A Detailed Investigation into the Effectiveness of Earcons

Stephen A. Brewster, Peter C. Wright and Alistair D. N. Edwards

HCI Group
Department of Computer Science
University of York
Heslington
York, Y01 5DD, UK.
Tel.: 0904 432765
sab@minster.york.ac.uk

ABSTRACT

A detailed experimental evaluation of earcons was carried out to see whether they are an effective means of communicating information in sound. An initial experiment showed that earcons were better than unstructured bursts of sound and that musical timbres were more effective than simple tones. Musicians were shown to be no better than non-musicians when using musical timbres. A second experiment was then carried out which improved upon some of the weaknesses of the pitches and rhythms used in Experiment 1 to give a significant improvement in recognition. From the results some guidelines were drawn up for designers to use when creating earcons. These experiments have formally shown that earcons are an effective method for communicating complex information in sound.

KEYWORDS

Auditory interfaces, earcons, sonification

INTRODUCTION

The use of non-speech audio at the graphical user-interface is becoming increasingly popular due to the potential benefits it offers. There are many reasons for this. In everyday life people communicate using all their senses, with information in one sensory modality being backed up by data from the others. When they come to use computers the interaction is restricted almost solely to the visual channel and this limitation can cause the interface to intrude into the task that the user is trying to perform. The aim of a multimedia interface is to make the interaction more natural and the interface more transparent by using different forms of input and output. Most current interfaces make little use of sound other than for beeps to indicate errors. It can be used to present information otherwise unavailable on a visual display for example mode information [10, 13, 14] or information that is hard to discern visually, such as multi-dimensional numerical data[5]. It is a useful complement to visual output because it can increase the amount of information communicated to the user or reduce the amount the user has to perceive through the visual channel.

In order to fully understand the design of auditory interfaces one should have some knowledge of psychoacoustics: the study of the perception of sound. This aims to describe the relationships between the characteristics of a sound which enters the ear and the sensations these produce within the auditory system. In order to create sounds which a listener is able to hear and differentiate, the range of human auditory perception must not be exceeded. Frysinger [5] says “The characterisation of human hearing is essential to auditory data representation because it defines the limits within which auditory display designs must operate if they are to be effective”. Moore [11] gives an overview of the field of psychoacoustics.

The work reported here is part of a research project looking at the best ways to integrate auditory and graphical information at the interface. The research aims to find the areas in an interface where the use of sound will be most beneficial and also what types of sounds are the most effective for communication. Sound will be used to present information where the visual interface breaks down and does not tell the user everything that they need to know.

Earcons and Auditory Icons

One major question that must be answered when creating an auditory interface is: What sounds should be used ? Brewster [3] outlines some of the different systems available. Gaver’s *auditory icons* have been used in several systems, such as the SonicFinder [6], SharedARK [7] and ARKola [8]. These use environmental sounds that have a semantic link with the object or action they represent. They have been shown to be an effective form of presenting information in sound.

One alternative, and previously untested, method of presenting auditory information are *earcons* [2, 15, 16]. Earcons are abstract, synthetic tones that can be used in structured combinations to create sound messages to represent parts of an interface. Blattner *et al.* define earcons as “non-verbal audio messages that are used in the computer/user interface to provide information to the user about some computer object, operation or interaction”. Earcons are composed of motives, which are short, rhythmic sequences of pitches with variable intensity, timbre and register.

Blattner *et al*. define a system of hierarchical earcons in their papers. Each earcon is a node on a tree and inherits all the properties of the earcons above it. Figure 1 shows a hierarchy of earcons. There is a maximum of five levels to the tree as there are five parameters that can be varied: rhythm, pitch, timbre, register and dynamics. In the diagram the top level of the tree is the family rhythm, in this case it is a sound representing error. This sound has rhythm but no pitch; the sounds used are clicks. The rhythmic structure of level one is inherited by level two but this time a second motive is added where pitches are put to the rhythm. At this level the timbre is a sine wave, which produces a ‘colourless’ sound. This is done so that at level three the timbre can be varied. At level three the pitch is also raised by a semitone to make it easier to differentiate from the pitches inherited from level two. Other levels can be created where register and dynamics are varied. To play the final earcon requires three separate motives to be played. In order to speed up the presentation of earcons this paper suggests that only the last motive need be played. This motive (the one labelled ‘triangle’ in Figure 1) contains all of the information needed. It has the rhythm of the error family, it has the pitch structure of the execution error and the timbre of an underflow error. If only the last motive is used then the length of the whole sound is greatly reduced. The experiments described in this paper test hierarchical earcons used in this manner to discover their effectiveness.

One other powerful feature of earcons is that they can be combined to produce complex audio messages. Earcons for a set of simple operations, such as ‘open’, ‘close’, ‘file’ and ‘program’, could be created. These could then be combined to produce, for example, earcons for ‘open file’ or ‘close program’.

Earcons provide a powerful method of sonification. They can be used for adding sound to both data and interfaces. Related items can be given related sounds, hierarchies of information can be represented. Complex messages made of sub-units can be built up. They are a powerful and flexible means of creating auditory messages.

Up to now, no extensive formal experiments have been conducted to see if earcons are an effective means of communicating information using sound. Jones & Furner [9] carried out a comparison between earcons, auditory icons and synthetic speech. Their results showed that subjects preferred earcons but were better able to associate auditory icons to commands. Their results were neither extensive nor detailed enough to give a full idea of whether earcons are useful or not. Barfield, Rosenberg & Levasseur [1] also carried out experiments where they used earcons to aid navigation through a menu hierarchy. The earcons they used were very simple, just decreasing in pitch as a subject moved down the hierarchy. These sounds did not fully exploit all the advantages that earcons offer and were also only a secondary part of the experiment. This paper seeks to discover how well complex earcons can be recalled and recognised.

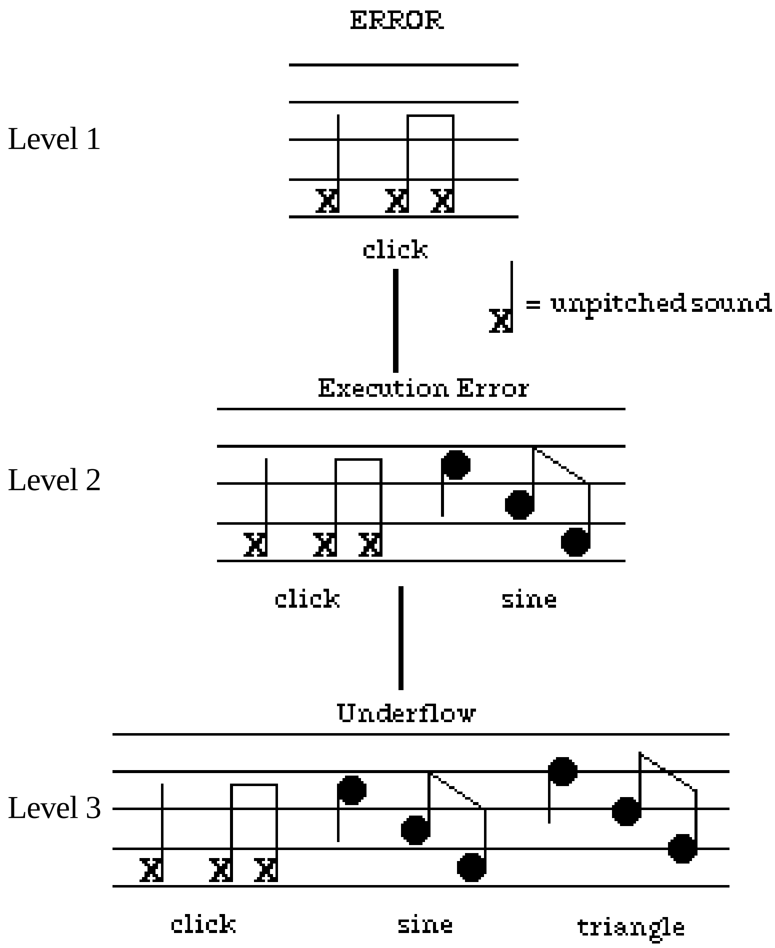

Figure 1. Hierarchical earcons (adapted from [2]).

The first experiment described attempts to discover if earcons are better than unstructured bursts of sound and tries to identify the best types of timbres to use to convey information. Blattner *et al.* suggest the use of simple timbres such as sine or square waves but psychoacoustics suggests that complex musical instrument timbres may be more effective [11]. The second experiment uses the results of the first to create new earcons to overcome some of the difficulties that came to light. Guidelines are then put forward for use when creating earcons.

The use of intensity as a parameter for manipulating earcons

Intensity is one of the parameters put forward by Blattner *et al.* for differentiating earcons. It is suggested here that care should be taken when using intensity. The intensity of a sound can be thought of as similar to the brightness on a video monitor. On a monitor the user can change the brightness of the display in response to the ambient light level. If the room is light then the brightness of the display will be increased so that the information on the screen can still be seen. If the room is dark then the brightness will be turned down so that the screen does not hurt the eyes. The volume control on a monitor acts in a similar way. If the room is noisy then the intensity will be increased to avoid masking. If the room is quiet then the intensity can be turned down to avoid irritation. If the sounds used vary widely in intensity then turning up the volume so that the quiet sounds can be heard will cause the loud sounds to become irritating. Conversely, turning down the loud sounds to a pleasant level may cause the quiet ones to fall below the threshold of hearing. One other important psychoacoustic factor is that changing the intensity of a sound can change the perceived pitch of the sound. Brewster [3] reports “Intensity also affects perceived pitch. At less than 2kHz an increase in intensity increases the perceived pitch. At 3kHz and over an increase in intensity decreases the perceived pitch”. These points indicate that intensities used in earcons should be kept within a narrow range and the overall control of intensity given to the user.

EXPERIMENT 1

Subjects

Thirty-six subjects, three groups of twelve, were used from the University of York. Seventeen of the subjects were musically trained (they could play an instrument and read music). They were randomly allocated to one of three groups so that there was an even mix of musicians and non-musicians (the simple tone group had only five musicians).

Sounds used

An experiment was designed to find out if structured sounds such as earcons were better than unstructured sounds for communicating information. Simple tones were compared with complex musical timbres. Rhythm and pitch were also tested as ways of differentiating earcons. According to Deutsch [4] rhythm is one of the most powerful methods for differentiating sound sources. The experiment also attempted to find out how well subjects could identify earcons individually and when played together in sequence. Figures 2 and 3 give the rhythm and pitch structures used in phase I and II of the experiment. The sounds were based around middle C. In each phase a dummy rhythm and dummy timbre were inserted into the testing phase. For example, the subject would hear a known rhythm but with a new timbre. This would test to see if subjects could recognise that the earcons had changed.

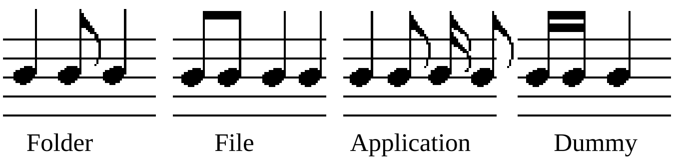

Figure 2: Rhythms used in phase I of Experiment 1

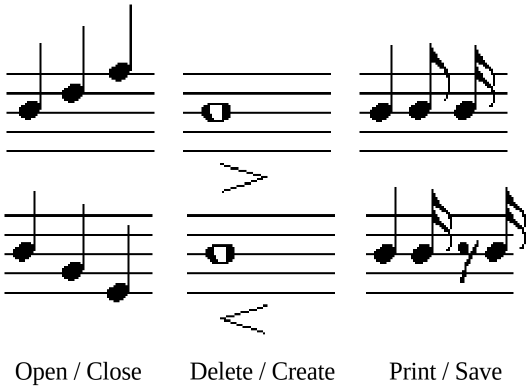

Figure 3: Rhythms used in phase II of Experiment 1

Three sets of sounds were created:

Musical Sounds: The first set were synthesised musical timbres: piano, brass, marimba and pan pipes. These were produced by a Roland D110 synthesiser. This set had rhythm information as shown in the figures above.

Simple Sounds: The second set were simple timbres: sine wave, square wave, sawtooth and a ‘complex’ wave (this was composed of a fundamental plus the first three harmonics. Each harmonic had one third of the intensity of the previous one). These sounds were created by SoundEdit on an Apple Macintosh. This set also had rhythm information as shown above.

Control Sounds: The third set had no rhythm information; these were just one second bursts of sound similar to normal system beeps. This set had timbres made up from the musical group.

Experimental design

As mentioned, three groups of twelve subjects were used. Each of the three groups heard different sound stimuli. The musical group heard the musical sounds described in the previous section. The simple group heard the simple sounds and the control group heard the control sounds. There were four phases to the experiment. In the first phase subjects heard sounds for icons. In the second they heard sounds for menus. In the third phase they were tested on the icon sounds from phase I again. In the last phase subjects were required to listen to two earcons played in sequence and give information about both sounds heard. Instructions were read from a prepared script.

Phase I

Training: The subjects were presented with the screen shown in Figure 4. Each of the objects on the display had a sound attached to it. The sounds were structured as follows. Each *family* of related items shared the same timbre. For example, the paint application, the paint folder and paint files all had the same instrument. Items of the same *type* shared the same rhythm. For example, all the applications had the same rhythm. Items in the *same* family and of the *same* type were differentiated by pitch. For example, the first Write file was C below middle C and the second Write file was G below that. In the control group no rhythm was information was given so types were also differentiated by pitch.

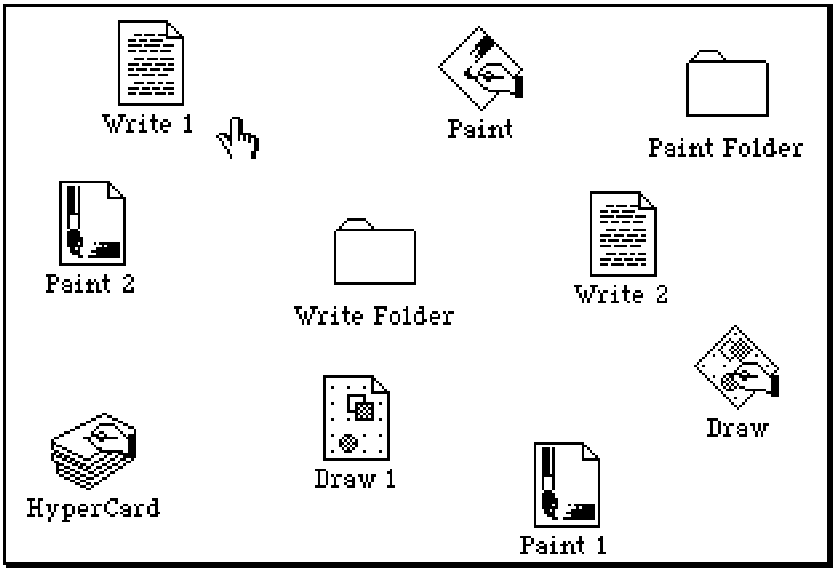

Figure 4: The Phase I icon screen

The icons were described to each subject. The relationships between types were described. For example, the relationship between applications was indicated by having icons with hands in the graphic. The relationships between families were described and all the members of each family pointed out. The subjects were then asked to learn the names of all the icons. When they thought they had done this they wrote them down. If they were not correct they were allowed more time to learn them. This meant that, at the end of the training, the subjects new the names of the icons present. The icons were then played one-at-a-time in random order to the subject for them to learn. The whole set of icons was played three times. [Sound example *A* gives the ‘Paint application', 'Write folder' and 'Draw file 1' earcons used in the Control, Simple and Musical groups respectively]

Testing: When testing the subjects the screen was cleared and a selection of the earcons were played back in random order. Before the sounds were played back it was indicated to the subject that they might hear some new sounds that they had not heard before. The subject had to supply what information they could about type, family and if it was a file then the number of the file. In this and all the phases the subject was allowed to hear each sound a maximum of three times. When scoring, a mark was given for each correct piece of information supplied.

Phase II

In this phase, earcons for menus were tested. Each *menu* had its own timbre and the *items* on each menu were differentiated by rhythm, pitch or intensity. The screen shown to the users to learn the earcons is given in Figure 5. The training was similar to phase I. The subjects were tested in the same way as before but this time had to supply information about menu and item. [Sound example *B* gives the 'Open', 'Save' and 'Undo' earcons used in the Control, Simple and Musical groups respectively]

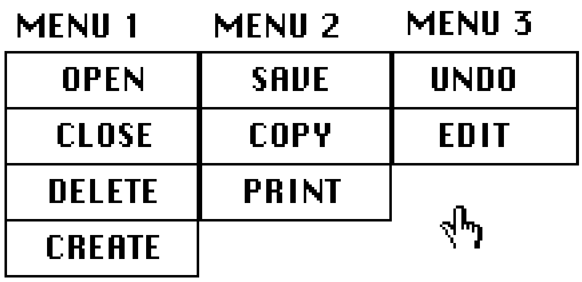

Figure 5: The Phase II menu screen

Phase III

This was a re-test of phase I but no further training time was given and the earcons were presented in a different order. This was to test if the subjects could remember the original set of earcons after having learned another similar set.

Phase IV

This was a combination of phases I and II. Again, no chance was given for the subjects to re-learn the earcons. The subjects were played two earcons, one followed by another, and asked to give what information they could about each sound they heard. The sounds they heard were from the previous phases and could be played in any order (i.e. it could be menu then icon, icon then menu, menu then menu or icon then icon). This would test to see what happened to the recognition of the earcons when played in sequence. A mark was given for any correct piece of information supplied. [Sound example *C* gives the 'Open Paint application' and 'Save Write folder' earcons used in the Control, Simple and Musical groups respectively]

RESULTS AND DISCUSSION OF EXPERIMENT 1

From Figures 6 and 7 it can be seen that the musical earcons came out best overall and in each of the phases. Unfortunately these differences did not reach statistical significance.

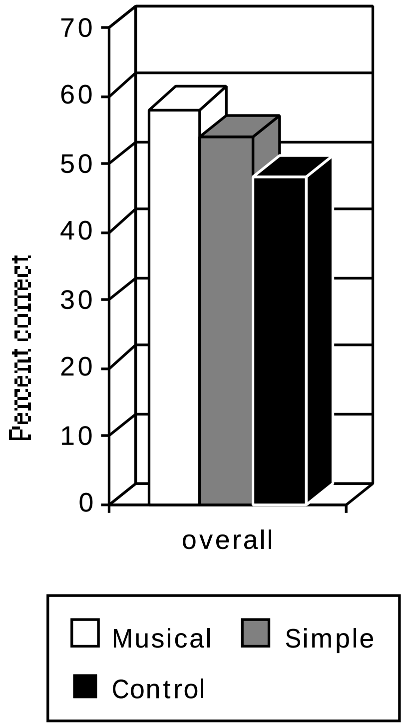

Figure 6: Overall scores in Experiment 1

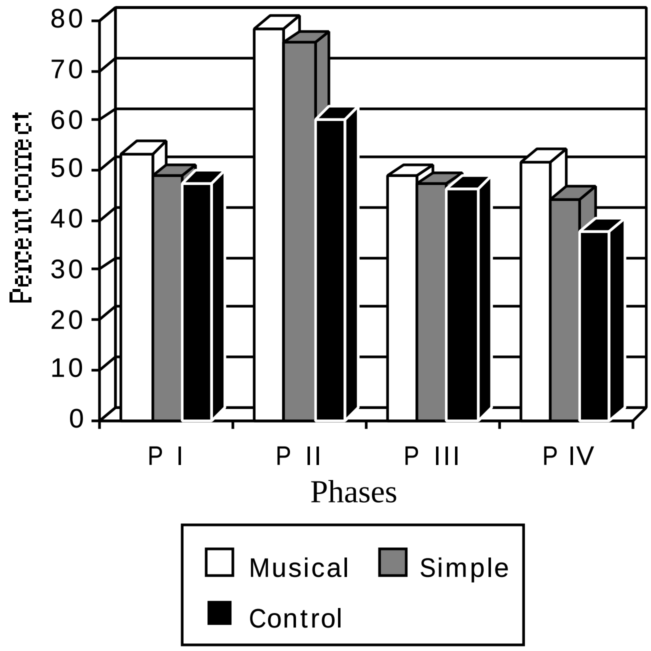

Figure 7: Breakdown of overall scores per phase for Experiment 1

Phase I

The breakdown of scores can be seen in Figure 8. A between-groups ANOVA was carried out on the family scores (family was differentiated by timbre) and showed a significant effect ($F(2,33)= 9.788$, $p<0.0005$). A Sheffe F-test showed that the family score in the musical group was significantly better than the simple group ($F(2,33) =6.613$, $p<0.05$). This indicates that the musical instrument timbres were more easily recognised than the simple tones proposed by Blattner *et al*.

There were no significant differences between the groups in terms of type (differentiated by rhythm). The control group should have performed the worst as they had no rhythm information. However, the results show that the simple and musical groups performed no better. Therefore, the rhythms used did not give any better performance over the straight bursts of sound. This indicates that the chosen rhythms were ineffective. The file scores should have been the same as all groups were differentiated by pitch. A wide variation in results occurred indicating that pitch alone is not an effective means of differentiation.

Within groups the type scores were significantly worse than the family scores for the musical and control groups ($T(11)=2.96$, $p<0.05$, $T(11)=3.55$, $p<0.05$ respectively). This again shows that the rhythms chosen were difficult to use. In the simple group the type score was significantly better than the family score. Again this could be because the simple sounds are hard to remember so that the scores were lower.

Phase II

The overall scores were significantly better than those for phase I (see Figure 7). An ANOVA on the overall scores showed a significant effect ($F(2,33)=5.182$, $p<.011$). This suggests that the rhythms used were more effective, as the timbres were similar to the previous phase. Figure 9 shows the simple and musical groups performed similarly which was to be expected as both used the same rhythms. A Sheffe F-test showed both were significantly better than the control group (musical vs. control $F(2,33)=6.278$, $p<0.05$, simple vs. control $F(2,33)= 8.089$, $p<0.05$). Again, this was to be expected as the control group had only pitch to differentiate items. This shows that if

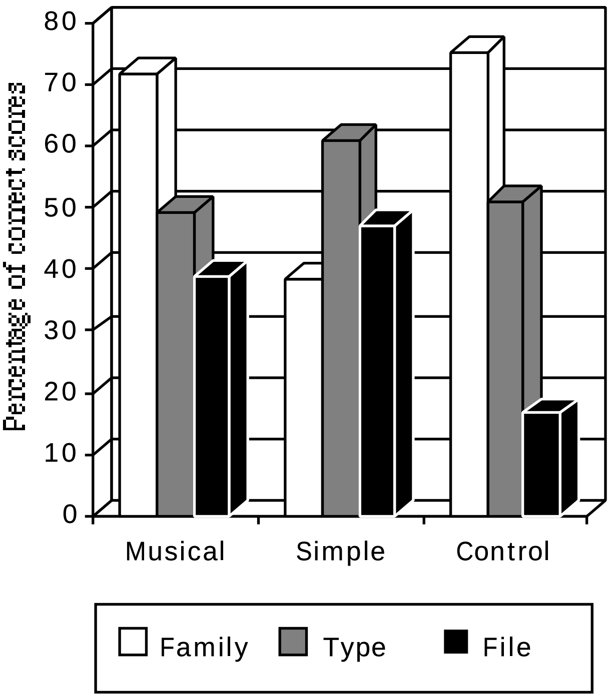

Figure 8: Breakdown of scores for phase I of Experiment 1

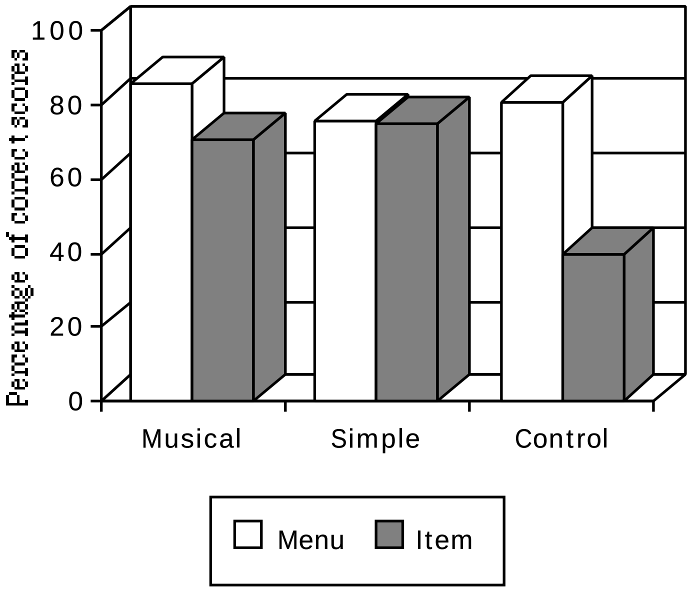

Figure 9: Breakdown of scores for phase II of Experiment 1

rhythms are chosen correctly then they can be very important in aiding recognition. It also shows that pitch alone is very difficult to use.

A Sheffe F-test showed that overall in phase II the musical group was significantly better than the control group ($F(2,33)=4.5$, $p<0.05$). This would indicate that the musical earcons used in this group were better than unstructured bursts of sound.

An ANOVA on the menu scores between the simple and musical groups showed an effect ($F(1,22)=3.684$, $p<0.68$). A Sheffe F-test showed that the musical instrument timbres just failed to reach significance over the simple tones ($F(1,22)=3.684$, $p<0.10$). A within-groups t-test showed that in the musical group the menu score (differentiated by timbre) was still significantly better than the item score ($T(11)=2.69$, $p<0.05$). This seems to indicate, once more, that timbre is a very important factor in the recognition of earcons.

Phase III

In this phase the earcons from phase I were tested again. A period of approximately 15 minutes had passed since the subjects learned the earcons. The scores in phase III were not significantly different to those in phase I (see Figure 7). This indicates that subjects managed to recall and remember the earcons from phase I even after learning the sounds for phase II, which were very similar. This seems to indicate that a subject’s memory for earcons is strong.

Phase IV

Figure 10 shows the scores in phase IV where combinations of earcons were tested. A within groups t-test showed that, in the musical group, the menu/item combination was significantly better than the family/type/file combination ($T(11)=2.58$, $p<0.05$). This mimics the results for the musical group from phases I and II. When comparing phase IV with the other phases performance was worse in all groups with the exception of type recognition by the musical group and family recognition by the simple group. This indicates that there is a problem when two earcons are

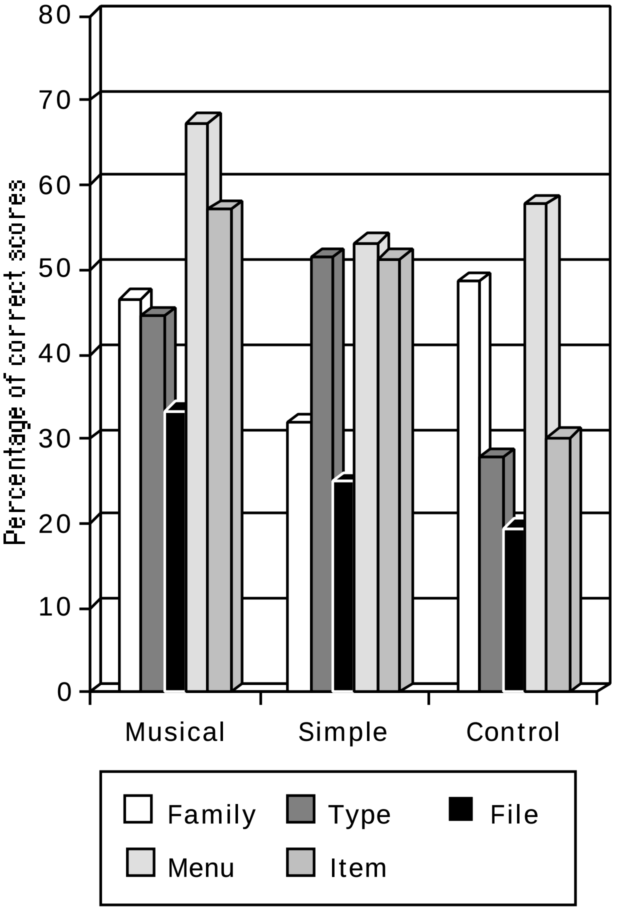

Figure 10: Breakdown of scores for phase IV of Experiment 1

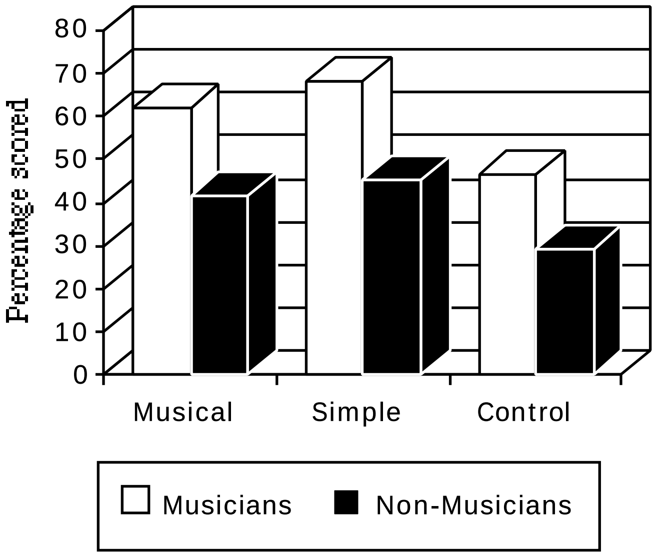

Figure 11: Scores of musicians and non-musicians in phase IV of Experiment 1

combined together. If the general perception of the icon sounds could be improved then this might raise the scores in phase IV

A between-groups ANOVA showed the only significant effect to be on item scores ($F(2,33)=4.044$, $p=0.0269$). A Sheffe F-test showed that the item scores in the musical group was significantly better than the control group ($F(2,33)=3.647$, $p<0.05$). Items were differentiated by combinations of rhythm, pitch and intensity in the musical group but the control group only used pitch. This indicates again that pitch alone is not a good way of differentiating but combining it with other variables makes for much better recognition rates.

When comparing the results of previous phases with the corresponding part of phase IV it can be seen that in the musical group all scores are significantly worse (except for the phase I type score which was not significantly different from the type score in phase IV). In the simple group all the phase IV scores were worse than the previous phases apart from the family score. In the control group all the scores were worse than the previous phases. This indicates that there is a problem when two earcons are put together.

Discussion of musicians / non-musicians

One important factor to consider is that of musical ability. Are earcons only usable by trained musicians or can non-musicians use them equally as effectively ? If the former were true then it would limit the effectiveness of earcons in the sonification of information.

Musical group

The earcons in the musical group from Experiment 1 were not statistically significantly better recognised by the musicians than the non-musicians. This means that a non-musical user of a system involving earcons would have no more difficulties than a musician. The only time a significant difference occurred was in phase IV where the musicians were better at identifying the file number ($T(10)=2.83$, $p<0.05$). This is due to musical training allowing the identification of individual pitches more accurately. The results from phase IV are shown in Figure 11.

Simple group

When the simple sounds were used the musicians were significantly better than the non-musicians with the phase I types and families (types $T(10)=3.27$, $p<0.05$, families $T(10)=2.26$, $p<0.05$). Again, in phase IV the musicians were better on type ($T(10)=3.09$, $p<0.05$). The problems with the rhythms in phase I have been discussed above. It seems that the musicians were able to use the difficult rhythm information better as they are more highly trained. In a similar manner the musicians were able to recognise the simple tone timbres which the non-musicians found hard to differentiate.

Control group

In phase IV the musicians did significantly better than the non-musicians with the menus ($T(10)= 2.49$, $p<0.05$), items ($T(10)=2.48$, $p<0.05$) and families ($T(10)=2.85$, $p<0.05$).

Overall

These results seem to indicate that if musical sounds were to be used for the earcons then musically untrained subjects would not be at a disadvantage to musicians. The only case where musicians proved to be better was with things differentiated by pitch alone. If simple tones or bursts of sound are used then musicians will have an advantage. Therefore, to create sounds that are usable by the general population musical earcons should be used.

Discussion of Dummy Earcons

In each phase dummy earcons were introduced in the testing stage to see if the subjects could recognise them. The musical and simple groups had both timbre and rhythm dummies. The control group had only timbre dummies as its earcons contained no rhythm information. An example of the dummy rhythm used in phase I is given in Figure 2.

Musical group

Recognition of dummy earcons in the musical group is shown in Figure 12. In phase I it can be seen that dummy timbre recognition was high (83% were recognised) but dummy rhythm recognition was low (only 8% were recognised). This again indicates that the musical timbres were easy to recognise but the rhythms chosen were hard, mirroring the results shown in Figure 8. Phase II recognition rates were high. This shows, as before, that if the rhythms are used carefully then high rates of recognition can be achieved. In phase III the recognition of the dummy rhythms increased to 50%. It is unclear why the subjects could recognise the dummies better after a period of time. The scores in phase IV mirror the scores shown in Figure 7. The overall rate of recognition was 52% not significantly different to the overall recognition to the genuine earcons. These results seem to indicate that subjects could recognise dummy earcons with the same level of accuracy as the genuine ones. The implication of this is that subjects can identify earcons not heard before which makes earcons a more robust means of communication.

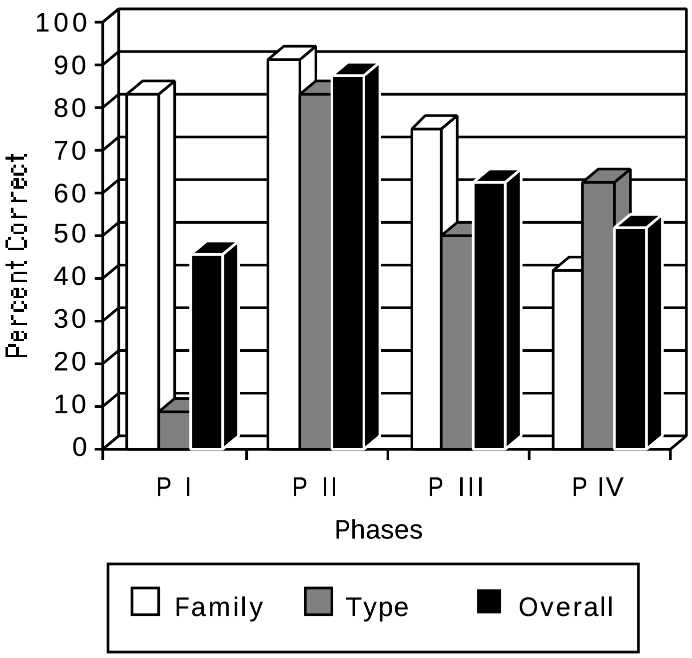

Figure 12: Recognition of dummy earcons in Experiment 1

Simple group

In phase I the simple group had a similar overall score to the musical group but with lower timbre scores. This matches the results from the genuine earcons. These results imply that, with simple earcons, users would find it difficult to identify sounds that they had not heard before. In all of the other phases this group had lower identification rates than the musical group.

Control group

This group had no dummy rhythms. This group had similar timbre scores to the musical group in phases I, II and III. In phases IV the timbre scores were the lowest (4.1%). This could be due to the overall difficulty of the control sounds, as shown in Figure 7.

Overall

This result reinforces the power of the musical earcons. Subjects found it easier to recognise the dummies in this group. This again implies that musical earcons are the most effective of the sounds tested.

Summary of Experiment 1

Some general conclusions can be drawn from this first experiment. It seemed that earcons were better than unstructured bursts of sound at communicating information under certain circumstances. The issue of how this advantage could be increased needed further examination. Similarly, the musical timbres came out better than the simple tones but often by only small amounts. Further work was needed to make them more effective. The results also indicated that rhythm must be looked at more closely. In phase I the rhythms were ineffective but in phase II they produced significantly better results. The reason for this needed to be ascertained. Finally, the difficulties in recognising combined earcons had to be reduced so that higher scores could be achieved.

EXPERIMENT 2

From the results of the first experiment it was clear that the recognition of the icon sounds was low when compared to the menu sounds and this could be affecting the score in phase IV. The icon sounds needed to be improved along the lines of the menu sounds.

Sounds Used

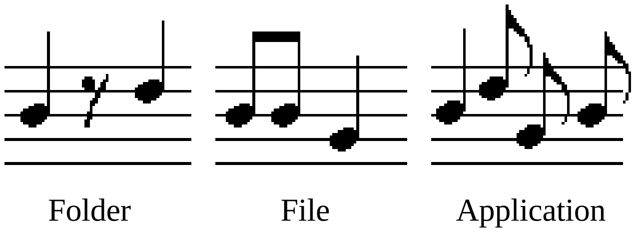

Figure 13: New phase I rhythms and pitch structures used in Experiment 2

The sounds were redesigned so that there were more gross differences between each one. This involved creating new rhythms for files, folders and applications each of which had a different number of notes. Each earcon was also given a more complex within-earcon pitch structure. Figure 13 shows the new rhythms and pitch structures for folder, file and application. No changes were made to the rhythms and pitch structures used for the phase II menu item sounds.

The use of timbre was also extended so that each family was given two timbres which would play simultaneously. The idea behind multi-timbral earcons was to allow greater differences between families; when changing from one family to another two timbres would change not just one. This created some problems in the design of the new earcons as great care had to be taken when selecting two timbres to go together so that they did not mask one-another.

Findings from research into the perception of sound were included into the experiment. Patterson [12] gives some limits for pitch and intensity ranges. This lead to a change in the use of register. In Experiment 1 all the icon sounds were based around middle C (261Hz). All the sounds were now in put into a higher register, for example the folder sounds were now made two octaves above middle C. In experiment 1, the ‘file 1’ earcons were an octave below middle C (130Hz) and the ‘file 2’s a G below that (98Hz). These frequencies were below the range suggested by Patterson and were very difficult to tell apart. In Experiment 2 the register of the ‘file 1’ earcons were three octaves above middle C (1046Hz) and the ‘file 2’s at middle C. These were now well within Patterson’s ranges.

In response to informal user comments from Experiment 1, a 0.1 second delay was inserted between the two earcons. Subjects had complained that they could not tell where one earcon stopped and the other started. [Sound example *D* gives the new earcons for ‘Paint application’, ‘Write folder’ and ‘Draw file 1’]

Method

The experiment was the same as the previous one in all phases but with the new sounds. A single group of a further twelve subjects was used. Subjects were chosen from the same population as before so that comparisons could be made with the previous results.

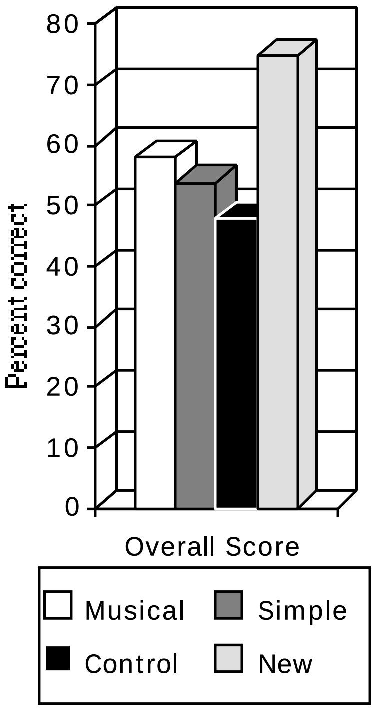

Figure 14: Percentage of overall scores with Experiment 2

RESULTS AND DISCUSSION OF EXPERIMENT 2

As can be seen from Figure 14, the new sounds performed much better than the previous ones. An ANOVA on the overall scores indicated a significant effect ($F(3,44)= 6.169$, $p<0.0014$). A Sheffe F-test showed that the new group was significantly better than the control group ($F(3,44)=5.426$, $p<0.05$) and the simple group ($F(3,44)= 3.613$, $p<0.05$). This implies that the new earcons were more effective than the ones used in the first experiment. Comparing the musical group (which was the best in all phases of Experiment 1) with the new group the level of recognition in phases I and III has been raised to that of phase II.

Phase I

The overall recognition rate in phase I was increased because of a very significantly better type score (differentiated by rhythm) in the new group ($F(1,22)= 26.677$, $p<0.05$). The scores increased from 49.1% in the musical group of experiment 1 to 86.6% in the new group (see Figure 15). This seems to indicate that the new rhythms were effective and very easily recognised.

The wider register range used to differentiate the files made a significant improvement over the previous experiment ($F(1,22)=4.829$, $p<0.05$). This indicates that using the higher pitches and greater differences in register made it easier for subjects to differentiate one from another. The general improvement in recognition in phase I brought the scores up to the level of the musical group in phase II of the previous experiment. This indicates that with more careful design of earcons recognition rates can be significantly improved.

Phases II and III

The scores in phase II were unchanged from the previous experiment as was expected. In phase III the scores were not significantly different to phase I, again indicating that the sounds are easily remembered.

Phase IV

In phase IV (combinations of earcons) the overall score of the new group just failed to reach significance over the musical group ($F(1,22)= 3.672$, $P<0.10$). However, the new earcons were significantly better than the musical ones from the previous experiment in terms of type ($F(1,22)=9.135$, $p<0.05$) and family ($F(1,22)=4.989$, $p<0.05$). Figure 16 indicates this. The menu and item scores were not different, as was expected, because the same earcons were used as in experiment 1. T-tests revealed that recall in phase IV was still slightly lower than the other phases. The overall phase I score of the new group was significantly better than the score in phase IV ($T(11)=3.02$, $p<0.05$).

The multi-timbral earcons made no difference in phase I. The family score for the new group was not significantly different to the score in the musical group. There were also no differences in phases II or III. However, in phase IV the recognition of icon family was significantly better than in the musical group ($F(1,22)=4.989$, $p<0.05$). A further analysis of the data showed that there was no significant difference between the phase I and phase IV scores in the new group. However, the phase IV score for the musical group was worse than phase I ($T(11)=4.983$, $p<0.05$). This indicates that there was a problem in the musical group that was overcome by the new sounds. It may have

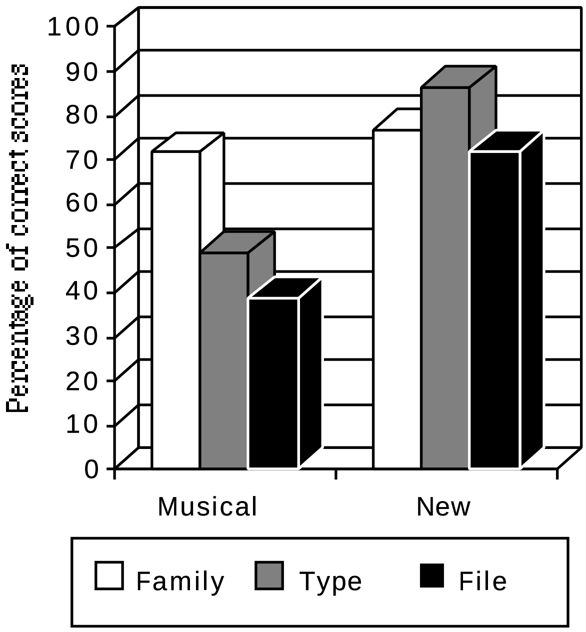

Figure 15: Breakdown of scores for phase I of Experiment 2 compared to the musical group from Experiment 1

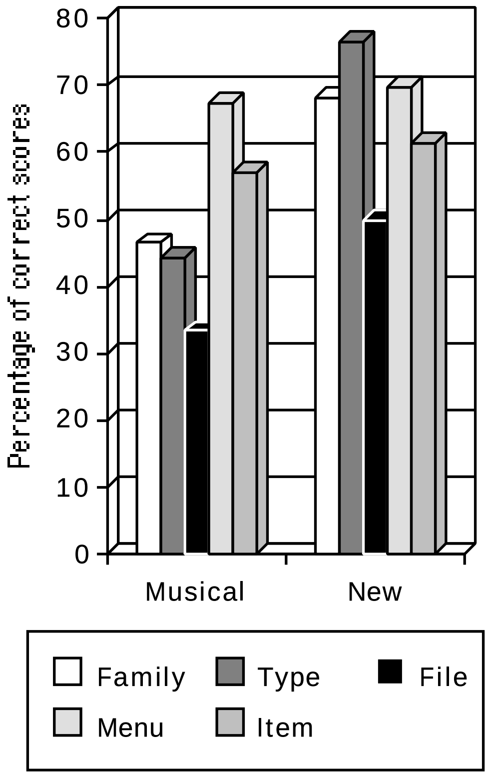

Figure 16: Breakdown of scores for phase IV of Experiment 2

been that in phases I, II and III only one timbre was heard and so it was clear to which group of earcons it belonged (icons sounds or menu sounds). When two earcons were played together it was no longer so clear as the timbre could be that of a menu sound or an icon sound. The greater differences between each of the families when using multi-timbral earcons may have overcome this.

Musicians and non-musicians

The results also show that there was no significant difference in performance between the musicians and non-musicians with the new sounds in Experiment 2. This seems to indicate that musical earcons are the most effective way of communicating complex information for general users.

GUIDELINES

From the results of the two experiments and studies of literature on psychoacoustics some guidelines have been drawn up for use when creating earcons. These should be used along with the more general guidelines given in [15, 16]. One overall result which came out of the work is that much larger differences than those suggested by Blattner *et al* must be used to ensure recognition. If there are only small, subtle changes between earcons then they are unlikely to be noticed by anyone but skilled musicians.

• *Timbre*: Use musical instrument timbres. Where possible use timbres with multiple harmonics. This helps perception and avoids masking. Timbres should be used that are subjectively easy to tell apart e.g. use ‘brass’ and ‘organ’ rather than ‘brass1’ and ‘brass2’.

• *Pitch*: Do not use pitch on its own unless there are very big differences between those used (see register below). Complex intra-earcon pitch structures are effective in differentiating earcons if used along with rhythm. Some suggested ranges for pitch are: Max.: 5kHz (four octaves above middle C) and Min.: 125Hz - 150Hz (an octave below middle C).

• *Register*: If this alone is to be used to differentiate earcons which are otherwise the same, then large differences should be used. Two or three octaves difference give good rates of recognition.

• *Rhythm*: Make them as different as possible. Putting different numbers of notes in each rhythm was very effective. Patterson [12] says that sounds are likely to be confused if the rhythms are similar even if there are large spectral differences. Small note lengths might not be noticed so do not use notes less than eighth notes or quavers. In the experiments described here these lasted 0.125 seconds.

• *Intensity*: Although intensity was not examined in this test some suggested ranges (from Patterson) are: Max.: 20dB above threshold and Min.: 10dB above threshold. Care must be taken in the use of intensity. The overall sound level will be under the control of the user of the system. Earcons should all be kept within a close range so that if the user changes the volume of the system no sound will be lost.

• *Combinations*: When playing earcons one after another use a gap between them so that users can tell where one finishes and the other starts. A delay of 0.1 seconds is adequate. If the above guidelines are followed for each of the earcons to be combined then recognition rates should be sufficient.

FUTURE WORK

No work was done in this paper to test the speed of presentation of earcons. The earcons took around 1.5 seconds to play. In a real application of earcons they would need to be presented so that they could keep up with activity in the interface. A further experiment would be needed to test the maximum speed of display attainable whilst still remaining recognisable. Work is now underway to investigate the presentation of earcons in parallel to speed up the rate of display.

The subjects only heard each of the earcons three times in the training parts of the experiment but reached 80% recognition rates. A more long term study would show what levels of recognition could be reached when subjects had more time to learn the sounds. Work is also needed to look at the intensity of presentation. In many existing systems the sounds are played much too loud and so become intrusive.

CONCLUSIONS

The results indicate that earcons are an effective means of communication. The work shown has experimentally demonstrated that earcons are better for presenting information than unstructured bursts of sound. This gives a formal basis for their use in future systems. Musical timbres for earcons proved to be more effective than the simple tones proposed by Blattner *et al.*. The subtle transformations suggested by Blattner have been shown to be too small to be recognised by subjects and that gross differences must be used if differentiation is to occur. The results of Experiment 1 indicated that earcons were effective but needed refinements. The results from Experiment 2 show that high levels of recognition can be achieved by careful use of pitch, rhythm and timbre. Multi-timbral earcons were put forward and shown to help recognition under some circumstances. A set of guidelines has been suggested, based on the results of the experiments, to help a designer of earcons make sure that they will be easily recognisable by listeners.

This research now means that there is a strong experimental basis to prove that earcons are effective. This work has shown that earcons can be individually recognised rather than recognition being based on hearing a relative change between two sounds. Earcons could therefore be used as landmarks in an auditory space where they give absolute information about events. Developers can create sonifications of data or interfaces that use earcons safe in the knowledge that they are a good means of communication.

ACKNOWLEDGEMENTS

We would like to thank all the subjects for participating in the experiment. Thanks also go to Andrew Monk for helping with the statistical analysis of the data.

REFERENCES

1. Barfield, W., Rosenberg, C. & Levasseur, G. The use of icons, earcons and commands in the design of an online hierarchical menu. *IEEE Transactions on Professional Communication* **34**(2) (1991): 101-108.

2. Blattner, M., Sumikawa, D. & Greenberg, R. Earcons and icons: Their structure and common design principles. *Human Computer Interaction* **4**(1) (1989): 11-44.

3. Brewster, S.A. *Providing a model for the use of sound in user interfaces* (Technical Report No. YCS 169). University of York, Department of Computer Science, 1992.

4. Deutsch, D. The processing of structured and unstructured tonal sequences. *Perception and Psychophysics* **28**(5) (1980): 381-389.

5. Frysinger, S.P. Applied research in auditory data representation. In *Proceedings of the SPIE vol 1259. Extracting meaning from complex data: processing, display, interaction*. 1990.

6. Gaver, W. The SonicFinder: an interface that uses auditory icons. *Human Computer Interaction* **4**(1) (1989): 67-94.

7. Gaver, W. & Smith, R. Auditory icons in large-scale collaborative environments. In *Interact'90*. edited by Diaper, D., 735-740. Elsevier Science Publishers B.V. (North Holland), 1990.

8. Gaver, W., Smith, R. & O'Shea, T. Effective sounds in complex systems: the ARKola simulation. In *CHI'91*. edited by Robertson, S., Olson, G. & Olson, J., 85-90, New Orleans: ACM Press, Addison-Wesley, 1991.

9. Jones, S.D. & Furner, S.M. The construction of audio icons and information cues for human-computer dialogues. In *Contemporary Ergonomics: Proceedings of the Ergonomics Society's 1989 Annual Conference*. 1989. Reading:

10. Monk, A. Mode Errors: A user-centered analysis and some preventative measures using keying-contingent sound. *International Journal of Man-Machine Studies* **24** (1986): 313-327.

11. Moore, B.C. *An Introduction to the Psychology of Hearing*. London: Academic Press. 1989, 1-10.

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

13. Sellen, A., Kurtenbach, G. & Buxton, W. The prevention of mode errors through sensory feedback. *Human Computer Interaction* **7** (1992): 141-164.

14. Sellen, A.J., Kurtenbach, G.P. & Buxton, W. The role of visual and kinesthetic feedback in the prevention of mode errors. In *Interact'90*. edited by Diaper, D., 667-673, Elsevier Science Publishers: Holland, 1990.

15. Sumikawa, D., Blattner, M., Joy, K. & Greenberg, R. *Guidelines for the syntactic design of audio cues in computer interfaces* (Technical Report No. UCRL 92925). Lawrence Livermore National Laboratory, 1986.

16. Sumikawa, D.A. *Guidelines for the integration of audio cues into computer user interfaces* (Technical Report No. UCRL 53656). Lawrence Livermore National Laboratory, 1985.