ADVANTAGES AND ISSUES WITH CONCURRENT AUDIO PRESENTATION AS PART OF AN AUDITORY DISPLAY

David K. McGookin, Stephen A. Brewster

Glasgow Interactive Systems Group,
Department of Computing Science,
University of Glasgow,
Glasgow, G12 8QQ
{mcgookdk,stephen}@dcs.gla.ac.uk
http://www.dcs.gla.ac.uk/~{mcgookdk, stephen}

ABSTRACT

This paper presents an overview of the concurrent presentation of sound as part of an auditory display. The reasons why designers would wish to use concurrent sound presentation (such as increasing communication bandwidth and allowing comparisons between data to be more easily made) are discussed and examples given. The paper then discusses the problems that can occur when sounds are concurrently presented before summarising and critically evaluating work by the authors on the identification of concurrently presented earcons.

1. INTRODUCTION

The field of auditory display has matured over the last few years to provide clear, evaluated methods, to allow designers to exploit the advantages that sound can bring to a human computer interface. Further advantages have emerged over the last few years such as spatialisation of sound which has lead to a number of researchers working with concurrent audio presentation, where more than one sound is presented to a user at a time, which may be of specific advantage with mobile computing devices such as mobile telephones and personal digital assistants (PDAs). Such devices are becoming increasingly more popular, with greater functionality constantly being added by manufacturers. The usability of these devices is, however, open to debate as mobile computing has several usability issues that need to be addressed. Notably, due to the form factor of mobile devices, the visual display space available is severely limited in comparison to other computing devices, such as the personal computer. For example, the Palm Tungsten PDA has a display of only 6x6 cm. The low resolutions of such displays also contribute to limiting the amount of data that can be useably presented. Also, because mobile computer users are likely to be on the move whilst using their device, they cannot devote their entire attention to the computing task. They must constantly monitor the environment for danger and react accordingly. This places further strain on the user's visual resource. The use of concurrent audio may provide a solution with this problem.

This paper discusses the ways in which concurrent audio has been used by auditory display designers, and the advantages that it provides. The disadvantages of concurrent audio are then discussed followed by a critical summary of work undertaken to identify and understand these problems as regards concurrent earcon identification. Finally conclusions are drawn and possible future work which may improve our understanding of concurrently presented sound is outlined.

2. WHY USE CONCURRENT AUDIO PRESENTATION?

There are several reasons why concurrent audio may be of use as part of an auditory display. In this section several advantages are outlined and work that seeks to exploit these advantages is discussed.

2.1. Increased Bandwidth of Communication

One advantage of concurrent auditory displays is their ability to increase the rate of information which can be presented to a user. This means that information can be pertinently delivered, without having to be delayed until other auditory messages have finished playing. Whilst auditory displays can be designed to associate a priority to each message delivered to a user [1], there may be several instances where messages are of equal importance to the system, and it is up to the user to determine which message is of greatest importance. Several systems such as the ARKola [2] bottling plant simulator, and Nomadic Radio [3], have used this technique to allow the user, rather than the system, to determine what data are and are not important, and therefore what should be attended to. Consider for example, the "diary in the sky" system by Walker *et al.* [4], which used a spatialised audio environment to encode the times of certain diary appointments. Sounds were consecutively presented according to their time of appointment. If a user remembered that they had something important scheduled at 9PM, but could not remember what it was, they would need to consecutively listen to all of the appointments up to 9PM before reaching the information required, which may take a long time (around eight seconds for four appointments). If all of the diary entries were presented in parallel, the user could make use of the cocktail party effect [5], to "tune in" to the 9PM appointment, reducing the time taken to locate the required information.

Such an application of concurrently presented audio may be of specific use with mobile computing interfaces which, as previously discussed, suffer from both small screen displays [6], and users not being able to constantly attend to that visual display in the same way as with a desktop computer [7]. Concurrent audio may allow for more information to be presented through audio and thus relieve some of the demands on a user's visual resource.

2.2. Faster Information Presentation

It may be possible for some types of sound used in auditory displays to be split into chunks, and then for those "chunks" to be presented in random order without having any impact on the meaning of the message to be communicated. If this is possible, chunks of the sound could be presented to the user in parallel, thereby decreasing the presentation time of the information. Such presentations may be advantageous if long sounds are used and their presentation must keep pace with a user's interaction in a computer interface. Whilst in many respects this advantage is similar to that described in Section 2.1 we believe that is may be useful to draw a distinction between presenting more information in less time and more information.

Brewster, Wright and Edwards [8] found that such a technique could be successfully employed to shorten the presentation time of compound earcons. They identified that for compound earcons which contained only two component earcons, playing each component in different ears, as well as introducing an octave difference in the registers (one component's pitch was approximately double that of the other's [9]) each of the component earcons was played in, did not significantly impact participants' ability to identify the earcons. Such presentation did however half the time taken to present the earcons to users. Unfortunately they did not investigate the impact of such presentation on the cognitive demands of users, or expand their work to earcons with more than two parts, which would benefit much more from a shortening of their presentation time.

2.3. Browsing Multiple Data

One of the drawbacks of auditory displays is their temporal nature [10]. This can make it difficult to make comparisons between multiple data in an auditory display, such as determining relationships between two sonified graphs [11]. Concurrently presenting data through audio however, has been shown to be an effective way to overcome the temporal issues of sound, and thus allows comparisons between data to be made more easily. Comparing data through sound in such a way is akin to the way in which we interpret the real world, and as such is a natural task to perform in an auditory display. As Blattner, Papp and Gilnert [12] note *"Our awareness and comprehension of the auditory world around us for the most part is done in parallel"*. In addition, the cocktail party effect [5] has long been an interesting problem for psychologists seeking to explain how the human auditory system works.

Brown *et al.* [11] investigated if users could identify key features of two concurrently presented sound graphs. Graphs were constructed by mapping the y-axis to the musical pitch of a MIDI acoustic grand piano timbre, and using cursor keys to navigate along the x-axis. As a user moved along the x-axis, the pitch representing the appropriate y-axis value was played. Each graph was presented to different ears, to avoid the sounds perceptually fusing together. Brown *et al.* found in comparison to serial presentation, where each graph was presented individually, that whilst concurrently presenting sound graphs did not have a significant effect on the accuracy of responses, it did significantly reduce the time taken to find intersection points between the graphs.

Fernström and Bannon [13] created a system to allow browsing in the Fleischmann Collection of traditional Irish music. Their Sonic Browser had a visual interface which allowed each musical composition to be graphically laid out in a starfield like display [14]. This allowed users to map the x and y axes of the visualisation to different data parameters. As a cursor was moved across the visualisation, the eight nearest musical compositions would be concurrently played in representative spatial locations around the user's head. The use of spatialisation, and the variations between the musical compositions, helping to avoid them being placed in the same auditory stream [15]. Sonic Browser does indicate that concurrently presenting multiple audio sources can have benefits over visual presentation since it is difficult to present music visually in a meaningful way to non-musicians [13]. Similar advantages are apparent in the concurrent presentation of speech based audio, such as Kobayashi and Schmandt's Dynamic Soundscape [16] which used 3D sound presentation to allow users to simultaneously browse and monitor multiple parts of the same audio recording.

Hankinson and Edwards [17] have used comparisons between concurrently presented compound earcons to provide information to users about the validity of computer interface operations. For example if a user tried to copy a printer, the copy and printer earcons would be concurrently presented but would be designed in such a way that they would sound inharmonic when presented, whereas the copy and file earcons would be harmonious when concurrently presented. This provided the user with information due to the comparison of the earcons, that would otherwise not be available.

The examples of this section have shown the ways in which concurrent auditory presentation has been used as part of an auditory display. This work points to key advantages that concurrent auditory presentation brings, and shows that new possibilites for auditory display exist when sounds are concurrently presented. Unfortunately, many of the systems discussed above have not been fully evaluated and therefore guidance is not available for designers as to what types of sound are the most effective and the problems and solutions of presenting sounds concurrently as part of an auditory display.

3. ISSUES WITH THE CONCURRENT PRESENTATION OF AUDIO

The previous section has shown some of the advantages that can be gained by the incorporation of concurrently presented sound as part of an auditory display, there are still practical issues and limitations that must be considered and require further research to allow the wide scale adoption of concurrent auditory displays. The major issue is that sounds which are concurrently presented may interfere with each other in undesirable ways. As Norman [18] notes on the interference of multiple warning alarms, *"they often conflict, and the resulting cacophony is distracting enough to hamper performance"*. Low level psychoacoustical interference such as Masking [19] or phenomenon such as the pitch of the missing fundamental [20] are relatively easy to avoid in concurrent auditory environments. However, there are other interactions that can occur between concurrently presented sound which can be harder and more complex to predict and control.

Auditory Scene Analysis [15] is the study of why we hear the world the way we do, in other words, how do we make sense of all of the distinctive sounds which reach our ears at any point in time. For example, if you are listening to a performance by a concert orchestra, and one player decided to perform a different composition, it will be quickly obvious that they are not playing along with the rest of the orchestra. Similarly if at a performance, someone's mobile telephone starts to ring, it is likely that you would not consider it as part of the performance. Conversely, the sound of a car driving along a road is composed of distinct sounds, such as the engine noise, the noise caused by the tyres on the road etc. Whilst it is possible to detect the surface on which the car is driving, whether the engine is petrol or diesel etc, you can still amalgamate the sounds and consider them as coming from a single car.

Auditory Scene Analysis therefore tries to explain how the multiple complicated sounds which reach our ears (the auditory scene) are categorised by the human auditory system into separate streams. Streams as defined by Bregman [15] (pp.10), are *"a perceptual unit that represents a physical happening"*, for example in the concert orchestra telephone example, there would be two streams, one for the concert orchestra and another for the mobile telephone ring. Note that the word stream is not interchangeable with the word sound, as a stream can be made up of more than one sound, such as the concert orchestra. Whilst some (experienced) listeners will be able to pick out individual instruments, most will only hear the overall composite sound formed from all of the individual instruments playing at the same time.

The reasons why the concert orchestra and mobile telephone are placed in separate streams are obvious; they sound very different from each other. The mobile telephone is likely to have a timbre formed from a sine or square wave generator, whereas the instruments that form the orchestra will have much richer timbres formed of many harmonics. The melodic components of both sounds are also likely to be quite different. In other words, it is the differences and similarities between sounds which determine whether they will be placed in the same or different streams. The greater the differences between the sounds, the more likely it is they will be placed in different streams. The greater the similarities between sounds the more likely they will be placed in the same stream.

It is out with the scope of this paper to provide a detailed overview of auditory scene analysis, however the interested reader is directed towards Bregman [15] and Deutsch [21]. As regards concurrent auditory displays, it is generally desirable to ensure that sounds which are concurrently presented are placed into different streams, to ensure that the information encoded within them can be determined. If sounds are placed in the same stream, or worse split into multiple separate streams it will become difficult, if not impossible, to hear the sounds as intended by the designer [22]. This is because it is both difficult to listen in detail to multiple stream simultaneously as well as make temporal comparisons between streams [15]. Unfortunately, although there is a large body of work which investigates which auditory properties influence the streaming process [15, 21], there is no over arching theory which will predict how a given arbitrary sound will be streamed by the auditory system. Whilst it can be considered that incorporating maximum differences between concurrently presented sounds along the dimensions shown to influence streaming will solve the problems of sounds interfering, this may not be a practical solution since all sounds which are used as part of an auditory display are designed to communicate information and arbitrary modification may destroy the mapping between sound and data. In the following section work which has been undertaken to investigate the specific problems of concurrent earcon identification and what can be undertaken to overcome identification problems is discussed.

4. CONCURRENTLY PRESENTING EARCONS

The previous section has indicated in general terms the problems that can occur when sounds are concurrently presented as part of an auditory display. However, how likely are these problems to occur practically and what can be done about them? In this section work which has been undertaken concerning the identification of concurrently presented earcons will be discussed and its implications outlined.

Whilst there exists a body of work which investigates the concurrent presentation of sounds as part of an auditory display (much of which was discussed in Section 2), the majority of this work remains unevaluated. Whilst some informal evaluations on such systems have shown a positive user reaction, such studies do not allow for the extraction of design guidelines for the sounds used. It is unclear therefore how designers can exploit the advantages of concurrent sound presentation whilst avoiding the problems discussed in Section 3. The work described in this section sought to address these issues as regards the concurrent identification of earcons. Earcons were chosen for these evaluations since they are largely underrepresented in the work described in Section 2. This makes it unclear how they would perform in concurrent presentation systems. Additionally due to the nature of some types of earcon, namely the transformational and hierarchical types [12], earcons from the same set are likely to be similar according to the discussion of auditory scene analysis from Section 3 making it likely that they will interfere with each other and be undesirably streamed by the auditory system which will complicate retrieving the information encoded by them. The investigation attempted to answer three distinct questions.

- How is earcon identification affected by the number concurrently presented?
- How can earcon identification be improved when earcons are non-spatially presented?
- How does spatial presentation affect concurrent earcon identification?

4.1. Earcon Identification versus Number Presented

In order to investigate how the identification of earcons is affected by the number of earcons concurrently presented, an experiment was performed that concurrently presented between one and four earcons taken from the same set, and asked participants to identify information encoded in each earcon (participants had earlier learned the mapping from a training session). Each of the earcons encoded three trinary parameters (each of the three parameters had three possible values), one parameter each in the timbre, melody and presentation register of the earcon. Participants' individual identification for each of these scores was also recorded. A graph showing how participants' identification of the earcons and their attributes was affected as the number of earcons concurrently presented was increased is shown in Figure 1.

As can be seen in Figure 1, as the number of earcons concurrently presented is increased, the proportion of those earcons and their attributes that are successfully identified by participants decreases. Identification performance for the number of earcons identified is similar to work by Brungart, Ericson and Simpson [23] who found a similar proportional decrease in identification when presenting concurrent similar CRM (Coordinate Response Measure) speech phrases. Whilst identification of individual earcon

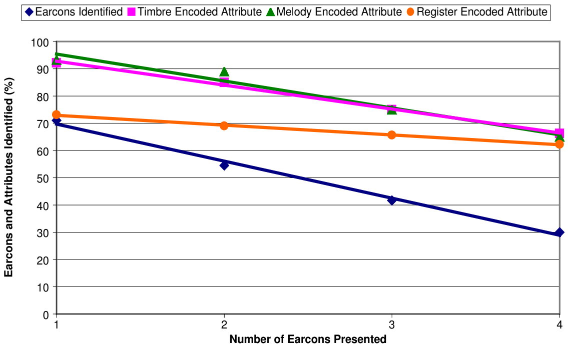

Figure 1: Graph showing best fit trend lines showing how the proportion of earcons and earcon attributes correctly identified when one, two, three and four earcons were concurrently presented.

attributes was greater than whole earcons, the same relatively steep trends were observed. The register encoded attribute is believed to have a flatter gradient due to modifications that were incorporated to avoid harmonic intervals between concurrently presented earcons and the general lower overall identification of register. Overall participant workload, measured via a set of NASA TLX [24] subjective workload scales showed a similar trend with an increased number of earcons requiring greater subjective work to identify them. Although there is an obviously greater amount of work required when identifying four earcons rather than one which may explain this result.

The outcome of this study must be that, at least when concurrently presenting earcons from the same set, that if even a small number are concurrently presented users may be unable to identify the earcons or the information encoded within them, and that ways should be sought to improve concurrent earcon identification.

4.2. Improving Non-Spatial Earcon Identification

To overcome the problems as discussed in the previous section a set of consistent modifications based on auditory scene analysis principles were applied to a set of earcons which were based on the guidelines for earcon construction by Brewster, Wright and Edwards [25]. As with the previous experiment, participants were presented with sets of earcons which encoded three data attributes. Unlike the previous experiment, in all cases four earcons were concurrently presented. Sets of earcons which contained auditory scene analysis modifications were compared to a set which did not. Modifications to the melody, timbre, synchronicity of onset and increasing participant familiarity were tried. However, only staggering the onsets of the earcons by 300ms and having a distinct timbre for each concurrently presented earcon improved the identification of either the number of identified earcons or the number of earcon attributes identified. Additionally, the magnitude of those improvement was small with an average of less that half an earcon improvement between the standard earcons and the best performing modified earcons. It may seem that these results are at odds with the discussion of auditory scene analysis from Section 3 however it is important to remember that when modifying sounds as part of an auditory display to make them more likely to be separately streamed, that there is some relationship between the sound and data to be communicated. This means that if sounds are changed too much, the relationship between the sound and the information encoded within it may be lost. This is a particular issue with earcons, where multiple bits of information may be encoded in each earcon, changing the timbre or pitch of the sound may cause this information to be lost. Therefore in modifying the timbre only small changes can be accommodated before the mapping will become lost. Due to the multidimensional nature of timbre [20] the point at which this becomes an issue is not clear, however in the work discussed here [26], concurrently presented timbres representing the same encoded attribute where presented with different musical timbres chosen from the same instrument group (according to the work of Rigas [27]). It may be however, that larger separations in timbre could be accommodated without loss of identification, this remains a possible future investigation.

Whilst the similarity of concurrently presented sounds is a problem for earcons due to those which are from the same set being similar to each other, other auditory display mapping techniques may be less likely to be similar to each other. For example, on the choosing of auditory icons for an auditory display, Mynatt [28] recommends sounds which are subjectively dissimilar to each other should be used since *"although the sounds of a copier and printer may be quite distinct, it may be difficult to correctly identify them when they are both used in the same interface."*. Further evidence pointing to the robustness of auditory icons is concurrent presentation situations is the ARKola bottling plant simulator system by Gaver [2]. Here two user had to collaborate to operate a simulated bottling plant and make as much profit as possible. Different processes were presented visually as well as by rhythmically repeating auditory icons. A small evaluation by Gaver found that users preferred the sounds over a purely visual display and were more likely to collaborate to solve problems in the environment. As with the work described here, Gaver did not spatialise the sounds and sometimes up to fourteen could be played together.

4.3. The Spatial Presentation of Concurrent Earcons

In many of the systems which have used concurrent auditory presentation, sound source spatialisation has been prominent. In many of these systems spatialisation is not only used to separate concurrently presented sounds and again contribute towards their separate streaming, but is also used in a mapping to some other information, usually time. Nomadic Radio [3] used the azimuthal angle that a sound was presented around the user's head to represent the time of arrival of messages and news. Kobayashi and Schmandt [16] used the same principle, mapping the timecode of an auditory recording to the azimuthal location of that recording around the user's head such that over time the recording moved around the head. Whilst spatialisation is an element of auditory scene analysis and can contribute to the separation of sources when concurrently presented, the effectiveness of this cue is dependant on the degree of similarity and the location of the sources which need to be separated. Best, van Schaik and Carlile [29] have identified that in order for two sounds to be identified in distinct locations there may need to be a 60^o azimuthal difference between them. Whilst this may mean that in some cases concurrent audio can be easily employed when used with spatialisation, when spatial location is mapped to some data parameter it will not be possible to arbitrary adjust the position of sounds to make them further apart as this would destroy the mapping between the location of the sound and the information encoded by it. In such situations, at least as regards earcons, identification may be closer to non-spatial presentation than with spatial presentation. Spatialisation has also been looked at by Brewster, Wright and Edwards [25]. As discussed in Section 2.2 they split compound earcons such that each earcon part was presented at the extremities of a stereo field. Unfortunately it is difficult to determine the specific impact of the separation as they also presented each part of the compound earcon in a different register as well as having each earcon part derived from a different set of earcons. Whilst they found that concurrent presentation did not lead to significantly lower identification of the earcons, it is difficult to identify the true impact of spatialisation on concurrent earcon presentation for the reasons discussed above.

McGookin and Brewster [26] identified that spatially presenting (using a generic head related transfer function (GHRTF) with active head tracking) earcons significantly improved identification over non-spatial presentation and incorporation of the finding of the previous section on using multiple timbres and staggering onsets provided further improvements in identification. However, as with the work discussed in Section 4.2 overall improvements in identification for concurrently presented earcons were not large. With both modifications of the previous section and maximum azimuthal spatialisation (90^o for four concurrently presented earcons) participants identified on average less than two of the four concurrently presented earcons. This again raises questions over the practicality of concurrently presented earcons.

4.4. Ecological Impact of Concurrently Presented Earcons

The work outlined in Sections 4.1 to 4.3 and further detailed in McGookin and Brewster [30, 31] has identified that concurrently presenting earcons from the same set made it difficult identify the information contained in those earcons. Whilst modifications including both spatialisation and modifications based on auditory scene analysis research [15] could improve identification, these improvements were not large and were of questionable real world usefulness. In order to consider these issues a more ecological experiment was carried out which involved participants answering questions about simulated diaries on a mobile device whilst walking a pre determined route around a set of traffic cones. Each diary consisted of five days, with each day containing four appointments, each appointment being represented with an earcon. Each day's appointments were concurrently presented and as with McGookin and Brewster's work [30, 26] four earcons were presented at a time. Two sets of earcons were compared, one set which was solely based on the guidelines of Brewster, Wright and Edwards [25] and another which was modified to incorporate the staggered onset and different timbres as discussed in Section 4.2 and were spatially presented as described in Section 4.3. Since the experiment was performed on a real mobile device a lower quality spatialisation system a lower quality spatialisation system had to be incorporated than that used in the work from Section 4.3. The standard spatialisation system for Microsoft DirectX [32], which is to use stereo panning with attenuation of higher frequencies for sounds at the back of the head was used which. Additionally a suitable active head tracking device was not available for use in the experiment. This may have affected the results. The results showed that participants using the modified earcons did not answer questions more accurately or walk faster than when using the unmodified earcons. Though this may be attributable to the poorer quality spatialisation system used in this work in comparison to that used in Section sec:thespatialpresentationofconcurrnetearcons. Participants did however rate the subjective workload required (measured via modified NASA TLX workload ratings) when using the modified earcons to be significantly lower than with non-modified earcons. This shows that modifications to earcons to improve their identification in concurrent presentation environments can provide some advantages in real work interfaces however, this advantage may not be practical in true real world scenarios.

5. DISCUSSION AND FUTURE WORK

This paper has discussed the advantages of the use of concurrent audio presentation as part of an auditory display. The ability to both increase the bandwidth of presentation and compare multiple data has been used by many designers as part of an auditory display. However few of these displays have been empirically evaluated and there are few guidelines which can be used by designers to design sounds which can be used in such specific situations. Such guidelines are important due to the discussion of Section 4 where sounds which are concurrently presented may interfere with each other in unpredictable ways. Whilst such problems may not, as exhibited in ARKola by Gaver, Smith and O'Shea [2], be prevalent for all types of sounds used in auditory displays, in cases where there are problems, these problems may easily outweigh the advantages of such displays. From the work of McGookin and Brewster [26, 31] as discussed in Section 4, earcons are one type of sound which is problematic to identify when concurrently presented. Additionally, modifications to the design and presentation of the earcons (designed to both reduce the interference between them without weakening the mappings between the earcons and the information encoded by them), failed to increase the identification of concurrently presented earcons to a level which could be regarded as being of practical usefulness. The work of McGookin and Brewster does however note that conventional earcons (based on the guidelines of Brewster, Wright and Edwards [25]) have problems when they are presented concurrently, and that more research should be undertaken to try to overcome these problems.

5.1. Limitations and Future Directions

There are several features which should be noted about the work of McGookin and Brewster that may impact the results that were obtained and would assist in understanding the identification of concurrent earcons.

All of the experiments undertaken by McGookin and Brewster encoded three data parameters in each earcon. As discussed in Section 4.2, the degree of modifications that can be applied to the earcons is constrained due to the need to preserve the mappings between data and sound. If fewer attributes were encoded in the earcons this would "free up" an auditory parameter which could incorporate greater changes between the earcons and make them less prone to interfering with each other. Additionally, how the number of earcons identified is affected by the number presented when those earcons incorporate modifications to make them stream more desirably has not been investigated. Other research [23] which has considered the relationship between the number of concurrently presented sound sources (in this case speech) found that in such situations although the trend shown in Figure 1 is still likely to occur, its gradient may be significantly flattened. It remains a feature of future work to identify the impact on earcon identification of these features.

Whilst these further proposed investigations may significantly improve the identification of earcons when concurrently presented, they may as found by the previously discussed work lead to significant, but not practically useful, increases in identification. In such an instance it may be that more radical approaches to improve the identification of concurrently presented earcons may be necessary, perhaps with a fundamental redesign of the earcons themselves. In any instance further research into the presentation of concurrent sounds as part of an auditory display would be of use and allow designers to truly exploit the advantages of concurrent audio presentation as discussed in Section 2.

6. CONCLUSIONS

This paper has sought to make explicit the reasons why auditory display designers may wish to consider the use of concurrent sound as part of an auditory display. These advantages have been demonstrated through examples and counter examples. Additionally it has been noted that many of these systems have not been fully evaluated making it difficult to extract design guidelines for future designers to use to exploit the advantages concurrent audio affords. This is particularly problematic due to the problems of concurrent sounds interfering with each other if they are sufficiently similar or spatially proximal to each other causing the information encoded by the sound to be lost. We then critically evaluated previous work by the authors to identify the problems of concurrently presented earcons. This work found that although the identification of concurrent earcons could be improved by modifications to their design and presentation, in order to preserve the mappings between sounds and data, these modifications were limited, leading to only small improvements in identification and being of questionable real world usefulness. Whilst further modifications can be undertaken to improve the identification of concurrently presented earcons further research is required if the true advantages of concurrent sound presentation are to be exploited.

7. REFERENCES

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

[2] W. W. Gaver, R. B. Smith, and T. O'Shea, "Effective sounds in complex systems: The arkola simulation," in *Proceedings of CHI 91*, New Orleans, 1991, vol. 1, pp. 85–90, ACM.

[3] N. Sawhney and C. Schmandt, "Nomadic radio: Speech audio interaction for contextual messaging in nomadic environments," *ACM Transactions on CHI*, vol. 7, no. 3, pp. 353–383, 2000.

[4] A. Walker, S. A. Brewster, D. McGookin, and A. Ng, "Diary in the sky: A spatial audio display for a mobile calendar.," in *proceedings of IHM-HCI 2001*, Lille, France, 2001, vol. 1, pp. 531–539, Springer.

[5] B. Arons, "A review of the cocktail party effect," *Journal of the American Voice I/O Society*, vol. 12, no. 7, pp. 35–50, 1992.

[6] A. Walker and S. A. Brewster, "Spatial audio in small screen device displays," *Personal Technologies*, vol. 4, no. 2, pp. 144–154, 2000.

[7] S. Holland, D. R. Morse, and H. Gedenryd, "Audiogps: Spatial audio navigation with a minimal attention interface," *Personal and Ubiquitous Computing*, vol. 6, no. 4, pp. 253–259, 2002.

[8] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "Parallel earcons: Reducing the length of audio messages," *International Journal of Human-Computer Studies*, vol. 43, no. 2, pp. 153–175, 1995.

[9] D. M. Randel, *Harvard Concise Dictionary of Music*, vol. 1, Harvard University Press, Cambridge, Massachusetts, 1 edition, 1978.

[10] G. Kramer, "An introduction to auditory display," in *Auditory Display: Sonification, Audification and Auditory Interfaces*, Gregory Kramer, Ed., pp. 1–77. Addison-Wesley, Reading, Massachusetts, 1994.

[11] L. Brown, S. Brewster, R. Ramloll, W. Yu, and B. Riedel, "Browsing modes for exploring sonified line graphs," in *Proceedings of BCS-HCI 2002*, London, UK, 2002, vol. 2, pp. 6–9, BCS.

[12] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and icons: Their structure and common design principles," *Human Computer Interaction*, vol. 4, no. 1, pp. 11–44, 1989.

[13] M. Ferstrom and L. Bannon, "Explorations in sonic browsing," in *Procedings of HCI 97*, Bristol, UK, 1997, vol. 1, pp. 117–131, Springer Verlag.

[14] C. Ahlberg and B. Shneiderman, "Visual information seeking using the filmfinder," in *Proceedings of CHI 94*, Boston, Massachusetts, 1994, vol. 2, pp. 433–434, ACM Press.

[15] A. S. Bregman, *Auditory Scene Analysis*, MIT Press, Cambridge, Massachusetts, 1994.

[16] M. Kobayashi and C. Schmandt, "Dynamic soundscape: mapping time to space for audio browsing," in *Proceedings of CHI 97*, Atlanta, Georgia, 1997, vol. 1, pp. 194–201, ACM Press.

[17] J. C. K. Hankinson and A. D. N. Edwards, "Designing earcons with musical grammars," *ACM SIGCAPH*, vol. 65, pp. 16–20, September 1999.

[18] D. A. Norman, *The Design of Everyday Things*, vol. 1, MIT Press, Boston, Massachusets, 1st edition, 1998.

[19] W. A. Yost, *Fundamentals of Hearing: An Introduction*, vol. 1, Academic press, Boston, 4th edition, 2000.

[20] B. C. J. Moore, *An Introduction to the psychology of hearing*, Academic Press, London, 4th edition, 1997.

[21] D. Deutsch, "Grouping mechanisms in music," in *The Psychology of Music*, Diana Deutsch, Ed., pp. 299–348. Academic Press, San Diego, 2nd edition, 1999.

[22] S. M. Williams, "Perceptual principles in sound grouping," in *Auditory Display: Sonification, Audification, and Auditory Interfaces*, Gregory Kramer, Ed., vol. 1, pp. 95–125. Addison-Wesley, Reading, Massachusetts, 1994.

[23] D. S. Brungart, M. A. Ericson, and B. D. Simpson, "Design considerations for improving the effectiveness of multitalker speech displays," in *Proceedings of ICAD 2002*, Kyoto, Japan, 2002, vol. 1, pp. 424–430, ICAD.

[24] S. Hart and L. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Human Mental Workload*, P Hancock and N Meshkati, Eds., vol. 1, pp. 139–183. North Holland, Amsterdam, 1988.

[25] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "Experimentally derived guidelines for the creation of earcons," in *Proccedings of BCS-HCI 95*, Huddersfield, UK, 1995, vol. 2, pp. 155–159, Springer.

[26] D. K. McGookin and S. A. Brewster, "Space the final frontearcon: The identification of concurrently presented earcons in a synthetic spatialised auditory environment," in *Proceedings of ICAD 2004*, Sydney, Australia, 2004, ICAD.

[27] D. I. Rigas, *Guidelines for Auditory Interface Design: An Empirical Investigation*, Ph.D. thesis, Loughborough University, 1996.

[28] E. D. Mynatt, "Designing with auditory icons," in *Proceedings of ICAD 94*, Santa Fe, New Mexico, 1994, pp. 109–119.

[29] V. Best, A. van Schaik, and S. Carlile, "Two-point discrimination in auditory displays," in *Proceedings of ICAD 2003*, Boston, Massachusetts, 2003, pp. 17–20, ICAD.

[30] D. K. McGookin and S. A. Brewster, "An investigation into the identification of concurrently presented earcons," in *Proceedings of ICAD 2003*, Boston, MA, 2003, pp. 42–46, ICAD.

[31] D. K. McGookin and S. A. Brewster, "Understanding concurrent earcons: Applying auditory scene analysis principles to concurrent earcon recognition," *ACM Transactions on Applied Perception*, vol. 1, no. 2, pp. 130–155, 2004.

[32] B. Bargen and P. Donnelly, *Inside DirectX*, vol. 1 of *Microsoft Programming Series*, Microsoft Press, Redmond, Washington, 1998.