Australian Government
Australian Transport Safety Bureau

ATSB TRANSPORT SAFETY REPORT
Aviation Safety Research Grant – B2005/0120

Design and Evaluation of Auditory Icons as Informative Warning Signals

Dr Catherine Stevens, Nathan Perry, Dr Mark Wiggins, Clare Howell
MARCS Auditory Laboratories
University of Western Sydney

August 2006

Australian Government

Australian Transport Safety Bureau

ATSB TRANSPORT SAFETY REPORT

AVIATION SAFETY RESEARCH GRANT B2005/0120

Design and Evaluation of Auditory Icons as Informative Warning Signals

Dr Catherine Stevens, Nathan Perry, Dr Mark Wiggins, Clare Howell
MARCS Auditory Laboratories
University of Western Sydney

August 2006

Published by: Australian Transport Safety Bureau
Postal address: PO Box 967, Civic Square ACT 2608
Office location: 15 Mort Street, Canberra City, Australian Capital Territory
Telephone: 1800 621 372; from overseas + 61 2 6274 6590
Facsimile: 02 6274 6474; from overseas + 61 2 6274 6474
E-mail: atsbinfo@atsb.gov.au
Internet: www.atsb.gov.au

Aviation Safety Research Grants Program

This report arose from work funded through a grant under the Australian Transport Safety Bureau's Aviation Safety Research Grants Program. The ATSB is an operationally independent bureau within the Australian Government Department of Transport and Regional Services. The program funds a number of one-off research projects selected on a competitive basis. The program aims to encourage researchers from a broad range of related disciplines to consider or to progress their own ideas in aviation safety research.

CONTENTS

DOCUMENT RETRIEVAL INFORMATION

Report No.	Publication date	No. of pages	ISBN
B2005/0120	August 2006	41	1 921092 87 4

Publication title

Design and Evaluation of Auditory Icons as Informative Warning Signals

Author(s)

Dr Catherine Stevens, Nathan Perry, Dr Mark Wiggins, Clare Howell

Organisation that prepared this document

MARCS Auditory Laboratories, University of Western Sydney
Locked Bag 1797, Penrith South DC 1797

Funded by

Australian Transport Safety Bureau
PO Box 967, Civic Square ACT 2608 Australia
www.atsb.gov.au

Abstract

Auditory icons – caricatures of everyday sounds – have the potential to convey information by non-verbal means quickly and accurately. Two experiments investigated the application of auditory icons as warning signals to the civil aviation cockpit environment. Warning signals that are iconic and that stand in a direct relation to the event being signalled, such as the sound of coughing to signal the presence of carbon monoxide, should convey information about the nature of the critical event as well as alerting the operator that there is a problem. By contrast, signals that are arbitrarily associated with an event, such as a beep to signal the presence of carbon monoxide, provide little information about the nature of the event. Speed and accuracy of recognition in response to these different types of warnings may also be influenced by modality (visual, auditory, auditory + visual) and by task demand (low, high). Experiment 1 investigated effects of signal iconicity (iconic, abstract), modality, and task demand on warning recognition speed and accuracy. One-hundred and seventy-eight participants completed a computer-based training session and test task that involved responding to warnings associated with nine critical events while completing low- and high-demand concurrent tasks. As hypothesized, fewer training trials were required to learn iconic warnings compared with abstract warnings. During the test phase, the effect of iconicity, as hypothesized, was influenced by modality and task demand. Bimodal (auditory + visual) warnings were recognized with the greatest consistency and accuracy. Auditory abstract warnings elicited slow reaction times and poor accuracy. Auditory iconic warnings, under conditions of high demand, evoked levels of accuracy comparable with bimodal warnings. Experiment 2 investigated recognition speed and accuracy in response to four auditory iconic and four abstract warnings in an Advanced Aviation Training Device. As hypothesized, accuracy was greater in response to auditory iconic than abstract warnings and recognition accuracy and reaction time were unaffected by level of flying experience. Reaction times in the Advanced Aviation Training Device were approximately 1 second. These initial experiments suggest that there is potential for the use of auditory iconic warnings and bimodal warnings as the means, not only to alert, but also inform pilots about the nature of a critical incident.

ACKNOWLEDGEMENTS

The authors acknowledge the funding support provided by the Australian Government, through the Australian Transport Safety Bureau's Aviation Safety Research Grant Program.

The authors would also like to thank those who volunteered to take part in the experiments.

ABBREVIATIONS AND GLOSSARY

Abbreviations

AATD	Advanced Aviation Training Device
ANOVA	Analysis of Variance
ATSB	Australian Transport Safety Bureau
A+V	Auditory plus visual presentation
AV	Auditory visual
M	Mean
ms	Milliseconds
RT	Reaction time
s	Seconds
SD	Standard deviation

Glossary

Abstract warning	A warning that stands in a non-conventional or arbitrary relation to the event with which it is paired, e.g., the sound of three high-pitched tones to signal the presence of dangerous levels of carbon monoxide.
Analysis of Variance	An analysis of variance or ANOVA is a statistical procedure used to investigate "main" effects of single independent variables on a dependent variable, and "interaction" effects involving interactions between two or more independent variables.
Attenson	An initial, non-verbal signal that is used to alert an operator. An attenson is often followed by a verbal or non-verbal warning that is specific to the condition that initiated the warning signal.
Auditory icon	A caricature of an everyday sound, e.g., dog bark, a steam train whistle, or the sound of a slamming door.
Denotative referent	The event, condition, or critical incident indicated by a signal.
Iconic warning	A warning that stands in some meaningful relation to the event with which it is paired, e.g., the sound or visual image of coughing to signal the presence of dangerous levels of carbon monoxide.

Referent	The target event, condition, or incident to which a signal refers.
Signal	A visual or auditory pattern that is linked directly (causally) or indirectly (metaphorically or symbolically) with a critical incident, event or condition.
Signal-referent relation	The type of relation or mapping between a signal and the event or critical incident to which the signal refers. In a direct relation, the signal is causally related to the event, e.g., the sound of fire (the signal) refers to the event of fire (the referent). In an indirect relation, there is no direct causal link between the signal and the event to which the signal refers. This might be because there is no sound associated with the event or it may be because the sound of the event is ambiguous. An indirect (metaphorical) relation might be the sound of a mosquito (signal) referring to the presence of a helicopter (the referent).
Sign referent	A sign referent is an intermediate step in a signal-referent relation when the referent is difficult to portray, e.g., the sound of wind blowing could be a sign referent that serves as a surrogate sound denoting the event of aircraft icing (the referent).
Spectral profile	The spectrum of a sound wave is the distribution in frequency of the magnitudes (and sometimes the phases) of the component vibrations of the wave. The spectrum is often expressed by graphing power, intensity, amplitude, or level as a function of frequency. Spectral profile refers to a particular set of frequency components and magnitudes that make up a particular sound wave.

EXECUTIVE SUMMARY

Auditory icons or caricatures of everyday sounds were examined as a means to convey vital information to operators about critical events in a civil aviation context.

Two experiments were conducted that investigated the effect of systematic manipulation of warning type (abstract, iconic), modality (auditory, visual, auditory + visual) and task demand (low, high) on warning recognition speed and accuracy. In the multi-factor laboratory controlled experiment (Experiment 1) iconic warnings were learned more quickly than abstract warnings. Under conditions of high cognitive load, auditory iconic and auditory + visual iconic warnings were equally recognizable. Under those conditions, auditory abstract warnings were poorly recognized. Although both iconic and abstract warnings were learned in Experiment 1, reaction time was generally slow (4 to 8 seconds).

When placed in a more ecologically valid aviation context, and when a smaller number of signals were used (four), auditory iconic warnings were more likely to be recognized than auditory abstract warnings (Experiment 2). Importantly, verbal responses in naming the critical event to which the auditory signal referred reduced to approximately 1 second. In the applied experiment, the results were unaffected by participants' level of flying experience.

Individual differences were observed in warning recognition speed and accuracy. In an actual training setting it may be necessary to tailor training for different operators. For example, to train to a criterion level of performance such that warnings are recognized with 100 per cent accuracy by the end of the training session. This should minimise the effect of individual differences during a flight task and is also likely to improve reaction time to auditory warnings during a flight phase

The two experiments suggest that auditory iconic and also bimodal auditory + visual signals, as previously recommended by the Australian Transport Safety Bureau in relation to cabin altitude, have potential as warning signals in the civil aviation environment. Future research needs to investigate their potential in increasingly complex and realistic aviation settings.

1 INTRODUCTION

Since 1999, there have been at least two occurrences in Australia where flight crew have failed to respond appropriately to a reduction in air pressure in the cabin/cockpit (ATSB Report: 199902928; ATSB Report: 200003771; CASA Discussion Paper DP 0102CS[1]). The only indication that the flight crew had of the depressurisation was a single warning light on the console. Accordingly, it was recommended by the Australian Transport Safety Bureau (ATSB) that aural warnings be fitted to operate in conjunction with the cabin alert warning system on all Beechcraft Super King Air aircraft and other applicable aircraft (DP 0102CS).

There is general consensus that, in the case of depressurisation, an auditory warning is more effective than a visual warning in effecting a response. Auditory warnings attract attention and allow events both inside and outside the operator's field of view to be monitored (Calhoun, Janson & Valencia, 1988; Gaver, 1989; McKinley & Ericson, 1997). Dynamically changing events are more readily represented in auditory than in visual displays (Gaver, 1989, 1993) and reaction times (RT) to visual signals are shortened when accompanied by an auditory warning signal (Stokes & Wickens, 1988).

Research has highlighted significant constraints in the design and use of abstract sounds as warnings. Alarms that blare, beep or ring, involve an arbitrary and abstract mapping between a signal and the event to which it refers. A litany of weaknesses of auditory alarms has been documented, including that they: are loud and repetitive and may mask other communication, annoy rather than inform, do not convey the urgency of a situation, may be too loud and elicit a startle response that interferes with the necessary reaction, go unrecognised 40 per cent of the time when there are more than seven or so different alarms, and require excessive training and retraining (Begault, 1994; Momtahan et al., 1993; Patterson, 1982). Accordingly, Patterson made detailed recommendations about how to select the frequency components of a warning sound's spectral profile to lessen the likelihood of masking by noise and other warnings. He also recommended the use of no more than four to nine warning signals. A system with four signals and two attensons that signalled different levels of urgency was regarded as the optimal design – the purpose of the attenson being to inform the operator of the type of warning and that more information is available.

Auditory icons or caricatures of everyday sounds (Gaver, 1994) circumvent a number of the problems associated with arbitrary alarms or abstract symbolic sounds. The recognition of auditory icons requires "everyday listening", which is the "experience of listening to determine the source itself" (Gaver, 1994, p. 419). Auditory icons can be short, are not easily masked, and are generally recognizable and distinguishable. They may involve either "direct" or "indirect" reference (Familant & Detweiler, 1993). In the former, there is only one referent involved in the signal-referent relation and this denotative referent is the event that is the target of the warning or message. An example would be to use the sound of machine gun fire to signal a firing machine gun. Indirect reference occurs when there are at least

1 The Civil Aviation Safety Authority Discussion Paper DP0102CS was released in response to the Australian Transport Safety Bureau's investigation into an incident involving a Beech 200 Super King Air aircraft on 21 June 1999, where the pilot in command lost consciousness due to hypoxia (ATSB Report: 199902928).

two referents. Typically, the signal refers to the denotative referent through an intermediate sign referent – sign referents in indirect relations serves as surrogates for denotative referents that may be difficult to portray. For example, on the desktop of a Macintosh computer, the file removal program (the denotative referent) is represented by the visual image of a trashcan (the sign referent). An example from the auditory domain is to use the sound of rattling trashcans (sign referent) to signal that a computer file has been successfully deleted (the denotative referent). Although indirect relations may require an additional cognitive step to associate target with referent, once learned, recognition speed and accuracy can be comparable with that of direct relations (Keller & Stevens, 2004).

When designing sets of auditory alarms for a particular context, there are open questions about the nature of the warning (abstract alarm or icon), the effectiveness of auditory, visual, and auditory plus visual (A+V) warnings, potential interactions between warning type, modality, operator expertise and task demand (low versus high cognitive load). The present study investigated these variables – warning, modality, load, and expertise – in two experiments.

Specifically, Experiment 1 conducted under controlled laboratory conditions investigated a range of event-warning mappings. Abstract warnings versus icons were compared and warnings presented under auditory alone, visual alone, and A+V conditions for a range of critical events that may occur during aircraft operations, including conflicting traffic, icing, cabin depressurisation, and the failure of the undercarriage to extend during final approach to landing. The measures of performance were the number of training trials needed to reach criterion, accuracy and RT. In Experiment 2, the most effective items identified from the speed and accuracy results in Experiment 1 were presented to qualified pilots in an aircraft simulator under conditions of low or high cognitive load. It was hypothesized that, in both Experiments 1 and 2, RT would be fastest and training trials fewest in response to iconic warnings compared with abstract warnings and when presented in the auditory or A+V conditions.

1.1 Recognition of auditory versus visual warnings

Auditory rather than visual warnings are often used to convey information in emergency situations, as auditory warnings are omnidirectional in nature (requiring less scanning of instruments by operators), and can reach further distances than can visual warnings (Doll & Folds, 1986). For these reasons, visual warnings do not play a primary role although they are still present. Visual warnings are used in cautions and advisories where the information can be conveyed to the operator without causing unnecessary alarm. This also allows auditory warnings to be kept to a minimum, increasing the likelihood that the operator will be able to remember the different meanings of the critical auditory warnings.

The most common visual warnings are written warning labels, as information can be conveyed easily and with little ambiguity (Edworthy & Adams, 1996). However, problems in written warning labels such as degradation of the warning label, and problems understanding the warning by people whose first language differs from that used in the warning, has led to a wider use of symbols as visual warnings. These visual icons allow a fast and effective communication of information without extensive training in the relation between signal and target. Visual icons are not subjected to the same sorts of problems associated with written warnings as they can be designed in such a way that they are recognised by people universally, are recognised at greater distances, can still be recognised after degradation, and can be recognised more quickly than written words.

The perceptual and cognitive processes involved in the recognition of auditory and visual warnings are likely to differ. Auditory warnings are temporal in that they unfold in time whereas visual warnings are spatial and immediately available to perception (Edworthy & Adams, 1996). It might be expected that when attention is focused solely on the warning, a visual warning signal will be recognised more quickly than an auditory signal. However, because visual warnings are spatial, it requires an operator to be looking at the control panel (for example), where the warning will be located at the time it is first displayed. Because auditory warnings are omnidirectional in nature they should be recognised significantly more quickly than visual warnings in situations where the operator is required to attend to more than one thing a time, as in many instances the operator may not be looking in the direction of a visual display. For this reason, it is hypothesised that in an experimental condition in which participants are performing a concurrent visually-based task, auditory warnings will be recognised more quickly than visual warnings. Bimodal, A+V warnings were also included to investigate their possible facilitatory effects on recognition speed and/or accuracy.

1.2 Warning recognition in high workload situations

Many situations require people to perform two tasks at once. For example, driving a car consists of a number of complex tasks such as steering the car, monitoring speed, and changing gears. The cognitive resources available, such as attention, processing effort, memory capacity, and communication channels have a significant impact on the level of performance achieved on each of the concurrent tasks (Norman & Bobrow, 1975). Although these cognitive processing resources are limited, they can be allocated. For example, a person may allocate attention between navigating a curve and monitoring speed on the speedometer (Wickens, 2002). Increasing the number of resources available to a task will usually result in greater performance on that task. However, this is only true up to the point where resources are exhausted, so there will eventually be deterioration in performance if a number of tasks are competing for the same available resources (Norman & Bobrow, 1975). However, not all tasks interfere with one another. For example, it has been shown that there is less interference when a visual and an auditory task are performed concurrently, than when two tasks are performed that involve the same modality (Wickens & Liu, 1988). Wickens (2002) has proposed a four-dimensional multiple resources model in which separate resources exist for each of the four resource categories. The four resource categories identified by Wickens (2002) are Modality (auditory vs. visual), Codes (spatial vs. verbal), Stages (perception vs. response), and Visual Channels (focal vs. ambient). Therefore, there should be greater interference between two concurrent tasks competing for the resources with the same categories, for example two visual tasks, than there will be for two tasks that use resources from different categories, for example, concurrent audio and visual tasks.

It was hypothesised that when performing a highly demanding task, recognition of abstract warnings is poorer than when the task is low in demand, because a high level of cognitive resources is needed to attend to the individual acoustic or spatial features of the warning. If it is the case, as theorised, that fewer cognitive resources are needed to interpret iconic warnings then there will be less interference with recognition of iconic warnings during a high demand task. It is also predicted that during the performance of a visually-based concurrent task, recognition of visual warnings will be poorer in comparison to auditory warnings. This difference should be more pronounced in the high workload condition.

2 EXPERIMENT 1: THE EFFECT OF ICONICITY, MODALITY, AND TASK DEMAND ON WARNING RECOGNITION SPEED AND ACCURACY

The aim of Experiment 1 was to investigate the effect of three factors on learning and recognizing warning signals. The first factor, iconicity, contrasted iconic and abstract warnings. The second factor, modality, consisted of three levels, auditory, visual, and A+V. The third factor, task demand, consisted of two levels, low and high. The first two variables were between-subjects factors. The dependent variables were learning rate measured as the number of trials taken to reach a criterion level of performance, warning recognition rate (accuracy), and warning recognition RT. It was hypothesized that:

1. On training trials, iconic warnings require fewer training trials than abstract warnings to reach a criterion level of performance;

2. Modality of warning signal affects accuracy and RT and interacts with task demand. Specifically, under high and low task demand, that (a) bimodal (A+V) warning signals elicit faster RT and greater recognition than unimodal warning signals; (b) auditory warning signals elicit faster RT and greater recognition than visual warning signals; (c) auditory-visual (bimodal) elicit faster RT and greater recognition than auditory (unimodal) warning signals; and (d) auditory-visual warning signals (bimodal) elicit faster RT and greater recognition than visual warning signals;

3. Participants respond more quickly and more accurately to the warning signals during the concurrent low demand dual task than during the high demand dual task (manipulation check); and

4. Under differing levels of demand (low and high), performance is influenced by both modality and iconicity. Specifically, it was expected that, with regard to accuracy and RT to warnings in both low and high demand concurrent tasks: (a) iconic warnings are more effective than abstract warnings; (b) auditory-visual iconic warnings are more effective than auditory icons; (c) auditory-visual icons are more effective than visual icons; (d) auditory icons are more effective than visual icons; (e) auditory-visual abstract warnings are more effective than auditory abstract warnings; (f) auditory-visual abstract warnings are more effective than visual abstract warnings; and (g) auditory abstract warnings are more effective than visual abstract warnings.

2.1 Method

2.1.1 Participants

Two-hundred and fifteen participants (168 females and 47 males), from the University of Western Sydney took part in the experiment for which they received course credit (mean age = 20.79 years, range 17 to 39 years). Thirty-three participants were excluded from analysis as they failed to reach criterion, ie they did not learn the signals, (specified by 28 correct responses in a row) before 12 blocks of presentations (excluding training trials). An additional four participants were excluded from analysis after data screening found them to be outliers.

One participant was excluded based on trials to criterion (Auditory Icon Easy, N = 121, z = 3.8) one on RT during learning phase (AV Abstract Easy RT = 12843 ms, z = 4.75), one on RT in the high demand dual task (Auditory Abstract Easy RT = 29335.75 ms, z = 4.013), and one on items incorrect in the learning phase (AV Abstract Hard, Items incorrect = 63, z = 3.67).

Following the exclusion of these participants, the abstract warning condition consisted of 89 participants (auditory = 29, visual = 32, A + V = 28) and the iconic condition consisted of 89 participants (auditory = 29, visual = 28, A + V = 32), a total of 178 participants.

2.1.2 Stimuli

From rating tasks conducted prior to Experiment 1, nine critical aviation events that could potentially lead to an accident were selected. Abstract auditory and visual warnings and iconic auditory and visual warnings were designed for each of these (table 1). The nine critical events were presented as 'clickable' buttons on a computer screen, equidistant from one another. Warnings were presented either visually within a square at the top of the computer screen for 1,000 ms, or auditorily, through built-in speakers, again lasting for 1,000 ms. Iconic warnings involved a relation between the warning and the event being signalled. For example, the sound of an elephant trumpeting was used to indicate that the aircraft is overweight (ie both are heavy). Abstract warnings contain no association between the warning and the event being signalled, for example, the sound of a series of arbitrary beeps. A mathematical addition task presented visually and concurrently, was designed to form low and high demand conditions. The low demand version consisted of three numbers all less than five presented in the middle of the computer screen, and the high demand task consisted of four numbers all greater than the number five. Participants were required to mentally add the numbers together and then say the answer aloud. The addition task was displayed on the screen for a period of 500 ms. Both low and high demand dual tasks consisted of a total of 36 additions. The order of low and high demand conditions was counterbalanced across participants to distribute serial order effects.

Table 1: Visual and auditory icons that attracted the highest signal-event ratings and used as stimuli in Experiment 1

Event	Visual icon with highest rating	Auditory icon with highest rating
Engine fire	Fire extinguisher	Fire engine siren
Low fuel	Petrol pump	Car failing to start
Conflicting traffic	Planes colliding	Car brakes screeching
Radio failure	Radio-cassette player	Radio static
Carbon monoxide	Skull and crossbones	Coughing
Ground proximity	Plane diving in mountain	Explosion
Electrical failure	Electricity symbol	Zapping sound
Aircraft icing	Snowflake	Cold wind
Aircraft overweight	Elephant	Elephant trumpeting

2.1.3 Equipment

The experiment was programmed in PowerLaboratory Version 1.0.3 (Chute & Westwall, 1996), and presented to participants on a PowerMacintosh 7300/200, with a 15-inch monitor and built-in speakers.

2.1.4 Procedure

Upon arrival, participants read an information sheet and signed a consent form (Appendix 1). They were provided with some context for each of the warnings by reading a Critical Aviation Events Information sheet. They were then trained on the relation between each of the warnings and the corresponding event. After this training, participants were tested as to how well they had learnt the association between warning and event, and were continually retrained until they reached 100 per cent accuracy on 28 trials. Once participants had reached this criterion level of performance they were required to perform a visual addition task in which numbers flashed on the screen and the participants were required to add these numbers together as fast and as accurately as possible, while still responding to the warnings when they were presented. The task took 30 to 40 minutes.

2.2 Results

Data consisted of a number of dependent variables obtained during the learning phase and the test phase under low and high task demand. The dependent variables were: trials to criterion (learning phase only); RT in ms (correct responses in learning and test phases); error rate in response to warnings (learning phase only); accuracy in response to warnings (test phase); and accuracy on concurrent arithmetic/addition task (test phase only). Data was analysed using separate Analyses of Variance (ANOVA) for each dependent variable, followed by planned comparisons. An ANOVA provides details of main effects (of single independent variables) and interaction effects (between two or more independent variables). Planned comparisons were used to analyse precise effects as specified in the hypotheses. Where multiple comparisons were conducted, the significance level was adjusted accordingly.

2.2.1 Hypothesis 1

It was hypothesised that iconic warnings would take significantly fewer training trials to reach criterion, and that RT to iconic warnings would be quicker during the learning phase.

Trials to criterion

A univariate ANOVA was conducted to test Hypothesis 1, that iconic warnings take significantly fewer training trials to reach criterion performance than abstract warnings. This hypothesis was supported $F(1, 176) = 88.973$, $p = .00.$, $\eta^2 = .336$. Iconic warnings took significantly fewer training trials to reach criterion with a mean of 39.83 ($SD = 15.56$) training trials compared with a mean of 104.42 ($SD = 62.69$) training trials required for abstract warnings.

Reaction time

A univariate ANOVA was conducted to test if RT in response to iconic warnings was significantly faster than to abstract warnings. The hypothesis was supported $F(1, 176) = 5.089$, $p = .017$, $\eta^2 = .032$. Participants responded significantly faster to iconic warnings ($M = 5021.399$ms, $SD = 721.033$) than to abstract warnings ($M = 5295.870$ ms, $SD = 796.395$).

Error rate

A univariate ANOVA was conducted to test the hypothesis that participants make fewer errors in response to iconic warnings than to abstract warnings. The hypothesis was supported $F(1, 176) = 105.170$, $p = .000$, $\eta^2 = .374$. Fewer errors were made in the learning phase in response to iconic warnings ($M = 1.96$, $SD = 2.927$) than to abstract warnings ($M = 19.45$, $SD = 15.825$).

2.2.2 Hypothesis 2

It was hypothesised that there is a main effect of modality across all dependent variables. Planned contrasts were used to test the hypothesis that the modality of the warning signals would elicit different patterns of performance across all dependent variables. Specifically, the hypotheses tested that: (a) bimodal warning signals elicit faster RT and greater recognition than unimodal warning signals, (b) auditory warning signals elicit faster RT and greater recognition than visual warning signals (c) auditory-visual (bimodal) elicit faster RT and greater recognition than auditory (unimodal) warning signals and (d) auditory-visual warning signals (bimodal) elicit faster RT and greater recognition than visual warning signals. Comparisons (c) and (d) were included to establish a pattern and to order the difficulty of the modalities. Alpha level was adjusted to .0125 to allow for multiple comparisons.

Learning phase

Trials to criterion

Participants needed significantly fewer warning signal presentations to reach criterion in the learning phase when warning signals were bimodal ($M = 58.40$, $SD = 34.404$) compared to when they were unimodal ($M = 79.48$, $SD = 55.70$), $F(1,175) = 6.443$, $p = .012$. A significant difference was also found between visual and auditory warning signals, $F(1,175) = 20.463$, $p = .000$. However, this effect was not in the predicted direction. Visual warnings took fewer trials to reach criterion, ($M = 57.667$, $SD = 33.470$) than auditory warnings ($M = 101.276$, $SD = 77.655$). A significant difference was found between auditory-visual and auditory warnings, $F(1,175) = 19.780$, $p = .000$, with auditory-visual signals requiring fewer training trials for participants to reach criterion than auditory warnings. The fourth prediction, that there would be a difference between auditory-visual warnings and visual warnings, was not supported, $F(1,175) = 0.006$, $p = .938$.

Reaction time

Response times were analysed for correct responses only using the same planned contrasts. Contrary to predictions, there was no significant difference in RT when comparing unimodal and bimodal warning signals, $F(1,175) = 3.625$, $p = .059$, and when comparing auditory-visual warnings and auditory warnings, $F(1,175) = 0.534$, $p = .082$. Significant differences were found in RT when comparing visual warning signals with either auditory-visual warnings, $F(1,175) = 16.354$, $p = .000$, or auditory warnings, $F(1,175) = 22.471$, $p = .000$; however, neither of these was in the expected direction. Reaction time to visual warnings was significantly faster ($M = 4773.597$, $SD = 540.292$) than for either auditory-visual ($M = 5403.815$, $SD = 853.266$) or auditory ($M = 5306.664$, $SD = 742.158$) warning signals.

Dual task

Recognition accuracy to warning signals during low demand task

Recognition of warning signals during the concurrent low demand task were analysed using the same comparisons as previously, with similar patterns to those observed in items to criterion across modalities. A significant difference was observed between bimodal and unimodal warnings, $F(1,175) = 9.984$, $p = .002$, with bimodal warnings evoking greater accuracy ($M = 8.383$, $SD = .922$) than unimodal warnings ($M = 7.85$, $SD = 1.129$). A significant difference was also evident between visual and auditory warning signals, $F(1,175) = 12.754$, $p = .000$, however, this trend was not in the predicted direction. Accuracy in response to visual warnings ($M = 8.2$, $SD = 1.070$) was higher than to auditory warnings ($M = 7.50$, $SD = 1.188$). A significant difference was found between auditory-visual and auditory warnings, $F(1,175) = 19.780$, $p = .000$, with auditory-visual signals requiring fewer training trials to reach criterion than auditory warnings. Once again, the fourth prediction, that there would be a difference between auditory-visual warnings and visual warnings, was not supported, $F(1,175) = 0.006$, $p = .938$, indicating that there is no significant difference between auditory-visual warnings and visual warnings.

Recognition accuracy to warning signals during high demand task

Recognition accuracy during the high demand dual task was analysed using the same comparisons employed previously. The first prediction, that accuracy to bimodal warning signals would be better than to unimodal signals was supported, $F(1,175) = 11.842$, $p = .001$. Accuracy to bimodal warnings ($M = 8.050$, $SD = 1.213$) was higher than to unimodal signals ($M = 7.3055$, $SD = 1.452$). Predictions (c), that auditory-visual warnings will be recognised with greater accuracy than auditory warnings, $F(1,175) = 8.296$, $p = .005$ and (d), that auditory-visual warnings will be recognised better then visual warnings, $F(1,175) = 9.471$, $p = .002$, were both supported. Auditory-visual warnings were recognised more accurately than auditory ($M = 7.33$, $SD = 1.431$) and visual warnings ($M = 7.28$, $SD = 1.439$). However, under the high demand dual task there was no significant difference between accuracy recognition of auditory warnings and visual warnings, $F(1,175) = .031$, $p = .860$.

Reaction time to warning signals during low demand dual task

The response times during the low demand dual task were analysed for correct responses only using the same planned contrasts that were used previously. Contrary to predictions, there was no significant difference in response RT when comparing unimodal and bimodal warning signals, $F(1,175) = .914$, $p = .340$, and when comparing auditory-visual warnings and visual warnings, $F(1,175) = 1.389$, $p = .240$. Significant differences were found in RT when comparing auditory warning signals with either auditory-visual warnings, $F(1,175) = 7.923$, $p = .005$, or visual warnings, $F(1,175) = 15.866$, $p = .000$, both in the predicted direction. Reaction time to auditory warnings was significantly slower ($M = 5163.105$, $SD = 1543.589$) than to both auditory-visual ($M = 4547.879$, $SD = 988.120$) or visual ($M = 4292.580$, $SD = 948.544$) warning signals.

Reaction time to warning signals during high demand dual task

The response times during the high demand dual task were analysed for correct responses only using the same planned contrasts as used previously. Contrary to predictions, there was no significant difference in response RT when comparing unimodal and bimodal warning signals, $F(1,175) = 1.726$, $p = .191$, and when comparing auditory-visual warnings and visual warnings, $F(1,175) = .069$, $p = .793$. Significant differences were found in RT when comparing auditory warning signals with either auditory-visual warnings, $F(1,175) = 6.366$, $p = .013$, or visual warnings, $F(1,175) = 7.748$, $p = .006$, both in the predicted direction. Reaction time to auditory warnings was significantly slower ($M = 7043.872$, $SD = 3143.623$) than to both auditory-visual ($M = 5868.173$, $SD = 1991.173$) or visual ($M = 5746.812$, $SD = 2340.998$) warning signals.

Summary of results – Hypothesis 2 (where > refers to 'better than')

Trials to criterion:	AV/Visual > Auditory
RT in learning phase:	Visual > Auditory/AV
Low demand recognition accuracy:	AV/Visual > Auditory
High demand recognition accuracy:	AV > Auditory/Visual
RT during dual task (both low and high):	AV/Visual > Auditory

2.2.3 Hypothesis 3

It was hypothesised that participants would respond more rapidly and more accurately to the warning signals during the concurrent low demand dual task than in the high demand dual task. Significant main effects were found for both accuracy, $F(1,175) = 23.979$, $p = 0.00$, and RT, $F(1,175) = 86.481$, $p = .000$. Warning recognition scores were higher ($M = 8.028$, $SD = 1.068$) and RTs faster ($M = 5667.879$, $SD = 1186.840$) during the concurrent low demand task than recognition accuracy ($M = 7.554$, $SD = 1.364$) and RTs ($M = 6219.617$, $SD = 1186.840$) recorded during the high demand dual task.

2.2.4 Hypothesis 4

Planned comparisons were used to test the hypotheses that under differing levels of demand (low and high), performance would be influenced by both modality and iconicity. Specifically, it was expected that with regard to accuracy and RT to warnings in both low and high demand concurrent tasks: (a) iconic warnings would be more effective than abstract warnings; (b) auditory-visual iconic warnings would be more effective than auditory icons; (c) auditory-visual icons would be more effective than visual icons; (d) auditory icons would be more effective than visual icons; (e) auditory-visual abstract warnings would be more effective than auditory abstract warnings; (f) auditory-visual abstract warnings would be more effective than visual abstract warnings; and (g) auditory abstract warnings would be more effective than visual abstract warnings. All comparisons were conducted with a Bonferroni adjusted alpha of .007 to allow for multiple comparisons.

Recognition accuracy during high demand dual task

Recognition accuracy during the high demand dual task was analysed using the same planned comparisons that were used to test Hypothesis 2. Significant differences were found between recognition accuracy during the high demand dual task when warning signals were visual iconic and when signals were either auditory-visual iconic, $F(1,172) = 33.643$, $p = .000$ or when warnings were auditory iconic, $F(1,172) = 15.359$, $p = .000$. Overall, mean accuracy was lower for warnings that were visual iconic ($M = 7.071$, $SD = 1.654$) than when warnings were either auditory-visual iconic ($M = 8.813$, $SD = .397$) or auditory iconic ($M = 8.276$, $SD = .882$). Contrary to predictions, there was no significant difference between recognition accuracy in response to auditory-visual iconic warnings and auditory iconic warnings, $F(1,172) = 3.256$, $p = .073$. There were significant differences between auditory abstract warnings and warnings that were auditory-visual abstract, $F(1,172) = 6.763$, $p = .010$ and warnings that were visual abstract, $F(1,172) = 13.419$, $p = .000$. Accuracy was significantly lower when warnings were auditory abstract ($M = 6.379$, $SD = 1.237$) than when signals were auditory-visual abstract ($M = 7.179$, $SD = 1.249$) and visual abstract ($M = 8.276$, $SD = 1.218$). No significant difference was observed between auditory-visual abstract and visual abstract warnings, $F(1,172) = .935$, $p = .335$ (table 2, figure 1)

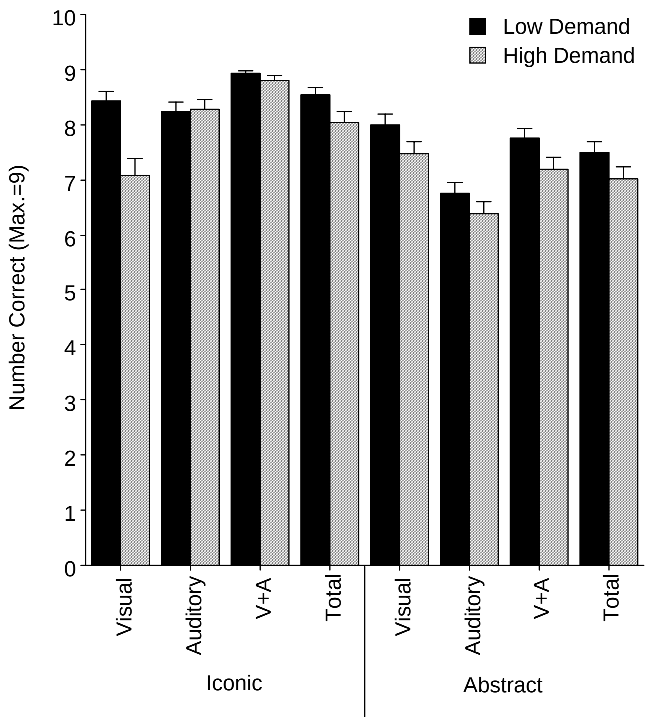
Table 2: Experiment 1 recognition scores (max. = 9) means, standard deviations and standard errors

		Low demand			High demand		
Iconicity	***Modality***	***M***	***SD***	***SE***	***M***	***SD***	***SE***
Iconic	Visual	8.43	.96	.18	7.07	1.66	.31
	Auditory	8.24	.87	.16	8.28	.88	.16
	A+V	8.94	.25	.04	8.81	.40	.07
	Total	8.54	.69	.13	8.05	.98	.18
Abstract	Visual	8.00	1.12	.20	7.47	1.22	.22
	Auditory	6.76	.98	.18	6.38	.1.24	.23
	A+V	7.75	1.01	.19	7.18	1.25	.24
	Total	7.50	1.04	.19	7.01	1.23	.23

Recognition accuracy during low demand dual task

Recognition accuracy during the low demand dual task was analysed using the same comparisons that were used to test performance during the high demand task. Significant differences were evident between recognition accuracy during the low demand dual task when warning signals were auditory-visual iconic and when the signals were either visual, $F(1,172) = 4.65$, $p = .032$ or when warnings were auditory iconic, $F(1,172) = 8.892$, $p = .003$. Overall, mean accuracy was higher for warnings that were auditory-visual iconic ($M = 8.938$, $SD = .246$) than when warnings were either visual iconic ($M = 8.429$, $SD = .959$) or auditory iconic ($M = 8.241$, $SD = .872$). Contrary to predictions, there were no significant differences between recognition accuracy to visual iconic warnings and auditory iconic warnings, $F(1,172) = .602$, $p = .439$. There were significant differences between auditory abstract warnings and warnings that were auditory-visual abstract, $F(1,172) = 16.887$, $p = .000$ and warnings that were visual abstract, $F(1,172) = 28.277$, $p = .000$. Accuracy was significantly lower when warnings were auditory abstract ($M = 6.759$, $SD = .988$) than when signals were auditory-visual abstract ($M = 7.750$, $SD = 1.005$) and visual abstract ($M = 8.000$, $SD = 1.136$). No significant difference was observed between auditory-visual abstract and visual abstract warnings, $F(1,172) = 1.126$, $p = .290$ (table 2, figure 1).

Figure 1: Experiment 1 - Mean accuracy as a function of iconicity, modality, and task demand



Reaction time during high demand dual task

Response times during the high demand dual task were analysed only for correct responses. The same planned contrasts were used as had been employed for the low demand task. Contrary to predictions, there were no significant differences in response RT across modalities in the iconic condition. However, significant differences were found in RT when comparing auditory warning signals with either auditory-visual warnings, $F(1,172) = 5.871$, $p = .016$, or visual warnings, $F(1,172) = 10.327$, $p = .002$, both in the predicted direction. Reaction time to auditory warnings was significantly slower ($M = 8189.648$, $SD = 3233.671$) than to both auditory-visual ($M = 6642.828$, $SD = 2214.704$) or visual ($M = 6204.466$, $SD = 2517.605$) warning signals. A significant difference was found between abstract and iconic warnings, $F(1,172) = 18.951$, $p = .000$ in the predicted direction (table 3, figure 2).

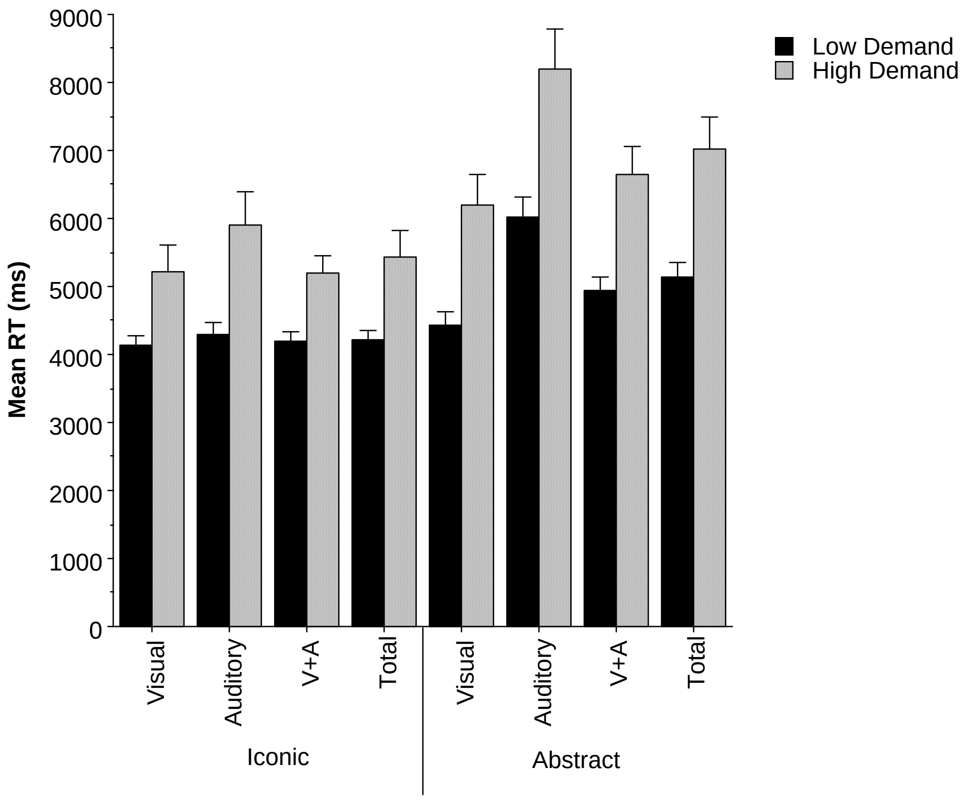
Table 3: Experiment 1 - RT (ms) means, standard deviations, standard errors

		Low demand			High demand		
Iconicity	***Modality***	***M***	***SD***	***SE***	***M***	***SD***	***SE***
Iconic	Visual	4138.79	686.15	129.67	5223.78	2041.31	385.77
	Auditory	4300.45	920.73	170.98	5898.10	2634.70	489.25
	V+A	4199.31	788.95	139.47	5190.35	1502.87	265.67
	Total	4212.85	789.61	146.70	5437.41	2059.63	380.23
Abstract	Visual	4427.16	1123.52	198.61	6204.65	2517.47	445.03
	Auditory	6025.76	1568.81	291.32	8189.65	3233.67	600.48
	V+A	4946.40	1053.67	199.13	6642.83	2214.60	418.52
	Total	5133.11	1248.67	229.69	7012.37	2655.25	488.01

Reaction time during low demand dual task

In the case of the low demand dual task, response times were analysed only for correct responses using the same planned contrasts that had been employed previously. Contrary to predictions, RT to the iconic warning signals during the low demand dual task did not differ significantly across modalities. No significant difference was evident between auditory-visual abstract and visual abstract warnings, $F(1,172) = 3.574$, $p = .060$. However, significant differences were found in RT when comparing auditory warning signals with both auditory-visual warnings, $F(1,172) = 14.732$, $p = .000$, and visual warnings, $F(1,172) = 34.511$, $p = .000$, both in the predicted direction. Reaction time to auditory warnings was significantly slower ($M = 6025.763$, $SD = 1568.815$) than to both auditory-visual ($M = 4946.395$, $SD = 1053.516$) and visual ($M = 4427.395$, $SD = 1123.516$) warning signals (table 3, figure 2).

Figure 2: Experiment 1 - mean RT as a function of iconicity, modality, and task demand



Performance on the addition task in the dual demand task

Participants made fewer errors during the low demand task ($M = 1.36$, $SD = 2.02$) than during the high demand task ($M = 24.03$, $SD = 10.46$).

2.3 Discussion – Experiment 1

The results of this multi-factor experiment demonstrate that auditory icons or environmental sounds are effective as warning signals in a civil aviation environment. Importantly, the efficacy of iconic signals is influenced by factors such as task demand and modality. Each hypothesis will now be discussed in turn and then general conclusions will be drawn.

The number of training trials needed to learn signal-event associations reflects the difficulty of the task. A relatively large number of training trials were required to learn abstract and iconic warning signals associated with nine different critical events. As hypothesized, fewer training trials were required to learn signal-event pairs when the signal was iconic. Reaction times to iconic warnings during training were, on average, also significantly faster than those recorded in response to abstract warning signals. The results are in keeping with the recommendation made by Patterson (1982) that operators should need to learn no more than nine (and preferably fewer) signals. It is also imperative that all signals are maximally distinguishable from one another. The higher error rate for abstract warnings during training suggests that there was confusion between similar abstract warnings. An overall recommendation is to limit the number of warnings in the range four to six and to ensure maximum discriminability between abstract signals in terms of pitch, tempo and timbre.

The use of a highly controlled laboratory-based experiment has enabled close scrutiny of important interactions between modality and cognitive load. Cognitive load was operationalised in Experiment 1 as a concurrent cognitive (mathematical addition) task with two levels, low and high. Hypothesis 2 demonstrated that the effectiveness of the modality of presentation (auditory, visual or A+V) interacts with high and low task demand. Bimodal (A+V) signals were learned, on average, with the least number of exposures; visual warnings required fewer training trials than auditory warnings. Auditory + visual and visual warnings were comparable in terms of the number of trials needed to reach criterion. Reaction times during training, on average, were shorter in response to visual than auditory or A+V warnings. Under high demand dual task conditions, bimodal warnings elicited significantly greater accuracy than unimodal warnings, A+V warnings were recognized with greater probability than auditory and visual warnings, and accuracy in response to auditory versus visual warnings did not differ.

As hypothesised, RTs to auditory warnings during the high demand task were significantly slower than those recorded in response to A+V and visual warnings. Under low demand conditions, A+V warnings elicited greatest accuracy, and accuracy in response to visual warnings was higher than for auditory warnings. Auditory + visual and visual warnings under low demand elicited similar levels of accuracy. Under low demand, RTs to auditory signals were slower than A+V and visual warning signals. Broadly speaking, bimodal signals are consistent in eliciting accurate responses under both high and low demand tasks. Under low demand conditions, response accuracy to visual warnings is also good. Importantly, as we shall see, these patterns of responding are further influenced by the iconicity of the warning signal.

The third hypothesis checked the reliability of the low versus high task demand manipulation. The manipulation of low and high demand concurrent tasks affected accuracy and RT in the expected direction. The paradigm developed here has application in future laboratory-based investigations of vigilance to auditory, visual and A+V tasks with varying cognitive load.

The final hypothesis investigated the complex interaction between the three independent variables, iconicity, modality and task demand. As we have seen, modality and task demand interact and the results of the multi-factor experiment also reveal that iconic warning signals are effective under certain circumstances. Under high demand conditions, visual iconic warning signals elicited significantly poorer recognition accuracy than A+V iconic or auditory iconic warnings. Iconic auditory warnings were recognized as well as A+V iconic warnings under high demand conditions.

Under high demand, and when warnings were abstract, accuracy was significantly greater in response to visual abstract and A+V abstract than auditory abstract warnings. Thus, there is an important interaction between iconicity and modality such that the auditory (and A+V) presentation is effective when signals are iconic but not when signals are abstract. Iconic warnings were also recognized more quickly than abstract warnings in the high demand condition. Under conditions of low demand, A+V iconic warnings were recognized with greater accuracy than auditory or visual iconic warnings. Under low demand, auditory and visual iconic warnings did not differ with respect to accuracy. Again, A+V and visual abstract warnings were recognized with greater accuracy than auditory abstract warnings under low demand. These results demonstrate that auditory icons are effective signals in situations that involve high cognitive load. Bimodal warnings are effective under both high and low demand. Importantly, accuracy and RTs in response to abstract warning signals is generally poor.

The three-way factorial design of Experiment 1 enabled scrutiny of complex interactions between variables. Auditory icons as warnings signals are recognized with greatest relative accuracy when task demand is high. Bimodal warnings perform well under a range of conditions. Most importantly, auditory abstract warnings lead to relatively poor performance under conditions of both low and high demand.

Reaction times add a vital piece of information to these observations. Reaction times recorded in response to all warning signals in this laboratory experiment were long, from 4 to 8 s. This length of response reflects the processing involved in warning recognition – the sound needs to be detected, the source of the sound identified (eg elephant trumpet), the association between sound and event brought to consciousness (overweight), and the response indicated by a button press.

Clearly, these steps require time and it is evident that these RTs prohibit the use of iconic signals where immediacy of response and appropriate action are of the essence. However, in situations where there is a need to inform the operator of a non-time-pressured event, iconic warnings may have a role. Importantly, recognition accuracy is particularly poor in response to abstract warnings and even with extensive training may not reach acceptable levels under either high or low demand conditions.

Future research could be undertaken to investigate ways in which response times to 1 s auditory iconic warnings and A+V iconic warnings could be made even more efficient. This may include the use of shorter sounds, attensons, and limited numbers of warnings. The latter idea – the use of a smaller set of sounds (four) – will be operationalised in Experiment 2.

3 EXPERIMENT 2: THE EFFECT OF ICONICITY, TASK DEMAND, AND EXPERTISE ON WARNING RECOGNITION SPEED AND ACCURACY IN AN ADVANCED AVIATION TRAINING DEVICE

The aim of Experiment 2 was to investigate the effectiveness of warning signals that differ in iconicity in the context of a more realistic aviation task. A small number of auditory iconic abstract warnings were used (four) and minimal training provided. The task was to pilot an aircraft simulator, maintaining an altitude of 4,500 feet and heading of 060 degrees. The flying scenario was completed during conditions of high turbulence (high demand) and low turbulence (low demand). The 2 x 2 experimental design consisted of the independent variables iconicity (icon, abstract) and demand (low, high) with both factors as within-subject variables. The order of low and high demand conditions was fully counterbalanced across the final sample. The dependent variables were accuracy and RT in naming the critical event to which an auditory signal referred. Flight experience (hours) was treated statistically as a covariate. It was hypothesised that auditory iconic warnings elicit greater accuracy and faster RTs than auditory abstract warnings under conditions of both low and high demand.

3.1 Method

3.1.1 Participants

The sample consisted of 10 pilots recruited from the Sydney, New South Wales area (age range: 20 to 58 years; mean age = 27.50 years, *SD* = 14.95). The average hours flight experience of the sample was 730.90 hours, *SD* = 1747.72 and average hours in command 501.20 hours, *SD* = 1370.55. Of the 10 participants, four held commercial licences, five held general flying progress tests and private licences, and one held a private and airline transport pilot licence. Three of the participants were also flight instructors. All participants had self-reported normal hearing and vision.

3.1.2 Stimuli

Four of the nine events and the associated auditory iconic and auditory abstract warnings from Experiment 1 were used. The critical events were: engine fire, carbon monoxide, ground proximity and low fuel. During the flight scenario, each of the four iconic and four abstract warnings sounds were presented once in a random order at intervals varying between 30 and 90s. The time intervals were random within an experimental session but were consistent across all experiment sessions. There were two different random orders of signals used in the low and high demand conditions.

3.1.3 Equipment

The simulator was a dual, Modular Flight Deck (MFD), Advanced Aviation Training Device (AATD) manufactured by Precision Flight Controls. Enclosed in fibreglass, the AATD is a two-place system that is capable of simulating a range of general aviation aircraft. In the case of the present study, the AATD was designed to simulate a Piper PA-28 Warrior aircraft. All of the instruments and controls that are normally available to pilots were available to participants. The instruments included an attitude indicator, a turn-bank indicator, and an airspeed indicator. The instruments were displayed through a 17 inch monitor embedded within the AATD. The external view was provided through a Panasonic Data Projector that displayed a 2 metre by 2 metre image immediately to the front of the pilot.

The AATD is integrated with X-Plane Version 8.30 software. The software is based on a realistic aircraft model and is designed to simulate, as accurately as possible, the behaviour of aircraft during flight. It also enables a significant level of flexibility in modifying a range of variables including turbulence, visibility, and wind direction and strength.

The training phase of the experiment was programmed in PowerLaboratory (Chute & Westall, 1996) and presented on a Macintosh G4 powerbook computer. The presentation of sounds during the flight was programmed in SuperLab 1.74 (Haxby, Parasuraman, Lalonde, & Abboud, 1993) and presented through stereo headphones to the pilot using the same powerbook computer. Pilots wore a small voice-key microphone on their lapel that registered their voice when they named a critical event signalled by an iconic or abstract sound. The voice-key registered the time of response and the RT was recorded to disk. The experimenter noted the verbal response manually.

3.1.4 Procedure

Upon arrival in the lab, pilots read an information sheet about the experiment and signed a consent form (Appendix 2). They then completed a training phase that consisted of a systematic presentation of iconic and abstract warning signals and the associated critical event. There was an iconic and abstract warning paired with each of the four events (eight different warning signals in total). Participants heard presentations of each of the signals and event pairs. Recognition accuracy was tested with the presentation of the signal alone and participants asked to indicate the event to which each auditory signal referred. There were 24 training/test trials.

The test phase used the AATD. Participants were instructed as follows: “You are operating a Piper Warrior (fixed pitch) aircraft with an indicated airspeed of 100 knots. Your goal is to maintain, to the best of your ability, 4,500 feet and a heading 060 degrees. The aircraft is currently set at this heading and altitude. Periodically, throughout the flight, you will hear a number of sounds. Please respond to these sounds as quickly as possible by saying aloud the nature of the event with which the sound is associated”. Each auditory signal was presented once in each test phase of the low and the high demand conditions. At the conclusion of the task, participants completed a short demographic information questionnaire (Appendix 3). The experiment took 45 minutes to complete.

3.2 Results

It was hypothesised that accuracy would be greater and RT faster in response to auditory iconic warnings compared with auditory abstract warnings. The hypothesis was supported in the training trials with significantly greater accuracy, $F(1,9) = 22.20$, $p = .002$, $\eta^2 = .74$, and faster RT, $F(1,9) = 29.98$, $p = .001$, $\eta^2 = .79$ recorded in response to auditory iconic than abstract warnings (see table 4). Reaction times to auditory iconic warnings were, on average, more than 2 s faster than those recorded in response to auditory abstract warnings.

Table 4: Experiment 2 - Training trial mean accuracy and RTs

	Mean	Standard Deviation
Items correct overall (/24)	19.00	3.20
Overall RT	5560.74	973.16
Icon correct (/12)	11.70	0.48
Icon RT	4990.71	968.56
Abstract correct (/12)	7.30	2.83
Abstract RT	7107.50	1863.08

The hypothesised effect of iconicity was also evident in accuracy and RT in response to signals while in the AATD. Auditory iconic warnings were recognized significantly more often than auditory abstract warnings under both low demand flight conditions $F(1,8) = 17.49$, $p = .003$, $\eta^2 = .69$, and under high demand flight conditions, $F(1,8) = 9.60$, $p = .02$, $\eta^2 = .55$ (table 5).

Table 5: Experiment 2 - Accuracy and RT (ms) means and standard deviations

	Low Demand		High Demand	
	M	***SD***	***M***	***SD***
Overall flight accuracy (/8)	5.90	0.99	6.00	1.16
Overall flight RT	989.70	538.43	1099.00	397.69
Icon				
Accuracy (/4)	3.90	0.32	3.70	0.48
RT	822.28	436.06	736.96	279.59
Abstract				
Accuracy (/4)	2.00	1.05	2.20	1.23
RT	1426.66	1198.29	1495.36	907.44

When pilots flew in either the low demand or high demand conditions, there was no effect of iconicity on RT (see table 5): low demand iconic versus abstract RT, $F(1,8) = 1.38$, $p = .25$, $\eta^2 = .15$; high demand iconic versus abstract RT, $F(1,8) = 3.11$, $p = .12$, $\eta^2 = .28$. There were also no interactions between iconicity and task demand. That is, there was no effect of low versus high demand on warning recognition accuracy or RT when responding to iconic or abstract warnings.

The above effects were unchanged by the inclusion of flying experience (hours) as a covariate.

The overall mean accuracy in the flight condition collapsed across levels of task demand was 11.80 out of a possible 16, *SD* = 1.75. The overall mean RT in the flight condition, collapsed across levels of task demand, was 967.30 ms, *SD* = 454.26.

The mean altitude in the low demand flight condition was 4,484.77 feet (*SD* = 20.26) and heading was 66.25 degrees (*SD* = 3.93). Mean altitude in the high demand flight condition was 4,445.70 feet (*SD* = 42.01) and mean heading 71.27 degrees (*SD* = 16.94).

3.3 Discussion – Experiment 2

Experiment 2 involved the investigation of auditory warning signal recognition speed and accuracy in the context of a realistic flight setting. The hypothesis that auditory iconic warnings are recognized significantly more quickly and accurately than auditory abstract warnings was upheld in the training trials. Greater accuracy was also maintained in response to iconic auditory warnings relative to abstract auditory warnings under both low and high demand conditions during a simulated flight task. Reaction time during the flight conditions was unaffected by iconicity. Of most importance is the improvement in both RTs and accuracy in this experiment relative to Experiment 1.

The use of a smaller set of signals (four rather than nine) and the use of a more realistic aviation task is the most likely explanation. Both the low and high cognitive load dual task conditions of Experiment 1 were particularly demanding. In Experiment 2, pilots were asked to maintain a particular altitude and heading and demand was varied in an ecologically valid way by introducing more or less turbulence. The smaller number of signals enabled higher recognition rates during training, and the need for fewer training trials (79 per cent accuracy after two exposures to each signal). Although the RTs during training were still high, once warning signals were presented to pilots in the AATD, verbal recognition of the associated incident took, on average, 1,000 ms. Thus, in an actual civil aviation setting, verbal response to the auditory warning signal was relatively fast.

It is also noteworthy that there are individual differences in warning recognition speed and accuracy. The large standard deviations in table 5, particularly those associated with RT and abstract warning recognition accuracy means, reflect variability within the sample. In an actual training setting it may be necessary to tailor training for different operators. For example, to train to a criterion level of performance such that warnings are recognized with 100 per cent accuracy at the end of the training session. This should minimise the effect of individual differences during a flight task and is also likely to improve RT to auditory warnings during a flight phase.

4 CONCLUSIONS

The results of two experiments have demonstrated the potential for the use of auditory icons as informative warning signals in a civil aviation context. Although further research, particularly in applied settings, is needed, the present results suggest that operators learn to associate meaningful environmental sounds with critical events with greater ease than learning to associate abstract beeps or bells with particular events. Reaction time to both iconic and abstract warnings during training was relatively slow. In a controlled laboratory task that included a highly demanding concurrent task, auditory abstract warnings were recognized relatively poorly while bimodal (A + V warnings) were recognized well. Auditory abstract warning recognition is poor when abstract auditory signals are similar in pitch, rhythm, timbre or tempo. In terms of accuracy, auditory iconic warnings are as effective as A+V warnings in high demand conditions.

In a more applied and realistic aviation context, marked improvements in performance were found for auditory warnings when the set of warnings was reduced to four auditory iconic and four auditory abstract warnings. Warning recognition accuracy during training was approximately 79 per cent and a level of 74 per cent warning recognition accuracy was maintained during the flight. Importantly, in the context of a simulated flight, the speed of recognition to the auditory warnings improved from 4 to 6 s down to 1 s. In Experiment 2, recognition speed and accuracy was unaffected by the participants' level of flying experience.

It is recommended that further research be conducted to investigate response speed and accuracy to auditory iconic and bimodal iconic warnings in the context of increasingly sophisticated and realistic flight scenarios. The results of two preliminary experiments that apply auditory icons to the civil aviation environment indicate that there is potential in using short caricatures of everyday sounds that inform operators about the nature of a critical incident.

5 REFERENCES

ATSB Report: 199902928. *Inflight Pilot Capacitation: Raytheon Super King Air 200, VH-OYA, 72km East of Edinburgh Aerodrome, SA, 21 June 1999*. Canberra: Australian Transport Safety Bureau.

ATSB Report: 200003771. *Pilot and Passenger Incapacitation: Beech Super King Air 200, VH-SKC, Wernadinga Station, Qld, 4 September 2000*. Canberra: Australian Transport Safety Bureau.

Begault, D. R. (1994). *3-D sound for virtual reality and multimedia.* Boston: AP Professional.

Calhoun, G., Janson, W. P., & Valencia, G. (1988). Effectiveness of three-dimensional auditory directional cues. *Proceedings of the 32nd Annual Meeting of the Human Factors Society* (pp. 68-72). Santa Monica, CA: Human Factors Society.

Chute, D. L., & Westall, R. F. (1996). Fifth generation research tools: Collaborative development with Powerlaboratory. *Behavior Research Methods, Instruments, & Computers, 28,* 311-314.

CASA. (2001). *Discussion Paper DP 0102CS Proposal for Aural Warning to Operate with Cabin Altitude Alert Warning Systems*. Canberra: Civil Aviation Safety Authority.

Doll, T. J., & Folds D. J. (1986). Auditory signals in military aircraft: ergonomic principles vs. practice. *Applied Ergonomics, 17,* 257-264.

Edworthy, J., & Adams. A. (1996). *Warning design: A research prospective.* London: Taylor & Francis.

Familant, M. E., & Detweiler, M. C. (1993). Iconic reference: Evolving perspectives and an organizing framework. *International Journal of Man-Machine Studies, 39,* 705-728.

Gaver, W. W. (1989). The Sonic Finder: an interface using auditory icons. *Human Computer Interaction, 4, 67*-94.

Gaver, W. W. (1993). What in the world do we hear: An ecological approach to auditory event perception. *Ecological Psychology, 5,* 1-29.

Gaver, W. W. (1994). Sound effects for computational worlds. *Proceedings of the Ergonomics Society/IEE Conference 'Designing Future Interaction'.* Warwick University, UK. April 8th. Loughborough, UK: The Ergonomics Society.

Haxby, J. V., Parasuraman, R., Lalonde, F., & Abboud, H. (1993). SuperLab: General-purpose Macintosh software for human experimental psychology and psychological testing. *Behavior Research Methods, Instruments, & Computers, 25,* 400-405.

Keller, P., & Stevens, C. (2004). Meaning from environmental sounds: Types of signal-referent relations and their effect on recognizing auditory icons. *Journal of Experimental Psychology: Applied, 10,* 3-12.

McKinley, R. L., & Ericson, M. A. (1997). Flight demonstration of a 3-D auditory display. In Gilkey, R. H., & Anderson, T. R. (Eds). *Binaural and spatial hearing in real and virtual environments* (pp. 683-699). Mahwah, NJ: Lawrence Erlbaum Associates.

Momtahan, K., Hetu, R., & Tansley, B. (1993). Audibility and identification of auditory alarms in the operating room and intensive care unit. *Ergonomics, 36,* 1159-1176.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7,* 44-64.

Patterson, R. D. (1982). *Guidelines for auditory warning systems on civil aircraft.* Paper No. 82017. London: Civil Aviation Authority.

Stokes, A. F., & Wickens, C. D. (1988). Aviation displays. In E. L. Wiener & D C Nagel (Eds.), *Human factors in aviation* (pp. 387-431). San Diego, CA: Academic Press.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science, 3,* 159-177.

Wickens, C. D., & Liu, Y. (1988). Codes and modalities in multiple resources: a success and a qualification. *Human Factors, 30,* 599-616.

6 APPENDIXES

6.1 Appendix 1 – Experiment 1: Information sheet and Consent Form

Dear Participant,

Research Project – Design and Evaluation of Auditory Icons for Civil Aviation: Experimental Investigation of Environmental Sounds as Informative Warning Signals

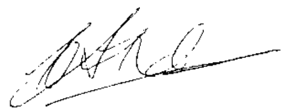
Thank you for signing up for this experiment. This research project investigates the design and evaluation of auditory icons as informative warning signals

The experiment will be conducted on a computer, and as a participant you will be trained to recognise and associate various events that may take place during a flight by a warning. You will then be asked to respond to these warnings and your reaction times and accuracy of responding will be measured.

If the task raises any concerns for you, please do not hesitate to contact the research leader A/Prof Catherine Stevens (02 9772 6324).

The results of these experiments will be presented as a report to the ATSB. A journal manuscript will be prepared and submitted to an international journal such as *Human Factors* or *Journal of Experimental Psychology: Applied.* The results will also be communicated to interested parties and the community in the form of a magazine article, newsletter item, and conference paper. As a participant you are welcome to view the results as they become available.

Your participation in this study is voluntary. You are free to withdraw consent and discontinue participation in the activity at any time without penalty. Any questions concerning this project can be directed to A/Prof Catherine Stevens of the School of Psychology, UWS Bankstown campus (phone: 9772 6324).



Catherine Stevens, PhD
Principal Researcher

NOTE: This study has been approved by the University of Western Sydney Human Research Ethics Committee. If you have any complaints or reservations about the ethical conduct of this research, you may contact the Ethics Committee through the Research Ethics Officers (tel: 02 4570 1136). Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome.

Ethics Protocol No: 05/126

Consent Form

Research Project – Design and Evaluation of Auditory Icons for Civil Aviation: Experimental Investigation of Environmental Sounds as Informative Warning Signals

I ___________________________ have read the information provided and any questions I have asked have been answered to my satisfaction. I agree to participate in this activity, realising that I may withdraw at any time. I agree that research data gathered for the study may be published.

______________________________________ ___________________________

Participant's signature Date

______________________________________ ___________________________

Investigator Date

6.2 Appendix 2 – Experiment 2: Information sheet and Consent Form

Dear Participant,

Research Project – Design and Evaluation of Auditory Icons for Civil Aviation: Experimental Investigation of Environmental Sounds as Informative Warning Signals

Thank you for signing up for this experiment. This research project investigates the design and evaluation of auditory icons as informative warning signals

The experiment will be conducted in a flight simulator, and as a participant you will be trained to recognise and associate various events that may take place during a flight by a warning. You will then be asked to respond to these warnings in a flight simulation scenario where your reaction times and accuracy of responding will be measured.

If the task raises any concerns for you, please do not hesitate to contact the research leader A/Prof Catherine Stevens (02 9772 6324).

The results of these experiments will be presented as a report to the ATSB. A journal manuscript will be prepared and submitted to an international journal such as *Human Factors* or *Journal of Experimental Psychology: Applied.* The results will also be communicated to interested parties and the community in the form of a magazine article, newsletter item, and conference paper. As a participant you are welcome to view the results as they become available.

Your participation in this study is voluntary. You are free to withdraw consent and discontinue participation in the activity at any time without penalty. Any questions concerning this project can be directed to A/Prof Catherine Stevens of the School of Psychology, UWS Bankstown campus (phone: 9772 6324).

Catherine Stevens, PhD
Principal Researcher

NOTE: This study has been approved by the University of Western Sydney Human Research Ethics Committee. If you have any complaints or reservations about the ethical conduct of this research, you may contact the Ethics Committee through the Research Ethics Officers (tel: 02 4570 1136). Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome.

Ethics Protocol No: 05/126

Consent Form

Research Project – Design and Evaluation of Auditory Icons for Civil Aviation: Experimental Investigation of Environmental Sounds as Informative Warning Signals

I ___________________________ have read the information provided and any questions I have asked have been answered to my satisfaction. I agree to participate in this activity, realising that I may withdraw at any time. I agree that research data gathered for the study may be published.

_______________________________________ ___________________________

Participant's signature Date

_______________________________________ ___________________________

Investigator Date

6.3 Appendix 3 – Experiment 2: Pilot demographic information

Please indicate your age (in years): ☐

Please indicate your gender: Male ☐ Female ☐

Please indicate which of the following licences that you hold:

- GFPT ☐
- Private ☐
- Commercial ☐
- ATPL ☐

Please indicate which of the following ratings that you hold:

- Instructor ☐
- Instrument ☐

Each of the following questions is related to your flying experience. Please estimate these figures as accurately as possible.

Number of hours(total) experience: ☐

Number of hours(total) as pilot in command: ☐

Number of hours(total) actual IFR experience: ☐

Number of cross-country hours experience (excluding training): ☐

Number of hours(total) during the previous 90 days: ☐

Number of cross-country hours during the previous 90 days: ☐