Brewster, S.A. (1997) Navigating telephone-based interfaces with earcons. In, Thimbleby, H. and O'Conaill, B. and Thomas, P.J., Eds. *12th British Computer Society HCI Conference, 12-15 August 1997*, pages pp. 39-56, Bristol, England.

http://eprints.gla.ac.uk/3250/

Navigating Telephone-Based Interfaces with Earcons

Stephen Brewster

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

Non-speech audio messages called *earcons* can provide powerful navigation cues in menu hierarchies. However, previous research on earcons has not addressed the particular problems of menus in telephone-based interfaces (TBI's) such as: Does the lower quality of sound in TBI's 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 excellent navigation cues for telephone-based interfaces.

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

1 Introduction

Previous research has shown that structured non-speech sounds can provide powerful navigation cues in menu hierarchies for non-visual interfaces (Brewster *et al.*, 1996a and Brewster *et al.*, 1996b). However, this work did not consider the particular problems of telephone-based interfaces (TBI's). This paper describes an experiment to investigate the ability of non-speech sounds to provide navigational cues in such interfaces.

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 (Maguire, 1996). The work described here will improve the usability of TBI's.

The telephone itself allows only a limited form of interaction. There is no graphical display (although some do have small LCD displays) so output is limited to speech and simple sounds. Users provide input via the keypad (although speech recognition is sometimes used). These techniques are limited and reduce the usability of telephone-based systems (Schumacher *et al.*, 1995).

One common problem when interacting in TBI's is that users get lost in the hierarchy of menus that they must go through to reach an option or function (Rosson, 1985 and Wolf *et al.*, 1995). For example, in a telephone banking system users might call their bank and navigate through a hierarchy of voice menus to find the service required. However, they may get lost in the menu structure before they get to the option they want. As Yankelovich *et al.* (1995, p 369) say: "These {telephone-based} interfaces, however, are often characterised by a labyrinth of invisible and tedious hierarchies which result when menu options outnumber telephone keys or when choices overload users' short-term memory". After analysing a similar TBI, Rosson (1985, p 251) concluded: "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. She suggested that this caused many of the usability 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 (Slowiaczek & Nusbaum, 1985). 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.

As a solution to the problems of navigation Rosson suggested using more speech. However, any extra speech is likely to make the problem worse instead of better. If navigation cues are given in speech then it will obstruct the information the user is trying to access. It is suggested here that speech cannot provide the necessary feedback. As an alternative, non-speech sounds can give the navigation cues (Brewster *et al.*, 1996a). A hierarchical system of sounds 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. The sounds could do this without interfering with the speech presenting information (just as one can simultaneously listen to the music and lyrics of a song). This is a similar approach to that taken by Stevens (1996) and Stevens *et al.* (1994) 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.

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 (Maguire,

1996 and Schumacher *et al.*, 1995) 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 (Brewster, 1994). Unlike speech, non-speech sound is universal; 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.

2   Earcons

The non-speech sounds used for this investigation were based around structured audio messages called *Earcons* (Blattner *et al.*, 1989 and Brewster, 1994). Earcons are abstract, musical tones that can be used in structured combinations to create sound messages to represent parts of an interface. Detailed investigations of earcons by Brewster *et al.* (1993) 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 (such as timbre, register, intensity, pitch and rhythm) *hierarchical earcons* can be created (Blattner *et al.*, 1989) which allow the representation of hierarchical structures.
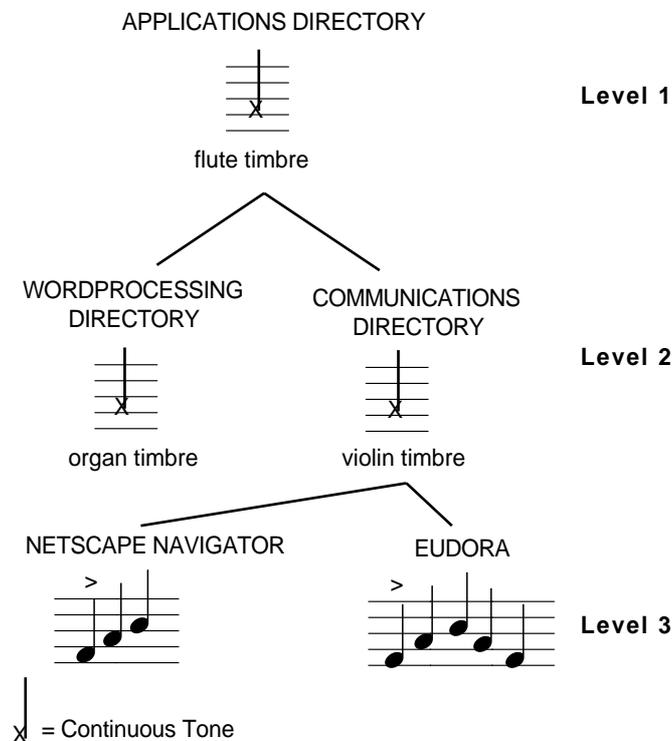

Figure 1: A hierarchy of earcons representing a family of applications.

Figure 1 shows a simple example hierarchy of earcons based on one possible family of applications. 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 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 telephone-based menu structures. The structure could be extended and users would not require retraining. Wolf *et al.* (1995) confirm the usefulness of this approach and suggest that this might be a solution to some of the problems they had with their combined voicemail and email speech interface system: "Replacing much of the text-to-speech feedback with brief distinctive earcons would make traversal of the mailbox more efficient".

2.1 *Previous attempts to use earcons to present hierarchy information*

There is little work on the use of sound to represent hierarchical structures. There has only been one previous study of the use of non-speech sounds for navigation information. This was performed by Brewster *et al.* (1996a,1996b). In this experiment we tried to represent a medium sized hierarchical structure of four levels and 25 nodes. Figure 2 shows the hierarchy used.

We created a set of 25 earcons on four levels based around fixed rules. These rules described the sound at each of the nodes. For example the rule for Level three used three different rhythms to indicate whether the node was a left, centre or right one (for more on the rules see below - in our current experiment Levels 1-3 of the hierarchy used the same rules as this previous experiment). Users were trained by the experimenter describing and playing each of the sounds. They were then given five minutes to use the system to learn for themselves.

Users had to identify their location by listening to an earcon and indicating their position in the hierarchy. Results showed that participants could identify their location with over 80% accuracy (see bar marked 'Control' in Figure 6), indicating that earcons were a powerful method of communicating hierarchy information. Participants were also tested to see if they could identify where previously unheard earcons would fit in the hierarchy. These new earcons were constructed using the same rules as the other earcons. If the users had learned the rules for earcon construction then they should have been able to work out their location from these new sounds. The results showed that they could do this with accuracy equal to the sounds they had heard before. These results showed that earcons were a robust and extensible method of communicating hierarchy information in sound. However, this work attempted a general solution for representing menu structures in sound and was not aimed particularly at TBI's.
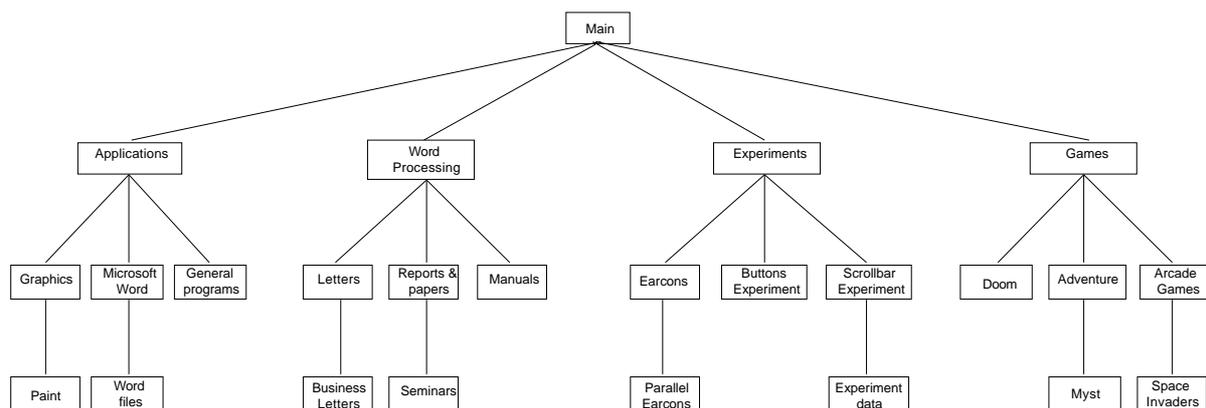
Figure 2: The hierarchy used by Brewster *et al.* (1996a) and in the current experiment.

2.2   Problems with the previous experiment

The experiment by Brewster *et al.* gave a strong indication that earcons could provide navigational cues in TBI's. However, there were important question left unanswered:

- The earcons used were of CD quality (16bit 44kHz) - 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 Brewster *et al's.* experiment 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 been done on this. 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 telephone interface because of the high cost to the provider of the service. Therefore an investigation is needed into alternative, more practical training techniques.

These five points motivated the research described in this paper. 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.

3   The experiment

The aim of the experiment described here was to investigate the five problems described above. Figure 2 shows the hierarchy used. This was the same as that used by Brewster *et al.*

(1996a) so there was consistency between the experiments and the results could be directly compared. The hierarchy had 25 nodes on four levels.

3.1 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 Brewster *et al.'s* 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 presentation 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.

3.2 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.

3.3 Sounds used

The earcons were based on those from Brewster *et al.* (1996a) to maintain consistency. They were designed using the guidelines proposed by Brewster *et al.* (1995) . The sounds were all played by HyperCard on an Apple Macintosh computer through a telephone handset (to simulate the output of a real telephone). The sounds were generated on a Yamaha TG100 sound synthesiser and recorded by a Macintosh at a sampling rate of 8bit 11kHz (sounds with this sampling rate played over the telephone handset were of telephone quality). Each sound played for 7.5 secs. The sounds used at each level of the hierarchy will now be described. For Levels 1 to 3 the sounds were the same as for Brewster *et al's* previous experiment except stereo position was not used. For more details on these sounds see the previous paper.

Level 1: For the top level of the hierarchy ('Main' in Figure 2) a constant sound with a flute timbre was used (see Table 1). It had a pitch of D_3 (261Hz).

Level 2: Each family was given a separate timbre and register. Table 1 shows these. Register was lowest on the left and highest on the right following the conventional musical pattern.

Nodes	Timbre	Register
Main	Flute	D_3
Applications	Electric organ	C_4
Word Processing	Violin	C_3
Experiments	Drum/synthesiser	C_2
Games	Trumpet	C_1

Table 1: The timbre and register for Levels 1 and 2 of the hierarchy.

The continuous sound was inherited from the Level 1 earcon but the instrument and pitch were changed.

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 3 shows the rhythms used. For example, from Figure 2 'Graphics', 'Letters', 'Earcons' and 'Doom' all had the left node rhythm. Each of these rhythmic groups repeated continuously once every 2.5 seconds. As Figure 3 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 (Brewster *et al.*, 1995).
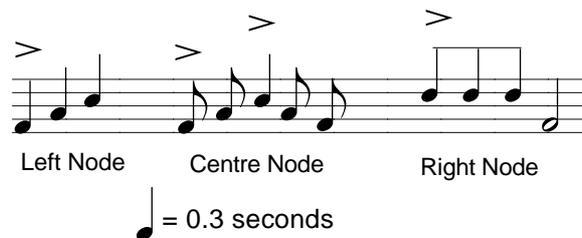

Figure 3: The rhythms used for Level 3 of the hierarchy.

At this level the earcons inherited timbre and register from Level 2. This meant, for example, that 'Graphics' used the left node rhythm described in Figure 3 and it was played with an electric organ timbre in the register of C_4.

Level 4: In the previous experiment by Brewster *et al.* (1996a) 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. 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 4). It was hoped that this would provide stronger information to the participants that they were on Level 4.
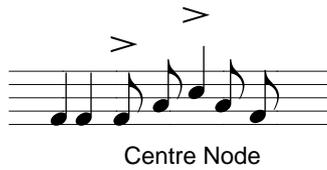
Centre Node

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

3.4 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 us 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 2.

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 2: 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 us to compare directly our results with those of Brewster *et al.* (1996a).

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 5 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.
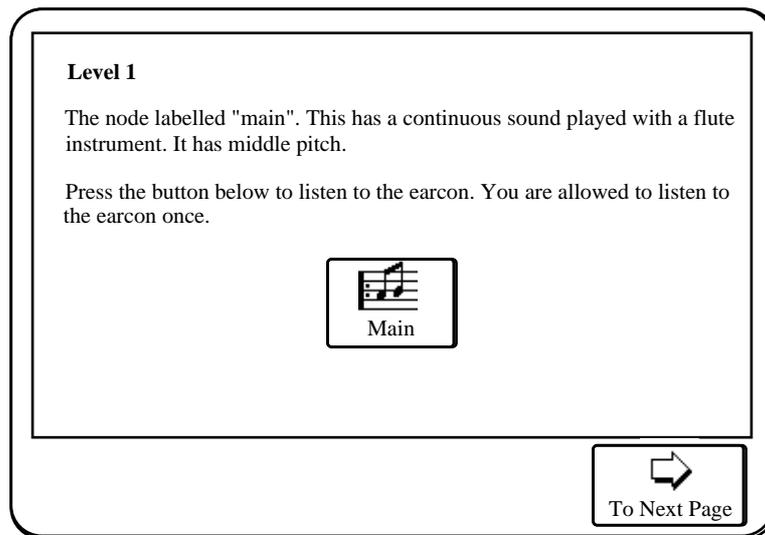
Figure 5: Training screen for the earcon 'Main' in Group B.

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 5. 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 the experiment started; 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)?

During the training participants could look at a map of the hierarchy (similar to Figure 2). 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.

3.5 Testing

The participants heard twelve earcons during testing. These were randomly selected from all of the sounds in the hierarchy (and were the same as those used in the previous experiment to maintain consistency). The same set of earcons was presented to each of the participants. An earcon was played and the participants then had to say where the it fitted into the hierarchy.

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.

4   Results

4.1   Comparison with previous results

The first comparison undertaken was to compare the results of Group A(1) with the previous results of Brewster *et al.* (here called the Control Group). 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 6. 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.
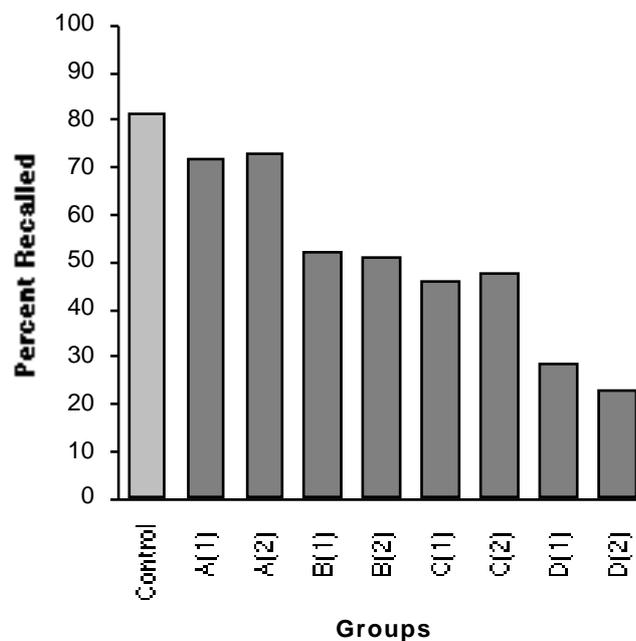

Figure 6: Recall rates for each of the groups for both presentation sessions. (1) shows results of the first testing session and (2) the second. The Control Group shows the results from the previous experiment by Brewster *et al.* (1996a). This group appears once because there was no re-testing in that experiment.

In Brewster *et al.'s* previous work, 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 Brewster *et al.'s* previous work 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.

4.2 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 6 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.

4.3 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 6 shows the 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.

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.

A more detailed analysis of the D(1) scores was undertaken to find out what information participants managed to extract in this condition. Results showed that participants got the level of the earcon correct 67.3% of the time (where 25% would occur by chance, so they were performing significantly better than chance).

5 Discussion

5.1 Comparison with previous results

Results from Group A(1) showed no significant differences to those of Brewster *et al.* (1996a). 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 (Wolf *et al.*, 1995 and Yankelovich *et al.*, 1995). After Brewster *et al.'s* previous work 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 show that the problems caused by sound quality can be overcome by better design of the earcons. In future experiments earcons at Levels 1-3 of the hierarchy will be improved 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. The user would listen to this sound and from it work out his/her location, therefore avoiding becoming lost in the hierarchy. This and our previous experiment have shown that earcons can provide good navigation cues. The next stage of this work is to incorporate earcons into a real telephone-based system and evaluate their effectiveness in real-world conditions.

5.2   Recall over time

The results here are the first to demonstrate the recall of earcons over time. In Brewster *et al.* (1992) and Brewster *et al.* (1993) 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.

5.3   The effects of training

Training type did have a significant effect on recall rates. Personal training by the experimenter gave the highest rates with the lowest coming from the purely textual training. The results showed that the only significant difference was between Groups A and D. However, there was much within-group variance which may have masked some potential differences. By looking at Figure 6 we can see three groupings 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 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' seems to help 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, even though the overall analysis presented in Figure 9 showed low recall rates, participants were able to extract some useful navigation information from the earcons. In fact, they were get the correct level in the hierarchy 67.3% of the time. 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 also be possible to increase recall by improving the design of the textual description. Remember 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 we might be able to improve rates (this would also apply to the other groups).

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

6   Conclusions

The experiment described here has answered many fundamental questions about the use of earcons in general and in particular about their application as navigation cues within telephone-based interfaces (TBI's). Brewster *et al.* (1996a) showed that earcons could be used as navigation cues but left many unanswered questions about their usefulness under the restricted conditions of the telephone. 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 system 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 do affect 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 indicate that an online tutorial plus a short period of free call time can enable users to reach high recall rates without much training cost. Now that these fundamental questions have been answered, designers of telephone services can use earcons to provide navigation cues to greatly enhance the usability of their systems.

Acknowledgements

References

Any references by Brewster (along with sound samples and further information) are available electronically from http://www.dcs.gla.ac.uk/~stephen/

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

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

Brewster, S.A., Raty, V.-P. & Kortekangas, A. (1996a). Earcons as a method of providing navigational cues in a menu hierarchy. In Sasse, Cunnigham & Winder (Eds.), *Proceedings of BCS HCI'96*, (pp. 169-183), London, UK: Springer.

Brewster, S.A., Raty, V.-P. & Kortekangas, A. (1996b). *Using earcons to provide navigational cues in a complex menu hierarchy* (Technical Report No. TR-1996-24). Department of Computing Science, University of Glasgow.

Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1992). A detailed investigation into the effectiveness of earcons. In Kramer (Ed.), *Proceedings of ICAD'92*, (pp. 471-498), Santa Fe Institute, Santa Fe: Addison-Wesley.

Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1993). An evaluation of earcons for use in auditory human-computer interfaces. In Ashlund, Mullet, Henderson, Hollnagel & White (Eds.), *Proceedings of ACM/IFIP INTERCHI'93*, (pp. 222-227), Amsterdam: ACM Press, Addison-Wesley.

Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1995). Experimentally derived guidelines for the creation of earcons. In Kirby, Dix & Finlay (Eds.), *Adjunct Proceedings of BCS HCI'95*, (pp. 155-159), Huddersfield, UK

Maguire, M. (1996). A human-factors study of telephone developments and convergence. *Contemporary Ergonomics*, 446-451.

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

Schumacher, R.M., Hardzinski, M.L. & Schwartz, A.L. (1995). Increasing the usability of interactive voice response systems. *Human Factors*, **37**, 251-264.

Slowiaczek, L.M. & Nusbaum, H.C. (1985). Effects of speech rate and pitch contour on the perception of synthetic speech. *Human Factors*, **27**, 701-712.

Stevens, R. (1996) *Principles for the Design of Auditory Interfaces to Present Complex Information to Blind people*. PhD Thesis, University of York, UK.

Stevens, R.D., Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1994). Providing an audio glance at algebra for blind readers. In Kramer & Smith (Eds.), *Proceedings of ICAD'94*, (pp. 21-30), Santa Fe Institute, Santa Fe: Santa Fe Institute.

Wolf, C., Koved, L. & Kunzinger, E. (1995). Ubiquitous Mail: Speech and graphical interfaces to an integrated voice/email mailbox. In Nordby, Helmersen, Gilmore & Arnesen (Eds.), *Proceedings of IFIP Interact'95*, (pp. 247-252), Lillehammer, Norway: Chapman & Hall.

Yankelovich, N., Levow, G. & Marx, M. (1995). Designing SpeechActs: Issues in speech user interfaces. In Katz, Mack & Marks (Eds.), *Proceedings of ACM CHI'95*, (pp. 369-376), Denver, Colorado: ACM Press, Addison-Wesley.