Understanding Concurrent Earcons: Applying Auditory Scene Analysis Principles to Concurrent Earcon Recognition

DAVID K. MCGOOKIN and STEPHEN A. BREWSTER
University of Glasgow

Two investigations into the identification of concurrently presented, structured sounds, called earcons were carried out. One of the experiments investigated how varying the number of concurrently presented earcons affected their identification. It was found that varying the number had a significant effect on the proportion of earcons identified. Reducing the number of concurrently presented earcons lead to a general increase in the proportion of presented earcons successfully identified. The second experiment investigated how modifying the earcons and their presentation, using techniques influenced by auditory scene analysis, affected earcon identification. It was found that both modifying the earcons such that each was presented with a unique timbre, and altering their presentation such that there was a 300 ms onset-to-onset time delay between each earcon were found to significantly increase identification. Guidelines were drawn from this work to assist future interface designers when incorporating concurrently presented earcons.

1. INTRODUCTION

Mobile computing devices are becoming increasingly more popular, with technologies such as Bluetooth and Wi-Fi enabling users to connect to computing systems and resources almost wherever they are. 3G technology has increased the communication bandwidth of mobile telephones, expanding the possible uses to which such devices can be put.

There are, however, several usability issues with these devices. First, in order for mobile devices to be mobile they must have a small form factor. The latest Palm PDA (personal digital assistant) computing devices have a visual display of only 6×6 cm, couple this with the low resolution of the display, and the amount of information that can be presented is severely restricted. Further, the mobile computing environment is very different from the comparatively safe office environment in which conventional personal computers (PCs) are used. In a mobile context a user must continuously be aware of their environment. If walking down the street a user must be aware of traffic and avoid walking into lampposts, litter bins, and other street users, as well as being aware of any other dangers. In an airport a user

Authors' address: David K. McGookin and Stephen A. Brewster, Department of Computing Science, University of Glasgow, 17 Lilybank Gardens, Glasgow G12 8QQ, UK; http://www.dcs.gla.ac.uk/~{*mcgookdk,stephen*}; email: {mcgookdk,stephen}@dcs.gla.ac.uk.

must monitor information boards to find out when and from where aircraft arrive and depart. These distractions force the user to continuously avert their gaze from the computing device and may cause them to miss important feedback or forget what they were doing.

One of the potential ways in which these problems can be overcome is through the use of auditory feedback to the user. Here, sound is used either to augment or replace visual feedback. Several studies have shown that the use of auditory feedback can improve interactions with mobile devices. Brewster [2002] has shown that the addition of simple nonspeech sounds to indicate whether buttons were successfully pressed in a PDA interface, allowed the size of the buttons to be reduced without any significant reduction in task performance. Because audio can be used to improve communication in mobile interfaces, and the uses to which mobile devices can be put is ever increasing, it may be that designers will want to present more information via the audio channel such that more than one item of auditory data will be concurrently presented to the user.

There are several examples of systems where designers have used concurrently presented audio to both increase the available communications bandwidth and present information faster to users. Nomadic Radio [Sawhney and Schmandt 2000] for example, used concurrent presentation of spatialized sounds in a mobile-diary-based context. Different types of audio were used to notify users of news stories, electronic mail messages, and telephone calls. Unfortunately no evaluation was carried out to determine the effectiveness of Nomadic Radio, although informal user comments were positive. Gaver et al. [1991] produced a collaborative cola bottling plant, which used concurrent audio presentation to inform the operators as to the status of the plant and the processes being carried out such as bottling, capping, and so on. Again while no formal evaluation was performed, informal user evaluations were positive.

Although there are advantages to concurrent audio presentation, there are also several disadvantages. Notably, concurrently presented sounds may interfere with each other, and there are unfortunately no guidelines for designers to use when creating interfaces that use concurrent audio presentation to design audio cues, which avoid unwanted interference. This makes it difficult to leverage the true potential of concurrent audio presentation. A previous study [McGookin and Brewster 2002] has identified that these lack of guidelines may be a particular problem for earcons [Blattner et al. 1989], short structured audio messages that can effectively communicate information in a human computer interface. Earcons can be formed from a "grammar," with earcons formed from this "grammar" being similar to each other, for example, sharing the same pitch, rhythmic structure, and so on. Because earcons are similar to each other, they are more likely to interfere when concurrently presented. This paper investigates the extent of these interferences and presents empirically evaluated guidelines for their reduction.

In order to do this we shall give a brief overview of the field of auditory display and introduce earcons, before explaining the problems of concurrent earcons through the use of auditory scene analysis (ASA). We will then show why solving the problems of concurrent earcon presentation are nontrivial, before describing original empirical research undertaken to combat those problems.

2. AUDITORY DISPLAY

2.1 What is an Auditory Display?

There is unfortunately, no explicit definition of what an auditory display is. However, based on the papers that have been published as part of the International Conference on Auditory Display (ICAD) [ICAD 2003], we can consider an auditory display to be the use of sound to communicate information about the state of a computing device to a user. Note that an auditory display is related, but not the same as an auditory interface. Auditory displays only concern themselves with the use of sound to communicate from the computing device to the user. Auditory interfaces may also use sound

(mainly through speech) to communicate from the user to the device. When using auditory display there are four main ways in which data can be encoded into audio (sonification, speech, auditory icons, and earcons). The method chosen to a large extent depends on the data to be communicated. These encoding strategies are described in the following sections.

2.2   Sonification

Sonification can be defined as *"a mapping of numerically represented relations in some domain under study to relations in an acoustic domain for the purposes of interpreting, understanding, or communicating relations in the domain under study"* [Scaletti 1994]. The technique of sonification has been heavily and successfully used in the auditory display of graphs on both mobile devices [Brewster and Murray 2000] and for the blind and visually impaired [Mansur 1985]. Concurrently presenting graphs has also been found to assist users in determining relationships between graphs [Brown et al. 2002]. Sonification has also been used to present information on computer program execution [Vickers and Alty 2000], the presentation of weather reports [Herman et al. 2003], as well as many other areas.

2.3   Speech

while sonification is good for showing trends in large data sets such as graphs, it is less useful for communicating absolute values. To communicate absolute data different techniques are required. The most common way to do this is through speech. Speech can either be synthesized or concatenated from audio recordings of human speakers. Speech has been used in many contexts, from screen readers for blind and visually impaired users such as JAWS [Freedom Scientific 2003], to telephone enquiry systems, airline cockpits, and subway announcement systems [BBC News 2002]. Its popularity in these contexts is most likely due to the meaning of the auditory messages being obvious. Unlike the other three forms of data encoding, it is likely that the way in which the data are mapped to sound will already be understood. That is, the user will have been able to interpret the language used since childhood. This can mean that speech is used in cases where other forms of auditory mapping would be better suited. As Petrie et al. [1998] notes on interfaces for the visually impaired using speech, *"Most interfaces for this group use synthetic speech to convey both the contents of the application and the interface element. This is potentially confusing, a relatively slow method to communicate information."* Speech may also have problems for the presentation of continuous data such as would visually be presented in a graph [Yu and Brewster 2003], which would be better displayed using sonification (see Section 2.2).

2.4   Auditory Icons

while speech can communicate absolute values, it is not without drawbacks, notably communicating information can be time consuming [Petrie et al. 1998]. In order to combat these issues other techniques have been developed. Auditory icons defined by Gaver [1997] as *"Everyday sounds mapped to computer events by analogy with everyday sound-producing events,"* are another popular way of mapping data to sound in an auditory display. Here, everyday familiar sounds are mapped onto computer events to which there is some obvious relationship. For example, sounds such as breaking glass and water pouring can be mapped onto the computer events "error" and "copying." Auditory icons have been used in several systems, such as the ARKola bottling plant [Gaver et al. 1991]. Here a collaborative bottling plant system used auditory icons to communicate information about the various processes involved in producing cola. Gaver et al.'s system could present multiple sounds concurrently to users which the authors claimed allowed users to more easily synchronize multiple machines, since the different sounds would be rhythmic with each other. Although no formal evaluation was performed to investigate the concurrent presentation of sounds, no apparent problems with them were found. This may be attributable to later produced guidelines for auditory icon design by Mynatt [1994] who stated

that dissimilar sounds should be used for auditory icons in case they were confused. As described later in Section 3, this may make them easier to identify when concurrently presented.

Although auditory icons can be easily understood, a lot of this ease comes from having appropriate and intuitive mappings between the sounds and what they represent in a computer interface. In many cases, it may be difficult to find sounds suitable to represent certain abstract events, such as a network connection going down, or a disk drive being full, and so on [Brewster 2002]. In addition, auditory icons can generally only communicate one or two bits of information as they are difficult to parameterize.

2.5 Earcons

while auditory icons are useful for communicating information where there is an intuitive link between the data and the sound used to represent it, they are ill suited to situations where there is no intuitive sound to represent the data. In such cases earcons can be used. Earcons were originally developed by Blattner et al. [1989] and are "*abstract, synthetic tones that can be used in structured combinations to create auditory messages*" [Brewster 1994]. Earcons can be used in all of the places where auditory icons can be used; however, whereas auditory icons rely on intuitive relationships between a data item and the sound used to represent it, earcons use an abstract mapping between a music-like sound and the data. This gives them the advantage of being able to represent any event or interaction in a computer interface; the disadvantage being that the association between sound and event must, at least initially, be explicitly learned. There are four types of earcon (one-element, compound, hierarchical, and transformational) [Blattner et al. 1989], which are described below.

2.5.1 *One-Element Earcons.* One-element earcons are the simplest type and can be used to communicate only one bit of information. These earcons may be only a single pitch or have rhythmic qualities. In either case the one-element earcon, unlike the other three types, cannot be further decomposed to yield more information. In many ways one-element earcons are like auditory icons except they use abstract sounds whose meaning must be learned as opposed to the intuitive meaning of auditory icons.

2.5.2 *Compound Earcons.* Compound earcons are formed by concatenating one-element, or indeed any other form, of earcon together to form more meaningful messages. In many ways they are analogous to forming a sentence out of words, where one-element earcons represent words, and compound earcons represent phrases. For example, three one-element earcons representing "save," "open," and "file" can form compound earcons by being played after each other to form earcons for the "open file" and "save file" operations [Brewster 1994].

2.5.3 *Hierarchical Earcons.* Hierarchical earcons are constructed around a "grammar," where each earcon is a node in a tree, and each node inherits all of the properties of the nodes above it in the tree. Hence an unpitched rhythm might represent an error, the next level will alter the pitch of that rhythm to represent the type of error and so on. This is summarized in Figure 1, taken from Blattner et al. [1989].

2.5.4 *Transformational Earcons.* Transformational earcons are similar to the hierarchical earcons described in Section 2.5.3 in that they are constructed around a "grammar." This has the advantage that instead of having to learn each individual earcon, such as with compound earcons (see Section 2.5.2), it is only necessary to learn the rules by which earcons are constructed in order to understand them. In the transformational earcon type, each auditory parameter such as timbre, pitch, rhythm, and so on, can be transformed or modified to change the meaning of an earcon. Hence, a low-pitched piano rhythm may represent an inexpensive roller-coaster theme park ride, but by changing the timbre to a violin, it now represents an inexpensive water ride. Hence, with this type of earcon each attribute of the data can be mapped to an individual auditory parameter. The common grammar is a strength of earcons as less
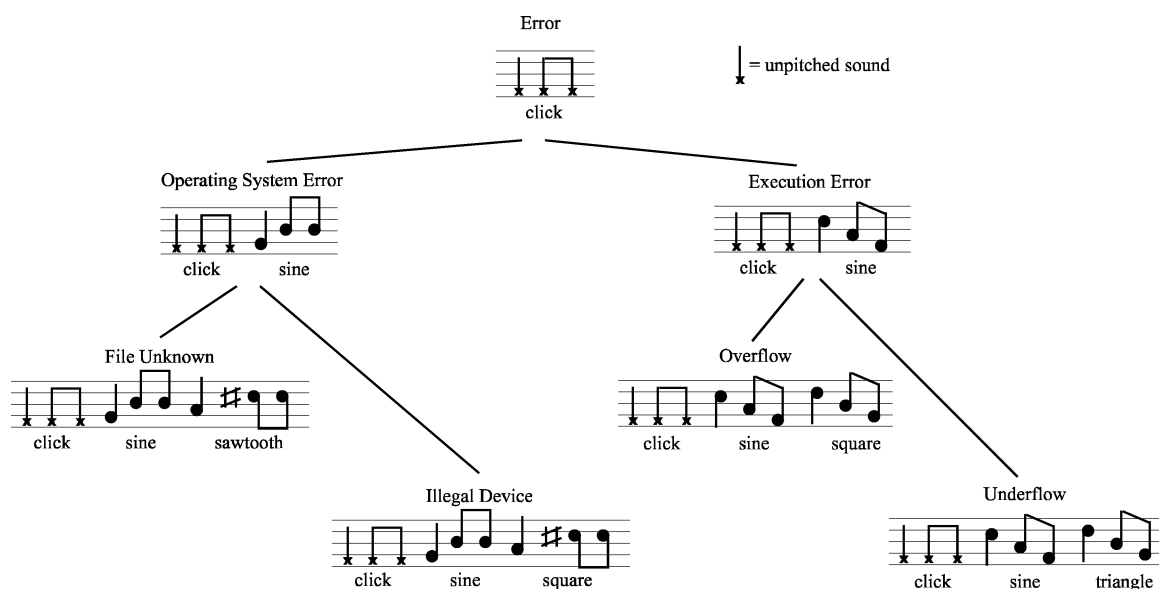
Fig. 1. An overview of the "grammar" used to construct a set of hierarchical earcons representing computer error messages. Taken from Blattner et al. [1989].

learning is required. Rather than having to learn a large number of arbitrary mappings and sounds, as with one-element earcons, it is only necessary to understand the grammar and the small number of different timbres, melodies, or registers used in the grammar to understand the earcons. Unfortunately the grammar is also a problem when earcons are concurrently presented, as those from the same grammar will be quite similar, sharing the same timbre, melody, or register. This, as will be explained in the following section, can cause problems for the interpretation of the earcons. Transformational earcons have been less studied, however, they share many similarities with hierarchical earcons and as Blattner et al. [1989] notes, their principles may be used to shorten hierarchical earcons for "expert users."

2.5.5 *Designing Earcons.* In the previous sections, we have outlined the various different kinds of earcons that can be used to convey data in an auditory display. However, how can earcons be designed to accurately and unambiguously convey the information that is encoded in them? For example, how different must two timbres be to be identifiable as representing different attributes, or how much should two pitches differ in order not to be confused? Brewster [1994] performed various experiments to identify how understandable earcons were and how they could be better designed. Although all of his work dealt with the compound and hierarchical earcon types, it is reasonable to consider it applicable to the two other earcon types because compound earcons are generally formed from one-element earcons and, as already stated in Section 2.5.4, transformational earcons are actually quite similar to hierarchical earcons. Brewster's work lead to a set of guidelines [Brewster et al. 1995], for how auditory attributes should be used in earcons to make them more easily understood by users. These guidelines are outlined below:

Timbre:　　Timbre, in the context of this paper, is defined as the synthesized musical instrument on which an earcon is played. Brewster states that musical instrument timbres that are subjectively easy to tell apart should be used. For example using MIDI instruments brass and organ instead of brass 1 and brass 2. He notes that it is important that care is taken

so that the timbres chosen can play at the registers required (see below), since not all instruments can play all pitches.

Rhythm: Rhythm is taken as the relative durations of notes and the gaps between them [Randel 1978]. Brewster states that rhythms used should be as different as possible. Using different numbers of notes can be effective at doing this. However, care must be taken to ensure that earcons are short enough to keep up with interactions in a computer interface. To ensure this, rhythms with more than six notes should not be used. Using tempo with rhythm is another effective way to improve the differentiation of earcons.

Pitch: Pitch is the perceived frequency of a sound. Hence higher frequency sounds are higher in pitch than low-frequency sounds [Moore 1997]. Brewster notes that the use of complex intra-earcon pitch structures can be effective in differentiating earcons if used with another attribute such as rhythm, thereby creating a melody for each earcon.

Register: Register is related to pitch, it involves retaining the relative pitch differences between consecutive notes but moving an entire sequence of notes up or down in pitch [Randel 1978]. Brewster notes that if identification of register is to be made then it should be used with care, with gross differences between the registers used.

3. AUDITORY SCENE ANALYSIS

3.1 What is ASA?

In the previous sections, we have introduced earcons and how they can be used to communicate data in an auditory display. The most powerful types of earcons are the hierarchical and transformational types. These are powerful because they are composed around a "grammar," meaning that each earcon will share several auditory attributes. For example, they may have the same timbre or melody. In this section, we shall describe research that explains why, when played concurrently, these earcons will interfere with each other.

ASA [Bregman 1994] is the study of how the multiple, complex waveforms that are detected by our auditory system are separated into meaningful representations. ASA is based on gestalts [Williams 1994], which states that the more similar (or different) two auditory sources are along a number of different dimensions, such as familiarity, continuation, and so on, the more likely it will be that they will be merged to be perceived as a single sound (or will be separated to be perceived as different sounds), that is, they will be placed in the same (or different) stream.

It is impossible to cover the entire field of ASA here; however, we will cover the material that is directly relevant to the original work described in this paper. For further background reading the reader is referred to Bregman [1994] and Deutsch [1999]. We will outline the relevant research to this work in gestalt categories derived from Williams [1994].

3.1.1 *Proximity*. Components of a sound that are close to each other are likely to be grouped together. This can occur in two main ways, through frequency proximity and temporal proximity.

Frequency Proximity: Frequency proximity has been shown to have an influence on streaming by several researchers, but most notably by van Noorden [1975], who performed several experiments, which showed that the greater the frequency difference between two alternating tones became, the more likely it was that the tones would be perceived as two separate streams, one stream of high-frequency tones and another of low-frequency tones. van Noorden further showed that if the presentation rates of the tones increased, the frequency difference required to separate the tones into separate streams was reduced, indicating that separating sounds in register will improve their identification when concurrently presented.

Temporal Proximity: Several experiments have shown how the perception of other auditory components can be altered by presenting audio sources at slightly different times. Rasch [1978], as described in Deutsch [1999], presented users with basic patterns that alternated between the subject's ears. He found that making the onset of the tones asynchronous allowed better discrimination of the higher and lower pitched tones. He found that after 30 ms it was possible to discriminate the tones as well as if they had been presented separately. Further evidence to the use of onset synchrony having an effect is shown by Darwin and Ciocca [1992]. They found that by mistiming a harmonic in a complex tone, and by moving the onset of this mistuned harmonic relative to the tone, the mistimed harmonic contributed less to the perceived pitch. When the harmonic was mistimed by 300 ms they found it made no contribution to the perceived pitch.

3.1.2 *Similarity*. Components that have similar attributes are more likely to be grouped together. Singh [1987] found that by presenting a pair of alternating tones, each of which had a different number of harmonic components, a component of timbre, separation of the tones into two separate streams was found to require a smaller difference in frequency between the two tones than if both tones had had the same harmonics. This indicates that presenting each sound with a separate timbre will improve identification when concurrently presented.

3.1.3 *Continuation*. Sounds that continue in a predictable way will be perceived as related. For example, tones that continually rise in frequency will be perceived as related. Heise and Miller [1951] found that if a sequence of tones on a frequency glide (sequences of tones which consistently rise or fall in pitch), and one tone differed too much from the others, it would stand out of the sequence. If sounds can be designed in such a way they may be more robustly identified when concurrently identified.

3.1.4 *Familiarity*. Sequences of sounds that are familiar are more likely to be separated from others, since familiar or known sounds that are embedded in a composite sound stream can be used as patterns with which to pick out the sounds embedded in the longer sequence. However, Bregman [1994] notes that if the components of the sound are somehow placed in different streams, perhaps due to frequency separation (see Section 3.1.1), it would become impossible to identify the existence of the sound. Training participants to identify sounds when they are concurrently presented may help to improve identification.

3.1.5 *Belongingness*. This is an important property where each component can form part of only one stream at any one time [Williams 1994]. It is analogous with the visual illusion of the vase and face. In this illusion it is possible to see either a vase or the profile of two faces looking at each other; however, it is impossible to see both at the same time. In this property the auditory system will work to resolve any conflicts that exists to come to a stable state [Bregman 1994], and having reached that state, the interpretation will remain fixed until it is no longer appropriate [Williams 1994].

3.2 Interdependence of Attributes

Although there is a large body of work on the factors that can cause streaming in composite sounds, there is less research on which attributes dominate over others in the streaming process. It is not clear how many of the factors described in the previous section relate. Instead it seems that the factors influencing ASA are dynamic, each exerting a gravitational like influence to pull an auditory scene into a reasonable consistent interpretation. There is therefore no definitive rule book which will take an auditory scene and determine how it will be perceived. This creates problems for auditory display designers who may wish to play multiple concurrent sounds for a user to interpret, as it is difficult to understand how the sounds will interact together.

4. EARCONS AND ASA

As already shown in Section 3.2, the playing of multiple sounds as part of an auditory display can have undesirable consequences, which by and large can be explained using ASA. These problems are even more of a issue for earcons (see Section 2.5), which use a "grammar" to map information to sound, as all instances of the same set of earcons will be similar, perhaps sharing the same pitch or melody. It is highly likely therefore that multiple concurrently playing earcons will fuse together and become unintelligible. Further, it is not possible to simply make instances of concurrently playing earcons arbitrarily different so that they will not interfere with each other, as this will destroy the "grammar," which makes earcons powerful communicating sounds [McGookin 2002]. This paper describes experiments where specific ASA principles are applied to earcon design to determine how many earcons can be concurrently attended to, and how ASA can be used to modify the design of earcons to work more robustly in concurrent presentation situations without destroying the "grammar" that makes them powerful communicating sounds. In the following sections, we will describe other research which has investigated the identification of concurrently presented audio and describe two experiments which seek to improve the concurrent identification of earcons taken from the same "grammar."

4.1 Related Work

Although the application of ASA to grammar-based earcons has not been previously researched, several researchers have proposed solutions to the problem of presenting concurrent auditory feedback to the user with other forms of sound.

Papp [1997] proposed that an audio server could be used to manage conflicting audio in a user interface. His "Computational Auditory Scene Synthesizer," acted like a controller of the computer's auditory output system. Applications would request that a particular item of auditory feedback was presented to the user and the audio server would decide, based on a set of heuristics that incorporated ASA features, what the impact of introducing the feedback would be based on the sounds already playing. In addition to just accepting or rejecting the request, the audio server could modify the sound to encourage it to stream separately, or present the feedback using a different method (an auditory icon instead of an earcon and so on). while Papp's work has a number of advantages, such as the designer of one auditory display not having to worry about the design of a different auditory display running as another application in a multitasking system, it does have a number of issues. First, as explained in Section 3.2, it is difficult to predict how different ASA factors influence each other, this is even harder for complex auditory sources because most ASA research is done on simple sinusoidal tones [Bregman 1994]. Papp performed no evaluation of his system on users to determine the validity of his criteria for playing or rejecting a sound in the interface. Further, the ability to modify a sound to make it stand out more may be undesirable for many types of sound. The problems with arbitrary modification of earcons have already been discussed (see Section 4), however, it can also be difficult for other auditory information types such as auditory icons (Section 2.4) where the modification of a sound can cause it to become difficult to identify what the sound represents, therefore causing problems for the user [Gaver 1993]. Finally, the ability of the server to dynamically change the type of audio presented based on what audio is currently being presented (perhaps from other applications) may be undesirable because the user could perform exactly the same interactions, in the same application, in the same order, and have different auditory feedback simply on the basis of other concurrently executing applications. This breaks Shneiderman's *strive for consistency* golden rule of interface design [Shneiderman 1998].

Brungart et al. [Brungart et al. 2002; Brungart and Simpson 2002] have investigated the identification of concurrent speakers in multitalker auditory environments. In their experiments, they looked at the identification of coordinate response measure (CRM) speech intelligibility tests. In CRM,

listeners hear one or more simultaneously presented phrases of the following type "*Ready,* **(Call Sign)**, *go to* **(color) (number)** *now*," where call sign is either "Baron," "Charlie," "Ringo," "Eagle," "Arrow," "Hopper," "Tiger," or "Laker," color is one of red, green, blue, or white and number is between one and eight. Brungart et al. [2002] performed several experiments and looked at both the number of competing talkers (number of messages being concurrently presented), and the timbre of the speaker's voice (all text spoken by the same speaker, by different sex speakers, and so on). They found that reducing the number of competing talkers had a positive effect on speech comprehension. They also found that in cases where there were different sex speakers (effectively modifying the timbre of the speaker's voice) speaking different texts, identification was on average 20% better than when all texts were being spoken by a same sex talker. Hence it is possible that timbre variations between concurrently presented sounds would encourage better discrimination.

5. NUMBER OF EARCONS VERSUS IDENTIFICATION

5.1 Motivation

While we have argued that concurrently presenting earcons causes them to interfere with each other and be difficult to uniquely identify, so far no studies have been undertaken to identify the extent to which identification of earcons is impaired by their concurrent presentation.

Studies of other auditory displays techniques, such as speech [Brungart and Simpson 2002] however, have shown that when the number of concurrent audio items is increased, the proportion of audio that can be identified is reduced. The experiment described here attempts to identify if such a trend exists in earcon identification, and the impact on earcon identification of increasing the number concurrently presented.

5.2 Procedure

Sixty four participants undertook the experiment, which was of a between groups design. All participants were aged between 18 and 24, and comprised of a mix of both males and females, and each was paid £5 on completion. The earcons used were of the transformational type and are described below. They were originally produced for the Dolphin multimodal focus and context system [McGookin and Brewster 2002]. Each earcon encoded three parameters of a theme park ride, the ride type (either roller coaster, static ride, or water ride), the intensity of the ride (either low intensity, medium intensity, or high intensity), and the cost of the ride (either low cost, medium cost or, high cost).

The main hypothesis of the experiment was that varying the number of concurrently presented earcons, would significantly alter the proportion of presented earcons, which could be successfully identified. The independent variable (IV) was the number of earcons concurrently presented, and the dependant variables (DVs) were the number of earcons, ride types, ride intensities, and ride costs successfully identified. Understanding how the number of earcon attributes correctly identified varies with the number presented is important, as although all attributes of an earcon are important, it may not always be the case that identifying all attributes of an earcon is always necessary. In addition, it was hypothesized that varying the number of earcons concurrently presented would significantly affect the subjective workload of participants. A set of modified NASA Task Load Index (TLX) scales [Hart and Staveland 1988] were used to collect data to test this hypothesis.

There were four conditions in the experiment: the one-earcon condition (where only one earcon was presented "concurrently"), the two-earcon condition (where two earcons were concurrently presented), the three-earcon condition (where three-earcons were concurrently presented), and the four-earcon condition (where four earcons were concurrently presented). The one-earcon condition, while not incorporating concurrent earcon presentation, since only one earcon was presented at a time, allowed a

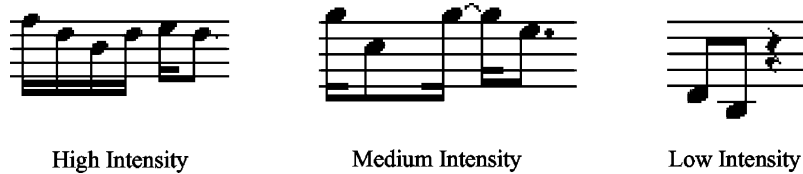High Intensity   Medium Intensity   Low Intensity

Fig. 2. Melodies used to represent High, medium, and low ride intensity theme park rides in the earcons.

general comparison of earcon identification to be made back to the work of Brewster [1994] who determined the identification levels of single-earcon presentation, and therefore a determination of the quality of the earcons used in this experiment to be made. Each condition in the experiment had two phases, a training phase and a testing phase. Both are described later in this section.

5.2.1 *Earcons Used.* The earcons used in the experiments were based on those designed and used by McGookin and Brewster [2002]. All of the earcons represented rides that may appear in an amusement/theme park. Earcons were of the transformational type and encoded three bits of information: the type of ride, the intensity of the ride, and the cost of the ride. How these values were encoded is described in the following subsections. All of the earcons were recorded from the output of a Roland super JV-1080 musical instrument digital interface (MIDI) synthesizer. Care was taken to ensure that the resultant recordings were of equally perceivable loudness to avoid low-level masking occurring [Moore 1997].

Type of Ride: There are three possible ride types: roller coasters, water rides, and static rides. The type of ride was encoded with timbre. According to Brewster's guidelines [Brewster et al. 1995], musical timbres that are subjectively easy to tell apart should be used. Therefore a grand piano (general MIDI instrument No. 000) was used to represent a roller coaster, a violin (general MIDI instrument No. 040) was used to represent a water ride, and a trumpet (general MIDI instrument No. 056) was used to represent a static ride. These instruments were chosen based on the subjective grouping work of Rigas [1996], where he categorized musical instruments based on how subjectively similar listeners rated them.

Intensity of Ride: The intensity of the ride can also be though of as how fast the ride is going. As with the type of ride (see Section 5.2.1), there are three possible values: low, medium, and high. The ride intensity attribute was mapped onto melody, which we regard as a rhythm, and a pitch structure for that rhythm. The guidelines of Brewster et al. [1995] were again employed by varying the number of notes in each rhythm, with two, four, and six notes used respectively for low, medium, and high ride intensities. The melodies for each of the ride intensities are given in Figure 2.

Cost of Ride: As with the intensity of the ride, the ride cost attribute can be either low, medium, or high. This attribute was mapped onto the register in which the earcon is played. Although Brewster's guidelines [Brewster et al. 1995] generally advise against register, we have used the gross differences between the registers that guidelines recommend, as well as staggering the notes for each melody in each register to avoid the same melodies having musically harmonic intervals with each other when played in different registers. Register was mapped in such a way that the low-cost rides were played in the lowest register, medium-cost rides were played in the next lowest cost register, and high-cost rides were played in the highest register. The registers used respectively were the octave of C4 for low cost, the octave of C5 for medium cost, and the octave of C6 for high cost.

5.2.2 *Training Phase.* The training phase involved participants reading a sheet, which described the grammar of the earcons used, followed by 10 min of self-guided training via a Web page interface,

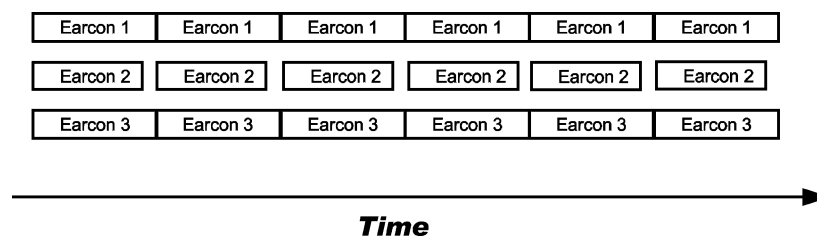| Earcon 1	Earcon 1	Earcon 1	Earcon 1	Earcon 1	Earcon 1
Earcon 2	Earcon 2	Earcon 2	Earcon 2	Earcon 2	Earcon 2
Earcon 3	Earcon 3	Earcon 3	Earcon 3	Earcon 3	Earcon 3

Time

Fig. 3.   An example of the presentation of concurrent earcons. Earcons are presented seven times concurrently, but synchronize their starts together. Taken from the three earcon condition.

where participants could listen individually to all possible earcons which could be composed from the earcon grammar. After this time participants were presented with three earcons independently, and were required to identify them without any form of assistance or feedback. If participants were unable to successfully identify the three earcons, they were given back the earcon grammar sheet and the Web page for a further 5 min of training, before retaking the test. If they were still unable to successfully identify the earcons, they took no further part in the experiment.

This training technique for earcons is similar to that used by Brewster [1998], who found that allowing participants to train by listening to the earcons used provided a significant improvement in task performance.

5.2.3 *Testing Phase.* The testing phase involved participants listening to 20 sets of concurrently presented earcons. The presentation order of each set was randomly selected for each participant to avoid order effects in the data. Each set of earcons was dioticly presented (monaural presentation to both ears) to participants seven times in rapid succession (see Figure 3 for an example of presentation from the three earcon condition), as Bregman [1994] notes that constant repetition helps to build up the effects of streaming. The number of earcons in each set was dependant on the experimental condition, with four, three, two, and one earcons used in the four-, three-, two-, and one-earcon conditions.

Participants had to identify the attributes of each earcon in a set and record what those attributes were in a dialogue box. See Figure 4 for an example of the dialogue box as used in the four-earcon condition. Participants were presented with the response dialogue box as the earcons started to be played, and were able to fill in the dialogue box as the earcons were being presented, or wait until the earcon presentations had ended before filling in responses. The experiment was of a forced choice design. Between successive sets of earcons, a mandatory rest break for participants of at least 4 s was introduced, since some research suggests that this is the period of time required by the auditory streaming mechanism to fully reset itself [Bregman 1994], thereby removing any possibility of one set of earcons influencing the perception of the next. Before the testing phase proper, each participant carried out a reduced version of the testing phase involving two sets of earcons, not used in the testing phase proper, to familiarize themselves with the experimental procedure.

5.3 Results

5.3.1 *Identified Earcons and Attributes.* To determine the number of earcons, ride types, ride intensities, and ride costs correctly identified by participants, the following method was used.

For each set of (one, two, three, or four) presented earcons, the set of earcons presented (SEP) and the set of participant responses to those earcons (SPR) were compared. If all parameters of an earcon (ride type, ride intensity, and ride cost) in the SPR matched an earcon in the SEP, and neither earcon had already been correctly identified (marked as allocated), the number of correctly identified earcons was increased by one, and both earcons were marked as allocated, that is, they had been correctly identified.

Fig. 4.   A screenshot of the dialogue box used by participants to fill in earcons identified. Taken from the four earcon condition.
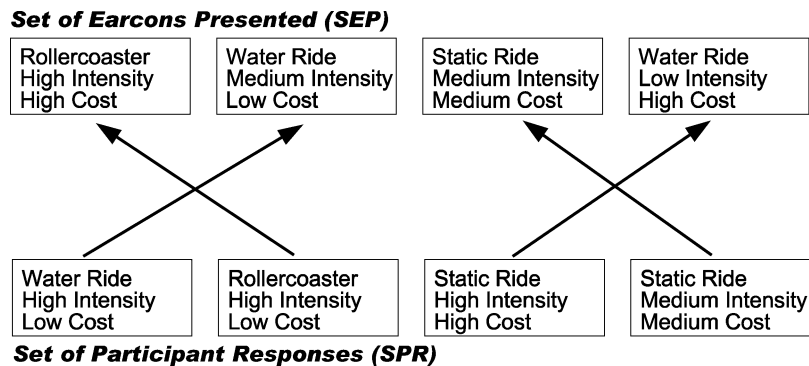

Fig. 5.   An example from the four earcon condition of how the set of participant responses (SPR) was mapped to the set of earcons presented (SEP) to determine the number of correctly identified earcons, ride types, ride intensities and ride costs.

Once the number of earcons which had been correctly identified had been determined, the number of correctly identified ride types, ride intensities, and ride costs were determined. All possible permutations of earcons from the SPR, which had not been fully correctly identified and thus were not allocated, were compared against the unallocated earcons from the SEP, and the attributes were compared. The permutation that yielded the highest overall number of correctly identified ride attributes, was used to determine the number of correctly identified ride types, ride intensities, and ride costs. An example mapping between the SEP and SPR, from the four-earcon condition is shown in Figure 5.

To guard against cases where in some conditions participants may have correctly identified all the earcons and therefore the number of correctly identified earcon attributes may be low, the number of correctly identified earcons was added onto the number of correctly identified ride attributes, since
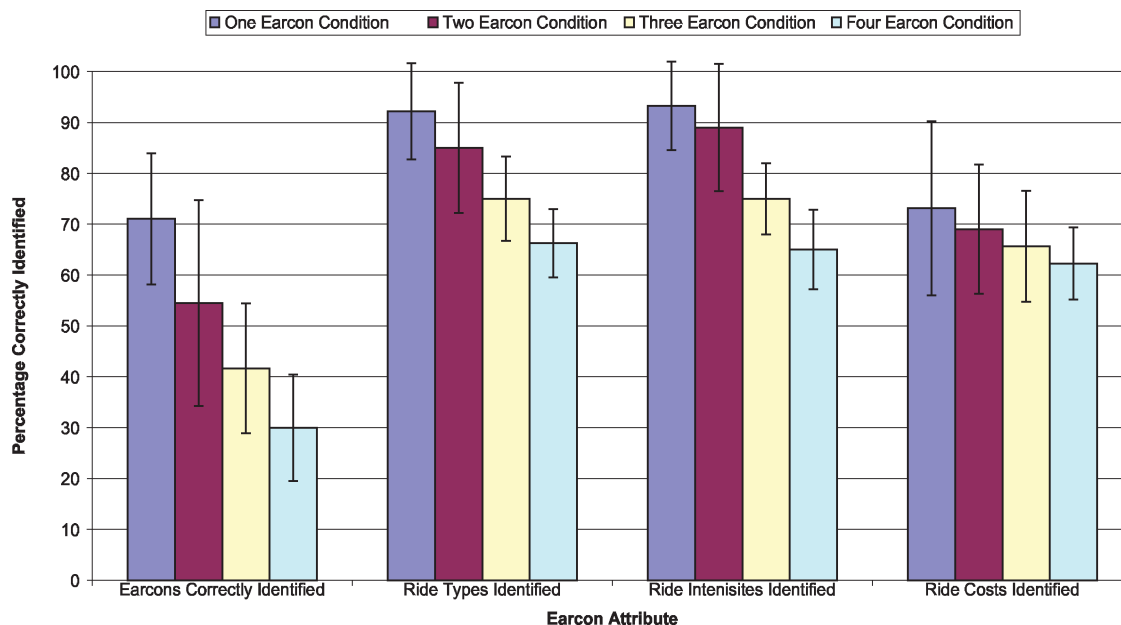
Fig. 6.   Graph showing the average proportion of earcons, ride types, ride intensities and ride costs correctly identified for the one, two, three, and four earcon conditions, shown with standard deviations.

a correctly identified earcon also represents a correctly identified ride type, ride intensity, and ride cost.

One issue with this experiment was that the number of earcons to be identified in each condition was different. Because of this, a direct numerical comparison between the earcons and their attributes correctly identified for different conditions would be unfair, and of limited value. Therefore, the average number of earcons, ride types, ride intensities, and ride costs identified by each participant were converted into percentages of the number of earcons that were concurrently presented. For example, in the three-earcon condition, if on average two earcons were correctly identified, the percentage was calculated as $(2/3) \times 100 = 66\%$. The average proportion of correctly identified earcons, ride types, ride intensities, and ride costs is presented graphically in Figure 6.

To determine if any of the differences shown in Figure 6 were statistically significant, four between groups, one-way analysis of variance (ANOVA) tests were carried out, one for the percentage of correctly identified earcons, and one for each of the percentage of earcon attributes (ride type, ride intensity, and ride cost) correctly identified. The ANOVA for percentage of earcons correctly identified was found to be significant ($F(3, 60) = 23.28, p < 0.001$). Post hoc Tukey HSD tests showed that earcons were significantly better identified in the one-earcon condition than in the two-, three-, and four-earcon conditions. The ANOVA for percentage of ride types identified was also found to be significant ($F(3, 60) = 22.29, p < 0.001$). Here, post hoc Tukey HSD tests showed that ride types were significantly better identified in the one- and two-earcon conditions than in the three- and four-earcon conditions and ride types in the three-earcon condition were significantly better identified than those in the four-earcon condition. For the percentage of ride intensities correctly identified, the ANOVA again showed significance ($F(3, 60) = 31.16, p < 0.001$). Post hoc Tukey HSD tests showed that the ride intensities were significantly better identified in the two-earcon condition than in the three-earcon condition, and ride intensities in the three-earcon condition were significantly better identified than those in the
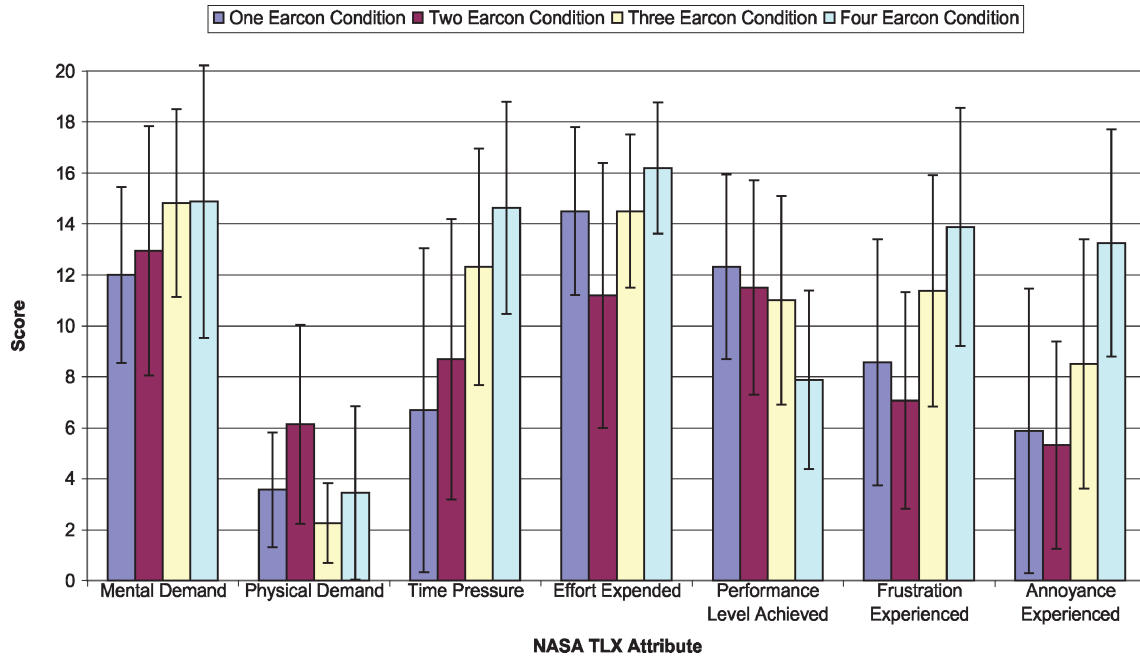
Fig. 7. Graph showing the mean values for NASA TLX workload data for the one-, two-, three-, and four-earcon conditions, shown with standard deviations.

four-earcon condition. The ANOVA for ride costs identified was not found to be significant ($F(3, 60) = 2.24$, $p = 0.093$).

5.3.2 *Workload Data*. In addition to collecting data about the number of earcons and their attributes that were correctly identified, participants also completed modified NASA TLX questionnaires for each condition (see Section 5.2). A graphical summary of these data is presented in Figure 7. Overall workload was calculated as a summation of each participant's individual workload attributes (excluding annoyance experienced). An ANOVA test on these data showed significance ($F(3, 60) = 5.96$, $p = 0.001$). Post hoc Tukey HSD tests identified that workload was significantly lower in the one- and two-earcon conditions than in the four-earcon condition ($p < 0.05$).

To determine which attributes contributed to this variation in workload between conditions, seven one-way ANOVA's, one for each NASA TLX attribute were carried out.

For the performance level achieved, the ANOVA showed significance ($F(2, 45) = 6.19$, $p = 0.004$). Post hoc Tukey HSD tests showed that participants had rated performance for the one-earcon and two-earcon conditions significantly higher than for the four-earcon condition ($p < 0.05$). For frustration experienced, the ANOVA again was significant ($F(2, 45) = 9.71$, $p < 0.001$). Post hoc Tukey HSD tests showed that participants had rated frustration significantly higher in the four-earcon condition than the two- and one-earcon conditions ($p < 0.05$). The ANOVA for effort expended was significant ($F(2, 45) = 6.94$, $p = 0.002$). Post hoc Tukey HSD tests showed that effort was rated significantly higher in the four-earcon condition than the two- and one-earcon conditions ($p < 0.05$). Effort was also rated significantly higher in the two-earcon condition than in the one-earcon condition. For the amount of annoyance experienced, the ANOVA was again significant ($F(3, 60) = 9.04$, $p < 0.001$). Post hoc Tukey HSD tests showed that annoyance for the three-, two-, and one-earcon conditions was rated significantly lower than the four-earcon condition ($p < 0.05$). For time pressure, the ANOVA again showed significance
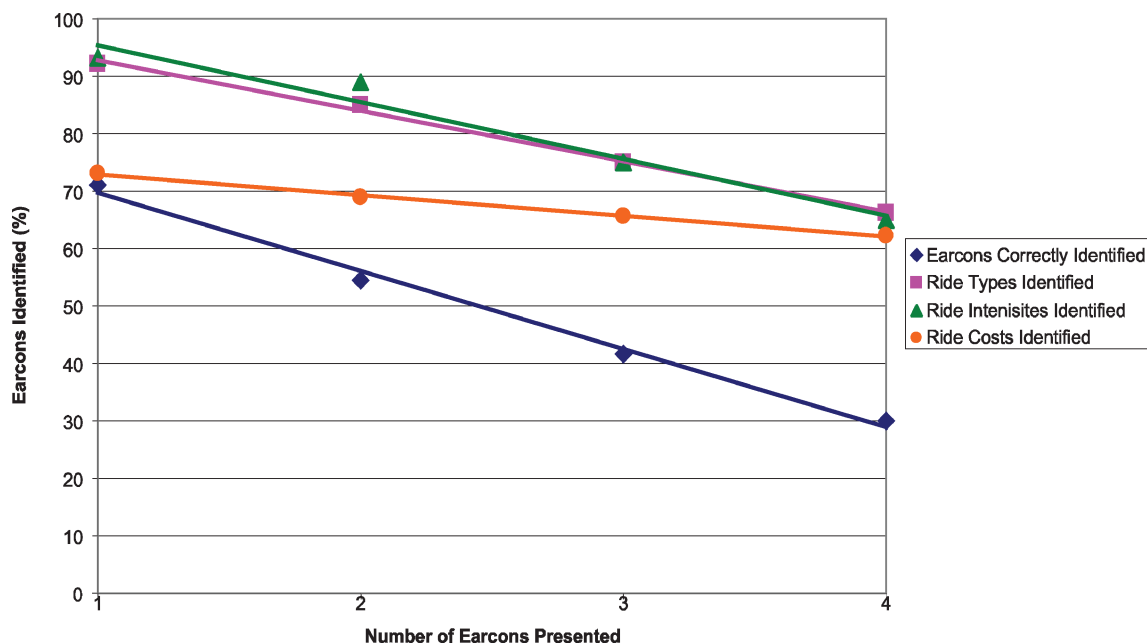
Fig. 8.   A graph comparing the relative performance of the one-, two-, three-, and four-earcon conditions showing trend lines for predicted performance.

$(F(2, 45) = 9.28, p < 0.001)$. Post hoc Tukey HSD tests showed that participants rated time pressure significantly higher in the four-earcon condition than the two- and one-earcon conditions ($p < 0.05$). Significantly higher time pressure was also recorded for the three-earcon condition over the one-earcon condition ($p < 0.05$). The ANOVAs for mental demand ($F(2, 45) = 1.66, p = 0.186$) and physical demand ($F(2, 45) = 3.47, p = 0.060$) were not significant.

5.4   Discussion

The results from the previous section have confirmed that trying to identify concurrently presented earcons is difficult, with on average only 30% of earcons correctly identified in the four-earcon condition. The results from the one-earcon condition showed around 70% accuracy in earcon identification. Although using a different experimental procedure, Brewster [1994] found similar levels of identification for individually presented earcons. The set of earcons used in this work can therefore be regarded as being of similar quality. Therefore, the problems with the concurrent identification of earcons cannot be attributed to a poorly designed earcon set, which causes poor individual earcon identification.

The results from Section 5.3 support the hypothesis that varying the number of concurrently presented earcons has a significant effect on a participant's ability to identify those earcons. The greater the number of earcons concurrently presented to participants, the lower the proportion of those earcons that can be successfully identified. This can be illustrated by the graph in Figure 8, which shows best-fit trend lines between the number of concurrently presented earcons and the percentage of earcons and earcon attributes correctly identified. The trend for the percentage of correctly identified earcons agrees with the work of Brungart et al. [2002] with a similarly sharp drop in performance as the number of concurrently presented values is increased.

Additionally, Figure 8 shows that while the percentage of correctly identified ride types and intensities closely follow the same trend as the percentage of correctly identified earcons, the percentage of correctly

identified ride costs has a much shallower gradient. This is believed to be a combination of the relatively poor performance of ride cost identification as well as the inharmonic register intervals between the three registers chosen to represent ride cost (see Section 5.2.1).

While the proportion of correctly identified earcons is greatly reduced as the number concurrently presented is increased, identification of individual earcon attributes is much higher, dropping to only 70% correct for ride types (timbre) and ride intensities (melody) when four earcons were concurrently presented. It may not always be the case that all parameters of an earcon are required to be identified to make use of the earcon, therefore, in real-world tasks the identification of individual attributes may be a more realistic metric of user performance. Additionally, the earcons used encoded three separate data attributes, this is around the maximum number that can be encoded in any one earcon, and no studies have been carried out which attempt to incorporate more than three data attributes into an earcon. The earcons used in this study can therefore be regarded as being of maximum complexity, and earcon identification can be regarded as being a "worst-case scenario." If the complexity of the earcons was reduced by removing one or more data attributes encoded in each earcon, identification performance may increase.

From the experiment described in this section the following guidelines for designers of auditory displays that use concurrent earcons can be drawn:

Guideline 1. Increasing the number of concurrently presented earcons significantly reduces the proportion of the earcons that can be successfully identified. Increasing the number of earcons concurrently presented can reduce correct identification from 70% to 30%. Great care should be taken when considering the amount of information users will need to extract from earcons when considering the number of earcons, which will be concurrently presented.

Guideline 2. If register is used to encode a data attribute, it may be beneficial to ensure that inharmonic intervals are used between concurrently presented registers. This is likely to reduce the impact on register identification when the number of concurrently presented earcons is increased.

6. BUILDING MORE ROBUST EARCONS

6.1 Motivation

As was discussed in Section 5.4, varying the number of concurrently playing earcons has a significant effect on the proportion of earcons that can be successfully identified by the user. This work, however, fails to take account of the specific interactions that can occur between concurrently playing earcons. For example, if two roller coasters of the same cost but different ride intensities are concurrently presented, the listener will have a limited amount of information (their a priori knowledge about earcon melodies from the earcon training phase) as to what the intensities of the two roller coaster rides will be. This is because the user will only be able to use the temporal differences between consecutive notes in the composite sound to separate the two earcons as all of the other attributes will be the same. To overcome this problem, we could simply play one earcon with a tuba timbre, instead of a piano timbre. However, recall from Section 4, that earcons are powerful because they are composed from a "grammar" and hence relationships exist between earcons from the same set. Changing the earcons to overcome the streaming problem just described would mean destroying this relationship. Therefore, a tension exists between retaining the relationships between earcons that make them powerful communicating sounds while making each individual earcon different enough to minimize the interactions that make it difficult to perceptually separate when played concurrently.

In this section, we apply some ASA modifications to the earcons described in Section 5.2.1 and evaluate if they are useful in improving identification of concurrently presented earcons. There are several points with which to note when trying to implement ASA research in the way we describe here. First, ASA

research by and large has tended to deal with pure sinusoidal tones or with complex tones composed of multiple harmonics of the same sinusoid. Earcons, as shown by Brewster [1994], are best identified when complex musical timbres are used. Also, ASA experiments tend to be composed of a participant listening to sounds, which will either be perceived as one or more auditory streams, an auditory parameter will be altered and the participant must indicate when the two streams merge into one, or when one stream perceptually segregates into two separate streams. In earcon identification, it is not simply sufficient to identify if one or two earcons are being presented, users must also be able to identify data that are encoded in each individual earcon. Bearing this in mind, the following sections describe the conditions of an experiment designed to apply the various ASA principles outlined in Section 3, to the design and presentation of earcons.

One feature of ASA, which may cause a significant improvement in earcon identification, is the use of 3D or spatialized sound. Here each earcon could be positioned in a different spatial location. This would fall into the gestalt category of proximity, where objects close to each other are grouped together [Williams 1994]. While spatialization of earcons could be argued as one of the more important features influencing how audio is grouped into streams as Bregman [1994] notes "*it would be a good bet that sounds emanating from the same spatial location have been created by the same source.*" We decided not to include this feature in the experiments described here because we believe there are several advantages in looking at 3D and non-3D features separately.

First, due to computational requirements and lack of appropriate hardware, many mobile devices do not have good quality spatial positioning ability. It may also be inconvenient for the user to wear headphones to use spatial feedback, for example, it is unlikely that a user would wear headphones to specifically interact with a mobile telephone menu. Also, even when using spatial positioning, it is difficult to know how far apart sound sources would need to be for them to be identifiable and distinct. In real-world scenarios it may not always be possible, even when using spatial positioning, to keep important audio objects apart (e.g., cartographic data). The use of spatialized presentation of concurrent earcons is, however, a topic that we are currently investigating due to the improvements in performance it may bring.

6.2 Methodology

Sixty-four participants between the ages of 18–24 undertook this experiment, which was of a between groups design. There were five conditions (the *original earcon set condition*, the *melody-altered earcon set condition*, the *multitimbre earcon set condition*, the *extended training condition*, and the *staggered onset condition*) in this experiment. Written consent was obtained from all participants prior to the start of the experiment, and each participant was paid £5 on completion. As with the experiment from Section 5, each condition consisted of a training phase and a testing phase, which are described in Section 6.2.1 and 6.2.2. The earcons used in this experiment were the same as those used in the previous experiment (see Section 5.2.1). Each earcon encoded three parameters of a theme park ride: the ride type (either roller coaster, static ride, or water ride), the intensity of the ride (either low intensity, medium intensity, or high intensity), and the cost of the ride (either low cost, medium cost, or high cost).

6.2.1 *Training Phase.* The training phase was largely the same as that used for the previous experiment. Participants were provided with a page describing the grammar from which the earcons for that particular condition were derived. When participants had read and understood the grammar, they were provided with a web page containing all of the earcons derived from the earcon grammar. Participants were allowed to listen to the earcons individually for 10 min. After this time participants were asked to independently, without any form of assistance, identify three individually presented earcons. If participants were not able to correctly identify the earcons, they were provided with the sheet describing the

earcon grammar and the web page containing the earcons for a further 5 min. After this time the test was carried out again. If participants were still unable to correctly identify the three earcons they took no further part in the experiment.

6.2.2 *Testing Phase.* The testing phase of this experiment was similar to the testing phase used for the experiment in Section 5. It comprised of participants listening to 20 sets of dioticaly, concurrently presented earcons, and trying to identify the attributes of each earcon presented. Unlike the previous experiment, each earcon set contained four earcons. The four earcons used in each set were randomly selected. The same 20 sets of concurrent earcons were used for all conditions, and were randomly presented to take account of order effects. Each earcon set was repeatedly played seven times and participants recorded the attributes of each presented earcon in a dialogue box (see Figure 4 for an example of the dialogue box used). As with the previous experiment, a mandatory 4 s break was introduced between successive sets of earcons to allow the ASA mechanism to reset [Bregman 1994]. Before carrying out the testing phase proper, participants carried out a shortened version involving two sets of earcons, which were not used in the experiment, to familiarize themselves with the experimental condition.

6.3 Conditions

6.3.1 *Original Earcon Set Condition.* This condition is the same as the four-earcon condition from the previous experiment described in Section 5. The earcons from Section 5.2.1 were used "as is." Because of this, the same data for the four-earcon condition in the previous experiment (Section 5) was used. This condition acted as a baseline with which to compare the modifications described in the other conditions to.

6.3.2 *Multitimbre Earcon Set Condition.* Although there is no universal definition of what timbre is [Gaver 1997], there is a body of research which shows that known elements of timbre can affect how tone sequences are perceived. Singh [1987] showed that two alternating tones (A and B), each containing a different number of harmonics, required a smaller frequency separation to be heard as two separate streams (one composed of A tones and another composed of B tones), than when both tones had the same number of harmonics. Also, as described in Section 4.1, Brungart et al. [2002] found that in a multitalker speech environment, having different sex speakers for each spoken test caused a 20% improvement in comprehension.

In this condition, whenever more than one earcon of a particular ride type was presented concurrently, each was presented with a different musical instrument from the same instrument group. Hence, if more than one roller coaster was presented at a time, one roller coaster would use an acoustic grand piano timbre (general MIDI instrument No. 000) and the other roller coaster would use an electric grand piano timbre (general MIDI instrument No. 002). Although one of the guidelines of good earcon design by Brewster et al. [1995], was that "*Musical instrument timbres that are subjectively easy to tell apart should be used,*" we do not require participants to be able to specifically tell the difference between an acoustic and electric piano, rather that the slight differences between them will help concurrently playing earcons of the same ride type to perceptually separate from each other significantly better without destroying the "grammar" of the earcon set being used. The groupings were derived from Rigas [1996]. However, due to the difficulties in getting three distinct groupings of instruments, two "folk" instruments and a marimba were used to represent the water ride type.

The earcons used in this condition incorporate the same "grammar" described in Section 5.2.1. The three MIDI timbres used for each of the ride types is given in Table I.

Table I. Timbres Used for the Rides in the Multitimbre Earcon Condition

Ride Type	Instrument	General MIDI Instrument No.
Roller coaster	Acoustic grand piano	000
Roller coaster	Electric acoustic piano	001
Roller coaster	Electric grand piano	002
Static ride	Tuba	058
Static ride	French horn	060
Static ride	Synth brass 1	062
Water ride	Marimba	012
Water ride	Shamisen	106
Water ride	Kalimba	108

High Intensity   Medium Intensity   Low Intensity

Fig. 9.   Melodies used for the earcons in the melody-altered earcon set condition.
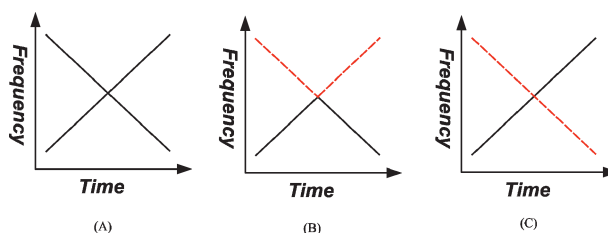

(A)   (B)   (C)

Fig. 10.   An example of two crossing frequency glides (A) and how they will be perceived by the auditory system if they are presented with the same timbre (B) and with different timbres (C).

6.3.3 *Melody-Altered Earcon Set Condition.* This condition used the earcons from Section 5.2.1, however, the melodies were modified such that successive notes in each melody consistently fell, stayed the same, or rose in pitch. The melodies for low, medium, and high cost rides are shown in Figure 9. As described in Section 3.1.3, sequences of tones that glide in this way were shown to provide better streaming than those which did not.

One problem with this approach is that if two separate melodies are played concurrently, they will cross over each other (see Figure 10(A)). Studies which have investigated this have shown that at the point of intersection of the two glides, grouping will occur by frequency (pitch) proximity, rather than by continuation [Tougas and Bregman 1985; Halpern 1977; Deutsch 1999]. The example glides in Figure 10(A) would therefore be perceived as two glides which bounce apart (Figure 10(B)), rather than two glides which cross (Figure 10(C)). Halpern [1977] identified that if the timbre of one of the glides was different, the glides would be heard as shown in Figure 10(C). However, it is unclear how different two timbres would need to be. It is therefore not possible to guarantee that if two earcons are concurrently presented their timbres will be sufficiently different; indeed they may have the same timbre (see Section 5.2.1). If the modifications to the timbre of the earcons previously discussed have a positive effect on earcon identification, then each concurrently presented earcon would have a unique timbre, which might be enough for frequency glides to stream separately. This issue is further discussed in Section 6.5.

6.3.4 *Extended Training Condition.* Although before performing any conditions of the experiment, participants were given training on the "grammar" of the earcons used and allowed to listen to them individually before the experiment, participants were not given any specific training on how to listen to concurrently presented earcons. On the basis of the familiarity gestalt principle discussed in Section 3.1.4, a tool was built that would allow participants to listen to two sets of four concurrently playing earcons (which were not used in the experiment proper) for them to better understand how adding and removing earcons affected the composite sound. This condition, therefore, was the same as the original earcon set condition (Section 6.3.1) with the addition of the concurrent training tool used prior to the experimental condition.

6.3.5 *Staggered Onset Condition.* As described in Section 3.1.1, introducing a time delay between the onsets of concurrently presented sounds can have an impact on how they are perceived. In this condition, although all four earcons were still presented together, a 300 ms onset-to-onset delay was introduced between the start of each earcon.

6.3.6 *Final Condition.* After the conditions previously outlined were performed, a preliminary analysis was carried out to identify which conditions caused significant improvements in earcon identification. These conditions were then combined to identify the composite improvement in identification. The conditions incorporated here were the multitimbre earcon set condition (see Section 6.3.2) and the staggered onset condition (see Section 6.3.5).

6.4 Results

6.4.1 *Identified Earcons and Attributes.* As with the previous experiment, participant's responses to the earcons presented were collected and the number of correctly identified ride types, ride intensities, ride costs, and earcons were determined using the method from Section 5.3.1. A preliminary analysis of the results revealed significant improvements in earcon attribute identification in the multitimbre earcon set condition, and the staggered onset earcon condition over the original earcon set condition. To determine the overall improvement in earcon identification, the features of these two conditions were combined, to yield the final condition, which was performed using the same method as the other conditions with a further 16 participants. The average number of correctly identified earcons, ride types, ride intensities, and ride costs for all six conditions are presented graphically in Figure 11.

To determine if any of the differences in the data presented in Figure 11 were statistically significant, four one-way ANOVA tests, one for each of the parameters, were carried out. The ANOVA for number of rides identified was found to be significant ($F(5, 90) = 7.12, p < 0.001$). Post hoc Tukey HSD tests showed that the staggered onset and final conditions were significantly better identified than the original earcon set condition. The ANOVA for the number of ride types identified was also significant ($F(5, 90) = 7.84, p < 0.001$). Post hoc Tukey HSD tests revealed that the multitimbre earcon set, the staggered onset, and the final conditions were significantly better identified than the original earcon set condition. For the number of ride intensities identified, the ANOVA was again significant ($F(5, 90) = 3.16, p = 0.011$). Post hoc Tukey HSD tests showed that the staggered onset and multitimbre earcon set conditions were significantly better identified than the melody-altered earcon set condition. The ANOVA for ride costs identified was not significant ($F(5, 90) = 0.31, p = 0.907$).

6.4.2 *Workload Data.* As with the results from the experiment of Section 5, modified NASA TLX workload ratings were recorded. The average score for each rating in each condition is presented graphically in Figure 12. To determine if workload significantly differed between conditions, overall workload was calculated as a summation of each participant's individual workload attributes (excluding annoyance experienced). An ANOVA on these data failed to show significance ($F(5, 90) = 0.69, p = 0.629$).
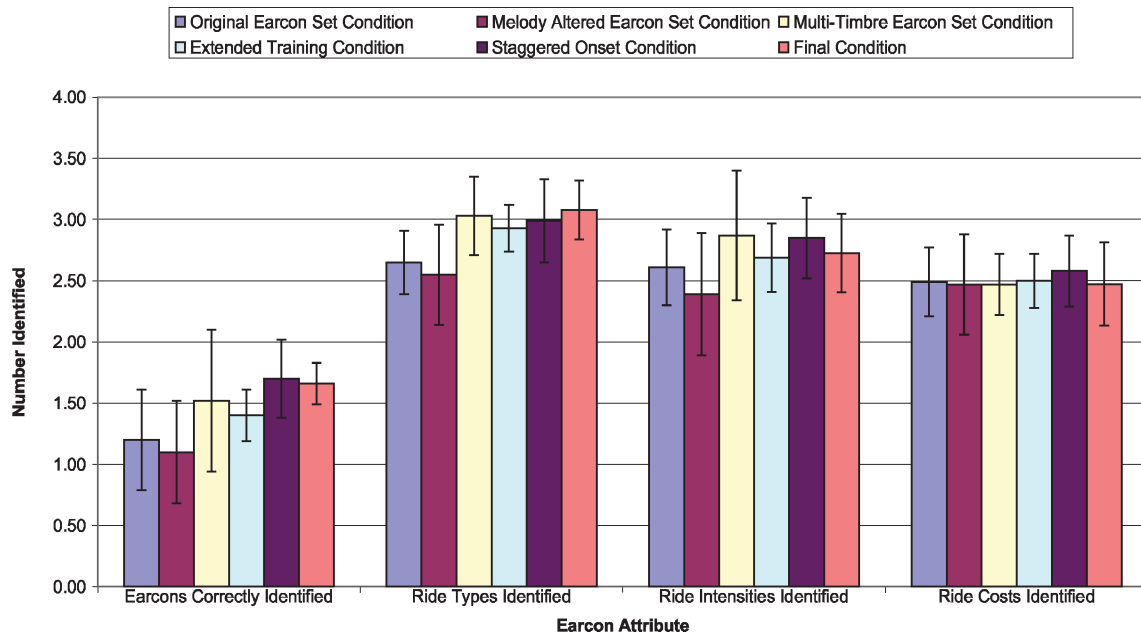
Fig. 11.   Graph showing the average number of earcons, number of correctly identified ride types, ride intensities, ride costs, and their standard deviations for the original earcon set, melody-altered earcon set, multitimbre earcon set, extended training, staggered onset, and final experimental conditions, shown with standard deviations.
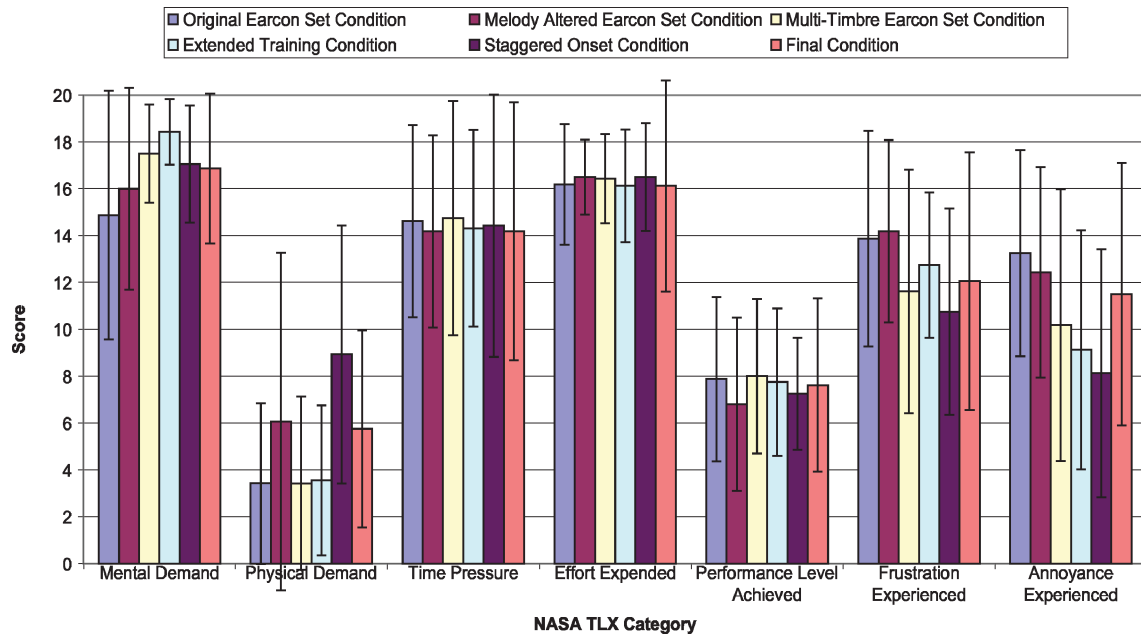

Fig. 12.   Graph showing the mean values for NASA TLX workload data with standard deviations for the original earcon set, melody-altered earcon set, multitimbre earcon set, extended training, staggered onset, and final experimental conditions, shown with standard deviations.

An additional ANOVA was carried out separately on the scores for annoyance experienced. This ANOVA showed significance ($F(5, 90) = 2.33$, $p = 0.049$). Post hoc Tukey HSD tests showed that the staggered onset condition was judged to be significantly less annoying than the original earcon set condition ($p < 0.05$).

6.5 Discussion

The results show that both the multitimbre earcon set condition and the staggered onset condition significantly improved the identification of earcon attributes over the original earcon set condition. The number of correctly identified ride types in each condition being significantly greater than the number correctly identified in the original earcon set condition. The number of earcons that were correctly identified in the staggered onset condition was also significantly greater than in the original earcon set condition.

The final condition which incorporated the staggered onset condition and multitimbre earcon set condition, these being the only conditions which significantly increased the identification of earcons or their attributes over the original earcon set condition, had significantly higher earcon and ride type identification than the original earcon set condition. However, there was no significant interaction between the final, staggered onset, and multitimbre earcon set conditions. Given that both the staggered onset and multitimbre earcon set conditions significantly improved the identification of ride type, they may both have solved the same earcon interaction problems (see earlier in the section).

While the modified NASA TLX workload ratings participants provided for each condition show few significant results, participants did rate the staggered onset condition to be significantly less annoying than the original earcon set condition. It is, however, unclear why the final condition failed to produce a significant result for this attribute.

The melody-altered earcon set condition did not show any significant improvement over the original earcon set condition, indeed the original earcon set condition outperformed the melody-altered earcon set condition in the number of earcons, ride types, ride intensities, and ride costs correctly identified. As was discussed in the section describing the melody altered earcon set, the perception of crossing frequency glides is difficult without further differences between the two glides in terms of their timbre. It was considered that in a large number of cases, the earcon sets presented that had different ride intensities would also vary in timbre, thereby assisting the correct interpretation of the melody [Halpern 1977]; or both earcons would be played in different registers, and as such would be inharmonically separated so that the melodies would not actually cross in frequency. If an improvement in the identification of earcons or their attributes had been seen, even if this was not significant, it would have indicated that melody alterations to concurrently presented earcons may have been useful in improving their identification if each earcon also had a unique timbre. Future investigations could have determined if applying the multitimbre earcon set features to guarantee more than just the melody differences between earcons would significantly improve earcon identification. However, given that the results appear to show that the melody-altered earcon set condition reduces correct identification of earcons and their attributes over the original earcon set condition, melody alteration of concurrently presented earcons may not assist in improving earcon identification. Further investigation into melody modifications to concurrent earcons therefore may be required. This result does however validate the work carried out in this paper, as it is not easy to predict exactly what interactions will occur when concurrent audio is presented.

The extended training condition did not show a significant improvement in the identification of concurrently presented earcons or their attributes. However, for the number of earcons, ride types, and ride intensities identified, identification performance is closer to the staggered onset and multitimbre earcon set condition than the original earcon set condition. It is unlikely that the lack of significant

results for the extended training condition can be attributed to participants not spending time using the concurrent earcon trainer. On average participants spent 13 min on the first training session, and 8 min on the second. Because there is no guidance on training participants to listen to concurrently presented earcons, it is possible that a redesign of the training tool may improve concurrent earcon identification. As with the melody-altered earcon set condition, further investigation into this issue is required.

Applying modifications to the design and presentation of the earcons, however, may make it acceptable for designers to increase the number of earcons concurrently presented. Other studies, which have investigated modifications to the design and presentation of concurrent sounds [Brungart et al. 2002], have shown that while increasing the number of concurrently presented sounds still causes a reduction in performance, that reduction is much less severe. Hence, if modifications to the design and presentation of earcons were carried out, the gradient in the graph from Figure 8 would be expected to be much flatter.

Additionally, the degree of modifications that could be applied without destroying the grammar of the earcons was constrained. That is, only small variations in timbre could be applied between concurrent earcons rather than gross differences. The degree of such limitation is partially due to the number of data attributes encoded within the earcon. If a data parameter is encoded by an auditory attribute, there is a limited amount of modification that can be undertaken on that audio attribute to allow concurrently presented earcons to be separately streamed. The earcons used in this work, as already stated, encode three data parameters and as such are a "worst-case scenario." If less data attributes are encoded in each earcon, this would free an audio attribute, such as timbre or register, which may allow greater changes between concurrently presented earcons without destroying the "grammar" of the earcons used, since the audio attribute would not be used as part of the earcon "grammar." This may lead to increased improvements in earcon identification, however, the amount of information encoded in each earcon would be significantly reduced.

The results of this experiment have allowed the guidelines derived from the previous experiment to be extended, and the following guidelines to be added:

Guideline 3. When timbre is used to encode a data parameter, each concurrently presented earcon should have a different timbre. The guidelines of Brewster et al. [1995], should be used to select timbres to encode different values of a data parameter, but if two earcons with the same timbre encoded value are to be concurrently presented, each should use a different musical timbre from the same instrument group. The work of Rigas [1996] can be used to determine distinct instrument groupings.

Guideline 4. Concurrently presenting earcons which start at the same time should be avoided. The introduction of at least a 300 ms onset-to-onset gap between the starts of concurrently presented earcons will make the earcons more identifiable to users.

7.   CONCLUSIONS

This paper has described work investigating the identification of concurrently presented earcons. It has shown that when concurrently presented, earcons interact with each other such that identifying the data encoded in each earcon is difficult. Additionally, the relationship between earcon identification and the proportion of those earcons that can be successfully identified has been identified. It was identified that by reducing the number of earcons concurrently presented, the proportion of those earcons that can be successfully identified is increased (see Figure 8).

These results on concurrent earcon identification lead to the obvious question: How can the design and presentation of earcons be modified to improve identification when they are concurrently presented? In doing this care must be taken. while modifying the earcons so that they stream separately might be

straightforward, doing so without destroying the "grammar" that affords many of the powerful features of earcons is much harder. It was identified through empirical experimental work that by staggering the onset of each earcon and presenting each with a different timbre significantly increased the number of earcons that could be identified. However, these improvements, while statistically significant, were not large. It is clear therefore that more work needs to be devoted to improving the identification of earcons when they are concurrently presented.

Out of this work the following guidelines for the presentation of concurrent earcons in nonspatialized environments have been identified:

Guideline 1. Increasing the number of concurrently presented earcons significantly reduces the proportion of the earcons that can be successfully identified. Increasing the number of earcons concurrently presented can reduce correct identification from 70% to 30%. Great care should be taken when considering the amount of information users will need to extract from earcons when considering the number of earcons which will be concurrently presented.

Guideline 2. If register is used to encode a data attribute, it may be beneficial to ensure that in harmonic intervals are used between earcons concurrently presented in different registers. This is likely to reduce the impact on register identification when the number of concurrently presented earcons is increased.

Guideline 3. When timbre is used to encode a data parameter, each concurrently presented earcon should have a different timbre. The guidelines of Brewster et al. [1995], should be used to select timbres to encode different values of a data parameter, but if two earcons with the same timbre encoded value are to be concurrently presented, each should use a different musical timbre from the same instrument group. The work of Rigas [1996] can be used to determine distinct instrument groupings.

Guideline 4. Concurrently presenting earcons which start at the same time should be avoided. The introduction of at least a 300 ms onset-to-onset gap between the starts of concurrently presented earcons will make the earcons more identifiable to users.

In conclusion, this paper has presented the first empirical study to investigate the identification of concurrently presented earcons, how their identification varies with the number presented and how they can be modified to be better identified. This work has lead to a set of guidelines that can be used by designers of future auditory displays to make more informed decisions when incorporating concurrently presented earcons into their systems.

REFERENCES

BBC NEWS. 2002. Available at http://news.bbc.co.uk/1/hi/sci/tech/1924144.stm.

BLATTNER, M. M., SUMIKAWA, D. A., AND GREENBERG, R. M. 1989. Earcons and icons: Their structure and common design principles. *Hum. Comput. Interact. 4*, 1, 11–44.

BREGMAN, A. S. 1994. *Auditory Scene Analysis*. MIT Press, Cambridge, MA.

BREWSTER, S. A. 1994. *Providing a Structured Method for Integrating Non-Speech Audio into Human–Computer Interfaces*. Ph.D. thesis, Department of Computer Science, University of York.

BREWSTER, S. A. 1998. Using non-speech sounds to provide navigation cues. *ACM Trans. CHI 5*, 2, 224–259.

BREWSTER, S. A. 2002. Overcoming the lack of screen space on mobile computers. *Pers. Ubiquitous Comput. 6*, 3, 188–205.

BREWSTER, S. A. AND MURRAY, R. 2000. Presenting dynamic information on mobile computers. *Pers. Technol. 2*, 4, 209–212.

BREWSTER, S. A., WRIGHT, P. C., AND EDWARDS, A. D. N. 1995. Experimentally derived guidelines for the creation of earcons. In *Proceedings of BCS-HCI '95*, vol. 2. Springer, Huddersfield, UK, 155–159.

BROWN, L., BREWSTER, S., RAMLOLL, R., YU, W., AND RIEDEL, B. 2002. Browsing modes for exploring sonified line graphs. In *Proceedings of BCS-HCI 2002*, vol. 2. BCS, London, UK, 6–9.

BRUNGART, D. AND SIMPSON, B. D. 2002. The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. *J. Acoust. Soc. Am. 112*, 2, 664–676.

BRUNGART, D. S., ERICSON, M. A., AND SIMPSON, B. D. 2002. Design considerations for improving the effectiveness of multitalker speech displays. In *Proceedings of ICAD 2002*, vol. 1. ICAD, Kyoto, Japan, 424–430.

DARWIN, C. J. AND CIOCCA, V. 1992. Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of mistuned component. *J. Acoust. Soc. Am. 91*, 1, 2281–3390.

DEUTSCH, D. 1999. Grouping mechanisms in music. In *The Psychology of Music*, 2nd ed., vol. 1, D. Deutsch, ed. Academic Press, San Diego, 299–348.

FREEDOM SCIENTIFIC. 2003. Available at http://www.freedomscientific.com.

GAVER, W., SMITH, R., AND O'SHEA, T. 1991. Effective sounds in complex systems: The ARKola simulation. In *Proceedings of CHI'91*, vol. 1. ACM Press, New Orleans, 85–90.

GAVER, W. W. 1993. Synthesizing auditory icons. In *Proceedings of INTERCHI'93*, vol. 1. ACM Press, Amsterdam, The Netherlands, 228–235.

GAVER, W. W. 1997. Auditory interfaces. In *Handbook of Human–Computer Interaction*, 2nd ed., vol.1, M. G. Helander, T. K. Landauer, and P. V. Prabhu, eds. Elsevier, Amsterdam, 1003–1041.

HALPERN, L. 1977. The effect of harmonic ratio relationships on auditory stream segregation. Tech. rep., Psychology Department, McGill University.

HART, S. AND STAVELAND, L. 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Human Mental Workload*, vol.1, P. Hancock and N. Meshkati, eds. North Holland, Amsterdam, 139–183.

HEISE, G. A. AND MILLER, G. A. 1951. An experimental study of auditory patterns. *Am. J. Psychol. 64*, 68–77.

HERMAN, T., DREES, J. M., AND RITTER, H. 2003. Broadcasting auditory weather reports—a pilot project. In *Proceedings of ICAD 2003*, vol. 1. ICAD, Boston, MA, 208–211.

ICAD. 2003. Available at http://www.icad.org.

MANSUR, D. L. 1985. *Graphs in Sound: A Numerical Data Analysis Method for the Blind*. M.Sc. thesis, University of California.

MCGOOKIN, D. K. 2002. The presentation of multiple earcons in a spatialised audio space. In *Proceedings of BCS-HCI 2002*, vol. 2. BCS, London, UK, 228–229.

MCGOOKIN, D. K. AND BREWSTER, S. A. 2002. Dolphin: The design and initial evaluation of multimodal focus and context. In *Proceedings of ICAD 2002*, vol. 1. ICAD, Kyoto, Japan, 181–186.

MOORE, B. C. J. 1997. *An Introduction to the Psychology of Hearing*, 4th ed. Academic Press, London.

MYNATT, E. D. 1994. Designing with auditory icons. In *Proceedings of ICAD '94*, vol. 1. Santa Fe, New Mexico, 109–119.

PAPP, A. L. 1997. *Presentation of Dynamically Overlapping Auditory Messages in User Interfaces*. Ph.D. thesis, University of California.

PETRIE, H., JOHNSON, V., FURNER, S., AND STROTHOTTE, T. 1998. Design lifecycles and wearable computers for users with disabilities. In *Proceedings of the First International Workshop of Human Computer Interaction with Mobile Devices*, vol. 1. GIST, Glasgow, UK.

RANDEL, D. M. 1978. *Harvard Concise Dictionary of Music*, 1 ed., vol. 1. Harvard University Press, Cambridge, Massachusetts.

RASCH, R. A. 1978. The perception of simultaneous notes such as in polyphonic music. *Acoustica 40*, 1, 22–33.

RIGAS, D. I. 1996. *Guidelines for Auditory Interface Design: An Empirical Investigation*. Ph.D. thesis, Loughborough University.

SAWHNEY, N. AND SCHMANDT, C. 2000. Nomadic radio: Speech & audio interaction for contextual messaging in nomadic environments. *ACM Trans. CHI 7*, 3, 353–383.

SCALETTI, C. 1994. Sound synthesis algorithms for auditory data representations. In *Auditory Display: Sonification, Audification, and Auditory Interfaces*, vol. 1, G. Kramer, ed. Adison-Wesley, Reading, MA, 223–251.

SHNEIDERMAN, B. 1998. *Designing the User Interface: Strategies for Effective Human–Computer Interaction*, 2nd ed., vol. 1. Addison-Wesley, Reading, MA.

SINGH, P. G. 1987. Perceptual organisation of complex tone sequences: A tradeoff between pitch and timbre. *J. Acoust. Soc. Am. 82*, 3, 886–899.

TOUGAS, Y. AND BREGMAN, A. S. 1985. The crossing of auditory streams. *J. Exp. Psychol. 11*, 788–798.

VAN NOORDEN, L. P. A. S. 1975. *Temporal Coherence in the Perception of Tone Sequences*. Ph.D. thesis, Institute for Perception Research, The Netherlands.

VICKERS, P. AND ALTY, J. L. 2000. Musical program auralisation: Empirical studies. In *Proceedings of ICAD 2000*, vol. 1. ICAD, Atlanta, GA, 157–166.

WILLIAMS, S. M. 1994. Perceptual principles in sound grouping. In *Auditory Display: Sonification, Audification, and Auditory Interfaces*, vol. 1, G. Kramer, ed. Addison-Wesley, Reading, MA, 95–125.

YU, W. AND BREWSTER, S. A. 2003. Evaluation of multimodal graphs for blind people. *Universal Access in the Information Society 2*, 2, 105–124.