Parallel Earcons: Reducing the Length of Audio Messages

STEPHEN A. BREWSTER[1], PETER C. WRIGHT[2] AND ALISTAIR D. N. EDWARDS[2]

[1]*VTT Information Technology, Tekniikantie 4 B, P.O. Box 1203, FIN-02044 VTT, Finland*

stephen.brewster@vtt.fi

[2]*HCI Group, Department of Computer Science, University of York, Heslington, York, Y01 5DD, UK.*

[pcw, alistair]@minster.york.ac.uk

This paper describes a method of presenting structured audio messages, *earcons*, in parallel so that they take less time to play and can better keep pace with interactions in a human-computer interface. The two component parts of a compound earcon are played in parallel so that the time taken is only that of a single part. An experiment was conducted to test the recall and recognition of parallel compound earcons as compared to serial compound earcons. Results showed that there are no differences in the rates of recognition between the two groups. Non-musicians are also shown to be equal in performance to musicians. Some extensions to the earcon creation guidelines of Brewster, Wright & Edwards (1992) are put forward based upon research into auditory stream segregation. Parallel earcons are shown to be an effective means of increasing the presentation rates of audio messages without compromising recognition rates.

1. Introduction

If non-speech audio feedback is to be used at the human-computer interface it must be able to keep pace with the interactions that occur. If it does not, and either the system has to wait for the sound to finish before continuing or the sound playing refers to an interaction that has completed, then it will not be effective. It will not provide the user with any advantage so there will be no reason to use it. Sound takes place sequentially in time. One way to reduce the length of time an audio message takes, so that it can keep pace, is to play its sequential component parts in parallel. This paper discusses an experiment that attempted to discover if this was an effective method of reducing the length of audio messages.

Using combined graphical and auditory information at the interface is a natural step forward. The two senses combine to give complementary information about the world in our everyday lives. The visual system gives us detailed data about a small area of focus whereas the auditory system provides general data from all around,

alerting us to things outside our peripheral vision. The combination of these two senses gives much of the information we need about our environment. Blattner and Dannenberg (1992) (pp xviii-xix) discuss some of the advantages of using this approach in multimedia/multimodal computer systems:

> "In our interaction with the world around us, we use many senses. Through each sense we interpret the external world using representations and organisations to accommodate that use. The senses enhance each other in various ways, adding synergies or further informational dimensions".

A multimodal interface that integrated information output to both senses could capitalise on the interdependence between them and present information in the most efficient and natural way possible. For example, such an interface might allow us to concentrate our visual attention to one task, perhaps editing a document, but then monitor the state of other tasks on our machine by using listening to the sounds they made.

There is a growing body of research that shows the combination of sound and graphics is a more effective means of communication than graphics alone. Brown, Newsome and Glinert (1989) performed visual search experiments using auditory or visual target cues. Their aim was to reduce visual workload by using multiple sensory modalities. The experiments they conducted showed that the auditory modality could, in some cases, be more effective than the visual one. Their results also indicated that the auditory modality was not as fast although they suggest that "Extracting information from an auditory cue is foreign to most people. It is possible that with a longer training session this difference would no longer exist". Their findings suggest that humans can extract complex information from sound and then act upon it showing that dual-mode interfaces can be effective.

Work by Perrott, Sadralobadi, Saberi and Strybel (1991) illustrated that providing auditory cues can help in the location of visual targets on a display. They used three-dimensional sound to indicate the position of a visual target. The target was sometimes the only stimulus on the display, at other times many distractors were present to make the target more difficult to locate. The target could lie within the central visual field or be outside it. Their results were favourable (p 389):

> "The presence of spatial information from the auditory channel can reduce the time required to locate and identify a visual target even when the target occurs within a restricted region around the initial line of gaze".

Even when the target was close to a participants' focus of visual attention they still located it more rapidly with a sound cue present. Perrott *et al.* go on to say "The advantage of providing auditory spatial information is particularly evident when a substantial shift in gaze is required in the presence of a cluttered visual field". It is often the case with complex graphical interfaces and large, multiple-monitor displays that such situations occur and this research indicates that sound would be very effective in increasing performance with these systems.

Gaver and colleagues have developed several systems that use non-speech sound in conjunction with graphics. In the ARKola system (Gaver, Smith and O'Shea 1991) sound was integrated into the interface to a production line. When the sound was present it helped participants co-operate more freely and have a greater understanding of what was going on in the production line as a whole than when there was no sound.

Brewster, Wright and Edwards (1994) describe an experiment where they tested a sonically-enhanced scrollbar against a standard visual one. Sounds were added to overcome some problems users face when operating standard scrollbars. Sound significantly reduced the time taken by participants on certain tasks, it reduced the mental workload required to perform the tasks and participants strongly preferred the sonically-enhanced scrollbar to the visual one. Brewster *et al.* also measured subjective annoyance. The sonically-enhanced scrollbar was not rated significantly differently to the visual one in this respect. This work again shows the effectiveness of sound at the interface.

In a similar experiment, Brewster, Wright, Dix and Edwards (1994) added sound to on-screen, graphical buttons again to overcome usability problems. In this case the problem was due to the small area of visual focus. Users can mis-hit graphical buttons and not notice. This problem is difficult to solve with extra graphical feedback because the users' attention shifts away from the button directly after it has been pressed (and on to the next task to be performed), so he/she will not be looking at the button when it could indicate an error. Brewster *et al.* used sound to overcome this problem because it can be heard from all around and the user does not have to concentrate attention on the output device to hear it. Brewster *et al.* showed that time to recover from such mis-hit errors could be significantly reduced. Participants strongly preferred the sonically-enhanced buttons over the standard visual ones. As above, annoyance was not increased by adding sound. Brewster (1994) describes, in detail, these and other systems that have benefited from the inclusion of sound.

The examples of auditory interfaces provided show that adding sound can be effective at improving usability. Complex, graphical information can be presented rapidly and if sound is going to accompany it then complex audio messages must also be presented rapidly. This paper describes an experiment to see if sounds could be designed to convey complex information quickly without reducing recognition rates. This paper does not suggest that sound be used to replace graphics at the interface. It just investigates a method by which can sound keep pace with interactions that occur.

2. Earcons

The method for presenting information in sound that will be used here is called *earcons*. Earcons are structured sequences of synthetic tones that can be used in different combinations to create complex audio messages (see Blattner, Sumikawa and Greenberg (1989) , Sumikawa (1985) and Sumikawa, Blattner, Joy and Greenberg (1986)). Blattner *et al.* describe earcons as (p 13) “non-verbal audio messages that are used in the computer/user interface to provide information to the user about some computer object, operation or interaction”. Earcons are composed of motives, which are short, rhythmic sequences of pitches with variable intensity, timbre and register.

A detailed investigation of earcons was undertaken by Brewster, Wright and Edwards (1992) and Brewster, Wright and Edwards (1993) . They showed earcons to be an effective means of communication complex information in sound. They experimentally tested the recall and recognition of different types of earcons. Results from their work showed that 80% accuracy in the recall of earcons could be achieved with careful design of the sounds. One of the most important issues that these experiments brought up was that the participants could not differentiate earcons if there were only small differences between them. Some of the manipulations put forward by Blattner *et al.* (1989) . were not large enough to allow participants to tell the earcons apart; there had to be gross differences between the sounds. Brewster *et al.* (1992) Brewster *et al.* also found that there were no differences in performance between musicians and non-musicians. We also proposed a set of guidelines that could be used in the creation of earcons. The research described here builds on this previous work to try and overcome some of the problems that still exist with earcons.

2.1 THE DRAWBACKS OF COMPOUND EARCONS

One of the most powerful features of earcons is that they can be combined to produce compound messages. Motives for a set of simple operations, such as ‘open’, ‘close’, ‘file’ and ‘program’ could be created. A compound earcon can then be created that gives a sound for ‘open file’ or ‘close program’ by simply concatenating the two motives (see the bottom, serial, earcon in Figure 1).

Interactions between the computer and the user tend to happen quickly and audio feedback must be able to keep up. The main drawback of compound, or *serial*, earcons as proposed by Blattner *et al.* (1989) is that they can take a long time to play (1.3 - 2.6 seconds in our previous experiments). Each motive lasts a particular length of time depending on its notes and the tempo and these are then combined to produce longer compound earcons. Compound earcons could be played more rapidly (at a faster tempo) to overcome this problem but then errors in

recognition may occur. Our previous experiments did not test the maximum speed of playback to find out at what point user's recognition of the earcons broke down.

One alternative method of overcoming this problem is to play the earcon at the same rate but pack the information more densely. This can be done by playing two earcons in *parallel* so that they only take the time of one to play. With parallel compound earcons the individual parameters can be left as they are for serial earcons but two sounds be played at the same time. Figure 1 gives an example of serial versus parallel compound earcons. This method has the advantage that the guidelines from Brewster *et al.* (1992) can still be applied but the disadvantage that it may be harder for the user to differentiate multiple sounds playing simultaneously. This paper describes an experiment to investigate parallel earcons in more detail.

As Blattner, Papp and Glinert (1992) say (p 448): "Our awareness and comprehension of the auditory world around us for the most part is done in parallel". This suggests that parallel earcons could use a natural ability of the human auditory system. Research into auditory attention has looked at some of the problems of presenting

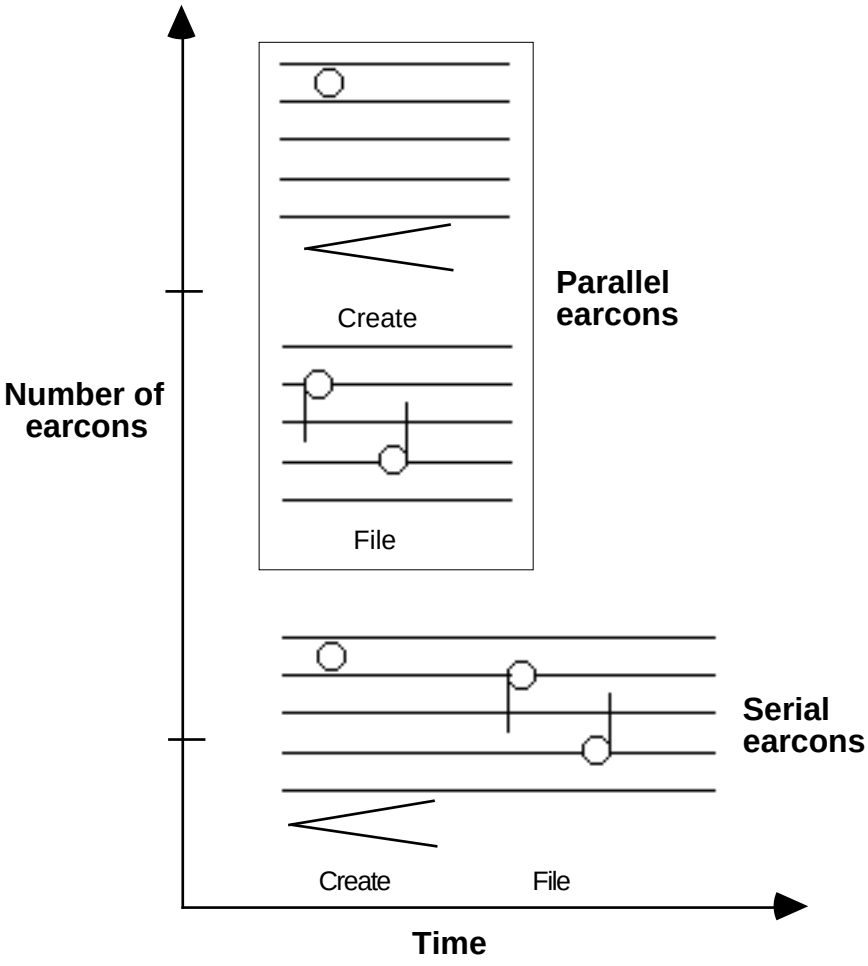

FIGURE 1: Serial and parallel compound earcons.

multiple sounds simultaneously. Gerth (1992) conducted several experiments to see if listeners could recognise changes in sounds when several were presented at once. As the density of sound (degree of polyphony or number of sounds playing simultaneously) increased recognition rates fell but remained at approximately the 90% correct level. Recognition rates did not fall significantly until three sounds were presented. The earcons discussed here are more complex than the sounds Gerth used so that combinations may be more difficult to recognise but only two are to be played in parallel. This may mean that two can be played in parallel without loss of recognition and perhaps four could be played if greater training was given. Blattner *et al.* (1992) have begun to investigate parallel earcons to give information about maps.

Sonnenwald, Gopinath, Haberman, Keese and Myers (1990) used parallel sound presentation to give feedback on parallel computations. They created a system, called InfoSound, that allowed the design of audio messages and their synchronisation with system events. In one example they describe how six part harmony was used to present six multiple concurrent processes. They do not describe any experiments to assess the effectiveness of the sounds or whether listeners could extract information about the processes. The sounds Sonnenwald *et al.* describe are, again, simpler than earcons. One problem they describe is that designers found it difficult to add sounds as they were not trained in music composition. Brewster *et al.* (1992) provided a set of guidelines to help interface designers without musical skills use sound. These guidelines will be enhanced to contain the results from the work described.

Parallel earcons use some of the attributes of the musical theory of *counterpoint.* It is defined by Scholes (1975) (pp 260-261) thus: “the combination of simultaneous voice-parts, each independent, but all conducing to a result of uniform coherent texture” (voice parts may include instrumental voices). In counterpoint individual instruments play separate musical lines which come together to make a musical whole. With parallel earcons, each component earcon is separate but the whole combined sound gives the meaning. This type of structure may give musicians an advantage over non-musicians that they never had with serial earcons.

3. Experiment

An experiment was designed to see if the recognition of parallel earcons was as accurate as that of serial earcons. There were three phases in the experiment. In the first participants learned earcons for objects (icons); in the next phase participants learned earcons for actions (menus); and in the final phase participants heard combined earcons made up of actions and objects. Figure 2 shows the format of the experiment. The work described in this paper seeks to discover how well earcons can be recalled and recognised. It does not suggest

that sounds used in this way should replace icons in the interface. Icons and menus were used because they provided a hierarchical structure that could be represented in sound.

Phases	Serial Group	Parallel Group
Phase I (train & test)	Object earcons	
Phase II (train & test)	Action earcons	
Phase III Presentation 1 (test only)	Serial compound earcons	Parallel compound earcons
Phase III Presentation 2 (test only)	Serial compound earcons	Parallel compound earcons

FIGURE 2: Format of the experiment.

3.1 PARTICIPANTS

Twenty-four participants were used, half were musically trained. They were split into two groups of twelve, half of the participants in each group being musicians. A participant was defined as being musically trained if she/he could play a musical instrument and read music. The participants were undergraduate and postgraduate students from the University of York.

3.2 SOUNDS USED

The sounds used were designed using the guidelines proposed by Brewster *et al.* (1992) . The musical instrument timbres used are shown in Figure 3. The phase I and some of the phase II rhythm, pitch and intensity structures are shown in Figure 4 and Figure 5. The sounds all lasted one second and were in 3/4 time. Care had to be taken in the timing of the earcons because two sounds would be played at the same time and they therefore had to be in time with each other, starting and finishing together, in order to sound pleasant.

The action earcons (phase II) were in the scale of D Major and began at D_4 (146Hz). For example ‘Open’ was D_4, F_4, A_4, the chord of D Major. The object earcons (phase I) were in the scale of C Major and began at C_2 (523Hz). For example ‘File’ was C_2, C_2 and C_2, E_2, G_2, the chord of C Major. Two different scales were used to help listeners separate the earcons when they heard them together. The octave separation was similar to techniques in music where a bass line and a lead line are used. This helped overcome octave perception problems that occur where it can be difficult to differentiate the same note played in different octaves (Deutsch, 1982) . If different base notes are used then this problem is reduced. To further help discrimination of sounds chords were used in the earcons for phase I but not phase II. Complex intensity structures were also used.

Timbre	Parallel & Serial Groups
Write	Piano
Paint	Brass
Spreadsheet	Pan Pipes
Menu 1	Marimba
Menu 2	Electric Organ
Menu 3	Cymbal

FIGURE 3: Timbres used in the experiment.

As the sounds were to be played in parallel to one group, care had to be taken so that each of the earcons could be heard as a separate sound source or *stream*. If each earcon was not heard as a separate stream then they would mix together and neither of the earcons would be distinguishable. Bregman (1990) and Williams (1992) propose some principles which can be used to ensure that sounds are grouped into separate streams, see them for more on each of the principles mentioned below.

♩ = 0.33 seconds giving a tempo of 180 beats per minute (bpm).

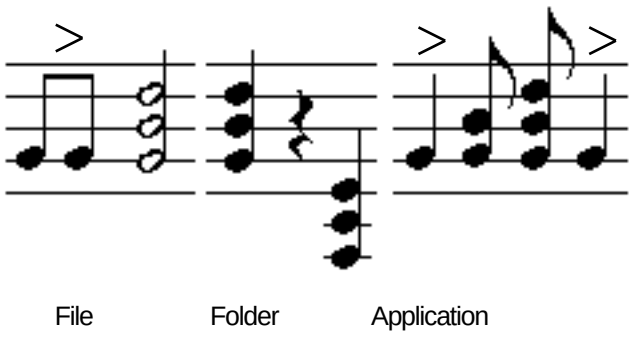

FIGURE 4: Rhythms, intensities and pitches used in phase I.

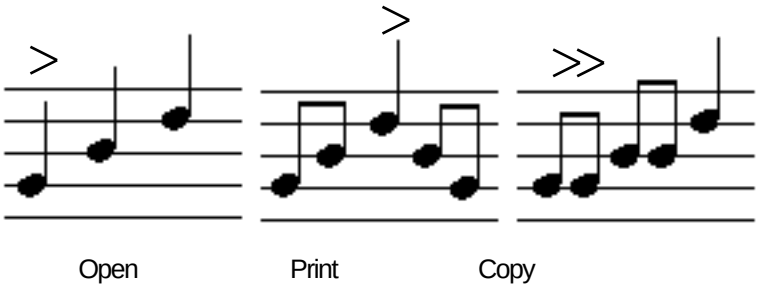

FIGURE 5: Rhythms, intensities and pitches used in phase II.

3.2.1 Similarity and dissimilarity

Components which share the same attributes will be perceived as related. Sounds will be grouped into a single source if they are similar, i.e. they have similar pitches, timbres, loudnesses and locations. All the pitches within an earcon were kept to the same octave. The action and object earcons were separated by two octaves to make sure the frequencies were dissimilar. To further increase the spectral dissimilarities, the object sounds used chords and the action sounds did not. Each of the earcons that could play together had a different timbre. This meant that the sounds had different spectral contents and amplitude and frequency modulations.

3.2.2 Proximity

Components which are close in time and frequency will be perceived as related. All the pitches used within an earcon were from the same octave so that there was frequency proximity. Chords also increased frequency proximity. Two octaves between the actions and objects helped separate them into different streams because there was a large frequency distance.

3.2.3 Coherence

Components of streams change in coherent ways. Components of a source tend to vary in a coherent manner, for example they change intensity and pitch together. An example of this is modulation, where frequency or intensity vary in a regular manner helping to differentiate one musical instrument from another. All the components of a single earcon varied in terms of pitch and intensity together. These variations were distinct from those in the other earcon that might have been playing at the same time. The common modulations of amplitude and frequency due to the timbres of the different sounds playing helped to make each a different stream.

3.2.4 Spatial location

Components originating from the same spatial location will be perceived as related. Buxton, Gaver and Bly (1991) suggest that spatially separating sound sources helps clarity by allowing the auditory system to focus on a single source from amongst many and lessens interference between sounds. Action sounds were presented on the right and object sounds on the left of the stereo space so that they would be heard as separate streams. This guideline is supported by research from Mayfield (reported in Gerth (1992)) where, without spatial separation of sources, recognition rates fell more rapidly as sound density increased than when there was stereo separation.

3.2.5 Rhythm

Rhythmic patterns tend to be perceived as sources. According to Deutsch (1986) , rhythm is one of the most powerful factors of pattern recognition. Handel (1989) suggests some methods for creating rhythmic groups. For the earcons used in the experiment two of these were used: *Intensity accentuation* (an accented note begins a group) and *Duration accentuation* (a long note ends a group).

These principles were used in both the serial and parallel earcons as they help to differentiate any two sounds. The only difference with the serial earcons was that 0.1 second delay was placed between the two sounds so that participants could more easily tell where one finished and the other started (Reich, 1980) .

The earcons for both groups were generated on a Roland D110 multi-timbral sound synthesiser and recorded by an Akai S950 digital sampler at a sampling rate of 48kHz for playback. The sounds were all played through a Yamaha DMP 11 digital mixer controlled, using MIDI, by an Apple Macintosh computer and presented using external loudspeakers.

3.3 EXPERIMENTAL DESIGN AND PROCEDURE

The design was very similar to that used by Brewster *et al.* (1992) . However, some changes were made based on the insights gained from our previous experiments. For example, in phase I here Write, Draw and Spreadsheet families were used. In the previous experiments, some participants commented that they had difficulty with the difference between ‘Draw’ and ‘Paint’ so it was decided to change ‘Draw’ to ‘Spreadsheet’ to avoid confusion. In phase II all the menus were three items long. This was changed from the previous experiments to avoid any indirect cueing that may have been given by different menu lengths.

Phases I and II were identical for both groups of participants. The purpose of these phases was to make sure the participants would recognise the earcons when used in phase III. In order to test the recognition of compound serial and parallel earcons in phase III any participant who did not reach a 65% recognition rate in both phase I and II was rejected. Only participants who had learned the individual earcons could be tested on the combined ones. The serial earcons could then be compared to the parallel ones to see if recognition rates varied. Instructions were read from a prepared script.

3.3.1 Phase I: Objects

Training: The participants were presented with the screen shown in Figure 6. Participants had to learn the names of all the icons. When they thought they had done this they wrote them down. If they were not correct

they were allowed more time to learn them. This meant that, at the end of the training the participants new the names of all the icons present.

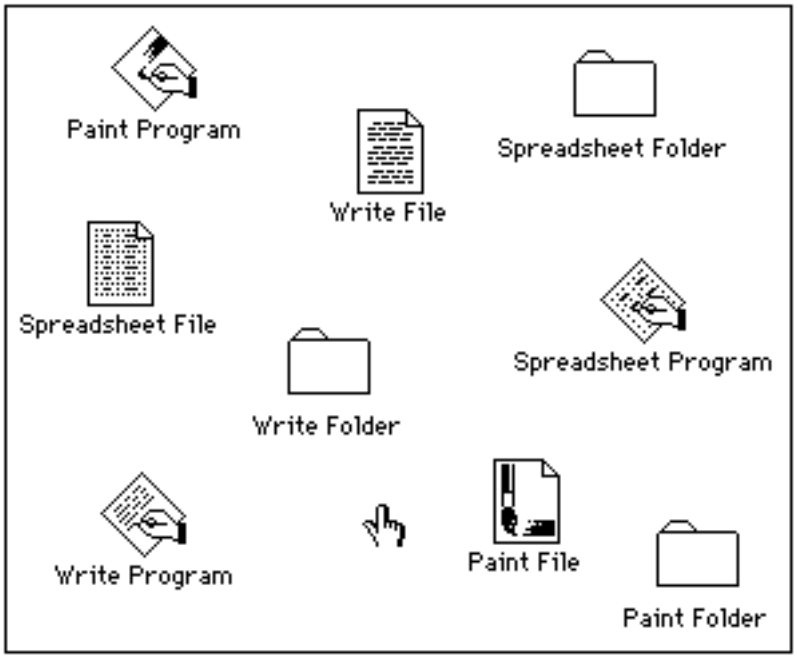

FIGURE 6: Phase I object screen.

Each of the objects on the display had a sound associated with it. The sounds were structured as follows: Each *family* of related items shared the same timbre. For example, the paint program, the paint folder and paint files all had the same timbre. Items of the same *type* shared the same rhythm. For example, all the programs had the same rhythm. These two different types of information allowed a unique sound to be created for each of the icons displayed. All of the information available graphically in the icons was available through sound in the earcons. The earcons were played one-at-a-time in random order to the participants for them to learn (in the same way as out previous experiments) and the whole set of sounds was played three times.

The random nature of the training made the sounds harder to learn. If, for example, paint file, paint folder and paint application had been played one after another then the relationships would have been easier to identify and it would have been easier for the subjects to form a model of the hierarchy used. Corcoran, Carpenter, Webster and Woodhead (1968) indicate that "…sounds that the trainee is likely to find difficult to distinguish should be presented alternately, not separated in time by other intervening sounds". However, doing it in the way described here gave a 'worst-case' learning situation. If the earcons could be learned under these conditions then they would be a robust method of communication. It would also simulate the unstructured training that a new user of a computer system might receive.

Testing: During testing the screen was cleared and the earcons were played back in a random order. The participant had to supply what information he/she could remember about type and family. When scoring, a mark was given for each correct piece of information supplied. Participants were allowed to hear any stimulus a maximum of three times. Nine questions were asked. In the testing of this and the other phases participants were not told the accuracy of their responses.

3.3.2 Phase II: Actions

In this phase earcons were created for actions. Each *menu* had its own timbre and the *items* on each menu were differentiated by rhythm, pitch or intensity. The menus were not designed to represent any existing system such as the Macintosh or Windows menu structures. This was done so that no group of users would be favoured. Participants were trained in the same way as before.

FIGURE 7: Phase II action screen.

The screen shown to participants to learn the earcons is given in Figure 7. The participants were tested in the same way as before but this time had to supply information about menu and item. Participants were allowed to hear any stimulus a maximum of three times. Nine questions were asked.

3.3.3 Phase III: Combinations

In the final phase participants heard combined earcons. Phases I and II prepared the participants for the main test in this phase. The parallel group heard parallel combined earcons and the serial group heard serial combined earcons made up of the sounds they heard in phases I and II. Before they were tested on phase III, participants were presented with three examples of the type of combined earcons they were about to hear.

In this phase of the experiment an object earcon was always played with an action one. In the serial case an action sound was followed by an object one, in the parallel case an object and an action were played together. Nine out of a possible set of 81 earcons were presented to the participants during this phase. Each combined earcon was played once and the participant was then instructed to give all the information he/she could about the family, type, menu and item of the stimulus heard. The stimulus was then presented again and the participant

could correct a previous answer or fill in any parts not recognised after the first presentation. This overcame a problem that occurred in our previous experiments where participants were allowed to hear a compound stimulus more than once before they gave their answer. This meant that participants could listen to the first part of the earcon on the first presentation and the second on the subsequent presentation. The way the current experiment was designed forced the participants to describe what they knew about the stimulus after one presentation. The second presentation was given to see what levels of recognition would be reached after greater exposure to the stimuli.

3.4 EXPERIMENTAL HYPOTHESES

The experiment attempted to find out if compound parallel earcons were as recognisable as serial ones. The main hypothesis for this experiment was that parallel earcons would be as recognisable as serial earcons and would thus reduce presentation time for audio messages. Listeners would not make more mistakes when listening to two complex stimuli. They could attend to and discriminate two complex sound sources at once because this is the way sounds are heard in the natural environment. On the first presentation of the earcons, recognition rates would be lower than on the second presentation. Hearing the sounds for the first time would test initial recognition rates. The second presentation would give higher rates, similar to those of prolonged exposure. Musicians would show no better performance than non-musicians. As demonstrated in our previous experiments, musical skill does not improve recognition of earcons. The overall recognition rates would be similar to those achieved in the previous experiments. The serial condition of this experiment was broadly the same as our previous ones so the results from this experiment would verify the results of the other.

4. Results and Discussion

Figure 8 shows the overall scores for each phase in the experiment. For the data analysis the main area of interest was the differences between the groups. A two-factor repeated-measures ANOVA was carried out between the groups across each of the phases. The results showed there was no main effect for group ($F(1,22)=0.07$, $p=0.801$), there was a main effect for phase ($F(3,22)=10.45$, $p=0.0001$) but no interaction between group and phase ($F(3,22)=0.77$, $p=0.515$). These results showed that there was no difference between the groups. This indicated that the parallel earcons were recognised as well as the serial ones confirming the main hypothesis of the experiment.

To find out where the main effect for phase occurred, post-ANOVA Tukey HSD tests were conducted for each group between each of the phases. The results showed that the only significant difference was between phase II and phase III(1) in both groups (Serial group II vs. III(1): $Q(22)=4.71$, $p=0.05$, Parallel group II vs. III(1): $Q(22)=6.47$, $p=0.01$). There were no other significant differences between the phases. Figure 8 shows that very high scores were achieved in phase II for both groups and the lowest scores for both groups were obtained in phase III(1).

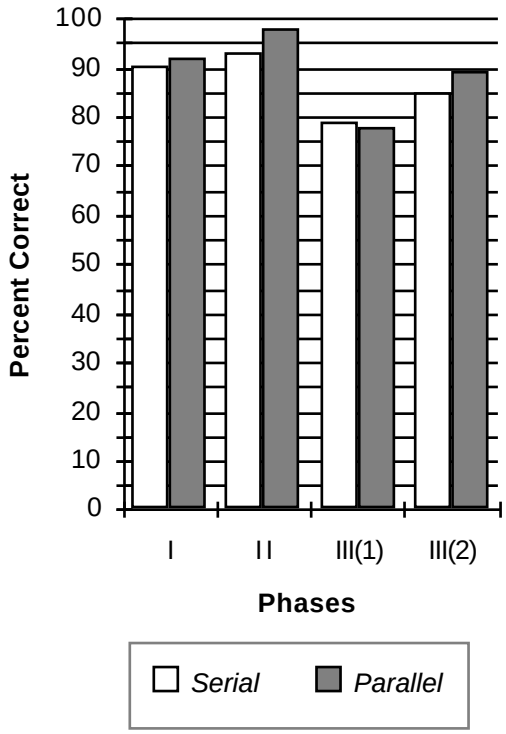

FIGURE 8: Overall scores per phase.

The phase III(2) scores were very close to the scores of phases I and II. These results show that 90% recognition rates can easily be achieved with carefully designed earcons. This work gives further evidence to show that earcons are an effective means of communicating complex information in sound. A more detailed analysis was undertaken to determine if the overall scores were hiding underlying differences between the groups.

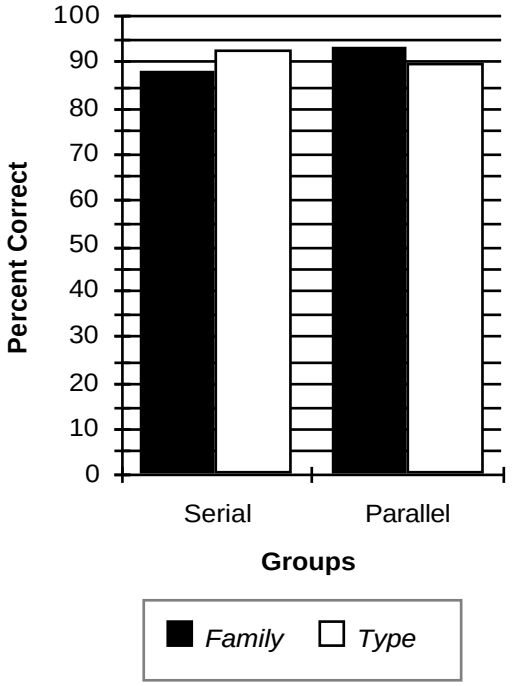

FIGURE 9: Breakdown of scores for phase I.

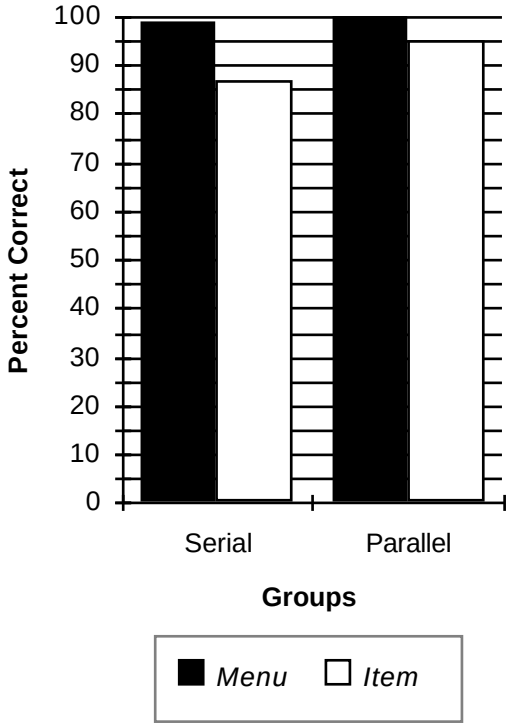

FIGURE 10: Breakdown of scores for phase II.

4.1 PHASES I AND II

Figure 9 and Figure 10 show the scores for phases I and II. It was expected that there would be no difference between the groups in phases I and II as they both received the same stimuli. The results showed no significant differences between the groups on any of the components.

The high rates of recognition achieved in these two training phases meant that the participants were well prepared for the main test in phase III. As mentioned above, any participant that did not reach 65% in any part of phase I or II was rejected. This lead to the rejection of three participants (see below for more details).

4.2 OVERALL PHASE III(1) AND III(2) RESULTS

Phase III tested the differences between serial and parallel compound earcons. The same stimuli were presented to the participants twice to investigate what would happen to the recognition rates. The stimuli were ones that the participants had been trained on in phases I and II. The first presentation was called III(1) and the second III(2). A two-factor repeated-measures ANOVA was carried out between the groups across the two presentations on the overall data. It showed no main effect for group ($F(1,22)=0.08$, $p=0.777$). This showed that

there was no difference in recognition rates between the serial and parallel earcons. It showed a very strong main effect for the repeated measure from presentation III(1) to III(2) ($F(1,22)=87.27$, $p=0.0001$). This indicated there was a significant increase in recognition from III(1) to III(2). There was also a significant interaction between group and presentation ($F(1,22)=6.04$, $p=0.022$).

To find out where the main effect for the repeated-measure occurred Tukey HSD tests were conducted on the data for each group across presentations. The results showed that in both groups phase III(2) was significantly better than III(1) (serial group: $Q(22)=6.902$, $p=0.01$, parallel group: $Q(22)=11.799$, $p=0.01$). This difference can be observed in Figure 11. These results show that the participants got significantly better when they heard the sounds a second time. This was investigated further to see where the interaction occurred. The differences between the III(1) and III(2) scores for both groups were calculated to see which group had increased the most. This was done by taking the phase III(1) score from the phase III(2) score for each group. A one-factor ANOVA was then used on these difference data. It indicated that the parallel group increased significantly more than the serial group ($F(1,22)=6.04$, $p=0.0223$). This can be seen in Figure 11.

4.2.1 Discussion

These results show that compound parallel earcons are as capable as compound serial earcons at communicating information. The recognition rates of both groups were not significantly different. This indicated that parallel earcons were an effective means of reducing the length of compound earcons without compromising recognition rates. Recognition rates were lower on the first presentation of the earcons, as was expected, but were still around 75%. On the second presentation, rates increased significantly to between 85% and 90%. This indicated that the more earcons were heard the better the recognition rates would be. This would be the situation if earcons were used in human-computer interfaces. The parallel group increased significantly more than the serial group from the first presentation to the second. However, as there were no overall differences in terms of group, this increase does not indicate that parallel earcons are more easily recognised.

The first presentation of the compound earcons was significantly worse than the recognition in phase II but by the second presentation there were no differences in recognition. This showed that the recognition rates of the combined earcons were as good as when the component earcons were heard individually.

4.3 DETAILED PHASE III RESULTS

A detailed examination of phase III was undertaken to investigate recognition of individual components of the earcons. The data are shown in Figure 11. A two-factor repeated-measures ANOVA was conducted on the two groups across the eight components of the two presentations. As expected from the results described above, there were no differences in recognition of the components in terms of groups ($F(1,22)=0.08$, $p=0.7776$). This showed that, for each of the components in both presentations, the groups did not differ significantly in recognition rates. There was a significant difference in terms of the components (the repeated measure) ($F(7,22)=9.33$, $p=0.0001$) but no interaction between group and components ($F(7,22)=1.16$, $p=0.3272$).

A significant difference in terms of repeated-measure was expected as the overall results showed that presentation III(1) was significantly worse than III(2). Tukey HSD tests were conducted to find out where the differences occurred. The significant results are shown in Figure 12 and Figure 13, any results not shown were not significant. For the serial group it can be seen that the menu scores in both presentations were significantly better than any of the other components. Menu in presentation III(2) was significantly better than item, file and type in presentation III(1). Menu in presentation III(1) was significantly better than item and family in the same presentation. There was no significant difference between the menu scores in III(1) and III(2). Figure 12 shows the high menu scores. The results for this group were similar to those in phase II versus phase I. There, menu was the best recognised of the components. Menu, differentiated by timbre, was a very powerful cue for the serial group. This confounds the definition of earcons given by Blattner *et al.* (1989) where it is suggested that pitch and rhythm are the most important factors for recognition.

In the parallel group a poor item score in phase III(1) accounted for the main differences between the presentations (see Figure 13). All of the phase III(2) components were better than III(1) item. Menu in III(1) was also better than item. It may be that when two earcons are heard in parallel the rhythms are harder to detect. Rhythm is a complex component and, if listeners were to switch their attention from it to listen to the timbre, for example, information could easily be lost. However, in the second presentation item score increased significantly over the first presentation. There were no longer any differences between the item score and any other component of III(2). Therefore, with greater exposure to earcons (as would occur if they were being used in a human-computer interface) problems of rhythm recognition would not be an issue.

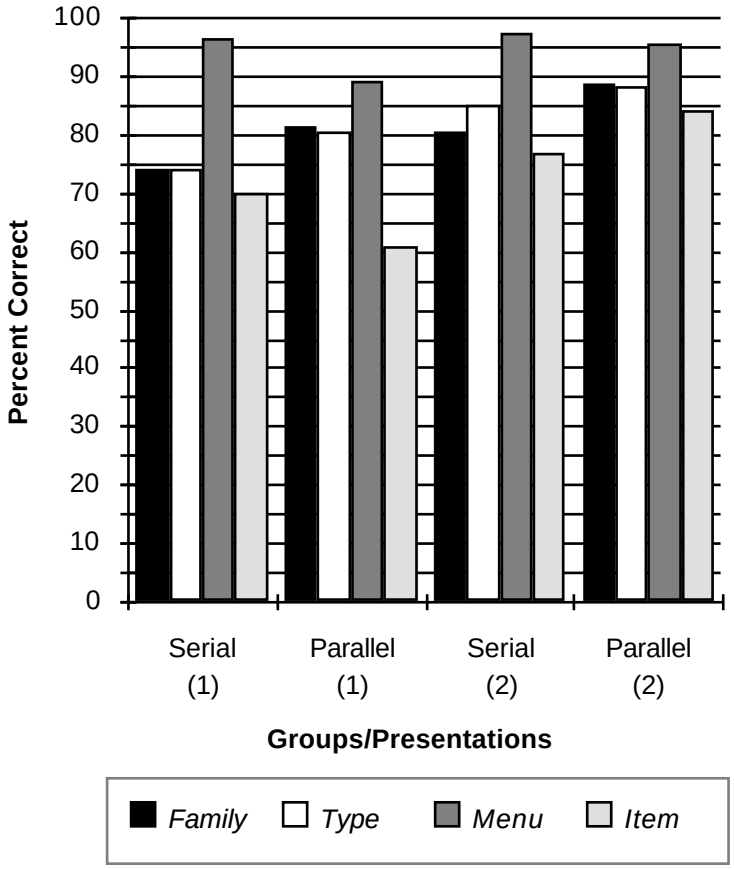

FIGURE 11: Breakdown of scores for phase III presentations 1 and 2.

Menu2	Menu1
Menu2 vs. Item1 Q(22)=6.114, p=0.01	Menu1 vs. Item1 Q(22)=5.684, p=0.05
Menu2 vs. Family1 Q(22)=5.255, p=0.05	Menu1 vs. Family1 Q(22)=4.825, p=0.05
Menu2 vs. Type1 Q(22)=5.052, p=0.05	

FIGURE 12: Serial group Tukey HSD tests showing significant differences between the components in phases III(1) and III(2). Menu1 = menu score in III(1), Menu2 = menu score in III(2).

Menu2	Family2	Menu1	Type2	Item2
Menu2 vs. Item1 Q(22)=7.781, p=0.01	Family2 vs. Item1 Q(22)=6.316, p=0.01	Menu1 vs. Item1 Q(22)=6.316, p=0.01	Type2 vs. Item1 Q(22)=6.114, p=0.01	Item2 vs. Item1 Q(22)=5.25, p=0.05

FIGURE 13: Parallel group Tukey HSD tests showing significant differences between the components in phases III(1) and III(2). Menu1 = menu score in III(1), Menu2 = menu score in III(2).

4.3.1 Discussion

Looking at performance on the individual components of phase III the overall results are confirmed. The second presentations of the earcons were better recognised than the first and there were no differences between the groups on any of the components. This again indicates that parallel earcons are as effective as serial ones at communicating information.

The scores show that both groups had problems identifying the item component. This led to the lowest score in phase III. Although by the second presentation the score in the parallel group was nearly 85%. It could be that type (the equivalent component to item in phase I) fared better than item as participants were able to use the structure information in the rhythms more effectively. There was less structure information available in the item sounds so that they were harder to remember. The item scores did increase significantly on the second presentation so, again, with practice participants can reach high levels of recognition. As the results described above suggest, the scores for item were not significantly worse than type (both of these were based on rhythm).

4.4 MUSICIANS AND NON-MUSICIANS

Brewster *et al.* (1992) showed that, for combined serial earcons, the performance of musicians was not significantly different to non-musicians. In this experiment musicians were again compared to non-musicians to see if they performed better with combined parallel earcons. It may be that parallel earcons were more easily recognised by musicians as they were used to listening to complex sounds playing in parallel. They might also have obtained higher scores on the first presentation due to their greater skill.

4.4.1 Results

The two groups were divided into four: Serial musicians, parallel musicians, serial non-musicians and parallel non-musicians. An overall two-factor repeated-measure ANOVA was performed on percentage data between the groups across the phases. It showed no main effect for group ($F(3,20)=0.65$, $p=0.5925$), it showed a main effect for phase ($F(3,20)=10.64$, $p=0.0001$) but no interaction between group and phase ($F(3,20)=1.07$, $p=0.3961$). The main effect for phase was expected because of the overall results described above. This showed that there were no differences between the musicians and non-musicians in any group on any of the phases. Therefore, musicians were not significantly better than non-musicians with parallel earcons.

These results confirm those of Brewster *et al.*: There were no differences between musicians and non-musicians in the recognition of serial earcons. There were also no differences in recognition of parallel earcons.

Musicians did not perform any better than non-musicians even with the more complex stimuli. The parallel group musicians did not reach higher recognition levels on the first presentation of the phase III earcons, as might have been expected due to their training. This seems to suggest that parallel earcons can be used effectively by those not skilled in music.

4.5 REJECTED PARTICIPANTS

As mentioned above, participants were rejected if they reached below 65% correct scores in either phase I or II after hearing each of the earcons three times in training. Nine questions were asked in each of phases I and II so any participant who got a score of less than 5.85 was rejected. Participants had to be able to recognise the individual component earcons before they could be tested on the combined ones. This lead to the rejection of three participants who did not reach the required level. These participants were all non-musicians and they all failed on phase II. The participants rejected obtained scores of 4, 4 and 5 in this part of the test. In phase II there was less structure information to help remember the sounds than in phase I which may have made it a harder test. In order to find out how much extra training each would need to reach the 65% level they were trained further. This involved going through the training and testing of the phase where the participant fell below the required level until they reached it. One of the participants reached the required level after one further training session, one participant required two sessions and one participant never reached the required level, even after three more training sessions. This seems to indicate that some users may have problems with earcons which could be due to tone-deafness. Some may require more training which could be done when the user initially came to the computer with an auditory interface. Some users may always have problems and this may be analogous to colour-blind users with coloured graphical interfaces. Moore (1989) suggests that tone-deafness is a misnomer (p 147): "…nearly everyone is able to judge that two tones are different in pitch when their frequency difference exceeds a certain amount". One common problem is a listener being unable to assign a direction to a pitch change; they can hear the tones are different but cannot say which one is higher in pitch. Moore says that very often this can be overcome by practice. The simple re-training we carried-out was not enough to overcome the problem. This is a very important point to be aware of when designing an auditory interface.

5. General Discussion

The results of phases I and II confirm those of Brewster *et al.* (1992) in that participants were able to recognise the individual earcons with a great degree of accuracy. The high rates of recognition of timbre (up to

almost 100% in phase II) again indicate that musical instrument timbres are very effective and users can easily identify them. The phase III results show that parallel compound earcons are as easily recognised as serial compound earcons. This means that parallel earcons are more effective in an auditory human-computer interface as they take only half the time to present to the user.

After prolonged exposure to the parallel earcons it is hoped that the participants would hear the two separate earcons as a single 'whole' earcon. For example, earcons for 'open' and 'write file' would coalesce and be heard as an earcon for 'open write file'. Listeners would become accustomed to the sounds and come to recognise the overall earcon. In order to test this another experiment would be needed which would train and test participants with much longer exposure to the earcons.

5.1 PARAMETERS FOR MANIPULATING EARCONS

There are five parameters that Blattner *et al.* (1989) propose can be manipulated to differentiate earcons. They suggest that rhythm and pitch are the primary (fixed) parameters and timbre, intensity (dynamics) and register are the secondary (variable) parameters.

Primary Parameters	Secondary Parameters
Rhythm	Pitch
Timbre	Intensity
Register	Stereo position
	Chords
	Effects (Echo, Chorus, Etc.)

FIGURE 14: The new parameters for manipulating earcons.

The results from our previous experiments indicated that timbre was much more important than suggested by Blattner *et al*. In those experiments (and this one) timbre was used to denote families of icons or menu items. The results showed that it played a much bigger role than that suggested by Blattner *et al*. Our previous experiments also showed that earcons differentiated by pitch alone were very difficult to discriminate. Register was shown to be more important so that big differences between earcons could be created. The earcons used in this current experiment were designed around the guidelines proposed in Brewster *et al.* (1992) and again high recognition rates were reached. It can therefore be argued that timbre and register, along with rhythm, are the primary parameters for creating the basic structure of a set of earcons. Secondary parameters, such as pitch, intensity, stereo position, chords and effects (such as echo or chorus) work together to help differentiate the earcons from each other (Figure 14 shows these parameters, the items in each column of the table are not ordered).

5.2 AUDITORY STREAMING AND EARCON GUIDELINES

The auditory streaming techniques, described above, that were used to differentiate the earcons proved to be effective. None of the participants in the parallel group complained they were unable to separate the two earcons. These techniques were also effective on the serial earcons. The guidelines from the previous chapter fit well with the general principles of auditory streaming. These principles suggest ways to make components in the same stream similar and dissimilar to components of other streams. The earcon guidelines also try to do this so that there are big enough differences between earcons that participants can recognise them individually.

The earcon guidelines, and the experiment described in this paper which used them, heavily stress *similarity* and *dissimilarity*. Each earcon family had a separate timbre so that items in the same family shared the same attributes. Chords were also used to give the object sounds different attributes to the action sounds. *Proximity* was also stressed in the guidelines. Related items were in the same octave and unrelated ones separated by two or more octaves. *Coherence* was important as each earcon had a different timbre with the attendant differences in modulation that brought about. *Spatial location* was not in the original guidelines but is another important method of differentiating earcons. Two locations were used in this experiment and a third central location could easily be added. MIDI allows at least 16 different stereo positions but it is unclear if each of these could easily be distinguished as the differences between them would be small. The work of Wenzel, Wightman and Foster (1988) has shown that three dimensions are possible in synthesised sound and this would provide more opportunities for recognisable locations. The methods described for creating rhythmic groups are an important addition to the guidelines as they make each of the rhythms into a whole and complete unit with a more defined start and end point.

The original earcon guidelines from the previous chapter have been shown to share some of the principles developed in auditory steam segregation research. This gives them a stronger foundation as auditory streaming research deals with some similar problems. Extensions have also been put forward to the guidelines to include spatial location and extended use of rhythm to create more complete rhythmic units. The guidelines now allow an interface designer to generate earcons that can be heard in parallel.

6. Future Work

The next step for this work would be to find out the maximum number of earcons that could be recognised in parallel. It may be that more can be recognised (as Gerth (1992) suggests) but that bigger differences between

them would be required and that larger and larger amounts of training needed to reach equivalent recognition rates.

One other aspect to consider is the workload required to recognise the parallel earcons. The parallel earcons might require more effort to recognise but this might not affect performance in low workload situations. When other tasks are being performed and more cognitive resources are in use then recognition rates might fall (see Hart and Wickens (1990) for more on this). Further work to compare the workload of parallel and serial earcons is needed. A similar experiment could be run again but workload measures recorded after each phase. See Brewster *et al.* (1994) for an example of this kind of evaluation of sonically-enhanced widgets using NASA Task Load Index workload measures (NASA Human Performance Research Group, 1987) .

7. Conclusions

The results of Brewster *et al.* (1992) showed that earcons could be recognised with a high degree of accuracy. A problem still remained that earcons took too much time to play and, if they were to be used in human-computer interfaces, might not be able to keep up with the pace of interaction. Slowing interactions down so that the sounds could keep up would be unacceptable so the sounds had to be able to communicate more rapidly. The experiment described here showed that the length of compound earcons could be reduced to half by playing the two components in parallel and the rates of recognition maintained. This means that displaying complex information in sound that can keep pace with interactions is possible. This research will allow a wider application of sound at the interface.

Combined parallel earcons were shown to be as effective as the combined serial earcons proposed by Blattner *et al.* (1989) . The results from this experiment confirm those of the previous ones (Brewster *et al.*, 1992) . The results are broadly the same, although there were some simplifications of the experimental design. Results from the first presentation of the earcons, phase III(1), showed that on a single presentation almost 80% correct scores could be achieved. The second presentation showed that recognition increased significantly when participants heard the sounds again. This indicated that, if earcons were used at the human-computer interface, then regular exposure would quickly lead to high levels of recognition. This level was achieved even though the training given was intentionally difficult.

Musicians have again been shown to be no better than non-musicians. This means that auditory interfaces will be usable by most users whatever their level of musical skill. The results of the re-training of rejected participants, however, showed that some users may always have problems using sound. This is similar to a

colour-blind person using an interface that depends heavily on colour. Some extensions have been put forward to the guidelines described in Brewster *et al.* (1992) based on research into auditory stream segregation. By using the guidelines a designer could create effective sounds for an interface.

This work has extended that described in Brewster *et al.* (1992) and shown that earcons are not only an effective means of communicating complex information in sound but that they can do it at a rate which can keep up with the pace of interaction in an interface. Humans process multiple sounds in parallel in their everyday world and earcons are able to exploit this to overcome rate of presentation problems at the interface. Earcons have now been shown to be a very effective method of communicating information in non-speech sound.

8. Acknowledgements

Thanks go to Ian Pitt for helping with the equipment used in the experiment and general advice on music. Thanks also to Saul Greenberg at the University of Calgary who helped generate some of the initial ideas for this work half-way up a mountain in Canada. This work was supported by SERC studentship 90310837 and ERCIM Research Fellowship 94-04. The work was undertaken whilst the first author was at York University, UK.

9. References

BLATTNER, M. & DANNENBERG, R.B., EDS. (1992). *Multimedia Interface Design*. Frontier Series, New York: ACM Press, Addison-Wesley.

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

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

BREGMAN, A.S. (1990). *Auditory Scene Analysis*. Cambridge, Massachusetts: MIT Press.

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

BREWSTER, S.A., WRIGHT, P.C., DIX, A. & EDWARDS, A.D.N. (1994). The sonic enhancement of graphical buttons. In Nordby, Helmersen, Gilmore & Arnesen (Eds.), *Proceedings of Interact'95*, (pp 43-48), Lillehammer, Norway: Chapman & Hall.

BREWSTER, S.A., WRIGHT, P.C. & EDWARDS, A.D.N. (1992). A detailed investigation into the effectiveness of earcons. In Kramer (Ed.), *Auditory display, sonification, audification and auditory interfaces. The Proceedings of the First International Conference on Auditory Display*, (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 INTERCHI'93*, (pp. 222-227), Amsterdam: ACM Press, Addison-Wesley.

BREWSTER, S.A., WRIGHT, P.C. & EDWARDS, A.D.N. (1994). The design and evaluation of an auditory-enhanced scrollbar. In Adelson, Dumais & Olson (Eds.), *Proceedings of CHI'94*, (pp. 173-179), Boston, Massachusetts: ACM Press, Addison-Wesley.

BROWN, M.L., NEWSOME, S.L. & GLINERT, E.P. (1989). An experiment into the use of auditory cues to reduce visual workload. In *Proceedings of CHI'89*, (pp. 339-346), Austin, Texas: ACM Press, Addison-Wesley.

BUXTON, W., GAVER, W. & BLY, S. (1991). Tutorial number 8: The use of non-speech audio at the interface. In *Proceedings of CHI'91*, New Orleans: ACM Press: Addison-Wesley.

CORCORAN, D., CARPENTER, A., WEBSTER, J. & WOODHEAD, M. (1968). Comparison of training techniques for complex sound identification. *Journal of the Acoustical Society of America*, **44**, 157-167.

DEUTSCH, D. (1982). *Psychology of music*. London: Academic Press.

DEUTSCH, D. (1986). Auditory pattern recognition, In Boff, Kaufman & Thomas, Eds., *Handbook of perception and human performance*. New York: Wiley.

GAVER, W., SMITH, R. & O'SHEA, T. (1991). Effective sounds in complex systems: The ARKola simulation. In Robertson, Olson & Olson (Eds.), *Proceedings of CHI'91*, (pp. 85-90), New Orleans: ACM Press, Addison-Wesley.

GERTH, J.M. (1992) *Performance based refinement of a synthetic auditory ambience: identifying and discriminating auditory sources*. PhD. Thesis, Georgia Institute of Technology.

HANDEL, S. (1989). *Listening: An introduction to the perception of auditory events.* Cambridge, Massachusetts: MIT Press.

HART, S.G. & WICKENS, C. (1990). Workload assessment and prediction, In Booher, Eds., *MANPRINT, an approach to systems integration*. New York: Van Nostrand Reinhold.

MOORE, B.C. (1989). *An Introduction to the Psychology of Hearing*. (2nd ed.) London: Academic Press.

NASA HUMAN PERFORMANCE RESEARCH GROUP. (1987). *Task Load Index (NASA-TLX) v1.0 computerised version* NASA Ames Research Centre.

PERROTT, D., SADRALOBADI, T., SABERI, K. & STRYBEL, T. (1991). Aurally aided visual search in the central visual field: Effects of visual load and visual enhancement of the target. *Human Factors*, **33**, 389-400.

REICH, S.S. (1980). Significance of pauses for speech perception. *Journal of Psycholinguistic Research*, **9**, 379-389.

SCHOLES, P.A. (1975). *The oxford companion to music*. (10th ed.) Oxford: Oxford University Press.

SONNENWALD, D.H., GOPINATH, B., HABERMAN, G.O., KEESE, W.M. & MYERS, J.S. (1990). InfoSound: An audio aid to program comprehension. *Proceedings of the 23rd Hawaii International Conference on System Sciences*, 541-546.

SUMIKAWA, D., BLATTNER, M., JOY, K. & GREENBERG, R. (1986). *Guidelines for the syntactic design of audio cues in computer interfaces* (Technical Report No. UCRL 92925). Lawrence Livermore National Laboratory.

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

WENZEL, E., WIGHTMAN, F. & FOSTER, S. (1988). Development of a 3D auditory display system. *SIGCHI Bulletin*, **20**, 52-57.

WILLIAMS, S. (1992). Perceptual principles in sound grouping. In Kramer (Ed.), *Auditory Display, sonification, audification and auditory interfaces. The Proceedings of the First International Conference on Auditory Display*, (pp. 95-126), Santa Fe Institute, Santa Fe: Addison-Wesley.