than four distinct attensons are identified, fewer trials will be required.

Two complementary measures of tone uniqueness exist. The first is the number of correctly identified candidate tones, or conversely, the number of missed alerts. However, this measure in itself may be insufficient to select the most distinct warning. It is not only important that operators correctly identify the candidate tone from the existing warnings, it is also important they do no identify an existing warnings as a candidate tone. This secondary measure, the number of "false positives" complements the simple accuracy measure and allows for a more thorough evaluation of the proposed audio alerts.

It is highly desirable to evaluate the uniqueness of the candidate warning tones under a variety of realistic ambient noise conditions. In addition to audio alerts, ambient noise samples were recorded during the June 5, 1998 CH 146 Griffon flight trials. Seven ambient noise conditions were recorded: low hover, doors open; low hover, doors closed; high hover, doors open; high hover, doors closed; nap of the earth (NOE) transit, doors closed; low level transit, 70 knots, doors open; low level transit, 70 knots, doors closed; high level transit, 70 knots, doors closed; high level transit, 70 knots, doors open; high level transit, 120 knots, doors closed; confined area hover, doors open. A subset of these conditions, representing scenarios in which wire strikes are most probable, would be appropriate for background noise during this experiment. Due to the increased probability of a wire strike during nap of the earth (NOE) flight, it should be selected as one of the ambient noise conditions used in the experiment. At least two additional noise conditions should be selected from the most demanding of the set while remaining representative of flight regimes in which pilots are likely to encounter wires. Based on the set of flight regimes recorded, it is anticipated that low level cruising with the doors open and one of the doors-open hovering conditions will be the worst case scenarios.

Ideally, each participant would hear all (four) candidate tones under all (three) ambient noise conditions. To reduce confusion on the part of the participant, all trials for one candidate tone should be grouped together. However, the order of the ambient noise conditions should be randomized for each candidate tone to prevent order effects. Similarly, the order in which candidate tones are presented should be randomized across participants. To minimize learning effects that may occur during the first block of trials for each candidate tone, a familiarity trial under a no ambient noise condition should be used to introduce the new alert. Additionally, during the learning phase, participants should receive feedback when they correctly identify the candidate tone. If each trial consists of thirty tones (candidate warning and distractors), and each participant hears all candidate tones under all ambient noise conditions, a participant would hear twenty-four blocks of tones. However, this does not include the familiarity trials: if those were included, each participant would hear twenty-eight blocks of tones.

If a complete randomized block design was implemented for this experiment, and all combinations were considered, 4! X 6! (17 280) participants would be required! This is clearly impractical. To reduce the number of participants required, a blocked random design or Latin Squares design is more appropriate. The candidate tones can be grouped in such a way so that each warning tone occurs in each order position and following each other tone only once. As an example, if the tones are labelled A-D, the following order of presentations ensures each tone is only heard in each position once: Group 1 -- ABCD; Group 2 -- BDAC; Group 3 -- CADB; Group 4 -- DCBA. (*Note:* This is only one of many possible groupings.) Grouping the candidate tones in this way reduces the number of participants to 4 X 6! (144). This is much more reasonable; however, it is possible to reduce the number further by stipulating that the two blocks with the same ambient noise condition be presented together. This requires 4 X 3! (24) participants. Finally, if the order of ambient noise conditions is blocked in a manner similar to the candidate tones, the number of participants can be reduced to 4 X 3 (12).

Even with the minimum number of participants described above, a parametric analysis between participants is possible. Several approaches may be used to analyze the experimental data. Distributions of the number of correct responses, misses, and false positives can be generated. Because of the design, a statistical analysis of the results requires a repeated measures analysis of variance. If a between subject analysis is desired, i.e. to test for learning or order effects, a two-way analysis of variance will suffice.

While there is much preliminary work to be performed before this experiment is undertaken, the design is straightforward. If desired, additional variables such as two-dimensional vs. three-dimensional sound, might also be explored during this study.

Simulator Experiment

Several questions may be addressed during a single study in the CH 146 Griffon training simulator. Perhaps the most important of these is the collection of performance data for different warning styles. However, because of time and financial constraints, it is not possible to collect information for visual displays. In addition to collecting data to evaluate the efficacy of voice directions, the use of three-dimensional audio in a realistic workload and environment can be examined. While other studies have been performed in simulators examining 3-D audio and target localization, they admittedly placed the operator in a low workload situation (Bronkhorst *et al.*, 1996). In addition to observing operator performance, data on aircraft responses can be collected from the simulator de-briefing station. Aircraft response times can then be combined with operator performance data, to estimate the minimum warning time that must be provided by ELBOWS. The final question to be examined during this experiment arose from comments received on the attitude survey. Several CF members reported that they felt audio warnings could be tuned out during periods of high auditory workload (Churchill, 1998, p. 35). This can be studied in a controlled manner in the simulator by varying the quantity of radio traffic and other audio alerts.

The advantage to performing at least one of the experimental studies in the training simulator is the ability to gather aircraft performance data in addition to human performance data. The debriefing station of the training simulator records aircraft information such as airspeed, altitude, attitude, heading, etc. This information can be used to calculate the time between initiation of the specified control input and the reaction of the aircraft. Data markers can be inserted at two points in the record for each warning: the first when the warning is presented and the second when the pilot initiates the control manoeuvre. This will allow independent examination of the pilot and aircraft reaction characteristics.

Because of the limited availability of the training simulator, the experiment must not interfere with normal, scheduled activities. Discussions with Captains Dave Scott and Sean Leonhardt suggest that one of four commonly used training scenarios would be appropriate for introducing ELBOWS alerts. This simulation is "flown" a maximum of four times a week. The flight generally is one-and-a-half to two hours in length, which suggests a maximum of eight warnings per trial.

A proposed design, which examines all of the factors but minimizes the number of trials and interruptions to the training simulation, is a 2x2x2 factorial design. Each of the three factors to be explored has two levels. These are:

1. Characteristic of audio warning: is the audio warning voice directions or an alerting tone?
2. Dimensionality of warning: is the warning presented using two- or three-dimensional sound?
3. Auditory workload: is the background radio traffic at a high or low level? Although more than two levels could be considered to investigate if auditory warnings can be tuned out, to minimize the number of trials, it is suggested that two levels are appropriate.

With eight conditions and a suggested maximum of eight warnings per flight, each combination can be presented once per training mission. Several issues, discussed below, must be addressed before the design is finalized.

The goal of this research is to quantify, in some way, the efficiency of different warning styles. Ideally, the measure used for this study could also be used for the evaluation of visual displays in order to offer a valid comparison between the warning styles. The obvious response measure is reaction time: how long does it take the pilot to react (measured by the initiation of a specific control response) after an alarm is presented? This measure is not without drawbacks. While reaction time is a good measure of the alerting characteristics of a warning, it does not evaluate the localization of obstacles. A complex reaction time, in which participants initiate a control maneuver to move the aircraft in the direction opposite to the obstacle, measures the reaction time of the pilot and the effectiveness of localization cues. To evaluate the claim that audio alerts can be tuned out during periods of high workload, counting the number of missed alerts during the high radio traffic condition will provide at least preliminary data. If the results suggest that the assertion made by operators is true, a more exhaustive investigation may be warranted.

Several issues must be addressed before the experimental design can be finalized. The first is how the three-dimensional warnings are presented to the two pilots. If both pilots are to hear three-dimensional warnings, a second head tracker will be required. A related issue is how pilots share the flying duties. As an example, in a SAR helicopter pilots switch flying duties every half-hour. Two approaches exist to control the problems associated with pilot switching. The first is to only present the warnings to one pilot, ensuring he/she is flying at the time the warnings is presented. However, the impact on the training mission is likely to be such that the training staff will not approve of this solution. The second alternative is to group the warnings on one of the three factors under investigation and present one group to each pilot. The groups of warnings do not need to be presented in blocks, e.g. pilot A hears 4 warnings and then pilot B hears 4 warnings, the order can be adjusted in real time based on flying duties.

The second issue that must be resolved before the experimental design is how to generate the warnings so that they are consistent with environmental features that would be linked to wire strikes. As an example, the warning should be triggered when flying across a highway (a situation where wires are likely to be present) as opposed to flying across the desert (a situation where wires are not likely to be present). Three methods can be used to accomplish this:

1. The experimenter rides in the fourth simulator seat for all trials, triggering the warnings at appropriate times. Because some of the training missions involve the flight engineer in addition to the pilots, there may not be a seat available for the experimenter. Written approval is required for an individual to stand in the simulator.
2. The trainer is instructed on how to generate the warnings and at what times those warnings should be generated.
3. The path that the aircrew is to fly is specified by the instructor during the pre-flight briefing. Presently, the crew is free to select their own path through the scenario. This freedom, while preferred by the training staff, may allow the crew to completely avoid areas in which obstacles such as wires may be found. If the flight path were specified by the trainer during the pre-flight briefing, it would be possible to direct the crew through situations in which obstacle strikes may occur.

No matter which method is used to control the variability in the terrain, the warnings must be generated in real time due to differences in speed, tasking, etc. Further input from the training staff at CFB Gagetown may help to resolve this issue.

Two problems must be controlled to obtain reliable results. The first is the variation in pilot workload that may occur. For reliability, the workload experienced by the pilots when ELBOWS alerts are generated must be comparable across all trials. If the warnings are generated in real time by any one of the measures described above, it will be straightforward to generate the warnings under similar workload conditions. The other issue to be addressed before the experiment is conducted is how to control for unexpected events, such as crashes or aircraft downed by enemy fire. Because of the intent to conduct the experiment on top of a normal training mission, it is possible that the aircraft can be shot down or crash due to equipment failures. How to control for this in the data analysis is an issue yet to be resolved.

A parametric statistical analysis is more rigorous and, consequently preferred for this experiment. The drawback to a parametric analysis is the large number of data points required. Drew *et al* (1996, p. 373) suggest a minimum of twelve to fourteen data points per variable for a parametric analysis, while six to eight data points are sufficient for a non-parametric analysis. In addition to providing sufficient data for the desired statistical analysis, the number of trials performed must also be adequate to ensure the data collected is random. Therefore, for the simulator experiment, a minimum of twelve trials are required in order to collect twelve data points for each warning style/dimension/audio workload condition. This fails to consider how to minimize bias and order effects. The most direct approach, a simple randomized design would require 8! (40 320) trials. As for the distinctiveness experiment, this is clearly impossible. However, if the warnings were grouped by dimensionality and the pilots heard the warnings in the presented same order, the number of trials could be reduced to 2! X 4! (48). To illustrate, Pilot A and Pilot B would both hear the warnings in the following order – voice warning, low noise; tone, high noise; tone, low noise; voice, high noise – but Pilot A would hear all of the warnings in two-dimensions, while Pilot B hears all of the warnings in three-dimensions. There are 4! possible ways to present the candidate warnings, if grouped by auditory dimension (or any one of the factors), which must then be multiplied by 2! to account for the assignment of auditory dimension to the two pilots. It may be possible to further reduce the number of trials required by grouping using another factor or considering a non-parametric analysis. If a parametric analysis is performed and the data points are grouped so that only the variable of interest is examined, i.e. interaction effects are not considered, an independent samples *t*-test is an appropriate measure. However, if interaction effects are to be evaluated, a two way ANOVA is required.

ACD Experiment

An experiment conducted using the Defence and Civil Institute for Environmental Medicine's (DCIEM) ACD will allow investigation of three research areas. The questions that may be explored are:

1. What is the optimal length of a voice warning?
2. Do variable length warnings influence operator reaction time or accuracy?
3. What are the characteristics of an optimal visual display?

There are many similarities between this experiment and the one to be conducted in the training simulator. As with the previous experiment, eight or nine warnings can be presented during a two hour trial. To compare the effectiveness of the different warning styles studied in this and the simulator experiment, the same measure, e.g. simple or complex reaction time, must be used. However, the design of the ACD experiment is more involved because of the potential for learning effects. If, as hypothesized, variable length warnings influence performance, presenting both styles of warnings to one participant may confound the experimental results (p. 3). There are four design approaches available to address the above questions:

1. During the first hour of a flight in the ACD, each participant hears four presentations of one of the candidate warnings. At the beginning of the second hour, the halfway point in the flight, the participant is provided with information instructing them about the change in experiment (and the now variable length of the warning). Four warnings will then be presented during the second hour, one presentation of each candidate warning (in random order). *Number of participants: 48.*
2. The warnings are presented in the reverse order described above. Participants would be randomly presented with each of the warning texts during the first hour of a flight. At the beginning of the second hour, the halfway point in the flight, the participant is provided with information instructing them about the change in experiment (and the now fixed length of the warning). *Number of participants: 48.*
3. The procedure is as described in Methods 1 and 2, but the order of presentation is randomized across participants. Half of the group would hear the fixed length warnings followed by the variable length, while the remaining half would hear the variable length warnings first. Number of participants: 48.
4. In this design, participants are only involved in one of the two conditions. One group of participants would hear each of the candidate warnings presented twice (in random order) over the course of the two hour flight. The other group of participants would hear four presentations of one of the candidate voice warnings during the first hour; and during the second hour each participant would then be presented with four warnings using one of the candidate visual displays. *Number of participants: 60.*

There are advantages and disadvantages to all of the designs; nonetheless, in spite of the increased complexity in the data analysis, the fourth design is preferred. The difficulty is increased due to the differing number of observations between conditions, as shown in Table 1. The significant advantage offered by the fourth design is that both voice and visual warnings can be investigated during the same study. There is also the advantage of eliminating possible learning effects between the fixed and variable length warning conditions.

Table 1: Number of Observations per Condition

Condition	Fixed Length	Variable Length
Voice Warning 1	48	24
Voice Warning 2	48	24
Voice Warning 3	48	24
Voice Warning 4	48	24
Visual Warning 1	64	-
Visual Warning 2	64	-
Visual Warning 3	64	-

A problem common to all designs for the ACD experiment is the large number of participants required and the two-hour duration of each trial. This is not an issue for the training simulator experiment because the data is to be collected during regular training missions; for this experiment, the participants must be recruited and brought to DCIEM in order to participate. As described in the simulator experiment design, ensuring consistent workload across participants and controlling for aircraft emergencies are issues to be resolved as the design is finalized.

References

Bliss, J.P. and Gilson, R.D. (1998) Emergency signal failure: implications and recommendations. *Ergonomics* 41(1): 57-72.

Bronkhorst, A.W., Veltman, J.A., and vanBreda, L. (1996) Application of a three-dimensional Auditory Display in a Flight Task. *Human Factors* 38(1): 23-33.

Chillery, J.A. and Collister, J. (1988a) Confusion Experiments on Auditory Warning Signals Developed for the Sea King Helicopter. Technical Memo FS(F) 688. Farnborough (Hants, England): Royal Aerospace Establishment.

Chillery, J.A. and Collister, J. (1988b) Perceived Urgency of Warning Signals Determined using a Forced Choice Pair Comparison Technique. Technical Memo MM 12. Farnborough (Hants, England): Royal Aerospace Establishment.

Churchill, L.L. (1998) Human Factors Analysis in Support of ELBOWS Interface Design. Mississauga: Daniel Edward Computing Solutions Inc.

Defence Evaluation and Research Agency (1997) A proposal for a UK standard for the Human Factors aspects of auditory cockpit warnings. Farnborough (Hampshire, England): Defence Evaluations and Research Agency.

Drew, C.J., Hardman, M.L, and Hart, A.W. (1996) *Designing and Conducting Research.* Needham Heights (Massachusetts): Allyn & Bacon.

Hellier, E.J., Edworthy, J., and Dennis, I. (1993) Improving Auditory Warning Design: Quantifying and Predicting the Effects of Different Warning Parameters on Perceived Urgency. *Human Factors* 35(4): 693-706.

Mack, I., Arrabito, R., and McFadden, S. (1997) Project Plan for Defence and Civil Institute of Environmental Medicine (DCIEM) participation in: Eye-safe Laser Based Obstacle Warning System (ELBOWS). Toronto: Defence and Civil Institute of Environmental Medicine.

McKinley, R.L., Erickson, M.A., and D'Angelo, W.R. (1994) 3-Dimensional Auditory Displays: Development, Applications, and Performance. *Aviation, Space, and Environmental Medicine* (May 1994): pp. A31-A38.

Newman, W.M. and Lamming, M.G. (1995) *Interactive System Design.* Don Mills (Ontario): Addison-Wesley Publishing Company.

Wickens, C.D. (1992) *Engineering Psychology and Human Performance, Second Edition.* New York: HarperCollins Publishers Inc.

S09688