Brewster, S.A. and Wright, P.C. and Edwards, A.D.N. (1993) An evaluation of earcons for use in auditory human-computer interfaces. In, Ashlund, S., Eds. *Conference on Human Factors in Computing Systems, 24-29 April 1993*, pages pp. 222-227, Amsterdam, Netherlands.

An Evaluation of Earcons for Use in Auditory Human-Computer Interfaces

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

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

ABSTRACT

An 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. A second experiment was then carried out which improved upon some of the weaknesses shown up in Experiment 1 to give a significant improvement in recognition. From the results of these experiments some guidelines were drawn up for use in the creation of earcons. Earcons have been shown to be an effective method for communicating information in a human-computer interface.

KEYWORDS

Auditory interfaces, earcons, sonification

INTRODUCTION

The use of non-speech audio at the user-interface is becoming increasingly popular due to the potential benefits it offers. It can be used to present information otherwise unavailable on a visual display for example mode information [9] or information that is hard to discern visually, such as multi-dimensional numerical data [4]. 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 receive through the visual channel. It makes use of the auditory system which is powerful but under-utilised in most current interfaces. There is also psychological evidence to suggest that sharing information across different sensory modalities can actually improve task performance (see [2] section 3.1). Having redundant information gives the user two chances of identifying the data; if they cannot remember what an icon looks like they may be able to remember what it sounds like. The foveal area of the retina (the part of greatest acuity) subtends an angle of only two degrees around the point of fixation [12]. Sound, on the other hand, can be heard from 360 degrees without the need to concentrate on an output device, thus providing greater flexibility. Sound is also good at capturing a user's attention whilst they are performing another task. Finally, the graphical interfaces used on many modern computers make them inaccessible to visually disabled users. Providing information in an auditory form could generally help solve this problem and allow visually disabled users the same facilities as the sighted.

This evaluation is part of a research project looking at the best ways to integrate audio and graphical interfaces. 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 communicating information.

One major question that must be answered when creating an auditory interface is: What sounds should be used? Brewster [2] outlines some of the different systems available. Gaver's *auditory icons* have been used in several systems, such as the SonicFinder [5], SharedARK [6] and ARKola [7]. These use environmental sounds that have a semantic link with the object they represent. They have been shown to be an effective form of presenting information in sound. One other important, and as yet untested, method of presenting auditory information is the system of *earcons* [1, 13, 14]. 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.

One of the most powerful features 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'.

As yet, no formal experiments have been conducted to see if earcons are an effective means of communicating information using sound. Jones & Furner [8] 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. This paper seeks to discover how well earcons can be recalled and recognized. It does not try to suggest uses for earcons in the interface. 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 (the study of the perception of sound) suggests that complex musical instrument timbres may be more effective [10]. The second experiment uses the results of the first to create new earcons to overcome some of the difficulties that came to light. Some guidelines are then put forward for use in the creation of earcons.

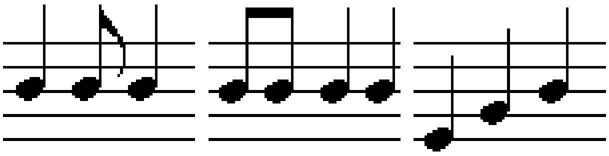

Figure 1: Rhythms and pitch structures for Folder, File and Open used in Experiment 1

EXPERIMENT 1

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 [3] rhythm is one of the most powerful methods for differentiating sound sources. Figure 1 gives some examples of the rhythms and pitch structures used for the different types of objects in the experiment. The experiment also attempted to find out how well subjects could identify earcons individually and when played together in sequence.

Three sets of sounds were created:

1. 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.

2. 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. This set also had rhythm information.

3. 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 previous two groups.

The sounds for all sets were all played through a Yamaha DMP 11 mixer controlled by an Apple Macintosh and presented using external loudspeakers.

Experimental Design

Three groups of twelve subjects were used. Half of the subjects in each group were musically trained. Each of the three groups heard different sound stimuli. The musical group heard set 1 described in the previous section. The simple group heard set 2 and the control group heard set 3. 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. Finally, the subjects were required to listen to two earcons played in sequence and give information about both sounds that were heard.

Phase I

The subjects were presented with the screen shown in Figure 2. 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 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 information was given so types were also differentiated by pitch. The icons were played one-at-a-time in sequence to the subjects for them to learn. The whole set of icons was played three times.

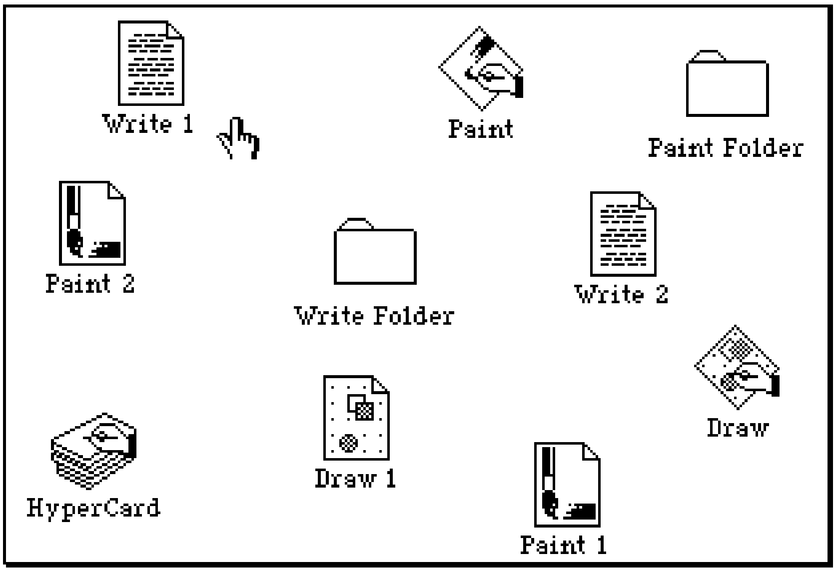

Figure 2: The Phase I icon screen

When testing the subjects the screen was cleared and some of the earcons were played back. The subject had to supply what information they could about type, family and number of the file of the earcon they heard. When scoring, a mark was given for each correct piece of information supplied.

Phase II

This time earcons were created for menus. 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 3. The subjects were tested in the same way as before but this time had to supply information about menu and item.

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 the subjects to see if they could remember the original set of earcons after having learned another 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

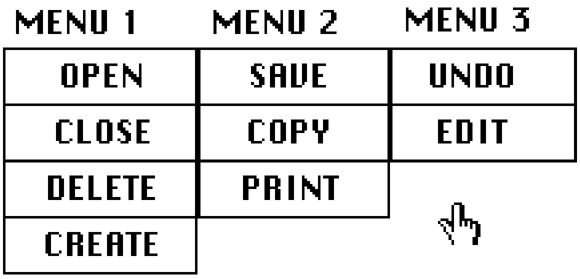

Figure 3: The Phase II menu screen

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.

Results and Discussion

From Figure 4 it can be seen that overall the musical earcons came out best in each phase. Unfortunately this difference was not statistically significant.

Phase I: 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). Therefore, the rhythms used did not give any better performance over a straight burst of sound for telling the types apart.

Phase II: The overall scores were significantly better than those for phase I. An ANOVA on the overall scores showed a significant effect ($F(2,33)=5.182$, $p<.011$). This suggests that the new rhythms were much more effective (as the timbres were similar). The simple and musical groups performed similarly which was to be expected as both used the same rhythms. 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 rhythms are used 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: The scores were not significantly different to those in phase I indicating that subjects managed to remember the earcons even after doing another very similar task. This implies that, after only a short period of learning time, subjects could remember the earcons. This has important implications as it seems that subjects will remember earcons, perhaps even as well as icons. Tests could be carried out to see if subjects can remember the earcons after longer periods of time.

Phase IV: A within groups t-test showed that, in the musical group, the menu/item combination was

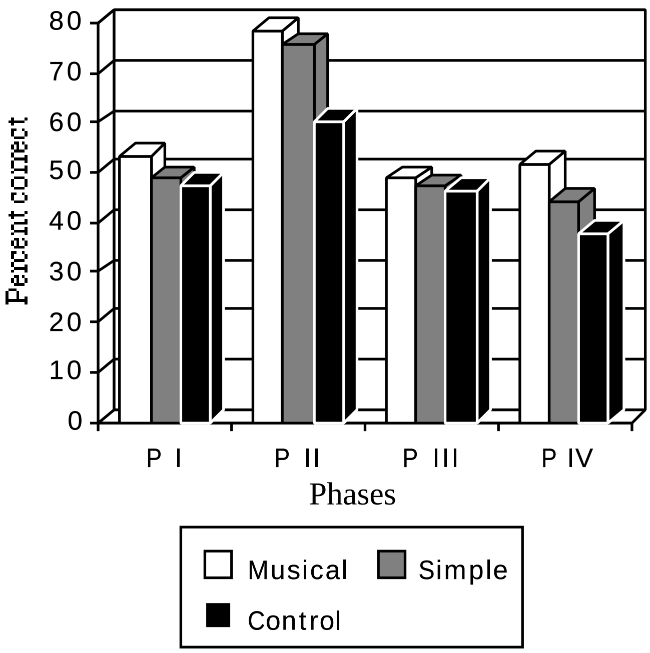

Figure 4: Breakdown of overall scores per phase for Experiment 1

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

Summary of Experiment 1: Some general conclusions can be drawn from this first experiment. It seems that earcons are better than unstructured bursts of sound at communicating information under certain circumstances. The issue of how this advantage can be increased needs further examination. Similarly, the musical timbres come out better than the simple tones but often by only small amounts. Further work is needed to make them more effective. The results also indicate 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 needs to be ascertained. Finally, the difficulties in recognising combined earcons must be reduced so that higher scores can 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 which achieved much higher recognition rates.

Sounds Used

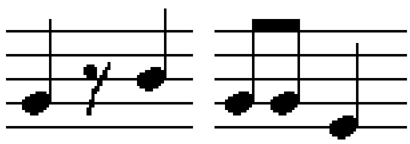

Figure 5: New rhythms for Folder and File in Experiment 2 (cf. Figure 1)

The sounds were redesigned so that there were more gross differences between each earcon. 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 5 shows the new rhythms and pitch structures for folder and file.

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. 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 [4] 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 [10] gives a detailed overview of the field of psychoacoustics and Patterson [11] includes 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. The first files were an octave below middle C (130Hz) and the second files 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 first files were three octaves above middle C (1046Hz) and the second files at middle C. These were now well within Patterson's ranges.

In response to informal user comments from Experiment 1 a delay was put between the two earcons. Subjects had complained that they could not tell where one earcon stopped and the other started. A 0.1 second delay was used.

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.

Results and Discussion

As can be seen from Figure 6, the new sounds performed

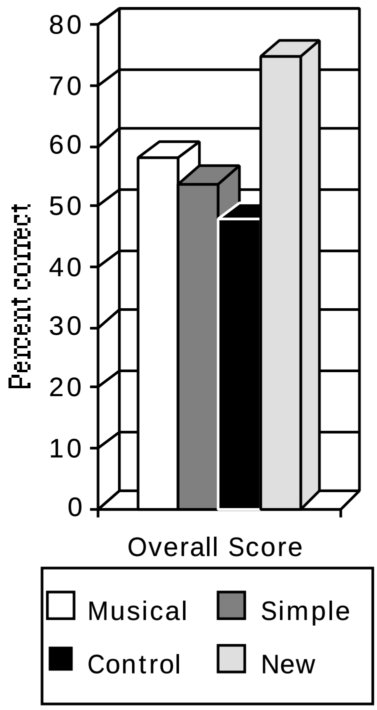

Figure 6: Percentage of overall scores with Experiment 2

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 we can see that the level of recognition in phases I and III has been raised to that of phase II (see Figure 7). However, t-tests revealed that 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 overall recognition rate in phase I was increased because of a very significantly better type score (differentiated by rhythm etc.) in the new group ($F(1,22)= 26.677$, $p<0.05$). The scores increased from 49.1% in the musical group to 86.6%. This seems to indicate that the new rhythms were effective and very easily recognised.

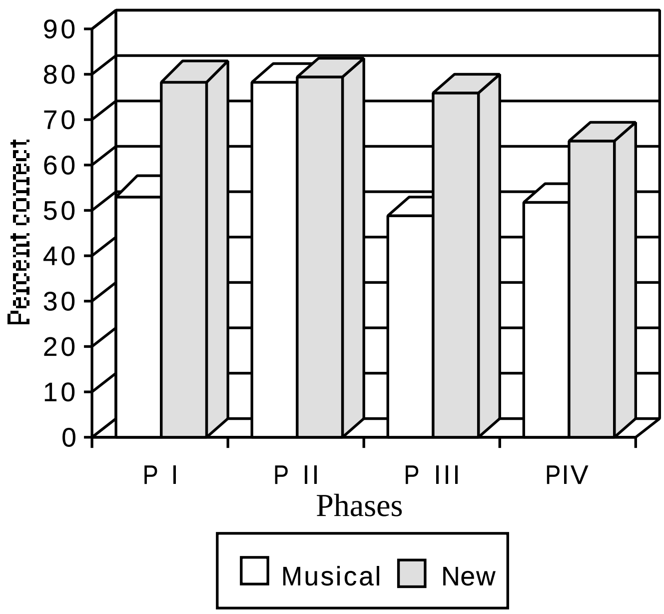

Figure 7: Breakdown of scores per phase with Experiment 2

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.

In phase IV 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 type and family scores were both significantly better than in the musical group (type: $F(1,22)=9.135$, $p<0.05$, family: $F(1,22)= 4.989$, $p<0.05$). This shows that bringing the icon sound scores up to the level of the menus increased the score in phase IV but there still seems to be a problem when combining two earcons.

The new use of pitch also seems to have been effective. In phase I the new group got significantly better recognition of the file earcons than the musical group ($F(1,22)=4.829$, $p<0.05$). This indicates that using the higher pitches and greater difference in register made it easier for subjects to differentiate one from another.

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 been that in phases I, II and III only one timbre was heard and so it was clear to which group of earcons it belong (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

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 ? The earcons in the musical group from Experiment 1 were, on the whole, no better recognised by the musicians than the non-musicians. This means that non-musical user of a system involving earcons would have no more difficulties than a musician. Problems occurred in the other two groups of Experiment 1. Musicians were better at types and families in the simple group and families, menus and items in the control group. The results also show that there is 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 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 in the creation of earcons. These should be used along with the more general guidelines given in [13, 14]. 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 synthesised 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. Three or more 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 (1982) 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 sec.
• *Intensity*: Although intensity was not examined in this test some suggested ranges (from [11]) 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. If any sound is too loud it may become annoying to the user and dominate the others. If any sound is too quiet then it may 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 research has been done to test the speed of presentation of earcons. The earcons took between 1 and 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 rate of presentation obtainable at which the earcons would maintain their high rates of recognition.

CONCLUSIONS

The results indicate that earcons are an effective means of communication. The work shown has demonstrated that earcons are better for presenting information than unstructured bursts of sound. 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. Developers can create interfaces that use them 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.

This work is supported by SERC studentship 90310837.

REFERENCES

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

2. Brewster, S.A. (1992). *Providing a model for the use of sound in user interfaces.* University of York Technical Report YCS 169, York, UK.

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

4. Frysinger, S.P. (1990). Applied research in auditory data representation. In D. Farrell (Ed.) *Extracting meaning from complex data: processing, display, interaction*, Proceeding of the SPIE, 1259, pp 130-139.

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

6. Gaver, W. & Smith, R. (1990). Auditory icons in large-scale collaborative environments. In D. Diaper *et al*. (Eds.) *Human Computer Interaction - INTERACT '90*, Elsevier Science Publishers B.V. (North Holland), pp 735-740.

7. Gaver, W., Smith, R. & O'Shea, T. (1991). Effective sounds in complex systems: the ARKola simulation. *CHI'91 Conference proceedings, Human Factors in Computing Systems, "Reaching through technology"* , New Orleans, pp 85-90, ACM Press: Addison-Wesley.

8. Jones, S.D. & Furner, S.M. (1989). The construction of audio icons and information cues for human-computer dialogues. *Contemporary Ergonomics, Proceedings of the Ergonomics Society's 1989 Annual Conference*, T. Megaw (Ed.)

9. Monk, A. (1986). Mode Errors: A user-centered analysis and some preventative measures using keying-contingent sound. *IJMMS*, 24, pp 313-327.

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

11. Patterson, R.D. (1982). *Guidelines for auditory warning systems on civil aircraft*, C.A.A. Paper 82017, Civil Aviation Authority, London.

12. Rayner, K. & Pollatsek, A. (1989). *The Psychology of Reading*, pp 9-10. Englewood Cliffs, New Jersey: Prentice-Hall International, Inc.

13. Sumikawa, D. (1985). *Guidelines for the integration of audio cues into computer user interfaces*, Lawrence Livermore National Laboratory Technical Report, UCRL 53656.

14. Sumikawa, D., Blattner, M., Joy, K. & Greenberg, R. (1986). *Guidelines for the syntactic design of audio cues in computer interfaces*, Lawrence Livermore National Laboratory Technical Report, UCRL 92925.