Representing Complex Hierarchies with Earcons

Stephen A. Brewster[1], Veli-Pekka Raty[2] & Atte Kortekangas[3]

[1]SINTEF DELAB,
N-7034 Trondheim,
Norway.
Tel:+47 73 59 42 52
stephen.brewster@delab.sintef.no

[2]VTT Information
Technology,
Kanslerinkatu 12 B,
Tampere, FIN-33101,
Finland
Tel: +358 31 316 3328
veli-pekka.raty@vtt.fi

[3]VTT Information
Technology,
Tekniikantie 4B,
Espoo, FIN-02044 VTT,
Finland.
Tel: +358 0 456 4311
atte.kortekangas@vtt.fi

ABSTRACT

This paper describes an experiment to discover if structured audio messages called *earcons* could provide navigational cues in a complex menu hierarchy. A hierarchy of 25 nodes and four levels was created with an earcon for each node. Rules were designed for the creation of the earcon at each node. The results showed that participants could recall over 80% of the earcons they heard, indicating that they are a good method of communicating hierarchy information. Participants were also tested to see if, by using the rules they had learned for the other earcons, they could identify where a previously unheard earcon would fit in the hierarchy. The results showed that they were able to do this very reliably. This indicates that earcons are a robust method of communicating complex hierarchy information in sound.

INTRODUCTION

This paper describes an experiment to investigate the power of non-speech audio for conveying navigational cues in a hierarchical structure. In some situations graphical feedback cannot be used to provide these cues. In completely auditory interactions such as telephone-based interfaces or interfaces for the visually disabled it is obviously impossible to use graphical cues. In other systems graphical feedback is available but the display may already be completely occupied by important information that extra graphical cues would hide. An example of this is an interface for people with speaking difficulties who need to access a graphical library of pictographic images to create the messages they want to communicate. This paper suggests that non-speech sounds can be used to provide the extra cues in both of these situations.

There is a growing body of research which indicates that the addition of non-speech sounds to human-computer interfaces can improve performance and increase usability [2, 4, 12]. Non-speech sound is an important means of communication in the everyday world around us and the benefits it offers should be taken advantage of at the interface It allows greater communication between the computer and the user. Sound is omni-directional so the user does not have to concentrate on a particular part of the display to preceive it. In fact he/she does not even have to be looking at the display at all. Sound is also attention grabbing and can be effectively used to indicate problems to users. It can work alongside synthetic speech in purely auditory interfaces or be integrated with graphical feedback. Often in graphical interfaces more and more information is displayed on-screen. This can result in overload and impotrant information may be missed. One way to overcome this problem is to use sound. Important visual information can be displayed on the screen and other information can be displayed in sound, reducing the overload of the visual sense. Brewster [4] showed that by adding sound to a graphical interface both the time taken to complete certain tasks and the time taken to recover from errors could be reduced.

Telephone-based interfaces

One important reason for using non-speech sound is to represent menu hierarchies in interfaces where visual feedback is not possible, for example telephone-based interfaces (phone banking) or interfaces for the visually disabled. In a telephone-based interface a user might call the bank and navigate through a hierarchy of voice menus to find the service required. One problem is that the user can get lost in the hierarchy. The communication channel is very limited so that little feedback can be given to the user about his/her current location. The more navigation information that is given in synthetic speech, the more it obstructs the information the user is trying to access.

Rosson [15] investigated such a hierarchical phone-based interface to give travel/visitor information. A user could call the system and move through the hierarchy to find information about the city of Austin, Texas such as the addresses and phone numbers of shops or restaurants. She describes one common problem with such systems (p 251): “It is important to note that the information needed to convey position in the hierarchy was implicit in the content of the utterances users heard”. Feedback confirming that one had moved from the top to the middle level of the hierarchy was available only by understanding a category/sub-category relationship. After hearing “Restaurants” and making a ‘Down’ move, the user might hear “Chinese” and would have to make the inference that a move to the lower level of the hierarchy had been made. She suggests:

> “This may have been the source of many of the problems, because the most common source of error appeared to arise from users’ failure to move laterally to an item once they had reached the appropriate level”.

Rosson suggested that one way to solve the problem could be to give extra speech feedback. For example “You have moved to the next item in the Chinese Restaurant list. It is…”. She suggested that this would make the interface appear slow and clumsy. This extra feedback could also be longer than the information being retrieved and so obscure it. Such feedback was rejected for this very reason by Stevens and colleagues [16, 17] when designing navigation cues in a system to provide synthetic speech access to mathematics. Rosson suggested another method: “More attractive possibilities are to increase the information implicit in the utterance itself, by systematically varying the syntax of the utterances at different levels, or by assigning a different “voice” to each level”.

There are problems with these two methods. Varying the syntax could make the voice messages harder to understand. It might result in complex syntax that again obscures the information being communicated. There are also only a few voices that could be used to indicate the different levels. The low quality of the telephone will reduce the listeners ability to differentiate many different voices. One final problem with this method is that of image. Perhaps a wider variety of voices could be created by using children’s or cartoon voices but these might not reflect the right image for the company with the telephone interface.

We suggest an alternative solution: Structured non-speech audio messages. These would provide a hierarchical system of sounds that could be used to represent the menu hierarchy. The sounds would play continuously in the background at each level giving information about where users were. They could listen to the current sound and from it work out their location in the hierarchy. This is a similar approach to that taken by Stevens and colleagues [16, 17] in the Mathtalk system. Here, non-speech sounds give the listener information about their location in a mathematical structure. They do this without interfering with the synthetic voice presenting the mathematics. The cues are also much shorter than an equivalent voice message. Speech and non-speech sounds are different media in the same way as text and graphics [14]. In the same way they can be used together to provide information. If they are carefully designed then they do not conflict.

Navigation in a communicator device

A second reason for investigating the use of sound to represent hierarchies is to present navigational cues in an interface for people who are speech-motor and/or language-cognitive impaired for TIDE ACCESS project 1001. The aim of this project is to create a mobile communication device. People will use the device to create messages they want to communicate and then play those messages via synthetic speech. These types of users often use pictographic languages (for example, Bliss) to communicate. The pictures representing words or actions and can be combined to create complex messages.

An experienced user of such a language may need access to a large number of symbols if he/she has a wide vocabulary. It is impossible to display all of the pictures on the screen at the same time. One way around this problem is to use a hierarchy of symbols. For example, at the top level there may be a picture representing a person. If this picture is selected then pictures of a man, woman and child would be displayed. If the user then selected ‘child’ different pictures representing children could be shown and the user would choose the one required. Thus a hierarchy of symbols is formed and this can lead to problems of getting lost similar to those described above.

Graphical feedback could be used to give navigational information, for example a map could show the current position in the hierarchy. However, this would take up valuable screen space and would also require the user to look at the map when he/she really wanted to look at the pictographic symbols. An alternative would be to use different colours to show the different levels. Unfortunately, colour is often used to show other groupings within the symbols such as nouns/verbs/adjectives. We suggest the use of sound. Non-speech sounds can be used to give information about one's location in the hierarchy without taking up screen-space needed for the pictographic symbols. The sounds would play continuously, but quietly, in the background. A user would be able to listen to the sounds and from them work out where he/she was in the hierarchy.

This is not only a problem in interfaces in interfaces for the disabled. Visual displays can only hold so much information. If an interface designer tries to display more and more information then some of it will not fit on the screen without hiding information already there. Blattner, Papp & Glinert [2] discuss this problem with computerised maps. They suggest that information can become hidden because of visual clutter: Only so much information can be displayed before the underlying map is obscured. If additional information must be displayed on a map, space must be allocated for it and eventually a saturation point will be reached where interference with the existing graphics and text cancels out any benefit from adding more information. Blattner *et al.* suggested that sound could be used to avoid these problems. Brewster [4] has also considered in detail the problems that occur due to information being hidden.

This paper describes an experiment to test non-speech sounds to see if they could convey the hierarchy information in the examples described above. If they cannot then the new approach suggested here would be ineffective.

EARCONS

The non-speech sounds to be used for this investigation are based around structured audio messages called *Earcons* [3, 4, 18]. Earcons are abstract, synthetic tones that can be used in structured combinations to create sound messages to represent parts of an interface. Detailed investigations of earcons by Brewster, Wright & Edwards [6, 7] showed that they are an effective means of communicating information in sound.

Earcons are constructed from motives. These are short rhythmic sequences that can be combined in different ways. The simplest method of combination is concatenation to produce *compound earcons*. By using more complex manipulations of the parameters of sound (timbre, register, intensity, pitch and rhythm) *hierarchical earcons* can be created [3]. These allow the representation of hierarchical structures. Sumikawa ([19], p 64) suggested three ways in which motives can be manipulated to create hierarchical earcons:

- *Repetition*: Exact restatement of a preceding motive and its parameters.
- *Variation*: Altering one or more of the variable parameters from the preceding motive (for example, rhythm, pitch, timbre, register or dynamics).
- *Contrast*: A decided difference in the pitch and/or rhythmic content from the preceding motive.

Research by Brewster, Wright & Edwards [6, 7] showed that some of the parameters indicated by Blattner were not effective. They suggested that, for example, pitch alone should not be used to differentiate earcons. One of their main findings was that using musical instrument timbres proved to be very effective at aiding recall of earcons. Guidelines for the creation of effective earcons were proposed by Brewster *et al.* [8]. These were used in the design of the earcons described in this paper.

Figure 1 shows a hierarchy of earcons using these types of manipulations (based on the guidelines from [8]). Each earcon is a node on a tree and inherits all the properties of the earcons above it. The different levels are created by manipulating the parameters of earcons (for example, rhythm, pitch, timbre, register, tempo, stereo position, effects and dynamics). In the diagram the top level of the tree is a neutral earcon representing a fictitious family of errors. It has a flute timbre (a 'colourless' instrument), a middle register and a central stereo position. The structure of the earcon from level one is inherited by level two and then changed. At level two the sound is still continuous but non-neutral timbres are used (in the figure organ and violin). Register is changed so that it matches a conventional musical layout (low register on the left, high on the right) and stereo position reflects the layout of the hierarchy, for example the node on the left has a left stereo position. At level three a rhythm is added to the earcon from level two to create a sound for a particular error. The rhythm is based on the timbre, register and stereo position from the level above. Other levels can be created by using other parameters such as tempo or effects.

An alternative method of presenting information in sound is Gaver's *Auditory Icons* [9, 10]. These are based around sounds sampled from the everyday environment (such as metallic or wooden sounds) rather than the musical sounds of earcons. For the creation of hierarchies

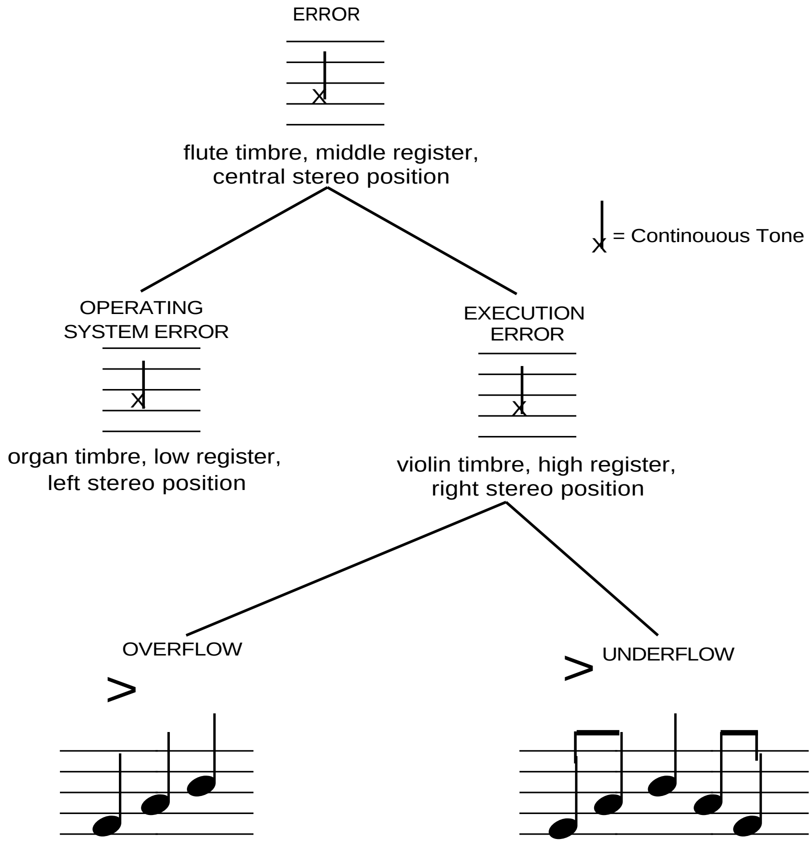

Figure 1: *A hierarchy of family earcons representing errors.*

earcons have the advantage that the sounds can be varied in a systematic way (as shown in the description above). Auditory icons are based on natural sounds and so cannot be as easily manipulated (for example, if an auditory icon is manipulated too much it may no longer sound like its natural equivalent). Gaver is working on this problem [11] to allow the controlled variation of auditory icons. However, at the present time it is not possible so we decidd to use earcons.

Previous attempts to use earcons to present hierarchy information

Barfield, Rosenberg & Levasseur [1] carried out experiments where they used earcons to aid navigation through a menu hierarchy. They say (p 102): "…the following study was done to determine if using sound to represent depth within the menu structure would assist users in recalling the level of a particular menu item". The earcons they used were very simple, just decreasing in pitch as a participant moved down the hierarchy. The sounds lasted half a second. They describe them thus (p 104): "…the tones were played with a harpsichord sound and started in the fifth octave of E corresponding to the main or top level of the menu and descended through B of the fourth octave".

These sounds did not fully exploit all the advantages offered by earcons (for example, they used neither rhythm nor timbre and did not exploit the highly structured nature of earcons) and did not improve user performance in the experimental task. Using pitch alone to differentiate the items was shown to be ineffective in the experiments of Brewster *et al.* If better earcons had been designed then advantages may have been found.

Brewster, Wright & Edwards and Brewster [4, 6, 7] also tested the ability of earcons to present hierarchical information in sound. In two detailed experiments they showed that, with careful design of the earcons, hierarchy information could be presented effectively. They also showed that some of the manipulations described by Blattner were too subtle to be noticed by listeners.

They used earcons to represent a small hierarchy of files, folders and applications and manipulated timbre, register, rhythm and pitch to create their hierarchy. Their results showed that 80% recall rates could be achieved for hierarchical earcons even with non-optimal training. This work indicated that earcons could be used to represent hierarchies. However, the hierarchy used by Brewster *et al.* was simple; only three levels and ten nodes. A bigger hierarchy must be tested before earcons can safely be used to represent complex hierarchies in sound.

THE EXPERIMENT

The aim of the experiment described here was to discover if a complex hierarchy could be represented by earcons. Figure 2 shows the hierarchy used. It had 25 nodes on four levels with four missing nodes on Level four (two of which are marked as A and B in Figure 2).

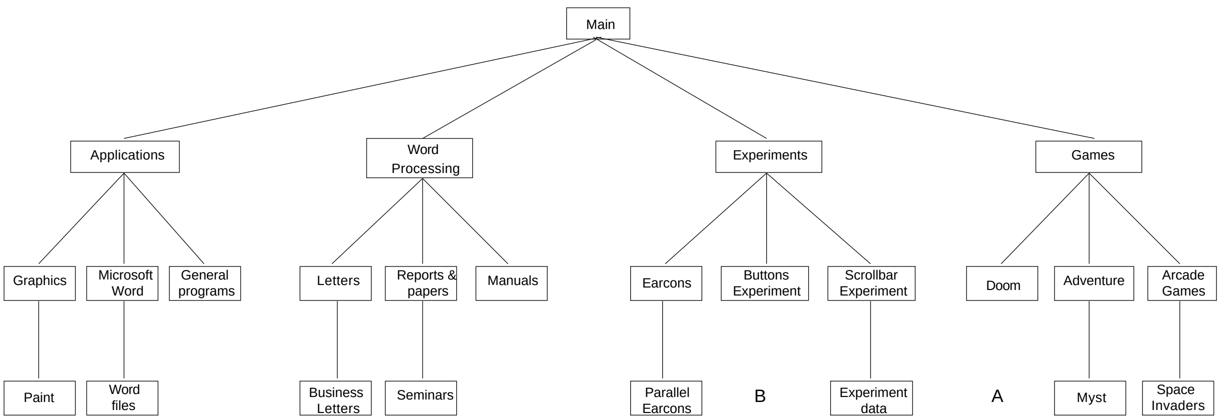

Figure 2: *The file-system hierarchy used in the experiment. A and B show the two new earcons presented to participants during testing.*

Hypotheses

The main hypothesis was that participants should be able to recall the position of a node in the hierarchy by the information contained in an earcon. If this was correct then high overall rates of recall would be expected.

Participants should also be able to listen to an earcon and position it in the hierarchy even if they have not heard it before by using the rules from which the earcons were constructed. This should be demonstrated by high rates of recognition when participants were presented with new earcons.

Participants

Twelve volunteer participants were used, both men and women. They were a mixture of students at Helsinki University of Technology and members of staff at VTT Information Technology. They were all familiar with computers and computer file systems.

Sounds used

The earcons were designed using the guidelines proposed by Brewster [8]. As suggested, the earcons were designed using timbre, register and spatial location for the main sub-groups in the hierarchy. The sounds were all played from HyperCard via MIDI on a Yamaha TG100 sound synthesiser and presented to participants via loudspeakers. The sounds used at each level of the hierarchy will now be described:

Level 1: For the top level of the hierarchy (‘Main’ in Figure 2) a constant sound with a flute timbre was used (see Figure 3). It had a central spatial location and a pitch of D_3 (261Hz). A flute timbre was used at it is a pure sound close to a ‘timbreless’ sinewave.

Level 2: At this level each family had a separate timbre, register and spatial location. Figure 3 shows the timbre, spatial location and register used for the earcons. Register was lowest on the left and highest on the right following the conventional musical pattern (for example, a piano keyboard). The stereo position of the earcons also moved from left to right mirroring their position in the hierarchy (see Figure 2).

The continuous sound was inherited from Level 1 but the instrument was changed as was its pitch and its stereo position. Three parameters were used so that if the listener could not remember which instrument went with which node he/she could still use register or stereo position. Register was used in conjunction with the other two parameters as the results of Brewster *et al.* [4] showed it was not effective on its own.

Nodes	Timbre	Stereo position	Register
Main	Flute	Centre	D_3
Applications	Electric organ	Far left	C_4
Word Processing	Violin	Centre left	C_3
Experiments	Drum/synthesiser	Centre right	C_2
Games	Trumpet	Far right	C_1

Figure 3*: The timbre, spatial location and register for Levels 1 and 2 of the hierarchy.*

Level 3: At this level rhythm was used to differentiate the nodes. Each left node had one rhythm, each centre node another rhythm and each right node another. Figure 4 shows the rhythms used. From Figure 2 'Graphics', 'Letters', 'Earcons' and 'Doom' all had the left node rhythm, 'Microsoft Word', 'Reports & Papers', 'Buttons Experiment' and 'Adventure' were centre nodes and 'General Programs', 'Manuals', 'Scrollbar Experiment' and 'Arcade Games' were right nodes. Each of these rhythmic groups repeated continuously once every 2.5 seconds. As Figure 4 shows, the first note in each group was accented. The last note of each group was also lengthened slightly. These two help make each group into a complete rhythmic unit (see [8]).

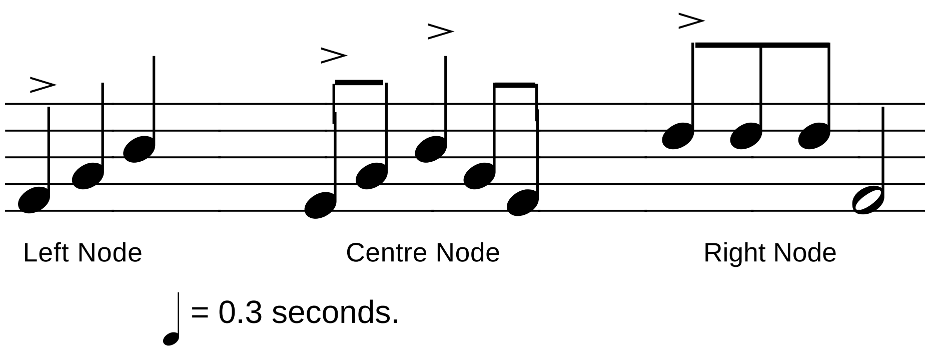

Figure 4*: The rhythms used for Levels 3 and 4 of the hierarchy.*

At this level the earcons inherited timbre, spatial location and register from Level 2. This meant, for example, that 'Graphics' used the left node rhythm described in Figure 4 and it was played with an electric organ timbre, on the left side of the stereo space and in the register of C_4. 'Letters' used the same rhythm but, in this case, the timbre was a violin, stereo position was centre left and the register was C_3.

Level 4: At this level a faster tempo was used to differentiate the items. The rhythmic units from Figure 4 now repeated once every second. In addition to this the effects reverb and chorus were applied to all of the earcons. These gave the earcons a much fuller sound. This time rhythm was inherited from Level 3. Each of the nodes in Level 4 used the same rhythm as its parent node but the earcons were repeated more frequently.

Experimental design and procedure

As shown in Figure 2 , the hierarchy was based on a computer file system. This was chosen as all of the participants were familiar with such hierarchies. This was an experiment to test the use of earcons to represent a hierarchy and a file system was a convenient hierarchy to use, this paper does not suggest that each directory and sub-directory in a real system should have a sound.

The hierarchy was constructed in a HyperCard stack. Figure 5 shows the screen of the top level of the hierarchy ('Main'). Each of the boxes in Figure 2 was a card in the stack. Buttons were provided for going up and down levels in the hierarchy and also for going left and right across the same level. As soon as a card was selected its sound started to play, it continued until another card was selected.

Training

The training was in two parts. In the first part the experimenter showed the participant each of nodes of the hierarchy in turn and played the associated earcon. The structure of the earcons at each level was fully explained. Both Brewster and Lucas [4, 13] have shown that explanation of the structure of earcons helps recall.

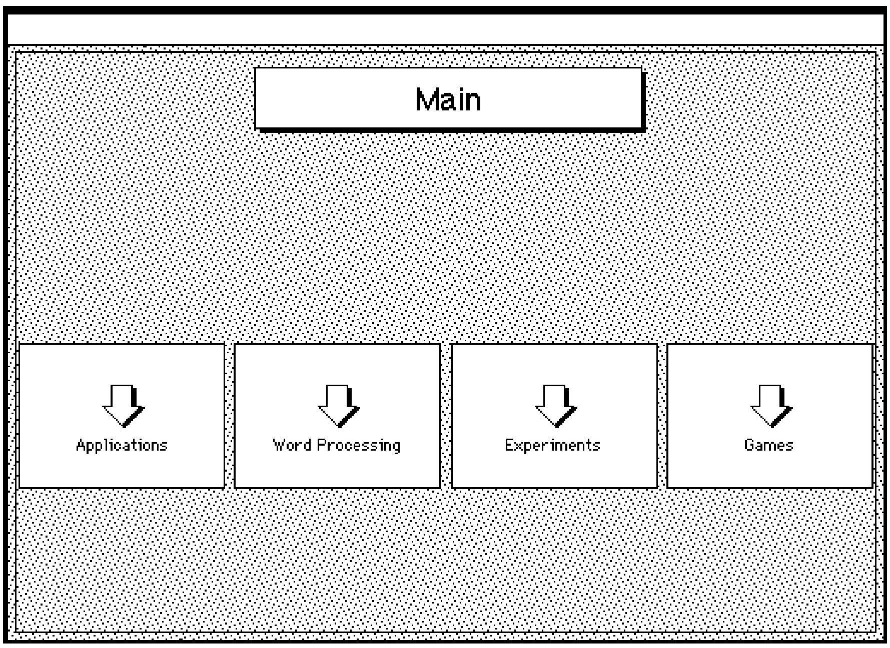

***Figure 5**: The top level of the hierarchy.*

In the second part of the training participants were given five minutes to learn the earcons by themselves. They could move around in the hierarchy using the buttons provided and listen to the sounds. The idea behind this type of training was that it would be similar to the kind users of a telephone interface might receive. They might get some spoken instructions describing the sounds in the system and then play with them themselves for a short time. For the experiment, longer training could have been given but this would then have been less like the actual application of such sounds.

During the training participants could look at a map of the hierarchy (similar to Figure 2 above). The aim of the experiment was not to test the participants' ability to learn the hierarchy but to test their ability to learn the earcons. Instructions were read from a prepared script.

Testing

The participants heard fourteen earcons during testing. These were randomly selected from all of the sounds in the hierarchy. The same earcons were presented to each of the participants. Twelve of the sounds were ones that participants had heard during the training. The last two earcons they heard were new ones (marked A and B in Figure 2). These were earcons for gaps in the hierarchy and were constructed using the same rules as the other earcons. Figure 6 shows which earcons represented which question.

Question	Earcon
Q1	Word processing
Q2	Space invaders
Q3	Experiments
Q4	Paint
Q5	Games
Q6	Doom
Q7	Word files
Q8	Parallel earcons
Q9	Graphics
Q10	Reports & papers
Q11	Business letters
Q12	Microsoft word
Q13	A
Q14	B

***Figure 6**: The earcons for each of the questions.*

During testing participants saw the screen shown in Figure 7. The could click on the button 'Play Sound' to hear three seconds of an earcon. They were allowed to do this only twice. They then had to chose where the earcon fitted into the hierarchy. The hierarchy was represented on the screen with buttons at each of the nodes. The participants had to click on the node they thought the earcon represented. None of the names of the nodes were included in the picture of the hierarchy to avoid any help they might have provided.

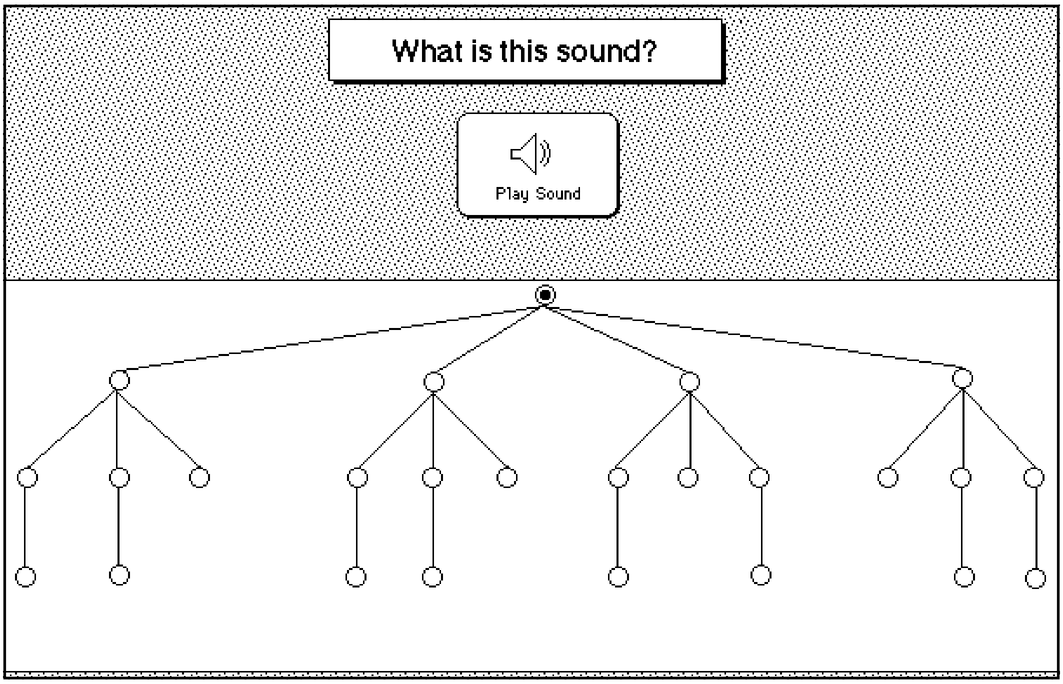

Figure 7: *The participant testing screen.*

RESULTS

Overall

The overall recall rate of the earcons was high: 81.5% of the earcons were correctly recalled. Figure 8 shows the percentage of correct answers for each question. An analysis was undertaken to find out if any of the questions were less well recalled than any others. A one-factor ANOVA was performed on the scores per question and it showed an effect ($F(13,154)=2.13$, $p=0.01$). In order to find out where the effect was (and so which were the worst-recalled earcons) *post hoc* Tukey HSD tests comparing each question were performed (see Figure 9). The three worst recalled earcons were from questions 2, 4 and 11. They were all from Level 4 of the hierarchy. The nodes were: ‘Space Invaders’, ‘Paint’ and ‘Business Letters’. However, the other Level 4 nodes ‘Word Files’ and ‘Parallel Earcons’ were significantly better recalled than these. ‘Paint’ was recalled worst of all.

Recall of components

Each of the participants’ answers were broken down to find where the problems occurred. There were three mistakes that could be made. Participants could mistake the family of an earcon: Whether it was from ‘Applications’, ‘Word Processing’, ‘Experiments’ or ‘Games’. Here they were mistaking the main sub-tree to which an earcon belonged. This was a fault with recall of timbre/register/stereo position. Participants could mistake the node an earcon referred to: Whether it was a left node, a centre node or a right node. This was a fault in the recall of rhythm. Finally, participants could mistake the level of an earcon: Whether it was from Level 1, 2, 3 or 4 of the hierarchy. This was a more complex mistake that was linked to the recall of both family and node. A mistake between Levels 1 and 2 indicated that participants had not recalled the timbre/register/stereo position information. correctly Between Levels 2 and 3 rhythm differences had not been recalled and between Levels 3 and 4 tempo/echo/chorus information had not been recalled.

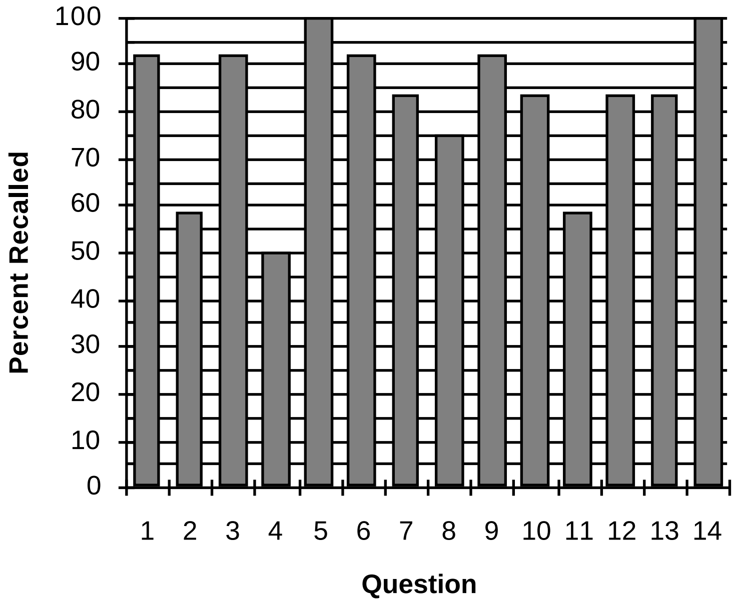

Figure 8*: Recall rates for each of the 14 questions.*

Questions	Question 4	Questions 2 & 11
q1	Q(154)=5, p=0.01	Q(154)=4, p=0.01
q2	Q(154)=1, p=0.01	--
q3	Q(154)=5, p=0.01	Q(154)=4, p=0.01
q4	--	--
q5	Q(154)=6, p=0.01	Q(154)=5, p=0.01
q6	Q(154)=5, p=0.01	Q(154)=4, p=0.01
q7	Q(154)=4, p=0.01	Q(154)=3, p=0.01
q8	Q(154)=3, p=0.01	Q(154)=2, p=0.01
q9	Q(154)=5, p=0.01	Q(154)=4, p=0.01
q10	Q(154)=4, p=0.01	Q(154)=3, p=0.01
q11	Q(154)=1, p=0.01	--
q12	Q(154)=4, p=0.01	Q(154)=3, p=0.01
q13	Q(154)=4, p=0.01	Q(154)=3, p=0.01
q14	Q(154)=6, p=0.01	Q(154)=5, p=0.01

Figure 9*: Tukey HSD results comparing the three worst questions. Question 4 was significantly worse than any other question. Questions 2 & 11 were worse than all except question 4.*

From the overall data obtained the scores were broken into three. If a participant got a question completely right he/she got three marks. If he/she got two parts right (for example, family and node, or node and level) then two marks were awarded. If only one part was correct then one mark was given. From this analysis it was possible to see where mistakes occurred. This was a similar analysis to that undertaken by Brewster [5] in previous earcon experiments. Figure 10 shows the results.

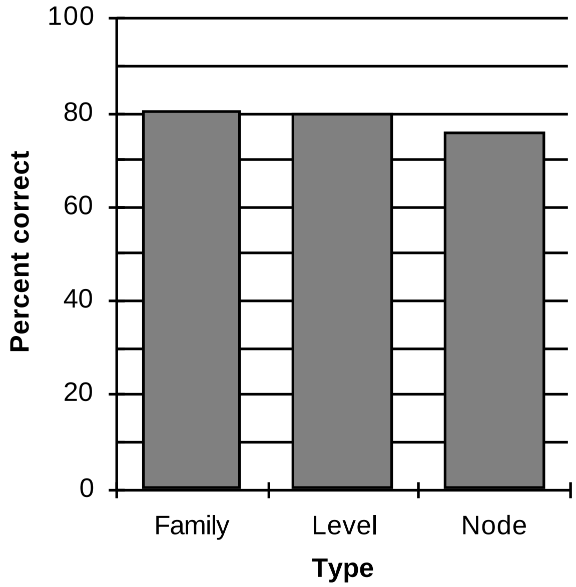

Figure 10*: Percentage of correctly recalled earcon components.*

There were no significant differences between the rates of recall of family, level and node. Participants recalled 80% of the family and level components and 76% of the node components. A one-factor ANOVA on the family, level and node data showed no significant difference between these (F(2,33)=1.27, p=0.291).

The component data for the three worst-recalled questions was examined in more detail to see if any common problems could be identified. Figure 11 shows the component scores. The table shows that there was no consistent problem causing the lower recall for these questions. In question 3 the worst recalled component was family, in question 5 it was level and node and in question 12 family and node.

Component/question	Family	Level	Node
Q2	75	83	92
Q4	100	67	67
Q11	83	92	83

Figure 11*: Percentage of correctly recalled components for the three worst recalled earcons.*

New earcons

As mentioned above, two new, previously unheard, earcons were presented to the participants during testing. These are marked by the letters A and B in Figure 2 . They were both from Level 4 and were constructed using the same rules as the rest of the earcons. By using the rules participants should have been able to identify where in the hierarchy the sounds belonged.

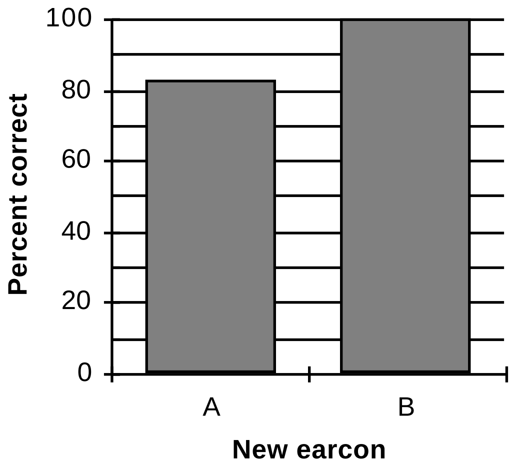

***Figure 12**: Recall rates for the new earcons.*

The new earcons were very well recognised (see Figure 12). Ten out of twelve participants recognised the earcon for A in Figure 2 and all the participants recognised the earcon for B. This indicates that the participants were able to use the rules to work out where an unheard earcon belonged.

DISCUSSION

Overall

The overall recall rate of the earcons was good. The participants recalled 81.5% of the earcons after being trained for only a short time. This indicates that earcons can convey complex hierarchical information in sound. The score is approximately the same as the rate of recall in phase I the second earcon experiment described by Brewster [4]. As mentioned above, the hierarchy used in that experiment was simpler (only 10 nodes) but the training and testing were more difficult (for example, the earcons were presented in a random order during training). In the experiment described here the training was more structured but the hierarchy was more complex, leading to similar recall rates.

The three worst recalled earcons were from Level 4 of the hierarchy. However, the other two earcons from Level 4 were significantly better recalled than these. The problem may have been that these earcons were at the bottom of the hierarchy so participants had to remember the most information to work out their location. They had to remember the manipulations at each of the previous levels to find out where they were. This left more room for mistakes than with the recall of earcons higher-up in the hierarchy where less had to be remembered. The manipulations chosen for Level 4 (tempo, chorus and echo) may not have been the best ones. Further research might suggest some more effective ones.

A detailed analysis of the data showed that there were no significant differences in the recall of the different component parts of the earcons. There was no common problem that was lowering the rate of recall. Longer training would have overcome this problem.

New earcons

The ability of the participants to identify the location of previously unheard earcons was good. They were able to use the rules for constructing the earcons to work out where an earcon belonged. However, as the new earcons were from one of each of the main families, or sub-trees, of the hierarchy the participants could use all of the rules to identify an earcon. For example, if they could only remember the timbre for the sub-tree they would still be able to identify the new earcon. This test could have been made more difficult by having two missing earcons from the same sub-tree. This would have made it harder to identify the missing one as the participants would have had fewer parameters to use. The result does show that participants were able to use the rules to locate the new earcons. This indicates that they are a very robust method of communicating hierarchy information in sound.

FUTURE WORK

These sounds will now be used in an interface to a mobile communicator device for the TIDE ACCESS project. The sounds will be played quietly, in the background, when the user moves through the hierarchy of pictographic symbols. The sounds will change when he/she moves up, down, left or right indicating position in the hierarchy.

For telephone-based interfaces this work will be extended in two ways. Firstly, the results have shown that high levels of recall can be gained from using earcons as described. However, in a telephone-based interface the quality of the sounds would be greatly reduced due to the limited bandwidth of telephone equipment. In this type of system there would also be no stereo effects available. The next stage of the research would be to conduct the experiment again but this time present the sounds as if they were coming from a telephone. This could be done by filtering the earcons to make them sound as if they were from a telephone or actually presenting them over the telephone to get the filtering. The results would then show what recall would be like under these more difficult conditions. The earcons could then be re-designed if necessary to maximise recall.

The second extension to the work would be to present the sounds to participants again after a period of time. If they were using the sounds in a real telephone-based interface they might use the system at irregular intervals (this would not be a problem for the communicator device as users would use it often). Testing again after a period of time would discover if listeners could remember the earcons and their structure after, for example, one week. If they could remember the structure of the sounds then they could still use them for navigation aids.

CONCLUSIONS

The aim of the experiment described here was to discover if earcons could provide navigational cues in a system of hierarchical menus. The results have shown that a complex hierarchy can be represented using earcons. 81.5% of the earcons were recalled after only a short amount of training. This indicates that earcons are a good way of providing navigational information for hierarchies. Listeners can hear an earcon and tell where it comes from in the hierarchy and hence where they are. Listeners could also recognise new earcons that had not been heard before by using the rules from which the earcons were constructed. This shows that earcons are a robust method of communication. This research shows that earcons are an effective way of providing hierarchy information in interfaces where graphical information cannot easily be used.

ACKNOWLEDGEMENTS

This research was supported by ERCIM Fellowship 94-04. It was undertaken whilst the first author was at VTT Information Technology in Finland. Thanks to Peter Wright at the University of York, UK for helping with the statistical analysis.

REFERENCES

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

2. Blattner, M., Papp, A. & Glinert, E. (1992). Sonic enhancements of two-dimensional graphic displays. In G. Kramer (Ed.), *Auditory Display, sonification, audification and auditory interfaces. The Proceedings of the First International Conference on Auditory Display*, Santa Fé Institute, Santa Fé: Addison-Wesley, pp. 447-470.

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

4. Brewster, S.A. (1994) *Providing a structured method for integrating non-speech audio into human-computer interfaces*. PhD Thesis, University of York, UK.

5. Brewster, S.A., Wright, P.C., Dix, A.J. & Edwards, A.D.N. (1994). The sonic enhancement of graphical buttons. *Accepted for publication at Interact'95,* Lillehammer, Norway.

6. Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1992). A detailed investigation into the effectiveness of earcons. In G. Kramer (Ed.), *Auditory display, sonification, audification and auditory interfaces. The Proceedings of the First International Conference on Auditory Display*, Santa Fe Institute, Santa Fe: Addison-Wesley, pp. 471-498.

7. Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1993). An evaluation of earcons for use in auditory human-computer interfaces. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel, & T. White (Eds.), *Proceedings of INTERCHI'93,* Amsterdam: ACM Press, Addison-Wesley, pp. 222-227.

8. Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1995). Experimentally derived guidelines for the creation of earcons. *Submitted to HCI'95,* Huddersfield, UK.

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

10. Gaver, W. (1992). Using and creating auditory icons. In G. Kramer (Ed.), *Auditory Display, sonification, audification and auditory interfaces. The Proceedings of the First International Conference on Auditory Display*, Santa Fe Institute, Santa Fe: Addison-Wesley, pp. 417-446.

11. Gaver, W. (1993). Synthesizing auditory icons. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel, & T. White (Eds.), *INTERCHI'93,* Amsterdam: ACM Press, Addison-Wesley, pp. 228-235.

12. Gaver, W., Smith, R. & O'Shea, T. (1991). Effective sounds in complex systems: The ARKola simulation. In S. Robertson, G. Olson, & J. Olson (Eds.), *Proceedings of CHI'91*, New Orleans: ACM Press, Addison-Wesley, pp. 85-90.

13. Lucas, P. (1994). An evaluation of the communicative ability of auditory icons and earcons. In G. Kramer (Ed.), *Accepted for publication in the proceedings of ICAD'94,* Santa Fe Institute, Santa Fe, NM.: Addison-Wesley.

14. Mayes, T. (1992). The 'M' word: Multimedia interfaces and their role in interactive learning systems. In A. D. N. Edwards & S. Holland (Eds.), *Multimedia Interface Design in Education*, pp. 1-22. Berlin: Springer-Verlag.

15. Rosson, M.B. (1985). Using synthetic speech for remote access to information. *Behaviour Research Methods, Instruments and Computers*, 17(2), pp. 250-252.

16. Stevens, R. (1995) *Principles for the design of auditory interfaces to present complex information to blind computer users*. PhD Thesis, University of York, UK.

17. Stevens, R.D., Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1994). Providing an audio glance at algebra for blind readers. *Accepted for publication in the proceedings of ICAD'94,* Santa Fe Institute, Santa Fe, NM: Addison-Wesley.

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

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