Using Non-Speech Sounds to Provide Navigation Cues

Stephen A. Brewster

Glasgow Interactive Systems Group
Department of Computing Science, University
of Glasgow, Glasgow, G12 8QQ, UK.
Tel: +44 (0)141 330 4966
Fax: +44 (0)141 330 4913
Email: stephen@dcs.gla.ac.uk
Web: www.dcs.gla.ac.uk/~stephen/

ABSTRACT

This paper describes three experiments that investigate the possibility of using structured non-speech audio messages called *earcons* to provide navigational cues in a menu hierarchy. A hierarchy of 27 nodes and four levels was created with an earcon for each node. Rules were defined for the creation of hierarchical earcons at each node. Participants had to identify their location in the hierarchy by listening to an earcon. Results of the first experiment showed that participants could identify their location with 81.5% accuracy, indicating that earcons were a powerful method of communicating hierarchy information. One proposed use for such navigation cues is in telephone-based interfaces (TBI's) where navigation is a problem. The first experiment did not address the particular problems of earcons in TBI's such as: Does the lower quality of sound over the telephone lower recall rates, can users remember earcons over a period of time and what effect does training type have on recall. An experiment was conducted and results showed that sound quality did lower the recall of earcons. However, redesign of the earcons overcame this problem with 73% recalled correctly. Participants could still recall earcons at this level after a week had passed. Training type also affected recall. With 'personal training' participants recalled 73% of the earcons but with purely textual training results were significantly lower. These results show that earcons can provide good navigation cues for TBI's. The final experiment used compound, rather than hierarchical, earcons to represent the hierarchy from the first experiment. Results showed that with sounds constructed in this way participants could recall 97% of the earcons. These experiments have developed our general understanding of earcons. A hierarchy three times larger than any previous created was tested and this was also the first test of the recall of earcons over time.

KEYWORDS

Earcons, telephone-based interfaces, auditory interfaces, non-speech audio, navigation.

1. INTRODUCTION

This paper describes three experiments to investigate the power of non-speech audio for conveying navigational cues in hierarchical structures. In some situations graphical feedback cannot be used to provide these cues. In completely auditory interactions such as telephone-based interfaces (TBI's) or interfaces for visually disabled people 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 pictographic symbols to create 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.

Sound has been successfully used to improve interaction in graphical human-computer interfaces [5, 11, 15]. It has not, however, been used before to provide navigation cues. The experiments described in this paper investigate the possibility of doing this. They also investigate the design of the sounds necessary to represent hierarchical structures. The approach taken was to investigate the recall of sounds representing a hierarchical structure. If users could recall the sounds then this would show that they could be used for navigation purposes, if users could not recall them then they would not be useful. It was decided to take this approach, rather then using the sounds immediately in a task-based study, so that it was certain the sounds could be used for navigation cues. If a task-based study had been undertaken then the problems with the sounds may have been confused with problems with the task itself. Using the approach described here will provide strong foundations for the use of earcons as navigation cues because any problems with the sounds will be clear.

The first experiment describes an initial study to discover if sounds could provide navigation cues in general hierarchical structures [8]. The second experiment focuses in on the problems faced when using sounds to provide navigational cues in TBI's and the final experiment investigates an alternative method of constructing the sounds to represent a hierarchy. Finally, a set of guidelines is presented that can be used by designers to add sounds to their systems.

1.1 Telephone-based interfaces

There were two main factors motivating this work. The first was the use of sound to provide navigation cues in non-visual interfaces. One important reason for using non-speech sound was to represent hierarchies in interfaces where visual feedback was not possible, for example TBI's (e.g. telephone banking or voicemail systems), interfaces for visually disabled people (file system hierarchies or electronic books) or other situations where screen space is limited but structured information must be presented (e.g. PDA's or wearable computers). TBI's are becoming an increasingly important method for interacting with computer systems. The telephone is an ubiquitous device and is many people's primary method of entry into the information infrastructure. Access to an increasing number of services is being offered over the telephone, such as voice-mail, electronic banking and even Web pages. The rapidly increasing use of mobile telephones means that people access these services at many different times and places. Telephones themselves are now also incorporating greater functionality (such as multi-party calling or call forwarding). The provision of this extra functionality may be rendered useless if usability issues are not considered [16]. The telephone itself allows only a limited form of interaction [18]. There is no graphical display (although some do have small LCD displays but these are not visible if the device is at the user's ear) so output is limited to speech and simple sounds. Users provide input via the keypad (although speech recognition is sometimes used).

In a TBI a user might call their bank and navigate through a hierarchy of voice menus to find the service required. One common problem when interacting in TBI's is navigation. Users get lost in the hierarchy of menus that they must go through to reach an option or function [17, 25]. As Yankelovich *et al.* [26] say (p 369): "These {telephone-based} interfaces, however, are often characterized by a labyrinth of invisible and tedious hierarchies which result when menu options outnumber telephone keys or when choices overload users' short-term memory". Rosson [17] investigated such a hierarchical TBI which gave travel/visitor information. A user could call the system and move through the hierarchy to find information such as addresses and phone numbers of restaurants. She describes one common problem (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 infer that a move to a lower level of the hierarchy had been made. She suggested that this may have been the source of many of the users problems.

Why is navigation information not given? Because it gets in the way of the information the user is trying to access with the TBI. The more navigation information that is given the more it obstructs the actual information the user is trying to get at. Speech is also serial and slow [19]. The lower quality of sound over the telephone system makes it hard to attend to more than one speaker at once, especially if the speech is synthesised or constructed from poorly concatenated samples. It is suggested here that many navigation problems occur because current TBI's are limited to using speech alone. Speech is forced to perform two tasks: Information and navigation. Designers choose to present information to users as that is what they are using the system to find. This means that navigation cues are not provided.

Rosson suggested that one way to solve the problem was to give extra speech feedback. For example "You have moved to the next item in the Chinese Restaurant list. It is…". She suggested, however, that this would make the interface appear slow and clumsy. This extra feedback might also be longer than the information being retrieved and so obscure it. Such feedback was rejected for this very reason by Stevens *et al.* [20, 21] when designing navigation cues in a system to provide non-visual access to mathematics. In this paper I propose an alternative solution: Structured non-speech audio messages. These would provide a hierarchical system of sounds that could be used to represent a menu hierarchy. The sounds would play continuously (but quietly) in the background at each level, giving location information. Users could listen to the current sound and from it work out their location in the hierarchy. The sounds would make explicit the differences moving from level to level or across the same level because the sounds would be related in different ways. This is a similar approach to that taken by Stevens *et al.* [20, 21] in the Mathtalk system. This system displays algebra to blind mathematicians. 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 in the same way as text and graphics. If carefully designed, they can also be used together to provide complementary information. Wolf *et al.* [25] also confirm the usefulness of this approach and suggest that this might be a solution to some of the navigation problems they had with their combined voicemail and email speech interface system (p 250): "Replacing much of the text-to-speech feedback with brief distinctive earcons would make traversal of the mailbox more efficient".

Little use has been made of structured non-speech sound in TBI's (apart from the standard dial tone, engaged tone, etc.). For example, guidelines for the design of TBI's [16, 18] include nothing about the use of non-speech sound. The use of such sounds, in addition to speech, will increase the bandwidth of communication between the system and the user, allowing a richer interaction. These 'multimedia' telephone interfaces will be more usable than their speech-only counterparts.

Sound has many advantages. For example, it is good for communicating information quickly [7]. Unlike speech, non-speech sound is not language dependant; the user is not tied to one language, which is important for the increased international use of computer systems. There is also great potential for the results of this work in other non-graphical interfaces such as those for visually disabled people and those where working conditions or protective clothing mean that a screen cannot be used. This paper will investigate some important questions about the use of non-speech sound for navigation cues in TBI's.

1.2 Navigation in a communicator device

The second reason for investigating the use of sound to represent hierarchies was to present navigational cues in an interface for people who are speech-motor and/or language-cognitive impaired for the TIDE ACCESS project. The aim of this project was to create a mobile communication device. People would use the device to create messages they wanted to communicate and then say these messages via synthetic speech. These types of users often use pictographic languages (for example, Bliss [4]) to communicate. The pictures represent words or actions and can be combined to create complex messages.

An experienced user with a wide vocabulary may need access to a large number of symbols. 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 but this can lead to users getting lost, in the same way as described above. Stifelman *et al.* [22] discovered a similar navigation problem in their *VoiceNotes* handheld note-taking system. They had to design the hierarchy of messages their system could contain, but avoid users getting lost (p 182) "Therefore, the database {of messages} was limited to a one-level hierarchy (a category of notes cannot contain a sub-category)". This was not possible in the system for the ACCESS project because we needed to present a large number (potentially hundreds) of symbols. An alternative method was necessary.

Graphical feedback could be used to give extra navigational information, for example a map could show the current position in the hierarchy. However, this would take up valuable screen space needed for the pictographic symbols and would also require the user to look at the map when he/she really wanted to look at the symbols. An alternative would be to use different colours to show the different levels. Unfortunately, colour is already used to show other groupings within the symbols such as nouns/verbs/adjectives. This paper suggests the use of sound; it can be used to give information about one's location in the hierarchy without taking up screen-space. Synthetic speech cannot be used as it would conflict with the message being created. However, non-speech sounds would not suffer from this problem.

Lack of screen space is not only a problem in interfaces for disabled people. Visual displays can only hold so much. 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 [5] discuss this problem with computerised maps. 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. Blattner *et al.* suggested that sound could be used to avoid these problems.

1.3 Earcons

The non-speech sounds to be used for this investigation are based around structured audio messages called *Earcons* [6, 7, 23]. Earcons are abstract, musical tones that can be used in structured combinations to create sound messages to represent parts of an interface. Investigations of earcons by Brewster, Wright & Edwards [9, 10] 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*. Simple motives can be created to represent various interaction components. These can then be concatenated to produce more complex messages. By using more complex manipulations of the parameters of sound (timbre, register, intensity, pitch and rhythm) *hierarchical earcons* can be created [6]. These allow the representation of hierarchical structures. Sumikawa ([24], 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.

Figure 1 shows a simple example hierarchy of earcons based on one possible family of applications (based on the earcon design guidelines from [12]). Each earcon is a node on a tree and inherits the properties of the earcon above it. The different levels are created by manipulating the parameters of earcons (for example, rhythm, pitch, timbre). In the diagram the top level of the tree is a neutral sounding earcon. It has a neutral flute timbre played continuously at middle C. The structure of the earcon from Level one is inherited by Level two and then changed. At Level two there is still a continuous flute sound but new timbres are added to play alongside it. At Level three a rhythm is added to the earcon from Level two to create a sound for a particular application. This rhythm is based on the timbre from the level above. In the case of Netscape Navigator there would be a continuous flute sound with a three note rhythm played on an organ accompanying it. Other levels can be created by using parameters such as tempo or effects.

Using earcons, this hierarchy is easily extensible. For example, to add another major category of applications all that is needed is a new timbre. To create a new type of communications application only a new rhythm is needed and it can be added to the existing hierarchy. Therefore earcons provide a very flexible system for representing hierarchical structures.

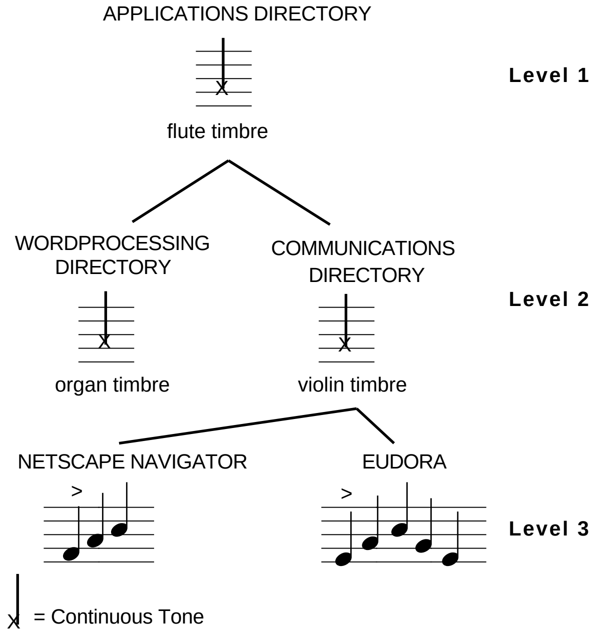


Figure 1: *A hierarchy of earcons representing a family of applications. The '>' symbols over a note indicate stress or accent.*

1.4 Previous attempts to use earcons to present hierarchy information

There has been little previous work on using earcons to represent hierarchies or to present navigation cues. Barfield, Rosenberg & Levasseur [3] 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* [7, 9, 10]. If better earcons had been designed then advantages may have been found.

Brewster, Wright & Edwards [7, 9, 10, 13] also tested the ability of earcons to present hierarchical information in a limited way. In three experiments they showed that, with careful design of earcons, a simple hierarchy could be presented effectively. They used earcons to represent a small hierarchy of files, folders and applications (see Figure 2). The relationships between the items was very simple. Things in the same family shared the same timbre, for example the paint program, paint folder and file all shared a brass instrument. Things of the same type shared the same rhythm, for example the paint, write and spreadsheet programs all shared the same three-note rhythm. In cases where items had the same family and type, register was used to separate them. 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 was simple with only nine nodes. It was possible that users could have just remembered the nine sounds (with short term memory capacity of 7±2 items). A bigger hierarchy must therefore be tested before earcons can safely be used to represent complex hierarchies. This work formed the basis for the sounds in Experiment 1 described below.

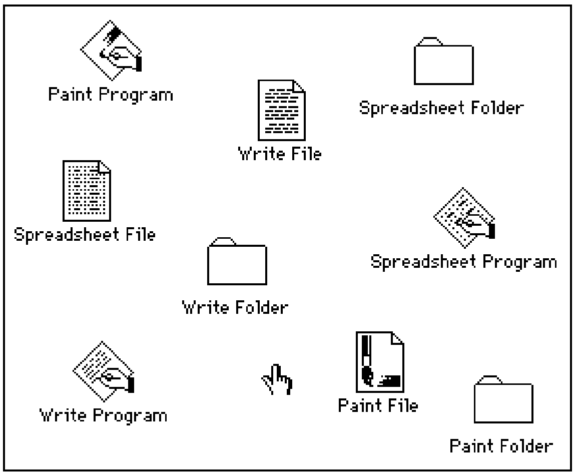


***Figure 2**: The hierarchy used by Brewster et al. [13].*

2. EXPERIMENT 1

The aim of this experiment was to discover if a hierarchy three times larger than that of Brewster *et al.* [13] could be represented by earcons. Figure 3 shows the hierarchy used. It had 25 nodes on four levels with four missing nodes on Level 4 (two of which are marked as A and B in Figure 3). This made a hierarchy of 27 nodes, three times larger than that tried previously. It was based on the structure of the file system on the author's computer and was similar to the one needed for the communicator device described above (this hierarchy is fairly small and there are TBI's which have more complex interfaces. However, my own telephone banking system has a less complex interface than this one. It was therefore decided that this hierarchy was a good place to start). It would allow the testing of earcons to represent such a structure. At this stage a general investigation of the ability of earcons to provide navigation cues was needed, therefore this experiment did not focus directly on TBI's. This came in the second experiment.

2.1 Hypotheses

The main hypothesis for this experiment 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.

2.2 Participants

Twelve volunteer participants were used. They were a mixture of students at Helsinki University of Technology and members of staff at VTT Information Technology in Helsinki. They were all familiar with computers and computer file systems. Previous research on earcons [7, 10, 13] showed that musical training did not affect their usability, therefore the musical skills of the participants used were not measured here. In these previous studies the affects of musical training were investigated in detail. The results showed that, if the earcons were designed using basic musical techniques with gross differences between each one, recall rates were not significantly different between musicians and non-musicians. Musicians were better than non-musicians when the sounds used were poorer (for example, sine waves and square waves, or rhythms that were very similar caused non-

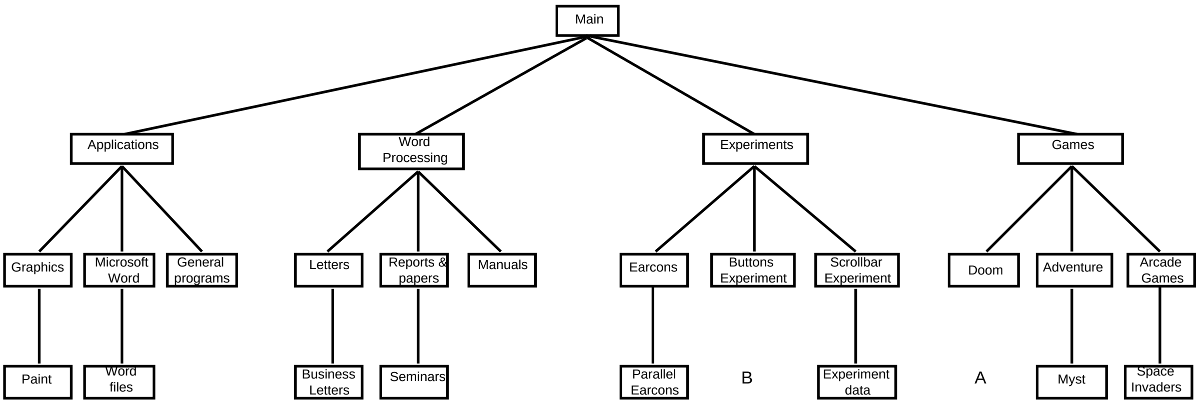


***Figure 3:** The file-system hierarchy used in the experiment. A and B show the two new earcons presented to participants during testing.*

musicians problems). The earcons here were designed to be usable by both groups, as would be needed in a TBI for the general population.

2.3 Sounds used

The earcons were designed using the guidelines proposed by Brewster *et al.* [12] and based on the simple hierarchy described in the previous section [13]. As suggested, the earcons were designed using timbre, register and spatial location for the main sub-groups in the hierarchy. The earcons were based on the 'musical' earcons described by Brewster *et al.* [7, 9, 10, 13]. These used standard musical techniques to gain significantly better recall than the simple tones initially proposed by Blattner *et al.* [6]. The earcons 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 3) a constant sound with a flute timbre was used (see Table 1). 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. The earcon was designed to be neutral sounding.

Level 2: At this level each family had a separate timbre, register and spatial location. Table 1 shows these. 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 3).

The continuous sound was inherited from the Level 1 earcon but the instrument, pitch and stereo position were changed. Three parameters were used so that if a 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 from Brewster [7] 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

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

Stereo position was used in the earcons even though it is not available in TBI's. This was done because it is available in interfaces for disabled people based on personal computers, such as the portable communicator device described above or in interfaces for the blind. For this exploratory experiment all of the sound manipulation techniques available were used to maximise usability.

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 3 '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 [12].

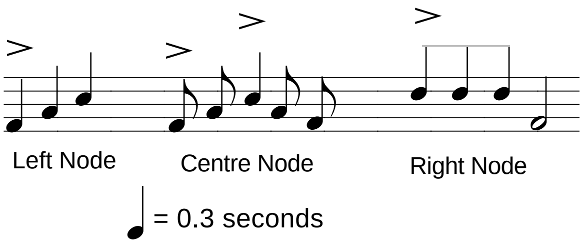


__Figure 4__: The rhythms used for Levels 3 and 4 of the hierarchy.

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: A faster tempo was used to differentiate the items (Blattner *et al.* [6] suggest this method for discriminating earcons). The rhythmic units from Figure 4 now repeated once every second. In addition to this, the effects of 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.

2.4 Experimental design and procedure

As shown in Figure 3, the hierarchy was based on a computer file system. This was an experiment to test the use of earcons to represent a hierarchy, this paper does not suggest that each directory and sub-directory in a real system should have a sound.

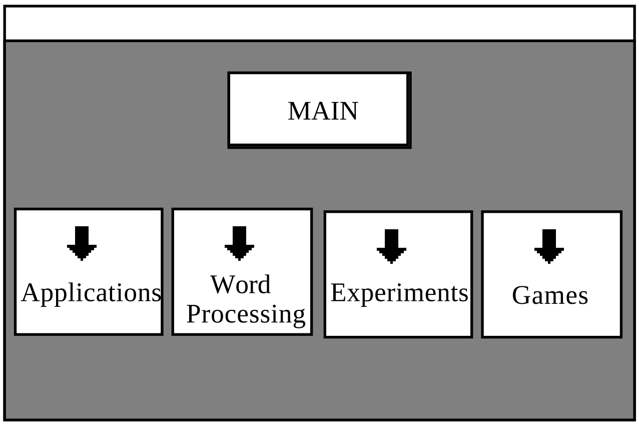


__Figure 5__: The top level of the hierarchy. The arrows show the direction of movement possible.

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 3 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 and continued until another card was selected.

2.5 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. This was done once only. The structure of the earcons at each level was fully explained.

In the second part of the training participants were given five minutes to learn the earcons by themselves, using the HyperCard stack, with no help from the experimenter. The training was short and simple to see if the rules by which the earcons created were obvious. Users of a communicator device of the type described above might only receive very simple training and have to learn the rest of the system by themselves. Users of a telephone-based system might get a short amount of training when they signed-up for a new telephone service and again have to learn it by themselves. The simple training used here would test these situations.

During the training participants could look at a map of the hierarchy (similar to Figure 3 above). The aim of the experiment was not to test the participants' abilities to learn hierarchies but to test their ability to learn the earcons. Instructions were read from a prepared script.

Question	Level	Node
Q1	2	Word processing
Q2	4	Space invaders
Q3	2	Experiments
Q4	4	Paint
Q5	2	Games
Q6	3	Doom
Q7	4	Word files
Q8	4	Parallel earcons
Q9	3	Graphics
Q10	3	Reports & papers
Q11	4	Business letters
Q12	3	Microsoft word
Q13	4	A
Q14	4	B

***Table 2**: The node and level in hierarchy for each of the questions. This is the order that the questions were presented to participants.*

2.6 Testing

The participants heard fourteen earcons during testing. These were randomly selected from all of the sounds in the hierarchy. The same set of earcons was presented to each of the participants. Twelve of the sounds were ones that participants had heard during the training. The last two earcons were new ones (marked A and B in Figure 3). These were earcons for gaps in the hierarchy and were constructed using the same rules as the other earcons. Participants were told that these were new earcons that they had not heard before representing the missing nodes in the hierarchy. Table 2 shows which earcon represented which question. An earcon was played and the participants then had to choose where the it fitted into the hierarchy. The hierarchy was represented on screen. Again, the aim of this exploratory experiment was to test the participants' knowledge of the earcons, not their ability to learn hierarchies. None of the names of the nodes were included in the picture of the hierarchy to avoid any help they might have provided.

3. RESULTS OF EXPERIMENT 1

3.1 Overall

The overall recall rate of all of the earcons was high: 81.5% of the earcons were correctly recalled. Figure 6 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 a significant main 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) Tukey HSD tests comparing each question were performed. 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'. 'Paint' was recalled worst of all (Paint versus space invaders: Q_{154}=1, p=0.01, space invaders versus parallel earcons: Q_{154}=2, p=0.01). However, the other Level 4 nodes 'Word Files' and 'Parallel Earcons' were significantly better recalled than these.

Table 3 shows a confusability matrix indicating which earcon was confused with which node for each of the questions. The new earcons (questions 13 and 14) are not included and are dealt with below. If only questions 1-12 are considered then the overall recall rate was 80%. The table shows that there were 29 errors in 144 answers. It also shows that 72.4% of these errors occurred in Level 4 of the hierarchy (there were only 2 errors on Level 2 and 6 on Level 3). This indicated that recall at this level was significantly worse than the other levels. The matrix shows that there were no common recall errors made. For example, it can be seen that 'Paint' (the worst recalled earcon) was only confused within the same sub-tree, or family, of earcons. Two participants mistook it for the earcon for 'Graphics' - they confused the tempo of the earcon. Two mistook it for 'Word Files' - they got the level correct but mistook the centre-node rhythm. Finally, two mistook it for 'Microsoft Word' - they confused the rhythm and the level. All of these mistakes were very close to the correct answer.

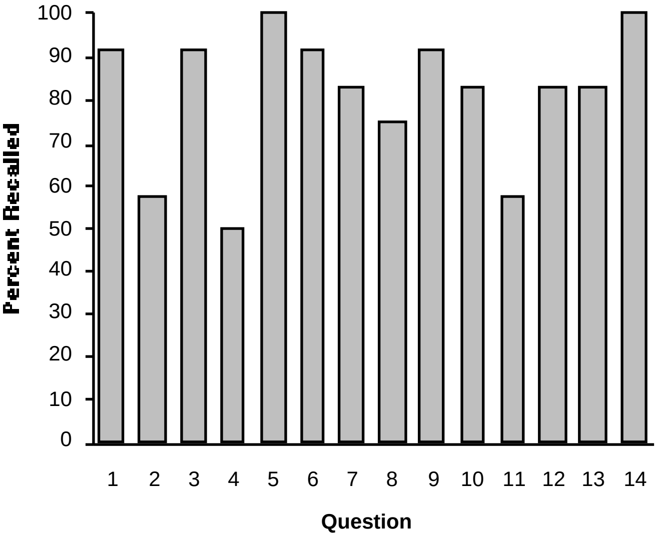


***Figure 6**: Recall rates for each of the 14 questions.*

3.2 Recall of components

A more detailed analysis of the results was undertaken. Each of the participants' answers were broken down to find exactly where the problems occurred. There were three mistakes that could be made. Participants could mistake the *family* of an earcon: For example, whether it was from 'Applications', 'Word Processing', 'Experiments' or 'Games'. They could also mistake the *node* an earcon referred to: For example, whether it was a left node, a centre node or a right node. Finally, participants could mistake the *level* of an earcon: Whether it was from Level 1, 2, 3 or 4. From the overall data obtained the scores were broken into three. If participants got a question completely right they received three marks, if they got two parts right then two marks were given etc. From this analysis it was possible to see where mistakes occurred.

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 were examined in more detail to see if any common problems could be identified. Table 4 shows the component scores. The table shows that there was no consistent problem causing the lower recall for these questions. In question 2 the worst recalled component was family, in question 4 it was level and node and in question 11 family and node.

	Participants												No. of
Question	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	errors
q1	1	1	1	1	1	main	1	1	1	1	1	1	1
q2	1	arcade games	1	exp data	1	adventure	1	1	exp data	1	exp data	1	5
q3	1	1	1	1	1	games	1	1	1	1	1	1	1
q4	word files	graphics	graphics	ms word	word files	ms word	1	1	1	1	1	1	6
q5	1	1	1	1	1	1	1	1	1	1	1	1	0
q6	1	1	1	1	1	1	adventure	1	1	1	1	1	1
q7	paint	1	1	1	1	1	1	1	seminars	1	1	1	2
q8	1	earcons	1	buttons xpt	1	myst	1	1	1	1	1	1	3
q9	1	1	1	1	1	1	ms word	1	1	1	1	1	1
q10	1	1	1	1	1	1	letters	seminars	1	1	1	1	2
q11	1	letters	paint	parallel earcons	1	seminars	seminars	1	1	1	1	1	5
q12	1	1	1	1	1	general progs	graphics	1	1	1	1	1	2
												total	29
												level 4 errors	21

***Table 3**: Confusability matrix for questions in Experiment 1. A 1 indicates a correct answer. The new earcons (question 13 and 14) are not included.*

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

***Table 4**: Percentage of correctly recalled components for the three worst recalled earcons (questions 2, 4 and 11).*

3.3 New earcons

As mentioned above, two new, previously unheard, earcons were presented to the participants during testing. The new earcons were very well recognised (see questions 13 and 14 in Figure 6): 91.5% were recognised correctly (where 25% would be expected by chance). Ten out of twelve participants recognised the earcon for A in Figure 3 and all the participants recognised the earcon for B. This indicates that the participants were able to recognise new earcons well. The new earcons were from Level 4 of the hierarchy. They were significantly better recalled than the earcons for questions 2, 4 and 11. However, there was no significant difference between these new earcons and questions 7 and 8 which were also from Level 4.

4. DISCUSSION OF EXPERIMENT 1

4.1 Overall

The overall recall rate of the earcons was good. This demonstrated that a listener could use earcons as navigation aids in a menu hierarchy. A listener would hear an earcon and from it be able to work out his/her location in the hierarchy and therefore avoid becoming lost. This was also possible after only a short amount of training. This shows that with careful design, earcons can easily be learned.

The results obtained here are approximately the same as the rate of recall in phase I of the second earcon experiment described by Brewster [7]. As mentioned above, the hierarchy used in that experiment was much simpler (only 9 nodes) but the training and testing were less efficient (for example, the earcons were presented in a random order during training in order to obtain baseline rates). In the experiment described here the training was more structured but the hierarchy was three times larger, 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. A detailed analysis of the results gave no clear indication of what might have caused the difficulty. The problem may have been that these earcons were at the bottom of the hierarchy so participants had to remember all of the earcon construction rules to work out their location. This left more opportunity 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) were perhaps not the most distinguishable so other methods need to be investigated.

4.2 New earcons

The ability of the participants to identify the location of previously unheard earcons was good (91.5%). They were able to use the rules for constructing the earcons to work out where a new earcon belonged. This showed that the listeners had learned the rules well. This research has shown that if new nodes were added to an existing hierarchy then listeners would not need to be retrained on the sounds, they could use their existing knowledge of the rules to understand the new earcons. This is a very promising result for the use of earcons as hierarchical navigation aids.

The new sounds were from one of each of the main families, or major sub-trees, of the hierarchy so the participants could use all of the earcon creation rules to identify them. There were also only four missing ones so that their choice was reduced (25% could be obtained by chance). For example, if they could only remember the timbre for the sub-tree they would still be able to identify the new earcon because each new earcon had a different timbre. This test could have been made more difficult by having two or more missing earcons from the same sub-tree. This would have made it harder to identify the missing one as there would be fewer differences between them. However, the result indicates that participants were able to use the rules to locate the new earcons. A more detailed analysis of recognition of new earcons is being undertaken.

5. EXPERIMENT 2

The first experiment gave a strong indication that earcons could provide navigational cues in hierarchies; listeners could recall an earcon with a high degree of accuracy and so from this they would know where they were in the hierarchy. However, there were important questions still to be answered if earcons were to be used for navigation cues in TBI’s:

- The earcons used in Experiment 1 were of CD quality (16 bit 44 kHz) - sounds played over the telephone are of much lower quality due to the narrow bandwidth of telephone equipment. This could have a significant affect on recall rates for the earcons.

- The earcons made use of stereo information to differentiate one from another. Again, stereo information is not available over the telephone so this might affect the usability of the earcons because position information is a very powerful cue.

- The sounds from Level 4 of the hierarchy were not recalled well, with 72% of all of the errors coming from this level. Tempo was used for this level and it may not have been a good indicator. Improvements were needed to reduce these errors.

- In Experiment 1 there was no measure of the ability of participants to recall the earcons over time. In fact, in none of the previous work on earcons has any investigation of this been done. It is important for TBI's because a user might not use the system frequently. He/she would have to be able to remember the earcons over time otherwise he/she would have to be retrained each time the system was used, which would be unacceptable.

- The training in the previous experiment was highly personal. The experimenter trained each participant individually. This would be impossible to do in a real TBI because of the high cost to the provider of the service. Therefore an investigation was needed into alternative, more practical training techniques.

These five problems motivated the second experiment. With answers to these questions the usefulness of earcons for TBI's would be established and our understanding of earcons in general would also be greatly enhanced. The same hierarchy was used as before so that the results would be comparable. However, the two new nodes (A and B) were not tested in this experiment.

5.1 Hypotheses

The first hypothesis was that the reduction in sound quality and lack of stereo information would reduce recall rates. With poor quality sounds it would be harder for participants to identify the earcons. This should be shown by comparison with the recall rates of the previous experiment.

Redesign of the Level 4 earcons should improve recall rates. In the previous experiment most errors came from this level. Redesign should result in increased recall rates in Level 4 as compared to the previous experiment.

Participants should be able to recall earcons equally as well a week after they were trained on them as they could when first trained due to the simplicity of the rules describing the earcons at each level. This would be shown by similar overall recall scores from testing session 1 to session 2.

Different training techniques should reduce the recall rates of the earcons with personal training giving the best recall rates and written training giving the worst rates.

5.2 Participants

Forty-eight volunteer participants were used, split into four groups of twelve. They were computer science students and staff from University of Glasgow. All were familiar with computers and computer file systems. None had taken part in the first experiment.

5.3 Sounds used

The earcons were based on those from Experiment 1 to maintain consistency. The sounds were all played by HyperCard on an Apple Macintosh computer through a telephone hand-set (to simulate the output of a real telephone). The sounds were generated on a Yamaha TG100 sound synthesiser and re-recorded by a Macintosh at telephone quality. Each sound played for 7.5 secs. For Levels 1 to 3 the earcons were the same as for the previous experiment except stereo position was not used and the quality was much lower.

In the previous experiment 72% of the recall errors came from Level 4 earcons. A faster tempo was used to differentiate the items at this level. The errors occurred because Level 4 was the bottom of the hierarchy so participants had to remember the most sound manipulations to work out their location. For the current experiment it was decided to try an alternative method of presenting the Level 4 information to make it clearer to participants. The Level 3 earcons were used again at Level 4 but a 0.3 sec. sitar note was played before each of the repeating rhythmic units. A sitar timbre was used as it sounded recognisable when played over the telephone and was also distinctly different to the other timbres used [1]. This acted as a Level 4 identifier to reinforce the switch to this level. The note was played once for a left node, twice for a centre and three times for a right node (see Figure 7). It was hoped that this would provide stronger information to the participants that they were on Level 4.

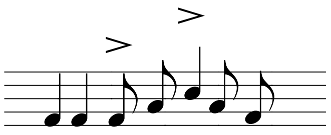


***Figure 7**: An example of a Level 4 centre node (for example Word files). The first two notes were played on the sitar.*

5.4 Training

One of the main aspects of this experiment was the investigation of different training techniques. In the previous experiment training was done by the experimenter. This would not be practical in a real world TBI due to cost. Therefore an investigation of alternative techniques and their effect on recall was needed. This would also give insight into the training of users to use sounds in other types of interfaces. Participants were randomly assigned to one of four groups. Each group investigated a different training technique. The techniques are summarised in Table 5.

Group / Training Type	Method	Sounds in Part 1	Part 2 Training
Group A	Personal	yes	yes
Group B	Online tutorial	yes	yes
Group C	Online tutorial	no	yes
Group D	Online tutorial	no	no

***Table 5**: The different training techniques used in the four groups.*

Group A: In this group training had two parts. For the first part, participants were given personal training by the experimenter. He showed the participant each of nodes of the hierarchy in turn and played the associated earcon. This was done once only. The structure of the earcons at each level was fully explained. In the second part of the training participants were given five minutes to learn the earcons by themselves with no help from the experimenter. They did this using the HyperCard stack. This was the 'best case' training. This condition allowed direct comparison with the results of the previous experiment where the training was exactly the same.

For a TBI this would be the equivalent of the telephone service provider sending a training officer to show new subscribers how to use the system. The subscribers would then get five minutes of free call time to try the system.

Group B: The training was the same as Group A except that the participants received an online tutorial explanation of the sounds rather than personal training. The tutorial fully explained the structure of the sounds and the participants listened to the sounds as they worked through the tutorial. Figure 8 shows an example of the online tutorial for this group.

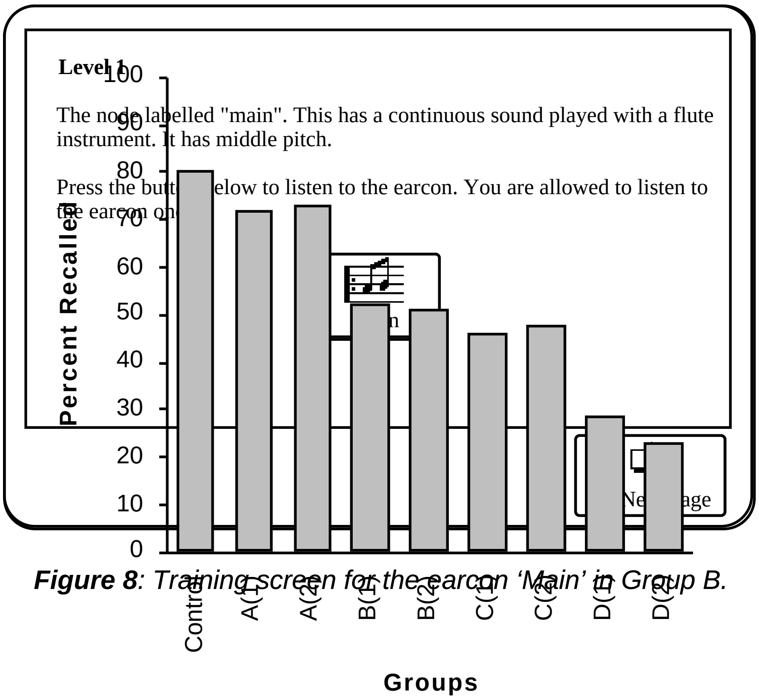
This type of training is equivalent to the telephone service provider giving a tutorial to their system on a training video or over the Web. Again the subscriber is allowed five minutes of free call time to try the system.

Group C: Training for this group was similar to Group B except that the online tutorial did not allow the participants to listen to the sounds. In all other respects the online training for this group was the same as that shown in Figure 8. Participants still heard the sounds in part two of the training.

For this group the training is equivalent to the service provider sending a training brochure to new subscribers with instructions on how to use their new system. Again they are given five minutes of free call time.

Group D: This final group was similar to Group C except that participants did not get part two of the training. This meant that the participants did not hear any of the sounds before testing; they just read once through a description of their structure. This was the 'worst case' training condition.

For this group the training is equivalent to the service provider sending a training brochure to new subscribers with instructions on how to use their new system. No free call time is given. The training in the four groups decreases in cost from Group A which is the most expensive to Group D which is the least. The important question to be answered was: What effect would this have on the recall rates of the earcons (and therefore their usability as navigation cues)? There is one other possible category where no training would be given on the



Level 1

The node labelled "main". This has a continuous sound played with a flute instrument. It has middle pitch.

Press the button below to listen to the earcon. You are allowed to listen to the earcon on

***Figure 8**: Training screen for the earcon 'Main' in Group B.*

***Figure 9**: Recall rates for each of the groups for both presentation sessions. (1) shows results of the first testing session and (2) the second. The Control Group shows the results from Experiment 1. This group appears once because there was no re-testing in that experiment.*

sounds at all. The user would call the system and not know what the sounds meant. This is currently being investigated in a separate study.

5.5 Testing

The participants heard the same set of twelve earcons during testing as in Experiment 1 (to maintain consistency). Testing was done in two sessions. The first presentation session was done directly after the training and the second a week later. No further training was given before the second testing session. This allowed an investigation of the ability of participants to recall the earcons over a period of time, which would be essential for the real-world use of earcons for navigation cues in TBI's.

6. RESULTS OF EXPERIMENT 2

6.1 Comparison with Experiment 1

The overall recall rate was good with 73% of the earcons recalled correctly in Group A(2) (see Figure 9). The first comparison undertaken was to compare the results of Group A(1) with the previous results (here called the Control Group - the value of 80% used here is the score from Experiment 1 not including questions 13 and 14 which were the new earcons and were not being tested here). Group A(1) had the same training method as the previous experiment. The only differences were the sound quality, lack of stereo information and construction of Level 4 earcons. The results are shown in the first two bars of Figure 9. A one-factor ANOVA showed no significant difference between the score from the previous experiment and the current one ($F_{1,22}$=1.11, p=0.301). The results showed that five participants in Group A(1) obtained scores equal to or higher than members of the Control Group.

In Experiment 1, 72% of the errors occurred with Level 4 earcons. In this experiment the Level 4 sounds had been redesigned to help reduce this error rate. Was the redesign successful? Examination of the data showed that Level 4 earcons now accounted for only 41% of the errors. This indicated that improvements to the Level 4 earcons had been successful. The Level 4 sounds became no harder to recall than any of the other levels.

Although the Level 4 error rate had been reduced the overall recall rate was not significantly different. The effect of lower sound quality and lack of stereo information was to distribute errors more uniformly throughout all levels. In Experiment 1, 28% of the errors came from Levels 1-3, in this experiment it was 59%. There were now significantly more errors in Levels 1-3 (T_{12}=4.4, p=0.0008). Therefore improvements in the Level 4 earcons offset the problems due to reduced sound quality and lack of stereo information. The end result was no significant difference between the recall rate in this and the previous experiment.

6.2 Recall over time

The next investigation undertaken was to compare the results of the participants after the first testing session with those after the second. This would indicate how well participants could remember earcons over time. Figure 9 shows the overall results for groups A-D, presentation sessions one and two. As can be seen from the figure the differences between the scores were small. T-tests showed no significant differences between any of the groups in presentation session one and two (for example D(1) vs. D(2) T_{11}=1.24, p=0.24. In D(2) four participants had equal or greater scores than in D(1)). This indicates that participants could recall earcons well over time.

6.3 The effects of training

One of the most important aspects of the experiment was to investigate the effect of training type on the recall of earcons. Again, Figure 9 shows the overall results. As discussed above, the Control and Group A used the same training techniques. There were no significant differences between these so analysis will concentrate on the differences between Groups A-D. There were also no significant differences between the scores obtained from presentation session one and two, therefore data from presentation one in each of the groups will be used to simplify analysis.

Figure 10 shows a breakdown of the recall rates for each of the questions in each of the four groups. A one-factor ANOVA between Groups A-D showed a significant main effect ($F_{3,44}$=5.3, p=0.003). In order to find out where the main effect occurred Tukey HSD tests were carried out between each of the groups. This analysis showed that the only significant difference was between Groups A and D (Q_{44}=43.5, p=0.01). This indicated that training type did have a significant effect on recall but only between Groups A and D. However, variance may have masked some differences between the groups (standard deviation for Group A: 25.6, B: 36.9, C: 23.2, D: 17.7).

The analysis showed that Group D had the lowest recall rates. A more detailed analysis of the D(1) scores was undertaken to find out what information participants managed to extract in this condition. An analysis of family, level and node components (as in section 3.2) showed that the level component was significantly better recalled than family or node ($F_{2,33}$=3.68, p=0.03): 45.5% of family components were recalled correctly, 67.3% of level components and 43% of node components. By chance participants would get 25% of the level components correct, so they were performing significantly better than chance. Out of 144 responses (12 participants x 12 questions) only 13% of responses had none of family, level or node correct, in all of the other cases either 1, 2 or all three components of the answer were correct (in this case 17% of responses would have nothing correct by chance, so participants were performing at approximately chance level).

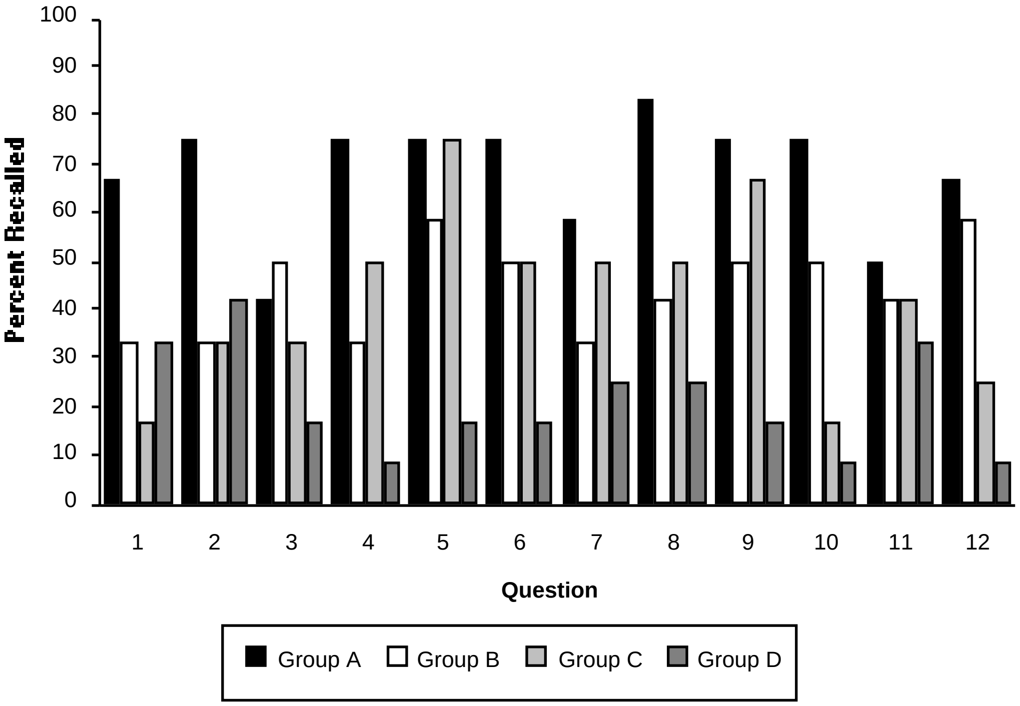


***Figure 10**: Recall rates for each of the questions for each of the groups in Experiment 2.*

7. DISCUSSION OF EXPERIMENT 2

7.1 Comparison with Experiment 1

Results from Group A(1) showed no significant differences to those of the previous experiment. Recall rates of 72% suggest that earcons can provide good navigation cues in telephone-based systems. Users of such systems can listen to an earcon and from it work out where they are in the hierarchy of menus. This then allows them to avoid becoming lost, one of the major problems in such systems [25, 26]. After Experiment 1 there was still a question about the ability of earcons to provide good navigation cues when the quality of the sounds were reduced to those of the telephone system. The results described here show that this is not a problem.

There was an increase in errors due to reduced quality of the sounds and the lack of stereo information. However, this was offset by the greatly improved recognition of earcons at the bottom of the hierarchy. These results showed that the problems caused by sound quality could be overcome by better design of the earcons. In future experiments earcons at Levels 1-3 of the hierarchy will be re-designed in the same way so that recall rates will hopefully again increase.

The results show that earcons can indicate position in a hierarchy of information very successfully. In a real system using earcons, a move to a new node would cause a new sound to play (in the background behind any speech). The user would listen to this earcon and from it work out his/her location, therefore avoiding becoming lost in the hierarchy. Now this has been established, the next stage will be to incorporate earcons into a real telephone-based system and evaluate their effectiveness in real-world conditions.

7.2 Recall over time

The results here are the first to demonstrate the recall of earcons over time. In Brewster *et al.* [9, 10] detailed investigations of earcons were undertaken. Brewster *et al.* tested recall over very short periods of time (approximately 15 - 20 mins. after training). The results showed there were no significant differences after this short period. However, only a small set of sounds (9 earcons) were tested and over a very short period of time. This was not characteristic of the use of earcons in everyday applications. The results presented here show that a large set of earcons can be recalled well over the period of one week. One reason for this is that the construction rules are simple and clear making them easy to remember. These results show that the participants understood the rules by which the earcons were constructed and could apply them again a week later. This is an important result for earcons in general but is particularly important for their use in TBI’s. It means that users of a TBI would not need to be retrained if they used the system infrequently.

7.3 The effects of training

Training type had a significant effect on recall rates. Personal training by the experimenter gave the highest rates with the lowest coming from purely textual training. The results showed that the only significant difference was between Groups A and D. However, variance may have masked some potential differences. By looking at Figure 9 three groupings can be seen in the results: The Control Group and Group A, Groups B and C, and finally Group D.

As expected, good results were obtained in the Control Group and Group A. Personal training was very effective. However, this type of training would be the most expensive for a telephone service provider. There was a 20% difference in recall from Group A to B (although in the experiment this was not significant) where the only difference was an online tutorial rather than personal training.

There was little difference between groups B and C which indicated that the use of sounds in part one of the training did not help recall. More important was to let users use the system themselves. This ‘active learning’ helped them remember the sounds better than reading about the sounds and then hearing them together.

Group D shows that if users cannot hear any of the sounds before they use the system then recall rates are likely to be poor. However, a detailed analysis of the Group D results showed that in 87% of responses participants got one or more of the family, level or node components correct. So, even though the overall analysis presented in Figure 9 shows low recall rates, participants were able to extract some useful navigation information from the earcons. According to Rosson (described above), many navigational problems came from mistaking switches to different levels in a hierarchy. Therefore, purely textual training can provide a reasonable solution to this problem. It may be possible to increase recall in Group D by improving the design of the textual description. Remember, also, that the participants were only allowed to read the training documentation once. If they were allowed to read over the document several times (which is more likely to happen in a real world use of earcons) then recall rates might be improved (this would also apply to the other groups as well). Further experiments are being conducted to investigate training further.

These results indicate that there was no significant difference between Groups A, B and C. This suggested that training of type C could be given and high recall rates achieved but with only a low training cost.

8. EXPERIMENT 3

In this final experiment a different method of representing hierarchies was investigated. In the previous two experiments hierarchical earcons were used for the sounds. At first sight this seems reasonable because they are, by their nature, able to present hierarchies. However, they do have problems. Making the hierarchy from Figure 3 wider is not too difficult as only a new instrument is needed for making the it wider at Level 2 or a new rhythm at Level 3, the rest of the attributes can be inherited as before. Going deeper than four levels may be difficult because once the parameters of timbre, rhythm, register and tempo have been used then there is nothing left to manipulate to create a new level. In Experiment 2 this was overcome by using a new timbre at Level 4. This method would soon lose its effectiveness (because there are only so many instruments that a user could remember) and it would not be possible to go much deeper. How can this problem be solved? It was decided that compound earcons [7, 10] could be used if the hierarchy was represented in a different way. The idea of a book was used, to represent the hierarchy where there are chapters, sections and subsections, for example Chapter 1, Section 1.1, Sub-section 1.1.1. Figure 11 shows how this was mapped on to the hierarchy.

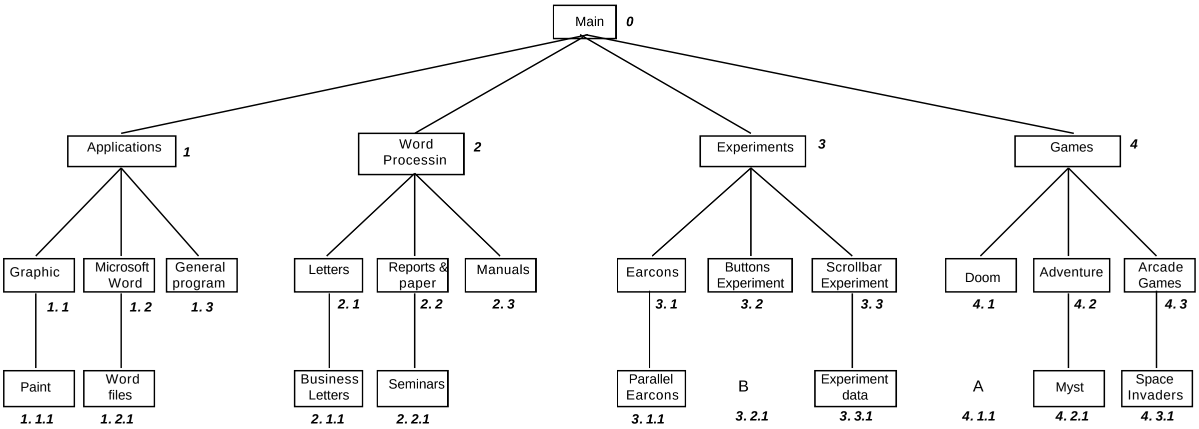


***Figure 11**: The hierarchy used in Experiment 3. The numbers by each node show the position in the hierarchy.*

A set of simple motives was created for each of the numbers and dot. The earcon for any node could then simply be constructed from the concatenation of these motives. This method of representing the hierarchy had the advantage that, with a complete set of motives to represent the numbers, any hierarchy of arbitrary size and depth could be created, overcoming the problem with hierarchical earcons. However, this method adds another level of indirection between the user and the hierarchy. In the previous experiments a user had to recognise a sound and then map that sound to the hierarchy. Using compound earcons the user would have to map the sound to a number and then the number to the position in the hierarchy. This could potentially make this method more difficult to use, therefore an experimental investigation was needed to assess the effectiveness of this method.

The same hierarchy was again used so that it would be possible to compare the results against those of the first experiment. A new set of 15 participants were used (students from the University of Glasgow), none of whom had been involved in any of the previous experiments. The ability of participants to recognise new, unheard earcons was again studied to see if users could recognise new earcons based on the rules for creating the sounds. To make a more realistic test, a whole new Level 2 family was created along with the sounds for A and B. The new family was for Internet applications and was given the number 5.

8.1 Hypotheses

The main hypothesis was that participants would be able to recall their position in the hierarchy with the same accuracy as Experiment 1. 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.

8.3 Sounds used

Simple motives were constructed to represent the numbers 0 - 4 (see Figure 11). These were just single notes (1 sec. duration) played at C_3 (261Hz). The sounds were high-quality (as in Experiment 1). They were created on a Yamaha TG100 synthesiser and played via HyperCard. Table 6 shows the instrument used for each number. These instruments were chosen because each one sounded distinctive and different from all of the others. Earcons for each node were created by concatenating these sound components.

Number	0/5	1	2	3	4	Dot
Instrument	Sitar	Piano	Orchestral hit	Bell	Flute	Marimba

***Table 6**: The instruments used for the numbers in Experiment 3.*

For numbers greater than 4 (of which only 5 was needed in this experiment, to represent the new family) the instruments were re-used but two notes were played. For example, 5 was two notes played on the sitar. For numbers greater than 9 the two motives would be added together, for example 10 was a piano followed by a sitar. This allowed the differentiation of 11 from 1.1 because 11 would not have the separator dot between the two piano notes.

Creating earcons in this way meant the participants did not have to remember more than seven rules. There were five sounds for 0-4, a rule indicating 5-9, and a sound for dot. This meant that the total number of rules was in the range of 7±2, making them as easy to remember as possible. Any more than this then users might not be able to remember them.

8.4 Training and testing

Training was the same as in Experiment 1 so that consistency was maintained and the results could be directly compared. Participants had the set of sounds described to them and then had five minutes to use the system themselves and learn the sounds.

During testing the same set of 14 questions was asked as in Experiment 1 (see Table 2). This included the new, unheard earcons for nodes A and B as shown in Figure 11. To make the test for the new earcons more realistic, in addition to A and B a whole new family was created. Participants were presented with the sound for number 5 (representing a new family of Internet applications) and told that it was for a new family at Level 2. Two earcons from within this family (5.3 and 5.1.1) were then presented to the participants. This meant they were asked 16 questions in total. The participants were thus using a hierarchy of 36 possible nodes if the new family was included. This would give more detailed information about participants’ abilities to recognise new, unheard earcons.

9. RESULTS OF EXPERIMENT 3

The overall recall rate of the trained earcons (questions 1-12) was very good, with 97% recalled correctly (see Figure 12). This was significantly better than the 80% recall rate of the trained earcons in Experiment 1 ($F_{1,22}$=13.89, p=0.001). The results indicate that compound earcons were a very successful method of presenting hierarchy information. The recognition rate of the new, unheard earcons was also very good with an average of 97% again recognised correctly (see questions 13-16 in Figure 12). This indicates that participants could recognise the new compound earcons well.

10. DISCUSSION OF EXPERIMENT 3

This new design for the earcons was significantly better than the hierarchical earcons in Experiment 1. This indicates that compound earcons can provide very usable navigation cues in non-visual hierarchies of information. The advantage of using this type of earcon is the possibility of creating arbitrarily sized hierarchies with great flexibility for changing the size and shape of the structure without re-training the users. This is of major benefit to TBI developers. However, the extra level of indirection between the sound and the thing it represents may make compound earcons harder to use, perhaps by increasing the workload in a task when they are used. The experiment here has shown that it is possible to represent hierarchies in this way and a task-based study is now needed to find out what affect they will have in a real situation. In further studies, NASA TLX workload measures will be used to investigate the subjective workload effects, as it is important to make sure that any increase in workload does not hinder the user with the rest of the task they are trying to perform (for example, remembering the amount of money in a bank account).

One other problem that compound earcons have is that they get longer the deeper one gets into the hierarchy. This may cause several problems. The first is that the user must listen to the full earcon before he/she gets the location information. If only part of the earcons is heard, for example the user is on node 1.1.1 but decides to move on to another node after only hearing 1.1, then he/she may become confused about the location. In addition, the longer the sound gets the harder it will be to recall. When there are many motives needed to make up the full earcon, the user may forget the initial motives when hearing the latter ones (this is know as the *recency effect* [2] - the term used to describe the enhanced recall of most recently presented items), again causing

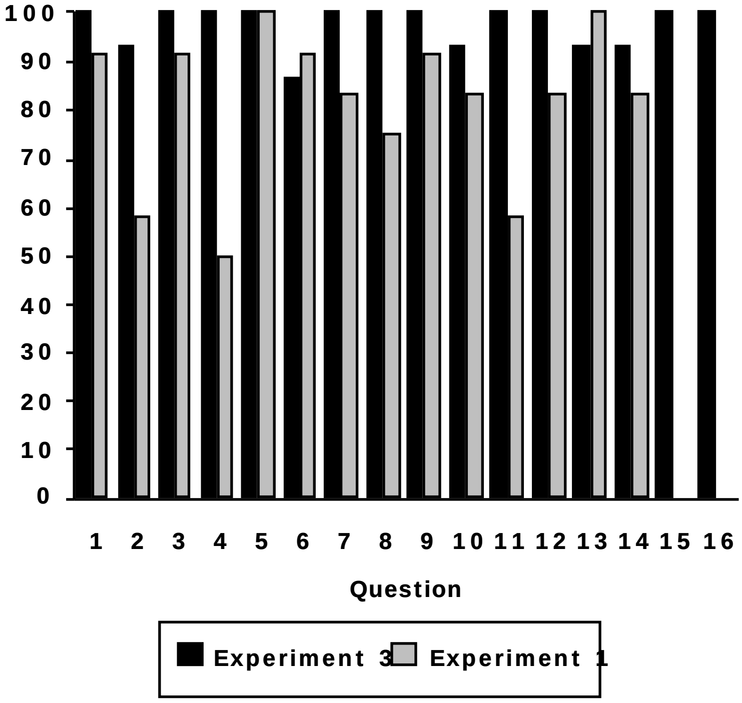


Figure 12*: The recall rates for the earcons in Experiment 3 compared to those in Experiment 1. Questions 1 -12 are the pre-trained earcons and 13-16 are the new earcons. In Experiment 1 only two new earcons (13 and 14) were tested therefore there are no questions 15 and 16 for that experiment.*

confusions in location. More investigations are needed to find out the maximum size of the hierarchy that could be represented with such earcons.

Part of the problem of the length of compound earcons has been addressed by *parallel earcons* [7, 13]. These allow the component parts of a compound earcon to be played concurrently, rather than sequentially. Results of this work showed that recall rates were not reduced when two motives were played together. The use of parallel earcons could perhaps speed up the presentation of the earcons representing the hierarchy. However, in the work on parallel earcons two separate earcons were played at the same time. In this work the order of presentation defines the node so the motives must be presented in the correct order. Further investigation is needed to use parallel earcons for navigational information in hierarchies.

11. OVERALL DISCUSSION

The three experiments have demonstrated the power of earcons for providing information in sound. In this section some general issues will be discussed.

One interesting piece of information came out of the first experiment about the design of the earcons. As mentioned, earcons are abstract so the sounds they are made from do not have any intuitive link to the thing they represent. However, some participants constructed meanings for the sounds which they reported in post-experiment debriefings. Several said that they felt that each group of nodes at Level 3 of the hierarchy (shown in Figure 4) was a triangular structure. The left node formed the left side (rising up towards the top), the centre node the top (rising a little, reaching the top, then falling a little) and the right node the right side (falling back towards the bottom). They said that this helped them to remember the earcons as it gave the rhythms a meaning - it was not just an abstract mapping of rhythm to node. The earcons were not constructed with this in mind. If listeners do try to make sense of the structure in this way then we should make use of this and design sounds that can be mapped to the structure, so improving recall rates in future earcon designs. Further experiments are needed to find out more about how this could be exploited and how the participants constructed their mappings.

If users are placing intuitive interpretations on the earcons it means that they become closer to Gaver’s *auditory icons* [14]. These are natural sounds that have an intuitive link to the object or action they represent. This is supposed to make them easier to learn. By careful construction of the earcons this advantage could be gained by earcons. In the continuum from abstract sounds (earcons) to representational sounds (auditory icons) these *representational* earcons fall in the centre - they have some abstract properties and some representational ones. The compound earcons of Experiment 3 lose this representational advantage because the sounds are completely abstract and can be combined in different ways. Even though they gained high rates of recall in the experiment here there may be situations where the more representational earcons from Experiment 1 might prove advantageous. Further research is needed to investigate this further.

There is subjective evidence to show that participants in Experiment 3 had little difficulty learning the motives for the numbers. As mentioned, participants were given five minutes to learn the sounds themselves after they had been trained by the experimenter. In Experiment 1 (which had the same training) the participants used the full five minutes to learn the sounds. In Experiment 3 most participants complained that five minutes was too much, that they knew the sounds and did not need to spend more time learning them (a fact also born out by the high rates of recall achieved). Because the rules describing the sounds were simple learning was very easy. As above, this is one of the advantages claimed for auditory icons. With careful design of the rules describing the earcons, auditory icons may not have an advantage over compound earcons, especially when a set of 36 sounds has to be learned. Again further investigation is needed on this. Applying the different training techniques from Experiment 2 to Experiment 3 would show, for example, what rates of recall could be achieved with purely textual training of compound earcons.

12. GUIDELINES FOR USING EARCONS FOR NAVIGATION CUES

Some guidelines for designers creating earcons can be drawn from the work reported here. These are in addition to the existing guidelines from Brewster *et al.* [12]. The main guidelines are:

Navigation cues: Earcons can provide good navigational cues in a menu hierarchy, for example a TBI. When designing the sounds start with a neutral sound to represent the root of the hierarchy. Use a combination of instrument and register at the next level and then rhythm at the next. These allow the structured to be expanded easily. If a new sub-tree is needed at Level 2 then all this is necessary is a new instrument. If expansion is needed at Level 3 then more rhythms can be created. For deeper levels then the cues must be made very obvious (because users have much to remember to work out their location when they get deep into the hierarchy), for example adding a level indicator instrument.

Tempo: This proved to be an ineffective way of indicating a new level of the hierarchy. Users were unable to differentiate between Levels 3 and 4 using tempo. It was not distinctive enough on its own. It also did not provide the possibility of expanding the hierarchy at Level 4 because there are not enough recognisable different tempos. Tempo could be used (like register in Experiment 1) alongside another parameter to increase the amount of difference between one level and another to aid recall.

Sound quality: Low quality sounds can be used for presenting earcons. However, reducing the quality will reduce recall rates so the earcons must be well designed to keep recall rates high.

Training: Allowing users time to learn earcons themselves ('active' learning) was very beneficial to recall rates and would be cost-effective for a telephone service provider. Providing both written instructions and free call time would give good recall rates without a high training cost.

Hierarchical and compound earcons: Hierarchical earcons have limitations which may make it difficult to represent deep structures. Compound earcons can provide an alternative way of doing this. One other advantage they have is that they are very simple to create. A set of motives to represent the numbers 0-9 and dot can be constructed. An earcon for any point in the hierarchy can then be created by simply concatenating the required motives. It is even possible for this to be done automatically, in a similar way to the British Directory Enquiries system generates a telephone number from a set of pre-recorded spoken numbers (samples of the individual numbers are concatenated to produce the required telephone number as required). A simple test of this was done when sound was added to the menu hierarchy of a software prototype of a mobile telephone. It proved simple to construct the sounds on the fly to represent the different menus.

13. FUTURE WORK

These earcons will now be used as they are in an interface to a mobile communicator device for the TIDE ACCESS project. The sounds will be played in the background when the user is in the hierarchy of pictographic symbols. The sounds will change when he/she moves up, down, left or right to a new node, indicating the new position in the hierarchy. The change in the earcon and a small increase in intensity will capture the listener's attention and indicate the new position. The earcon will then fade back to a lower intensity and recede into the background of the listener's consciousness. The user will only attend to (or 'tune-in') the sound again if he/she gets lost and wants navigational information. In this way the feedback will not be annoying for users [7].

The next stage for work in TBI's is to add the sounds to a real system and perform task-based studies. Further tests will extend the breadth and depth of the hierarchies represented and will investigate different training techniques. One important next step will be to investigate if users trained on one hierarchy can navigate around a different structure if the sounds are constructed using the same rules (without further training).

Studying real tasks will also allow the investigation of the affects that the sounds might have on the interaction as a whole. It could be possible that the sounds might affect the intelligibility of the speech. Therefore, the best ways of combining speech and non-speech sounds must be considered. For example, the sounds could be played at the same time as the speech, before it or after it. Studying the use of earcons in real telephone-based tasks will also be important to assess the level of recall needed to ensure the sounds are useful. Experiment 2 showed that even with limited training level information could be extracted from the earcons, but at what rate of recall are the sounds ineffective? In the area of speech recognition it is generally recognised that around a 95% recall rate is needed for the recognition to be effective and usable. There has been no research so far on what rate is required of the earcons.

14. CONCLUSIONS

The aim of the work described in this paper was to discover if earcons could be used to represent hierarchical structures. This is important for interfaces where graphics cannot be used or where the screen is already full. The particular problem addressed was the common one of users getting lost in hierarchies of information. The experiments also investigated some fundamental questions about earcons themselves, such as the size of hierarchy that could be represented, recall rates over time and how to construct the sounds to represent the hierarchy.

The first experiment investigated the ability of earcons to represent a 27 node hierarchy. A recall rate of 81.5% was achieved after only a short period of training, indicating that earcons could represent such a structure. Listeners could listen to an earcon and tell where it came from in the hierarchy and hence their location. Listeners could also recognise new earcons that had not been heard before by using the rules from which the other earcons were constructed. This shows that users could easily learn those rules. Therefore, earcons are a robust and extensible method of representing hierarchies.

The second experiment answered fundamental questions about the use of earcons as navigation cues within telephone-based interfaces (TBI's). The research in this paper has shown that reductions in the quality of sound that occur with telephone systems can be offset by improvements in the design of earcons, thus making earcons a good method for providing navigation cues in TBI’s.

This research was also the first investigation into the memorability of earcons over time. This is important for earcons in general and also for their use in TBI’s. If users do not use the TBI frequently then they must be able to remember the sounds in order to use them as navigation cues. Results here showed that there was no difference in the recall of earcons a week after their first presentation. This shows that they are a robust method of presenting navigation information.

Results showed that training techniques affected the recall rates of earcons. Training techniques are a cost to the provider of a telephone service. The provider must ensure that users can use the sounds whilst minimising the amount spent on training. Results here indicated that an online tutorial plus a short period of free call time can enable users to reach high recall rates without much training cost. Even if just a textual description of the sounds is given then listeners could still extract some useful navigation information.

The final experiment showed that by using compound earcons rather than hierarchical earcons to represent the hierarchy recall rates could be significantly increased, with 97% recalled correctly with only a small amount of training. This shows that compound earcons could represent a hierarchical structure well. Listeners could again recognise new, unheard earcons equally as well as the pre-trained ones. One reason for this was that there were only seven rules defining the sounds for any node. Now that these fundamental questions about earcons have been answered, designers of systems who need to present navigation information in a non-visual way can use earcons to greatly enhance the usability of their systems.

ACKNOWLEDGEMENTS

Part of this research was supported by ERCIM Research Fellowship 94-04 and EPSRC grant GR/L79212. Thanks to Atte Kortekangas and Veli-Pekka Raty at VTT, and to Peter Wright at the University of York, UK for helping with some statistical analysis. Many thanks to Ayotunde Aboaba who helped greatly in this project as part of his MSc. in Information Technology at Glasgow and also to Adrian Capriotti, Cordy Hall and Phil Gray. Thanks also to the reviewers for their insightful comments.

REFERENCES

1. Aboaba, A. *Using sound in telephone-based interfaces*. MSc IT Thesis, University of Glasgow, 1996.
2. Baddeley, A. *Human Memory: Theory and Practice*. Lawrence Erlbaum Associates, London, 1990.
3. Barfield, W., Rosenberg, C. and 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.
4. Baumgart, D., Johnson, J. and Helmstetter, E. *Augmentative and Alternative Communication Systems for Persons with Moderate and Severe Disabilities*. Paul Brookes Publishing Co., Baltimore, Maryland, 1990.
5. Blattner, M., Papp, A. and Glinert, E. Sonic enhancements of two-dimensional graphic displays. In *Proceedings of ICAD'92* (Santa Fe Institute, Santa Fe) Addison-Wesley, 1992, pp. 447-470.
6. Blattner, M., Sumikawa, D. and Greenberg, R. Earcons and icons: Their structure and common design principles. *Human Computer Interaction* 4, 1 (1989), 11-44.
7. Brewster, S.A. *Providing a structured method for integrating non-speech audio into human-computer interfaces*. PhD Thesis, University of York, UK, 1994.
8. Brewster, S.A., Raty, V.-P. and Kortekangas, A. Earcons as a method of providing navigational cues in a menu hierarchy. In *Proceedings of BCS HCI'96* (London, UK) Springer, 1996, pp. 169-183.
9. Brewster, S.A., Wright, P.C. and Edwards, A.D.N. A detailed investigation into the effectiveness of earcons. In *Proceedings of ICAD'92* (Santa Fe Institute, Santa Fe) Addison-Wesley, 1992, pp. 471-498.
10. Brewster, S.A., Wright, P.C. and Edwards, A.D.N. An evaluation of earcons for use in auditory human-computer interfaces. In *Proceedings of ACM/IFIP INTERCHI'93* (Amsterdam) ACM Press, Addison-Wesley, 1993, pp. 222-227.
11. Brewster, S.A., Wright, P.C. and Edwards, A.D.N. The design and evaluation of an auditory-enhanced scrollbar. In *Proceedings of ACM CHI'94* (Boston, MA) ACM Press, Addison-Wesley, 1994, pp. 173-179.
12. Brewster, S.A., Wright, P.C. and Edwards, A.D.N. Experimentally derived guidelines for the creation of earcons. In *Adjunct Proceedings of BCS HCI'95* (Huddersfield, UK), 1995, pp. 155-159.
13. Brewster, S.A., Wright, P.C. and Edwards, A.D.N. Parallel earcons: Reducing the length of audio messages. *International Journal of Human-Computer Studies* 43, 2 (1995), 153-175.
14. Gaver, W. The SonicFinder: An interface that uses auditory icons. *Human Computer Interaction* 4, 1 (1989), 67-94.
15. Gaver, W., Smith, R. and O'Shea, T. Effective sounds in complex systems: The ARKola simulation. In *Proceedings of ACM CHI'91* (New Orleans) ACM Press, Addison-Wesley, 1991, pp. 85-90.
16. Maguire, M. A human-factors study of telephone developments and convergence. *Contemporary Ergonomics* (1996), 446-451.

17. Rosson, M.B. Using synthetic speech for remote access to information. *Behaviour Research Methods, Instruments and Computers* 17, 2 (1985), 250-252.
18. Schumacher, R.M., Hardzinski, M.L. and Schwartz, A.L. Increasing the usability of interactive voice response systems. *Human Factors* 37, 2 (1995), 251-264.
19. Slowiaczek, L.M. and Nusbaum, H.C. Effects of speech rate and pitch contour on the perception of synthetic speech. *Human Factors* 27, 6 (1985), 701-712.
20. Stevens, R. *Principles for the Design of Auditory Interfaces to Present Complex Information to Blind people*. PhD Thesis, University of York, UK, 1996.
21. Stevens, R.D., Brewster, S.A., Wright, P.C. and Edwards, A.D.N. Providing an audio glance at algebra for blind readers. In *Proceedings of ICAD'94* (Santa Fe Institute, Santa Fe) Santa Fe Institute, 1994, pp. 21-30.
22. Stifelman, L., Arons, B., Schmandt, C. and Hulteen, E. VoiceNotes: A Speech Interface for a Hand-Held Voice Notetaker. In *Proceedings of ACM/IFIP INTERCHI'93* (Amsterdam) ACM Press, Addison-Wesley, 1993, pp. 179-186.
23. Sumikawa, D., Blattner, M., Joy, K. and Greenberg, R. *Guidelines for the syntactic design of audio cues in computer interfaces*. Lawrence Livermore National Laboratory, 1986, Technical Report, UCRL 92925.
24. Sumikawa, D.A. *Guidelines for the integration of audio cues into computer user interfaces*. Lawrence Livermore National Laboratory, 1985, Technical Report, UCRL 53656.
25. Wolf, C., Koved, L. and Kunzinger, E. Ubiquitous Mail: Speech and graphical interfaces to an integrated voice/email mailbox. In *Proceedings of IFIP Interact'95* (Lillehammer, Norway) Chapman & Hall, 1995, pp. 247-252.
26. Yankelovich, N., Levow, G. and Marx, M. Designing SpeechActs: Issues in speech user interfaces. In *Proceedings of ACM CHI'95* (Denver, Colorado) ACM Press, Addison-Wesley, 1995, pp. 369-376.