UNIVERSITY
of
GLASGOW

Understanding and Improving the Identification of Concurrently Presented Earcons

David Kerr McGookin

Submitted for the degree of Doctor of Philosophy
July, 2004

Abstract

The use of sound to communicate information as part of a user interface has been an active research area for several years. Research has shown that sound can be concurrently presented to users to increase the bandwidth and rate of data presentation. However, when sounds are concurrently presented, they may interfere with each other, such that determining the data encoded in the sound becomes difficult. Modifications to the sounds can help to avoid such interference, but due to the nature of the sounds the impact of the modifications may be constrained.

This thesis investigates such interaction with concurrently presented earcons. One experiment investigates how the identification of earcons is affected by the number concurrently presented. It was found that increasing the number of earcons concurrently presented lead to a significant decrease in the proportion of earcons and their attributes successfully identified by participants. With identification falling from 70% correct for one presented earcon to 30% for four concurrently presented earcons.

A second experiment identified how modifications to the design and presentation of concurrently presented earcons affected their identification. It was found that presenting each earcon with a unique timbre as well as introducing an onset-to-onset delay of at least 300ms caused a significant improvement in earcon identification, and the timbre encoded attribute of earcons. However overall identification levels remained low at around 30%.

Two further experiments investigated the impact of spatialisation on concurrent earcon identification. They showed that spatial presentation of earcons which did not incorporate the findings of the previous experiment significantly improved identification of earcons and the register encoded earcon attribute, over earcons that were not spatially presented but did incorporate the findings of the previous experiment. Another experiment showed that spatial presentation of earcons which incorporated the unique timbre and 300ms onset-to-onset modifications significantly improved the identification of the timbre encoded earcon attribute, although overall identification remained low.

These four experiments yielded a set of guidelines for concurrent earcon presentation. Due to the nature of those experiments however, a further experiment was conducted to determine the impact of the guidelines on more ecologically valid tasks. A set of modified and unmodified earcons which represented entries in a mobile diary system were compared. Overall task accuracy remained low, although participants rated the modified earcons to require significantly less subjective workload.

Contents

1 Introduction **17**

1.1 Motivation of the Thesis . . . 17

1.2 Aims of the Thesis . . . 19

1.2.1 Thesis Statement . . . 19

1.3 Contents of the Thesis . . . 20

2 Basics of Sound **22**

2.1 Introduction . . . 22

2.2 Introduction to Sound . . . 22

2.3 The Perception of Sound . . . 23

2.3.1 The Limits of Human Hearing . . . 23

2.3.2 Auditory Attribute Perception . . . 24

2.4 Spatial Audio . . . 28

2.4.1 Binaural Cues . . . 28

2.4.2 Localisation Performance . . . 30

2.4.3 Elevation . . . 31

2.4.4 Distance Perception . . . 31

2.5 Spatial Sound Synthesis . . . 32

2.5.1 Head Related Transfer Functions . . . 32

2.5.2 Generalised Head Related Transfer Functions . . . 32

2.6 Auditory Scene Analysis . . . 33

2.6.1 Introduction to Auditory Scene Analysis . . . 33

2.6.2 Determining Differences in Auditory Scene Analysis . . . 34

2.6.3 Other Issues with Auditory Scene Analysis . . . 38

2.7 Conclusions . . . 39

3 Auditory Display **40**

3.1 What is an Auditory Display? . . . 40

3.2 Advantages and Disadvantages of Auditory Display 40
3.2.1 Advantages of Auditory Display 40
3.2.2 Disadvantages with Auditory Displays 43
3.3 Mapping Data to Sound 45
3.3.1 Sonification 46
3.3.2 Speech 47
3.3.3 Auditory Icons 48
3.3.4 Earcons 50
3.4 Use of 3D Sound in Auditory Display 54
3.5 Conclusions 59

4 Issues Involving Concurrent Audio Presentation 60
4.1 Introduction 60
4.2 Why use Concurrent Audio Presentation? 60
4.2.1 Increased Bandwidth of Communication 60
4.2.2 Faster Information Presentation 61
4.2.3 Browsing Multiple Data 61
4.3 Problems with Concurrent Audio Presentation 63
4.4 Dolphin 64
4.4.1 Overview 64
4.4.2 Cluttered Audio Spaces 65
4.5 Evaluation of Dolphin 67
4.5.1 Procedure 67
4.5.2 Results 74
4.5.3 Discussion 75
4.5.4 Dolphin Conclusions 76
4.6 Related Work 77
4.6.1 Gerth 78
4.6.2 Multi-Talker Speech 79
4.6.3 The Computational Auditory Scene Synthesizer 81
4.7 Research Questions 82
4.8 Conclusions 84

5 Number of Earcons versus Identification 86
5.1 Introduction 86
5.2 Methodology 87
5.2.1 Issues With The Experimental Design 87

5.2.2 Procedure . . . 88
5.3 Results . . . 92
5.3.1 Identified Earcons and Attributes . . . 92
5.3.2 Workload Data . . . 94
5.4 Discussion . . . 95
5.4.1 Guidelines for Concurrent Earcon Presentation . . . 97
5.5 Conclusions . . . 97

6 Engineering More Robust Earcons **99**
6.1 Introduction . . . 99
6.1.1 Earcons and Auditory Scene Analysis . . . 100
6.1.2 Applying Auditory Scene Analysis to Earcons . . . 101
6.1.3 Why Not Spatial Audio? . . . 105
6.2 Methodology . . . 106
6.2.1 Training Phase . . . 106
6.2.2 Testing Phase . . . 107
6.2.3 Conditions . . . 107
6.3 Results . . . 110
6.3.1 Identified Earcons and Attributes . . . 110
6.3.2 Workload Data . . . 112
6.4 Discussion . . . 112
6.5 Conclusions . . . 116

7 The Impact of Spatialisation on Concurrent Earcon Identification **117**
7.1 Introduction . . . 117
7.1.1 Previous Spatial Research . . . 117
7.2 Spatial versus Non-Spatial Earcons . . . 118
7.2.1 Motivation . . . 118
7.2.2 Methodology . . . 119
7.2.3 Results . . . 121
7.2.4 Discussion . . . 123
7.3 Spatial versus Revised Spatial Earcons . . . 124
7.3.1 Motivation . . . 124
7.3.2 Methodology . . . 125
7.3.3 Results . . . 126
7.3.4 Discussion . . . 128
7.4 Conclusions . . . 129

8 An Ecological Investigation into Concurrent Earcons 131
8.1 Introduction . . . 131
8.2 Methodology . . . 132
8.2.1 Outline of Evaluation Domain . . . 132
8.2.2 Outline of Experimental Task . . . 134
8.3 Procedure . . . 134
8.3.1 Introduction . . . 134
8.3.2 Baseline Measuring Phase . . . 135
8.3.3 Training Phase . . . 136
8.3.4 Testing Phase . . . 137
8.3.5 Earcons Used . . . 139
8.3.6 Non-Spatial Condition . . . 140
8.3.7 Spatial Condition . . . 140
8.4 Results . . . 141
8.4.1 Correctly Answered Questions . . . 141
8.4.2 PPWS . . . 141
8.4.3 Modified NASA TLX Workload Ratings . . . 142
8.5 Discussion . . . 143
8.6 Conclusions . . . 145

9 Conclusions 147
9.1 Summary of the Thesis . . . 147
9.1.1 Literature Review and Motivation . . . 147
9.1.2 Original Thesis Work . . . 148
9.2 Contributions of the Thesis . . . 150
9.3 Limitations of the Thesis . . . 155
9.4 Future Work . . . 156
9.4.1 Spatial Separation versus Identification . . . 157
9.4.2 Reducing Earcon Parameters . . . 157
9.4.3 Identification in Mixed Concurrent Auditory Environments . . . 158
9.5 Final Remarks . . . 158

A Glossary 159

B Dolphin Evaluation - Training Sheets 160

C Dolphin Evaluation - Raw Data **167**
C.1 Average Time to Create a Tour . . . 167
C.2 Percentage Optimal Path Length . . . 167
C.3 Tours with Missing or Added Rides . . . 168
C.4 Modified NASA TLX Workload Ratings . . . 168

D Condition Grouping For Non-Spatial Experiments **171**

E Number of Earcons versus Identification - Training Sheets **172**

F Number of Earcons versus Identification - Raw Data **178**
F.1 One Earcon Condition . . . 178
F.1.1 Earcons and Earcon Attributes Identified . . . 178
F.1.2 Modified NASA TLX Workload Ratings . . . 179
F.2 Two Earcon Condition . . . 181
F.2.1 Earcons and Earcon Attributes Identified . . . 181
F.2.2 Modified NASA TLX Workload Ratings . . . 183
F.3 Three Earcon Condition . . . 185
F.3.1 Earcons and Earcon Attributes Identified . . . 185
F.3.2 Modified NASA TLX Workload Ratings . . . 186
F.4 Four Earcon Condition . . . 188
F.4.1 Earcons and Earcon Attributes Identified . . . 188
F.4.2 Modified NASA TLX Workload Ratings . . . 190

G Engineering More Robust Earcons - Training Sheets **193**

H Engineering More Robust Earcons - Raw Data **199**
H.1 Original Earcon Set Condition . . . 199
H.2 Melody Altered Earcon Set Condition . . . 199
H.2.1 Earcons and Earcon Attributes Identified . . . 199
H.2.2 Modified NASA TLX Workload Ratings . . . 200
H.3 Multi-Timbre Earcon Set Condition . . . 202
H.3.1 Earcons and Earcon Attributes Identified . . . 202
H.3.2 Modified NASA TLX Workload Ratings . . . 204
H.4 Extended Training Condition . . . 206
H.4.1 Earcons and Earcon Attributes Identified . . . 206
H.4.2 Modified NASA TLX Workload Ratings . . . 207
H.5 Staggered Onset Condition . . . 209

H.5.1 Earcons and Earcon Attributes Identified 209
H.5.2 Modified NASA TLX Workload Ratings 211
H.6 Final Condition 213
H.6.1 Earcons and Earcon Attributes Identified 213
H.6.2 Modified NASA TLX Workload Ratings 214

I Spatial Experiments - Training Sheets **217**

J Spatial Experiments - Raw Data **229**
J.1 Spatial versus Non-Spatial Conditions 229
J.1.1 Earcons and Earcon Attributes Identified 229
J.1.2 Modified NASA TLX Workload Ratings 230
J.2 Spatial(2) versus Revised Spatial Conditions 233
J.2.1 Earcons and Earcon Attributes Identified 233
J.2.2 Modified NASA TLX Workload Ratings 234

K Diary Experiment - Training Sheets **237**

L Diary Experiment - Raw Data **244**
L.1 Type S & M Question Scores 244
L.2 PPWS Scores 245
L.3 Modified NASA TLX Workload Ratings 245

M Companion CD-ROM **248**

List of Figures

2.1 Graph of a sinusoidal waveform illustrating the basic components of a sound (frequency, amplitude and phase). Adapted from Moore [88]. 23
2.2 Graph showing how detectability of a source at a particular frequency is dependent on the intensity with which it is presented. The Minimum Audible Field (MAF) shows the minimum intensity with which a particular frequency needs to be presented to be heard by a listener. The other lines show equal loudness contours, relationships between intensity and frequency where the sounds are judged to be equally loud, see Section 2.3.2. Taken from Gelfand [56]. 24
2.3 A figure showing a chromatic scale, and the names of the notes involved. The lowest pitch is the lowest E and the highest is the top F. Adapted from Taylor [135]. 25
2.4 A diagram showing how the azimuth (direction) and elevation (vertical position) of a sound relate to a listener. 28
2.5 Sound takes longer to reach each ear, creating an interaural time delay (ITD) between a sound arriving at the left and right ear. Adapted from Gelfand [56]. 29
2.6 Diagram showing how high frequency sounds cast a shadow over one ear due to their inability to bend around the head. 30
2.7 Diagram showing a cone of confusion, an area around a listener's ear where sound is poorly localised and the minimum audible angle (MAA) is large. 30
2.8 Overview of the effect of presentation speed on streaming. Taken from van Noorden [139]. (A) When tones A and B are relatively close in frequency the listener hears a galloping melody. However at larger frequency separations (B), the listener will perceive two distinct streams, one of high frequency tones, another of lower frequency tones. 36
2.9 A visual example of the principle of belongingness. Whilst it is possible to see a vase or the profile of two faces, it is difficult to see both at the same time. 37

2.10 A graphical illustration of the role of attention in the work of van Noorden. The graph shows that there are two regions where the tone sequences are too different or similar to be able to be separated. However, there is a large area in which participants could apply attention to hear either precept. Slightly modified from van Noorden [139]. 38

3.1 An overview of the “grammar” used to construct a set of hierarchical earcons representing computer error messages. Taken from Blattner *et al.* [13]. 51

3.2 A shortening of hierarchical earcons from Figure 3.1, showing the similarities between the transformational and hierarchical earcon types. Taken from Blattner *et al.* [13]. 53

3.3 A diagrammatic representation of the mapping techniques that can be used between data and sound for both auditory icons and earcons. 55

3.4 An illustration of Dynamic Soundscape showing how audio was mapped to time. Taken from Kobayashi and Schmandt [67]. 57

3.5 An illustration of Dynamic Soundscape showing a new speaker, the original speaker is represented as grey to indicate that its volume has been reduced. Taken from Kobayashi and Schmandt [67]. 58

4.1 An overview of the Dolphin system showing how the focus and context were mapped to different modalities. 65

4.2 An example of priority zones used in Dolphin. 66

4.3 Example of an audio space with three Earcons, labelled A, B and C. 67

4.4 Example of the audio space from Figure 4.3, shown after the user has scrolled the visual display, and thus moved the focus. 68

4.5 Example of the display from the standard scrolling condition, which is also an example of the visual interface (focus) from Dolphin. 69

4.6 Melodies used to represent high, medium and low intensity theme park rides. 70

4.7 Examples of icons used to represent theme park rides. 71

4.8 An example theme park map used in the experiment. Note the size of the map has been substantially reduced from the size used in the experiment. 72

4.9 An example question as used in the experiment. 73

4.10 Graph showing the average time taken to complete a tour in the standard scrolling and Dolphin conditions, showing standard deviations. 75

4.11 Graph showing the average percentage optimal path length tours for the standard scrolling and Dolphin conditions. Shown with standard deviations. 76

4.12 Graph showing the average number of tours either missing rides or incorporating incorrect rides into a tour in the standard scrolling and Dolphin conditions, showing standard deviations. 77

4.13 Graph showing the mean values for NASA TLX workload data for the standard scrolling and Dolphin conditions. Shown with standard deviations. 78
4.14 Graph showing the percentage of correct performances on a coordinate response measure (CRM) task with 0, 1, 2 and 3 competing talkers. Taken from Brungart, Ericson and Simpson [29]. 80

5.1 A diagram showing the stages involved in the identification of concurrent earcons. Participants must first recognise and separate both perceptual and cognitive processes, the earcons contained within the composite sound. Participants must then use cognitive processes to recall the mapping between earcon sound attributes and the labels associated with those sound attributes. 88
5.2 An example of the presentation of concurrent earcons. Showing how each set of earcons were concurrently presented seven times. Taken from the three earcon condition. 91
5.3 A screenshot of the dialog used by participants to fill in earcons identified. Taken from the four earcon condition. 91
5.4 An example from the four earcon condition of how the set of participant responses (SPR) were mapped to the set of earcons presented (SEP) in order to determine the number of correctly identified earcons, ride types, ride intensities and ride costs. 92
5.5 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. 93
5.6 Graph showing the mean values for NASA TLX workload data for the one, two, three and four earcon conditions. 94
5.7 Graph showing best fit trend lines for the one, two, three and four earcon conditions, showing how the proportion of earcons and earcon attributes correctly identified is affected by the number of concurrently presented earcons. 96

6.1 Two melodies representing, (A) high intensity, and (B) medium intensity, high cost theme park rides. (C) The same two earcons concurrently presented with the same, and (D) with different timbres (denoted by note colour). Note how the addition of timbre differences makes the earcons easier to separate. 99
6.2 Example of an upward frequency glide. 103
6.3 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). 104
6.4 Examples of the melodies used for high, medium and low intensity theme park rides as used in the melody altered earcon condition. 109

6.5 An example of concurrent earcon presentation from the staggered onset earcon condition. The four earcons are still concurrently presented seven times, however a 300ms onset-to-onset delay between the start of each earcon is introduced. 110
6.6 A screenshot of the concurrent earcon training tool, designed to give participants an opportunity to understand how earcons interfere with each other when concurrently presented. . 111
6.7 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, multi-timbre earcon set, extended training, staggered onset and final experimental conditions. 112
6.8 Graph showing the mean values for NASA TLX workload data with standard deviations for the original earcon set, melody altered earcon set, multi-timbre earcon set, extended training, staggered onset and final experimental conditions. Shown with standard deviations. 113

7.1 An illustration showing how the earcons used in the spatial condition were located relative to the listener's head. 121
7.2 Graph showing the average number of earcons, number of correctly identified ride types, ride intensities and ride costs for the non-spatial and original spatial experimental conditions. Shown with standard deviations 122
7.3 Graph showing the mean values for NASA TLX workload data for the non-spatial and original spatial experimental conditions. Shown with standard deviations. 123
7.4 Graph showing the average number of earcons, number of correctly identified ride types, ride intensities and ride costs for the second original spatial and revised spatial experimental conditions. Shown with standard deviations. 127
7.5 Graph showing the mean values for NASA TLX workload data for the second original spatial and revised spatial experimental conditions. Shown with standard deviations. . . . 128

8.1 An example of the week at a glance view from the standard Palm date book application. . . 133
8.2 An example of the route walked by participants whilst carrying out the experiment. 136
8.3 An example of the visual display of the experiment. Participants could switch between days by using the "<<<<<" and ">>>>>" buttons. 137
8.4 An example of a question dialogue from the experiment, where only one answer was required.138
8.5 An example of a question dialogue from the experiment, where more than one was required. 138
8.6 An example of how the Xybernaut device was worn by participants and (insert) the device itself. 139
8.7 Melodies used to represent mildly important, very important and essential appointments. . 140
8.8 Graph showing the mean number of type S and type M questions correctly identified in the experiment. Shown with standard deviations. 142

8.9 Graph showing participants' mean percentage preferred walking speed (PPWS) recorded for each condition. Shown with standard deviations. 143

8.10 Graph showing mean results from the modified NASA TLX questionnaires. Shown with standard deviations. 144

List of Tables

2.1 Groupings of MIDI synthesised instruments based on similarity. Taken from Rigas [118]. . 27

3.1 Table describing the main guidelines of earcon design from Brewster, Wright and Edwards (BWE) [25]. 54

4.1 Table showing the procedure for the two groups of participants undertaking the experiment. 68

4.2 Table showing the attributes and their values encoded in the earcons and icons used in the experiment. 69

4.3 Table showing the attributes and descriptions of NASA TLX workload questionnaires used in the experiment. Modified from Hart and Staveland [60]. 74

6.1 Table showing the timbres used to represent ride types in the multi-timbre earcon set condition. 108

7.1 Table showing the procedure for the two groups of participants undertaking the spatial versus non-spatial experiment. 119

7.2 Table showing the procedure for the two groups of participants undertaking the spatial versus revised spatial experiment 126

8.1 Table showing the order in which each group of participants undertook the experiment. . . 135

8.2 Table showing the timbres used to represent appointment types in the spatial condition. . . 141

Declaration

The work on Dolphin as described in Section 4.4 has been published as McGookin and Brewster at both IHM-HCI 2001 [79] and ICAD 2002 [80]. The work contained in Chapters 5 and 6 has been published as McGookin and Brewster at ICAD 2003 [81] and in ACM Transactions on Applied Perception [84]. The work contained in Chapter 7 has been published at ICAD 2004 [83]. The guidelines for concurrent earcon presentation as described in Chapter 8 have been published at BCS HCI 2004 [82].

This thesis exploits only the parts of those publications directly attributable to the first author.

Acknowledgements

Firstly, I would like to thank Steve Brewster for his good humour, motivation and eternal optimism. Especially at times where I had run out of all three! He has been more than is required, or could be hoped from a supervisor, and for that I regard myself as being extremely fortunate and grateful.

I would also like to thank my past and present second supervisors, Ashley Walker, Matthew Chalmers and especially Helen Purchase. They have all provided invaluable advice and second opinions on thoughts and ideas.

Thanks must also go to the Multimodal Interaction Group at Glasgow, both past and present, for a stimulating working environment (and politely suffering my continual moaning whilst writing up!). Special thanks must go to Steven Wall, Georgios Marentakis and Lorna Brown for reading sections of this thesis, as well as Tony Stockman from QMUL who read the whole thing and provided many invaluable comments. Additional thanks to Lorna for piloting every experiment in this thesis at least twice!

Finally, my parents and brother, for their love, support (in more ways than is possible to describe) and for occasionally asking how my "project" was getting on. I love you all, even if I don't always say it very well.

Chapter 1

Introduction

1.1 Motivation of the Thesis

How sound can be integrated into a human computer interface, and how it can be effectively used to present information to users has been an active area of research for over fifteen years [68]. Such auditory displays have been shown to offer significant advantages over interfaces which use a purely visual display [20, 9, 3, 2], or in some cases have allowed systems to be created which have no visual interface component at all [28, 41, 43, 44]. Sound can be used to reduce the number of errors when users are interacting with small screen visual displays such as with personal digital assistant (PDA's) [20], and may also increase the available display space on such mobile devices [22]. Sound may also be used to reduce users' visual demands in "eyes busy" tasks [140, 62], and provide effective access to computers for visually impaired users [41, 91].

Sound is however, not without its problems. The lower spatial acuity of the human auditory system, in comparison to the visual perceptual system [68], limits the bandwidth of communication in auditory displays. In most work on auditory displays, only one auditory source (which will be assumed to represent a data source) is presented at any one time. This is unlike the rich, parallel presentation that occurs in graphical user interfaces, where multiple icons, pictures, user interface widgets and textual information can be simultaneously presented to a user. Such presentation allowing users to browse large information spaces more quickly, and keep demands on short term memory low by being able to quickly switch between different information. For example, when undertaking a writing task, a user may wish to have a document processor, bibliography manager and electronic summary notes all visible on screen at the same time, thereby exploiting the spatial acuity of the human visual system. The auditory system however, is highly temporal and does not posses the degree of spatial acuity and separation available with the visual system. Audio is more suited to data which can be encoded in short auditory sounds, such as with Earcons [16] and Auditory Icons [52], or with temporally changing data (or data which can be easily mapped to the temporal domain, and be presented quickly such as graphs [76]). Due to the slowness of speech based sound, non-speech sound mapping techniques such as sonification, earcons and auditory icons are, in many cases, better suited when presenting information. However, even when using these methods, the communication bandwidth of audio is limited in comparison to a visual display.

There are an increasing number of situations where it would be desirable to increase the bandwidth of the auditory channel for information presentation. Mobile computing devices, such as the HP iPAQ have capabilities that rival desktop computers of just a few years ago, but in a form factor that allows them to be fitted in a pocket. The power of mobile telephones has allowed them to become much more than just devices for making telephone calls; applications such as Web browsers and diaries are commonly found in today's mobile telephones. However, for such devices to be mobile, the size of their visual display is limited, restricting the bandwidth of communication available to present information to users. Additionally, if applications are being used on the move, the user is attempting to use the device in an "eyes busy" situation and as such must also attend to the environment and be aware of obstacles or danger. Users cannot therefore constantly look at the device. To leverage the full power of such devices and the advantages of having contact and email information available on small personal devices, new forms of human computer interaction must be developed. Increasing the bandwidth of data communication in an Auditory Display can also assist with applications for blind and visually impaired people, where sound based applications would be able to more closely, and easily, replicate the expressability of a visual interface, allowing for less demanding interaction on the part of the user, and a richer communications channel for designers to exploit.

One way in which the bandwidth of auditory displays can be increased is through the concurrent presentation of sound. In other words, instead of presenting information one sound at a time, present multiple sounds at the same time, reducing the time to present information, as well as allowing real time comparison of multiple data [27].

Several systems which use concurrent audio presentation have been developed over the last few years, such as Gaver, Smith and O'Shea's ARKola bottling plant simulator [55] (see Section 3.3.3) and Sawhney and Schamndt's Nomadic Radio [124] personal notification diary system (see Section 3.4). Kobayashi and Schmandt [67] have developed a speech based recording browser (see Section 3.4) which uses concurrent presentation to monitor different parts of the recording. Fernström and Bannon [46] (see Section 4.2.3) have created a music browsing system, which uses concurrent audio presentation to assist in browsing musical compositions. Such systems allowing comparison between multiple auditory sources to be made, something which is more difficult to do when only one auditory source is presented at a time.

Although there have been several systems which use concurrent sound presentation, and those systems seem to be effective, there has been little formal evaluation to determine the extent of their effectiveness, or potential issues with such concurrent presentation of sound. An outcome of this is that there are no guidelines for designers to use when designing auditory displays which incorporate concurrently presented audio, making it difficult to fully exploit the advantages such presentation affords. Work by Gerth [57], Papp [104] and Brungart, Ericson and Simpson [29], indicates that when sounds are concurrently presented they may interfere with each other and thus any data which are encoded within those sounds becomes difficult to identify. Such a hypothesis is supported by Auditory Scene Analysis [14], a branch of psychology which investigates the perception of concurrently presented sounds. This creates problems for designers who wish to employ the advantages of concurrent presentation in auditory displays as it may be difficult to design sounds which both do not interfere with each other when concurrently presented, and can effectively communicate information. Whilst it may be possible to modify the design and presentation of sounds to reduce the likelihood of them interfering either each other, auditory displays implicitly assume there is some mapping between data and the attributes of the sound used to represent the data. It may not be possible to modify a sound enough to reduce its interference, whilst retaining intact the mapping between data and sound. This is particularly a problem for certain types of earcon, short structured audio messages which can be used to communicate information to a user in an auditory display [13]. Earcons are formed from a "grammar" and as such individual earcons may be very similar, meaning that they both interfere with each other, as well as be difficult to modify to avoid such interference. This creates significant problems when trying to use earcons in concurrent presentation situations. This thesis investigates such issues with concurrently presented earcons, and determines ways in which earcons can be redesigned to reduce the impact of interferences between them.

1.2 Aims of the Thesis

1.2.1 Thesis Statement

Identifying concurrently presented earcons where those earcons are constructed from a complex "grammar" is difficult. Whilst modifications can be undertaken to significantly improve earcon identification, these modifications are constrained due to the need to preserve the mapping between data and sound.

This statement will be defended through work which seeks to answer the following four research questions:

RQ1 What is the effect on earcon identification of varying the number of earcons which are concurrently presented?

RQ2 How can concurrently presented earcons be more robustly designed and presented in order to increase the number which can be identified?

RQ3 What is the impact of presenting concurrent earcons in a spatialised auditory environment?

RQ4 How much do modifications to the design and presentation of concurrently presented earcons affect performance in "real world" tasks?

These research questions seek to determine two key aspects of concurrent earcon presentation: how well are earcons identified when concurrently presented, and how can the design and presentation of those earcons be changed to improve their identification. This work is novel in that no previous study has attempted an extensive investigation to identify how earcon identification varies by changing the number concurrently presented. Whilst such a trend has been identified for other forms of audio (such as speech [29, 31]), one has not been identified for earcons. Further, no previous study has empirically investigated how interactions between earcons can be reduced whilst not destroying the “grammar” that connects the earcons (and provides many of their advantages). The answers to these research questions will provide a set of guidelines for concurrent earcon presentation which can be used by future auditory display designers to make informed decisions when incorporating concurrent earcons into a human computer interface.

1.3 Contents of the Thesis

The argument of this thesis is divided into chapters in the following way:

Chapter 2 *“Basics of Sound”* Provides an overview of audio essential to understand the original and novel work presented in this thesis. This chapter starts by introducing sound and its basic dimensions. The chapter follows this with a discussion of various attributes of sound that will be later used to build earcons. How these attributes can be understood in terms of the basic dimensions of sound, and as such how psychological work on concurrent sound identification can be applied to earcon design is discussed. The chapter concludes with a discussion of the localisation of sound in three dimensions (3D) as well as how concurrently presented sounds may be interpreted by the human auditory system.

Chapter 3 *“Auditory Display”* Provides an overview of auditory displays, or how the work on sound discussed in Chapter 2 can be applied in a human computer interface and communicate information effectively to users. The chapter starts by discussing the advantages and disadvantages of using sound in such a way. Following this, a taxonomy of the main four methods by which data can be encoded in sound, and the relative advantages and disadvantages of each is outlined. The chapter concludes with the advantages of the use of 3D, or spatialised sound, as part of an auditory display.

Chapter 4 *“Issues Involving Concurrent Audio Presentation”* Follows from Chapter 2 and 3 by discussing concurrent audio presentation in more detail. The advantages for using concurrent audio presentation are discussed. The disadvantages, including the interference of concurrently presented sounds are then outlined. A study which motivated this research, and illustrates how the problems of concurrent sound presentation can easily outweigh the advantages is discussed. The chapter follows by discussing the limited research that has been carried out on concurrent audio as part of an auditory display, before discussing the research questions posed in Section 1.2.

Chapter 5 *“Number of Earcons versus Identification”* Describes a novel experiment which provides answers for RQ1, *“What is the effect on earcon identification of varying the number of earcons which are concurrently presented?”*. The experiment compares the identification of 1, 2, 3 and 4 concurrently presented earcons, and identifies how earcon identification is affected when the number of concurrently presented earcons is varied. Guidelines for the concurrent presentation of earcons are identified and presented.

Chapter 6 "*Engineering More Robust Earcons"* Describes a novel experiment which provides answers for RQ2, *"How can concurrently presented earcons be more robustly designed and presented in order to increase the number which can be identified?"*. The experiment compares the identification of four concurrently presented earcons, when various modifications to the design and presentation of the earcons, based on auditory scene analysis research [14] have been undertaken. Guidelines for the concurrent presentation of earcons are identified and presented.

Chapter 7 *"The Impact of Spatialisation on Concurrent Earcon Identification"* Describes two novel experiments which provide answers for RQ2, *"How can concurrently presented earcons be more robustly designed and presented in order to increase the number which can be identified?"*, and RQ3, *"What is the impact of presenting concurrent earcons in a spatialised auditory environment?"*. The first experiment compares the identification of earcons presented in the same spatial location which incorporated the guidelines identified from the work of Chapter 6, with spatial presentation of earcons not incorporating the guidelines ("base" earcons). The second experiment compares the identification of four concurrent earcons that both incorporate the guidelines identified from the experiment in Chapter 6, and are presented in distinct spatial locations, with spatially presented "base" earcons. Guidelines for the concurrent presentation of earcons from these experiments are identified and presented.

Chapter 8 "*An Ecological Investigation into Concurrent Earcons"* The experiments from Chapters 5, 6 and 7 are of an abstract nature. The novel experiment of this chapter seeks to provide answers for RQ4, *"How much do modifications to the design and presentation of concurrently presented earcons affect performance in "real world" tasks?"*, and ground the guidelines from Chapters 5, 6 and 7 in a more ecological context. The experiment of this chapter compares performance in a mobile diary browsing task when diary entries are represented by non-spatially presented "base" earcons, and by spatially presented earcons which incorporate the guidelines of concurrent earcon presentation from Chapters 5, 6 and 7. The results of the experiment are discussed and directions for future work are outlined.

Chapter 9 "*Conclusions"* Summarises the work from the previous chapters and relates this back to the research questions from Section 1.2, discussing to what extent those four research questions have been answered. The limitations of the thesis, and how these might be resolved are described. Future directions for further investigations based on the results of the work described in the thesis are also outlined.

Chapter 2

Basics of Sound

2.1 Introduction

The previous chapter introduced the main research focus of this thesis, namely the investigation of concurrently presented structured sounds called earcons. Before discussing the novel work of this thesis in detail, it is important to understand a little about sound, what it is and how it is interpreted by human listeners, so that a background can be provided to understand the issues and solutions in the identification of concurrently presented earcons.

This chapter starts by describing the physical, objective components of simple sounds, before discussing how such components are interpreted by the human auditory system to form much richer auditory attributes such as those used in the construction of earcons. This is important as the work which will be applied later in the thesis to improve the identification of concurrently presented earcons, was undertaken with the lower level attributes of sound. The discussion then turns to two features of sound which will become of great significance through the remainder of the thesis: the interpretation of multiple concurrently presented sounds (see Chapters 6 and 7), and how the location of sound sources can be determined by the human auditory system (see Chapter 7).

2.2 Introduction to Sound

Sound, as defined by Moore [88], *"originates from the motion or vibration of an object. This motion is impressed upon the surrounding medium (usually air) as a pattern of changes in pressure"*. Sound consists, at its most basic level, of three objectively measurable components. Frequency is the number of times a waveform repeats itself in a given amount of time. This is usually measured in Hertz (Hz), where 1Hz is equivalent to 1 complete playing of the waveform each second. Amplitude relates to the size of the pressure increase or decrease from the mean pressure. It can crudely be considered a metric of loudness, although as will be explained in Section 2.3, things are slightly more complex. Finally, the phase of a sound is *"the proportion of the cycle through which the wave has advanced in relation to some fixed point in time"* [88]. These three features are illustrated on a sinusoidal waveform in Figure 2.1.

The intensity of a sound is the pressure per unit squared, generally measured in Newtons per meter

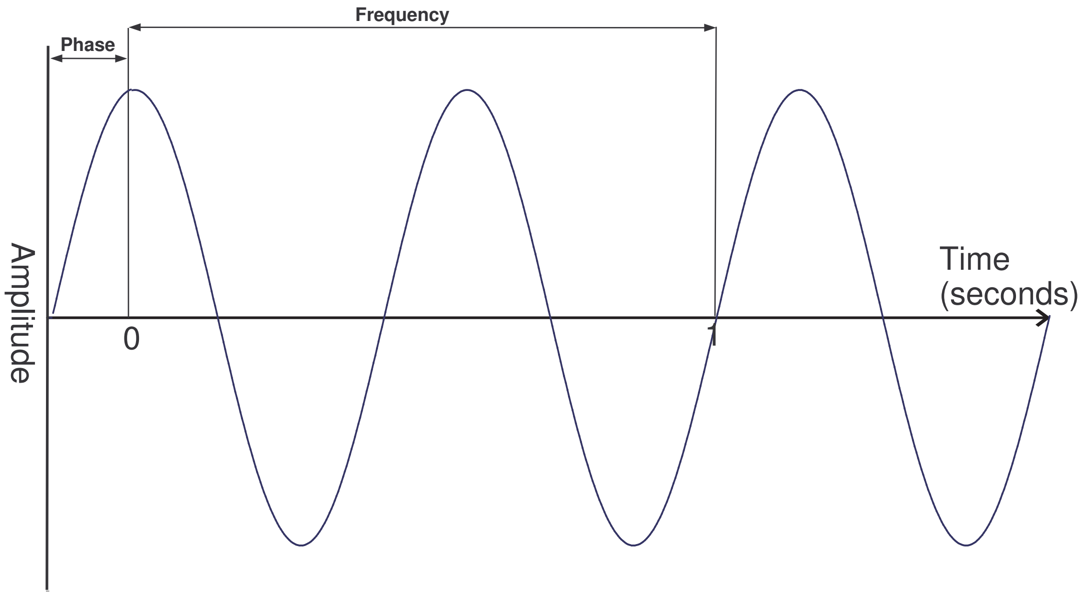

Figure 2.1: Graph of a sinusoidal waveform illustrating the basic components of a sound (frequency, amplitude and phase). Adapted from Moore [88].

squared, and is proportional to the square of the amplitude. Because of the large range of intensities the human auditory system can detect (the largest intensity is around 10^{14} that of the smallest), a ratio scale is commonly used, the units of which are decibels (dB) [152]. The decibel scale is a logarithmic scale of intensity relationships. Therefore if a sound is described as having an intensity of 1.3dB, it in fact means that the sound has an intensity (approximately) 20 times greater than that of some reference pressure. In order to simplify comparing different intensities measured in decibels, standard reference pressure levels have been defined. Sound Pressure Level (SPL) is the most common, using a reference pressure equivalent to the hearing threshold of a 1kHz sinusoidal tone [88]. See Section 2.3.2 for a discussion on the relationship between intensity and frequency.

2.3 The Perception of Sound

2.3.1 The Limits of Human Hearing

Whilst sound can be objectively described in mathematical terms, it is generally more useful to consider how it is perceived by the human auditory system. Humans can hear sound with frequencies in the range of 20Hz to 20kHz [56]. However with the onset of age, the upper limit on hearing tends to reduce to about 15kHz [88]. To hear all of these various frequencies, the intensity with which they are presented needs to be altered, with lower frequency sounds being presented with a greater intensity than higher frequency sounds. The point at which a particular frequency is presented with enough intensity to be audible is known as the

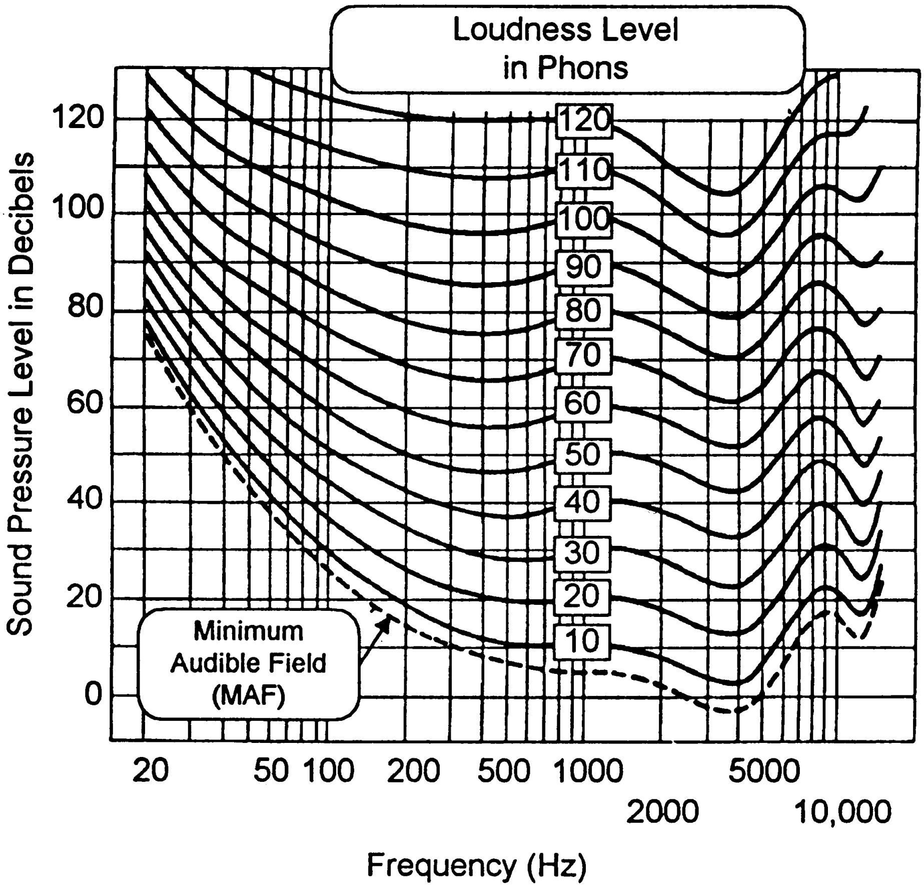

Figure 2.2: Graph showing how detectability of a source at a particular frequency is dependent on the intensity with which it is presented. The Minimum Audible Field (MAF) shows the minimum intensity with which a particular frequency needs to be presented to be heard by a listener. The other lines show equal loudness contours, relationships between intensity and frequency where the sounds are judged to be equally loud, see Section 2.3.2. Taken from Gelfand [56].

minimum audible field (MAF) [56]. A graph showing how the MAF changes with respect to the frequency and intensity of a sound is shown in Figure 2.2.

2.3.2 Auditory Attribute Perception

Whilst sound has physical characteristics (see Section 2.2), it has other properties that exist because of its interaction with the human auditory system. This section will describe the three main characteristics of this interaction which are used in the construction of earcons, and are therefore relevant to this thesis: pitch, timbre and loudness. How these relate to the basic attributes of sound, and therefore how the work described later in Section 2.6 can be applied to the design of earcons is also discussed.

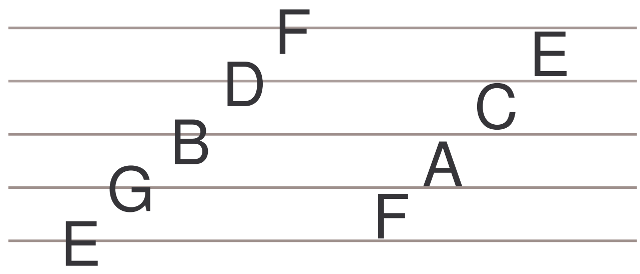

Figure 2.3: A figure showing a chromatic scale, and the names of the notes involved. The lowest pitch is the lowest E and the highest is the top F. Adapted from Taylor [135].

Pitch

Pitch, defined by the American Standards Association [6] as *"that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale extending from low to high"*, is the subjective equivalent of frequency. Indeed, the pitch of a sound is largely related to its frequency [88]. Whilst most listeners can identify the relative pitch between two sounds, or judge general classes of pitch, where there are large differences between values, e.g. low, medium or high, very few listeners can identify absolutely the pitch of a sound without the use of a reference tone. This ability known as perfect pitch appears to be present in less than 0.01% of the population [122]. This means that when pitch is used to encode information it is unlikely to be useful unless large differences between different pitch values are used. Pitch perception also suffers from the "pitch of the missing fundamental" issue [152]. Here, if a listener is presented with tones that are harmonics (a component of a complex tone whose frequency is an integral multiple of the fundamental frequency of that complex [88]) of some frequency, they will perceive a sound with a pitch of the fundamental frequency, even when that frequency is not present in the sound. For example, presenting the frequencies of 400Hz, 800Hz, 1000Hz etc., which are harmonics of a 200Hz tone will cause the listener to hear the same pitch as a separately presented 200Hz tone [88]. Therefore care should be taken to avoid such a phenomenon when concurrent sounds are used as part of an auditory display. Since earcons as will be described in Section 3.3.4 are *"abstract, synthetic tones that can be used in structured combinations to create auditory messages"* [16], it is most appropriate to think in terms of musical pitch. In this thesis a standard musical chromatic scale will be used to denote pitch [116]. An example musical stave showing the notes and their positions as used in this thesis, is given in Figure 2.3. This thesis also discusses register which is regarded to be a particular pitch class [116], or a scale which is referenced to a particular pitch. For example C4 represents has a frequency of around 261Hz, where as C5 has a frequency of around 523Hz [63].

Timbre

Timbre is perhaps one of the most obvious qualities of a sound, yet is difficult to properly define [88]. The American National Standard Institute (ANSI) define timbre as *"that attribute of auditory sensation in terms of which a listener can judge two sounds similarly presented and having the same loudness and pitch are dissimilar"* [6]. In other words timbre is composed of those features of a sound which do not have any other name, as Bregman notes *"The problem with timbre is that it is the name for an ill-defined wastebasket category"* [14]. He offers a further definition of timbre as *"we do not know how to define timbre but it is not loudness and is not pitch"*.

Whilst all the components of timbre are not fully understood, there is research which indicates some components. The number of harmonics, and relative amplitudes for each of those harmonics has a strong influence on timbre [39]. In work on the pitch of the missing fundamental problem, as described in the previous section, although no variation in pitch was found between a tone of a certain frequency and a tone formed from only the harmonics of that frequency, listeners reported that the timbre of the tones were different [88].

Although the number of harmonics and amplitude for each of those harmonics has a strong influence on timbre they can not explain timbre fully. As Deutch notes [39] *"a saxophone remains a saxophone whether it is heard over a distortion-ridden pocket-sized transistor radio or directly in a concert hall."*. Work by Stumpf as reported by Risset and Wessel [119], showed that removing the attack portion of notes played by certain instruments significantly impaired the ability of listeners to identify the instruments used. Again, as reported in Risset and Wessel [119], work by George showed that by playing a recording of a piano backwards gave a non-piano like quality to the sound, in spite of the sound having the same frequencies and amplitudes whichever way the recording was played. Indeed, more expensive wavetable music synthesisers, where each sound is produced from a mathematical manipulation of a pre-stored musical recording, store different recordings for the attack and decay, in order to produce a more realistic sound [20]. This is important to note as when sounds are used in human computer interfaces, and the timbre of those sounds has been synthetically generated, the timbre may not sound like the instrument that it synthesises.

Because timbre is not fully understood and is multi-dimensional, it is difficult to quantitively determine the similarity or dissimilarity between two timbres in the same way as with pitch or loudness [88]. This presents a problem for auditory display designers in attempting to choose sounds which will effectively communicate information and not be confused. Fortunately Rigas [118] has carried out experimental studies in order to categorise MIDI (Musical Instrument Digital Interface) [85] sounds in groups based on their subjective similarity. He presented listeners separately with tunes of 8 notes played on 23 different synthesised musical instruments and asked them to write down the name of the instrument that played the tune. He found that pianos, organs, xylophones and drums were particularly well identified by listeners. In a further study he presented listeners with a list of five named instruments (Piano, Guitar, Drums, Violin, Saxophone, Flute and Harp). Listeners were then played a sound of one of the instruments and had to select

Group	Instruments
Piano	piano, harp, guitar celesta, xylophone
Organ	organ, harmonica
Wind	trumpet, French horn, tuba, trombone, saxophone
Woodwind	clarinet, English horn, pan pipes, piccolo, oboe, bassoon, flute
Strings	violin, cello, harp

Table 2.1: Groupings of MIDI synthesised instruments based on similarity. Taken from Rigas [118].

which one they heard. Rigas found that in all cases recognition of the sounds was high with over 80% correct responses for each instrument except the harp which had only 30% correct responses. These results highlight the problems of cheap MIDI synthesisers not accurately representing sounds, as Rigas believes the MIDI synthesizer used did not accurately synthesise the harp timbre. Rigas proposed a grouping of timbre families, the members of which were unlikely to be confused with members of another family. His grouping is described in Table 2.1.

Whilst the similarity groupings of Rigas [118] may not be generally applicable since he used a relatively cheap wavetable synthesiser [20] to generate the sounds, they do provide guidance on the use of musical timbres in auditory displays. Therefore although it is not possible to fully explain timbre, it is possible at least musically, to provide general comparisons of similarity, which can be used in the design of earcons. In the remainder of this thesis whenever timbre is mentioned it will refer to musical timbre (the musical instrument used to play a sound).

Loudness

Loudness, defined as *"that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from soft to loud"* [152], is (as with pitch) subjective. Loudness is measured on the phon scale, where 1 phon is the loudness of a 1000Hz tone presented with the intensity of 1dB SPL [152]. As is implied by the minimum audible levels of sound is Section 2.3.1, loudness varies in terms of both frequency and intensity. Figure 2.2 presents a graph of equal loudness contours, showing the relationship between frequency and intensity required to equate in phones.

One of the issues surrounding the concurrent presentation of audio with different degrees of loudness is that the sounds may obscure or mask each other. Masking is *"the process by which the threshold of audibility for one sound is raised by the presence of another (masking) sound"* [6]. The most common form of masking is simultaneous masking, which occurs when two sounds are concurrently presented, where one sound is presented with such intensity, commonly called the masker, that the other sound, commonly called the maskee [56], cannot be detected [88]. In order for masking to occur, not only does the intensity of the masker need to exceed that of the maskee but it must also contain the same frequencies. Research by Fletcher [47] has shown that as the bandwidth of the masker is increased, the intensity that the maskee needs to be presented at to remain audible increases, but only up to a certain point beyond which increasing the

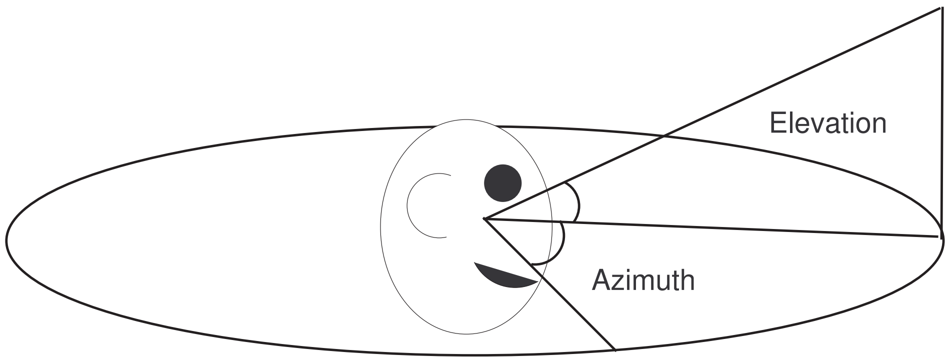

Figure 2.4: A diagram showing how the azimuth (direction) and elevation (vertical position) of a sound relate to a listener.

bandwidth of the masker has no effect on the amount of masking. The frequency range at which the amount of masking does not increase is known as the critical bandwidth (CB). It is believed that the human auditory system contains a number of filters each of which deals with a critical band [88]. If the masker and maskee fall into different critical bands, due to frequency variations, then masking will not occur. This indicates that using broadband sounds, which contain a range of frequencies, will be more robust to masking than pure tones when used in an auditory display if sounds are to be presented concurrently. It is also essential when using concurrent audio presentation to ensure that sounds are of equal perceivable loudness as possible to reduce the possibility of masking [88].

2.4 Spatial Audio

Whilst the previous sections have described how the human auditory system interacts with basic sound, it misses one important characteristic of human hearing: our ability to localise or identify the location in space of a sound source. In this section, how the human auditory system localises sound will be outlined, and the limitations of current “off the shelf” spatialisation technologies will be discussed.

2.4.1 Binaural Cues

Our ability to localise sound depends on several factors, the most important of which are the differences between the sounds reaching our left and right ears. These binaural cues [88] incorporate variations in both the timing and the spectra of sound between the ears to determine the azimuth (direction) and elevation (vertical height) of a sound source (see Figure 2.4).

The interaural time difference (ITD) (also known as the interaural phase difference (IPD)) uses phase differences (see Section 2.2) between the sounds arriving at the left and right ears to determine sound source location [88]. Because a listener’s ears are on opposite sides of the head, unless a sound is directly in front or behind the listener, it will take longer to reach one ear than the other. There will therefore be a time delay

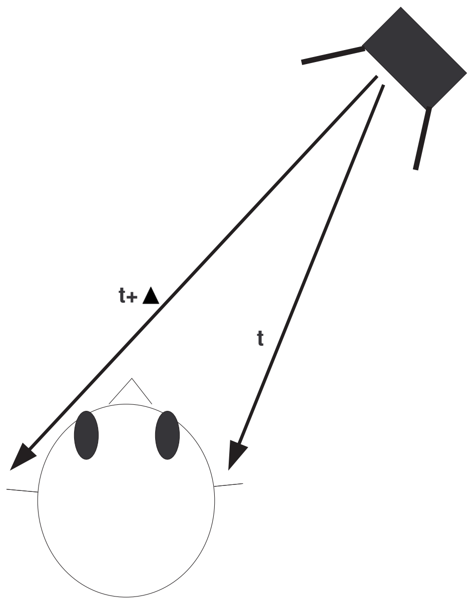

Figure 2.5: Sound takes longer to reach each ear, creating an interaural time delay (ITD) between a sound arriving at the left and right ear. Adapted from Gelfand [56].

between reaching the ears that can be used by the auditory system to identify the azimuth of a sound (see Figure 2.5).

The interaural intensity difference (IID) [8], sometimes called the interaural level difference (ILD) [152] uses intensity variations in the spectra of the sounds reaching each ear to determine azimuth. In order to reach the ears sound will have to move, or diffract, around the head. The ability of sound to do this depends on the frequency of the sound. Higher frequency sounds, bend less well than lower frequency sounds and hence the head casts an “auditory shadow” over the ear causing the intensity of the sound reaching it to be reduced (see Figure 2.6) [88].

The binaural cues are sometimes called Duplex Theory [8], as they are interrelated. The ITD tends to be more effective at localising lower frequency sounds, below about 1500Hz, since above this frequency, the phase difference between the ears is a multiple of the distance between the ears and as such becomes ambiguous. Similarly, the IID becomes useful for relatively high frequencies, above 1500Hz, when head shadowing becomes noticeable [152]. Taking these into account it seems that listeners are worst at localising sound in the range of 1500Hz. This should be taken into account by auditory display designers who choose to use spatial presentation, as sounds around that frequency may be poorly localised. The use of broadband stimuli, i.e. using a range of different frequencies instead of pure tones tend to be more accurately localised and may help to reduce this issue [89].

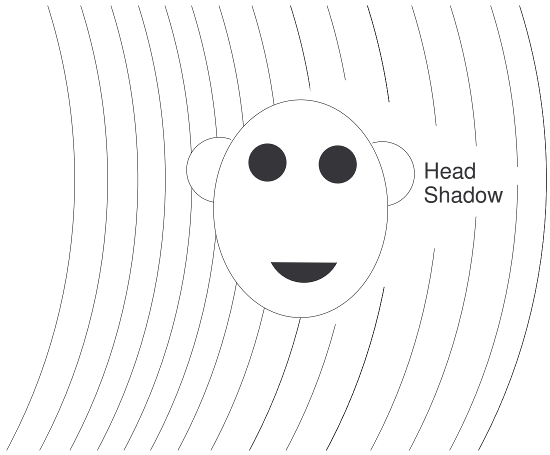

Figure 2.6: Diagram showing how high frequency sounds cast a shadow over one ear due to their inability to bend around the head.

2.4.2 Localisation Performance

The localisation of sound is also affected by the relative location of the sound source to the listener's head. For example, sounds which are directly in front of the head are more accurately localised than those presented to the side of the head. The minimum audible angle (MAA) is the minimum amount by which two sequentially presented sounds need to be separated in azimuth for a listener to determine that they originate from different positions [56]. Work undertaken on the MAA shows that for sounds directly in front of the listener only 1-2^o of separation is required for listeners to determine the sounds as being located in different positions. This is due to the relatively large change in interaural cues for movements in this region. In comparison, the MAA for sounds located around the listener's ears dramatically increases, due to the small relative changes in the binaural cues when sound sources are moved. This leads to the so called cones of confusions, see Figure 2.7, within which the MAA is extremely large.

As this thesis deals with concurrent auditory presentation, another important feature is the concurrent minimum audible angle (CMAA) [110]. The CMAA denotes the amount of spatial separation required

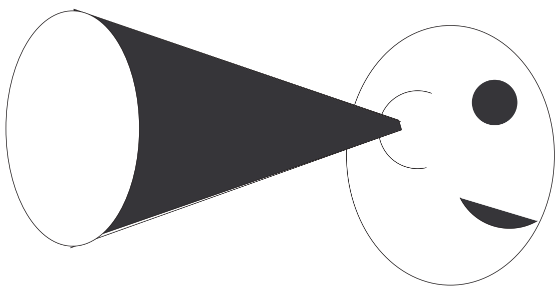

Figure 2.7: Diagram showing a cone of confusion, an area around a listener's ear where sound is poorly localised and the minimum audible angle (MAA) is large.

between two concurrently presented sounds for them to be determined as coming from different positions. As with the MAA, the CMAA is much smaller for positions directly in front of the listener than for those around the listener's ears. However, the CMAA is also affected by spectral differences between the sounds, such as timbre, pitch etc. [56]. The degree of similarity between the sounds can cause wide variation in the CMAA. For example, Best, van Schaik and Carlile [10] have found for sounds which are spectrally the same, the CMAA could be up to 60^o. It is important therefore that when using concurrent auditory stimuli and spatial audio, especially when those stimuli are similar, that large azimuthal separations between the auditory sources is maintained to reap the full benefits of spatialisation. However, as will be discussed in the next chapter, this may not always be possible.

2.4.3 Elevation

The perception of elevation, or vertical position, of a sound source (see Figure 2.4) is determined principally by the interaction of the sound wave with the head and pinnae (outer ear) which creates spectral cues that can be used by the auditory system to determine elevation. The pinnae is a highly directionally dependent filter [88], and works best at frequencies above 6kHz [126], as well as with broadband sounds containing a range of frequencies [152]. The interaction with the pinnae causes certain frequencies in the sound to be attenuated more than others, the principle being similar to that used in the IID. Depending on the elevation of a sound source, the amount of attenuation to different frequencies varies [56]. These cues can therefore be used by the auditory system to determine sound source elevation.

2.4.4 Distance Perception

The ability of listeners to determine the distance of an auditory source depends on both auditory and non-auditory features [154]. As would be expected, the intensity of a sound, and thus its perceived loudness, is reduced the further the listener is from the sound source due to absorption by the air [152]. This can be used by a listener to identify the distance of an auditory stimulus. However to be effective, the listener would need to be familiar with the object generating the auditory source [126]. For example, if a bumble bee and a jet aeroplane were presented at the same loudness and listeners were asked to judge which was closer, it is likely that the bumble bee would be judged as closer since a jet aeroplane is generally much louder than a bumble bee (at a given distance). Moore [88] notes that this cue is also effective to judge the relative distances of concurrently presented sources where such familiarity with the sound may not be necessary.

In addition, the ability of sound to reflect off surfaces (reverberation), can be used to judge source distance [126]. The reverberant to direct sound ratio defines the proportion of direct and reverberant sound reaching the ears. As a listener moves further away from a sound source, the intensity of direct sound, sound coming directly from the sound source, drops, whilst the intensity of reverberant sound remains constant, hence distance can be determined [8]. Some studies however, note that whilst reverberation can assist in the perception of distance, it can have a negative impact on azimuthal localisation, particularly in cases where there are multiple sources, or the ability to extract information from the sound source is required [126]. This is important for auditory display designers as if complex auditory messages (see Section 3.3.4) are to be identified by the user, incorporating reverberation may impair users' ability to identify that information.

2.5 Spatial Sound Synthesis

Whilst the issues outlined in Section 2.4 have been known about for many years, and the ability to spatialise sounds in limited environments for psychoacoustics research has been available for some time, it is only recently that general purpose computing tools have been available to auditory display designers for sound spatialisation [8]. Tools and technologies such as Microsoft DirectX [7] and the Open AL project [101], combined with sophisticated audio hardware such as the Creative Audigy [35] and the Turtle Beach Santa Cruz [137], are now found in many of the personal computers (PC's) sold today, making spatial presentation of sound a practical tool for Auditory Display designers to employ.

2.5.1 Head Related Transfer Functions

The ability to create such synthetic spatial environments is due to a head related transfer function (HRTF). An HRTF modifies (by applying duplex theory (see Section 2.4.1) as well as spectral cues (see Section 2.4.3)) a sound source such that the listener's auditory system perceives the sound to be coming from some position in space [146]. HRTFs provide the advantage of being able to construct auditory environments that can be easily changed and reconfigured without having to physically move the sound sources (e.g. repositioning loud speakers), allowing virtual spatialised auditory environments to be created. In addition, these sound environments can be presented to listeners over headphones or two shoulder mounted speakers, which makes it feasible to create sound environments that can be used in a mobile context [124].

The binaural and spectral cues used in an HRTF can be measured by placing microphones in a listener's ear canal, or by using a Kemar dummy, with pinnae shaped closely to those of the listener [8]. Sounds of varying frequency are played from fixed locations around the user, and compared with the in ear recordings to create finite impulse response (FIR) filters. These can then be mathematically modified to approximate FIR filters from locations where recordings were not made, to provide all round spatialisation ability [146].

2.5.2 Generalised Head Related Transfer Functions

Whilst the use of HRTFs can provide comparable localisation accuracy to free field listening [149], they are not without practical problems for auditory display designers. HRTFs are both time consuming and costly to produce, requiring a listener to be available so that a "customised" HRTF can be created for them [127]. Whilst "off the shelf" spatialisation hardware that can support customised HRTFs exist such as the Lake Huron [71], such technology is expensive.

To overcome these problems, auditory display designers can use generalised HRTFs (GHRTF), which incorporate the same cues found in the customised HRTF (see Section 2.5.1), except an “average” head and pinnae are used to calculate the FIRs [8]. This is the type of HRTF that is embodied on most current PC sound cards. Studies carried out on generalised HRTFs have found that whilst azimuth localisation is comparable to individualised HRTFs, elevation accuracy is degraded [148, 147]. In addition, the number of front back confusions is also increased when using generalised HRTFs since a listener’s pinnae are highly individual to that particular listener [127]. However this issue can somewhat be reduced by attaching a location and orientation sensing device to a listener’s head and using it to allow a listener to reorientate themself in the auditory environment, thereby providing more dynamic cues to determine orientation [126, 67]. There has also been a limited amount of research to determine if users can be trained to improve their localisation performance when using generalised HRTFs [145]. Work by Zotkin *et al.* [155] has investigated the selection of an HRTF based on measurement of the listener’s pinnae. Here listeners’ pinnae are matched using a number of comparative measurements to the pinnae of listeners from whom the HRTFs were measured. The HRTF of the pinnae which offers the closest match to the listener’s pinnae is then used as the HRTF for that listener. This work is at an early stage however, and it remains unclear how effective it will be.

The limitations of generalised HRTFs mean that when spatialised auditory displays are used, they will be most effective if sounds are positioned only in azimuth, with elevation not being used to separate sources or encode information. The use of a tracking device to dynamically respatialise the environment in response to listener’s head movements should also increase the effectiveness of the display. Such dynamic respatialisation allows for “cones of confusion” areas to be remapped in front of the listener, thereby allowing some of the problems with the MAA and CMMA to be overcome [74].

2.6 Auditory Scene Analysis

2.6.1 Introduction to Auditory Scene Analysis

In the previous sections, sound, from its physical properties to its interpretation by the human auditory system has been described. One feature of sound which was not considered in that discussion, which is essential to this thesis, is the interpretation of multiple concurrently presented sounds. Whilst the phenomena of masking (see Section 2.3.2), the CMAA (see Section 2.4.2) and the harmonics of a sound contributing to its timbre (see Section 2.3.2), can be considered as the interaction of concurrently presented sounds, there are interactions which can occur between sounds, that are of a more complex nature. In this section the study of these complex interactions, Auditory Scene Analysis (ASA), is discussed.

Auditory Scene Analysis 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 [14] (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. As will be shown in Chapter 4, trying to identify sounds in a human computer interface which are concurrently presented can be difficult if those sounds are (undesirably) placed in the same stream.

2.6.2 Determining Differences in Auditory Scene Analysis

Whilst it is simple to provide examples of streaming in everyday life, understanding how sounds are grouped into streams, and determining what auditory factors influence that grouping, and how those factors are related, is much more difficult. Indeed Auditory Scene Analysis is a very active area of research within psychology. Since Auditory Scene Analysis is not (yet) fully understood, it is not possible to analyse a composite sound and then determine how it will be streamed by the auditory scene analysis system.

Most research into auditory scene analysis has involved the investigation of one particular auditory attribute, and investigated how changing it affects the streaming of an auditory source. The results of such work show that streaming is much related to visual gestalt psychology [150], and as such can be classified in various gestalt categories. Modifying sounds along such dimensions can influence how they are streamed by the ASA system. Whilst it is impossible in this thesis to provide a complete overview of all ASA research, several examples of work are given to illustrate the principles of each category, and research which has been directly applied to improve the identification of concurrently presented earcons is discussed in Section 6.1.1. For a much more comprehensive overview of Auditory Scene Analysis, the reader is referred to Bregman [14] and Deutsch [39]. The following sections describe the various gestalt categories (taken from Williams [150]) and how they relate to auditory streaming.

Similarity

Components which have similar attributes are more likely to be grouped together. Singh [129] found that by presenting a pair of alternating tones, each of which had a different number of harmonic components (a key feature of Timbre, see Section 2.3.2), separation of the tones into two separate streams was found to require a smaller difference in fundamental frequency between the two tones than if both tones had the same harmonics. Additionally, work by Gerth (described further in Section 4.6.1) found that the presentation of concurrently presented melodies, each with a different timbre significantly improved melody identification over all melodies being presented with the same timbre.

Proximity

Components of a sound which are close to each other are likely to be grouped together. This can occur in three main ways: frequency proximity, temporal proximity and spatial proximity.

Frequency Proximity : Frequency proximity has been shown to have an influence on streaming by several researchers, but most notably by van Noorden [139], 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 (see Figure 2.8). 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.

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 [117] as described in Deutsch [39] presented users with basic patterns of high and low pitched tones which 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, so much so that after 30 msec 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 [37]. They found that by mistiming an harmonic in a complex tone, and by moving the onset of this mistimed harmonic relative to the tone, the mistimed harmonic contributed less to the perceived pitch of the complex. When the harmonic was mistimed by 300ms they found it made no contribution to perceived pitch.

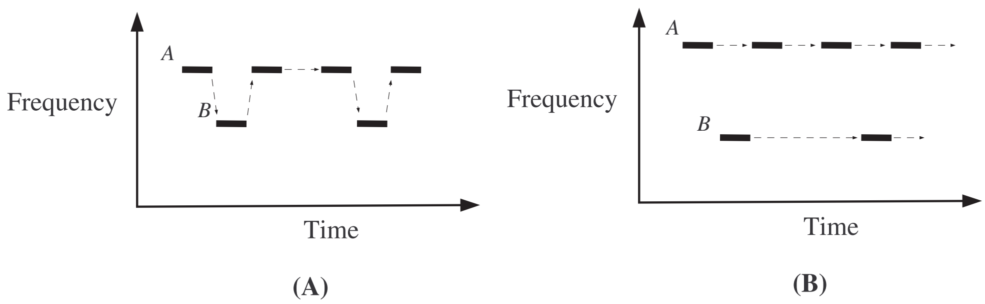

Figure 2.8: Overview of the effect of presentation speed on streaming. Taken from van Noorden [139]. (A) When tones A and B are relatively close in frequency the listener hears a galloping melody. However at larger frequency separations (B), the listener will perceive two distinct streams, one of high frequency tones, another of lower frequency tones.

Spatial Proximity : From Section 2.4, it was shown that the human auditory system is capable of determining the direction of a sound source in space. Such information is also used to assist in the auditory streaming process [5]. As Bregman [14] notes *"It would be a good bet that sounds emanating from the same location have been created by the same sound source"*. Work undertaken in the study of concurrently presented speech, and the separation of multiple talkers by spatial location has been found to improve the identification of those talkers [29, 31, 30].

Good Continuation

Sounds which continue to change in predictable ways will be more likely to be grouped together. Heisse and Miller [61] identified that if a sequence of tones which continually rose or fell in frequency was presented (a frequency glide), and one tone in the sequence was sufficiently altered in frequency, it would stand out and be separately streamed. Good continuation can also assist in cases where a sound may be partly masked. Bregman [14] identified that if a frequency glide was presented to a user which was partially masked by a broadband frequency noise (see Section 2.3.2), participants would identify one auditory stream, with the tone sequence continuing through the noise; instead of two tone sequences separated by a broadband noise.

Belongingness

This is an important property where each component can form part of only one stream at any one time [150]. 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, however it is impossible to see both at the same time (see Figure 2.9 for an example). In this property the auditory system will work to resolve any conflicts to come to a stable state [14], and having reached it, the interpretation will remain fixed until it is no longer appropriate [150]. This may mean that when sounds are concurrently presented, components of one sound may influence those of another and force those sounds to be inappropriately streamed when used in an Auditory Display.

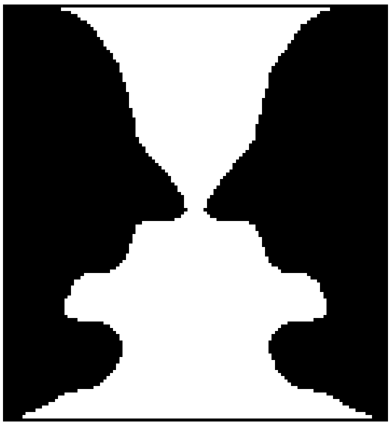

Figure 2.9: A visual example of the principle of belongingness. Whilst it is possible to see a vase or the profile of two faces, it is difficult to see both at the same time.

Familiarity

Sounds which are familiar are likely to be picked out more easily from compound sounds. This is similar to the way that a familiar voice will stand out more at a party from the other competing talkers and conversations [5]. Dowling [40], for example, found that the identification of melodies in an embedded sequence of other sounds was improved dependent on participants' prior familiarity with the melody. This indicates that learning of sounds which need to be concurrently identified is important and may improve identification.

Articulation

There is evidence to support the conclusion that Auditory Scene Analysis requires a certain amount of attention to operate. One example being the so called cocktail party effect [5, 134] which describes the phenomenon where guests at parties are able to selectively attend to different simultaneously occurring conversations (auditory streams), and switch between them based on the amount of attention given, although other factors such as those described in the previous section also have an impact. van Noorden [139] as described earlier, investigated the effect of tempo and frequency separation on tonal sequences. In his experiments participants were played a sequence of tones, which were modified in terms of the difference between their frequencies, and the repetition rate of the tones (tempo). Participants were asked two different questions. One question asked them to indicate when they could no longer hear the tonal sequence as a coherent stream, and another question which asked participants to indicate when they could include all of the tones in the same auditory stream. A graph highlighting his results is shown in Figure 2.10.

van Noorden found two limits in his experimental work which he termed the temporal coherence boundary (TCB), which is the limit above which it was not possible for participants to hear the tonal sequence as a single stream, and the fission boundary (FB), below which participants were unable to separate the tonal sequence into separate streams. The area between these boundaries allowed participants to hear either one or two streams dependent on the attention given. van Noorden's work also shows that if the sound sources

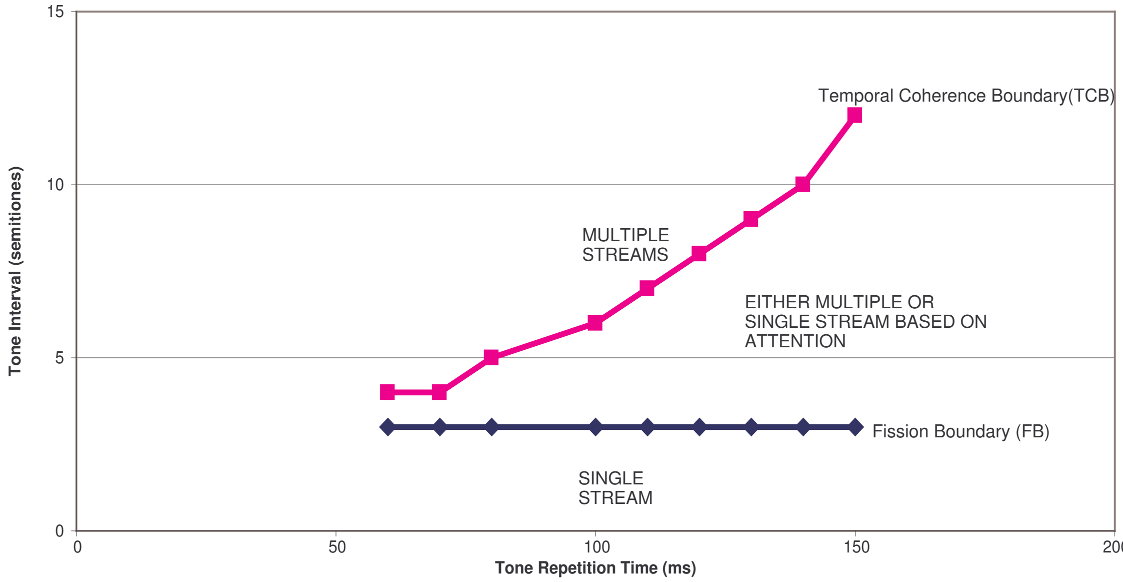

Figure 2.10: A graphical illustration of the role of attention in the work of van Noorden. The graph shows that there are two regions where the tone sequences are too different or similar to be able to be separated. However, there is a large area in which participants could apply attention to hear either precept. Slightly modified from van Noorden [139].

are sufficiently similar or different, the effect will be so strong that participants will not be able, even when trying, to control how the auditory scene is interpreted. This indicates that if sounds which are concurrently presented are desired to be streamed separately, incorporating the maximum differences between the different sounds will increase the streaming effect and listeners will require less attention to separate the different auditory sources.

2.6.3 Other Issues with Auditory Scene Analysis

Whilst all research pertaining to ASA can be fitted into one of the gestalt categories previously mentioned, and as such modifications along those dimensions will affect the perception of sound, there is still no clear understanding as to the interaction of different auditory features in the creation of an auditory stream. Do small changes to sound along some dimension dominate over larger changes in some other dimension? In many respects, the factors influencing ASA are dynamic, each exerting a gravitational like influence to pull an auditory scene into a reasonably 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.

In addition, most ASA research is undertaken on objectively measurable sound dimensions such as frequency (see Section 2.2). When higher order sound dimensions are used, these tend to be limited in some way, e.g. Singh [129] used sounds with a varying number of harmonics to investigate the impact of timbre in stream segregation. Therefore it is not possible to directly apply such ASA research to sounds with richer perceptual attributes (see Section 2.3) and assume that the same effect will exist. The influence of such features must be explicitly determined.

Because of these issues ASA can not be arbitrarily applied to sounds in order to influence their streaming, and as will be shown in Chapter 6, the sounds whose streaming is to be influenced may not allow such arbitrary application anyway. ASA research does however offer a framework to understand interactions between sounds, and to consider how the streaming of specific sounds can be influenced. This is particularly important when concurrently presented sounds are used as part of an Auditory Display since, as will be described in the next chapter, such Auditory Displays incorporate a mapping between sounds and the data they represent. If each sound is not placed into a separate stream by the ASA system, then it may become very difficult to extract the information encoded in the sounds. This, as will be discussed in Chapter 4, is a specific problem for Earcons.

2.7 Conclusions

In this chapter, sound and its interpretation by the human auditory system has been discussed. The objective physical components (frequency, amplitude etc.) from which sound is generated have been described. The interpretation of sounds by the human auditory system, which leads to further more subjective qualities of sound such as timbre, pitch etc. has been discussed. How the physical and subjective properties of sound relate, which is important in the context of this thesis in order to understand how research on concurrent sounds can be applied to improve concurrent earcon identification, has been discussed. The localisation of sound sources, and how the cues which allow this can be synthesised to create virtual acoustic displays has been described. The limitations of such virtual displays have also been discussed. Finally, the interactions which can occur between sounds when more than one is concurrently presented have been discussed. Understanding these issues is important as this thesis investigates the identification of concurrently presented earcons. Earcons are built from the attributes of sound discussed in Section 2.3, and due to their similarity along dimensions of pitch, timbre etc., and the discussion of Auditory Scene Analysis from Section 2.6, are likely to be streamed in an undesirable way when concurrently presented. The thesis will use the work on ASA and spatialisation, to redesign and present the earcons to encourage them to stream in more desirable ways. The discussions of this chapter allows for the understanding of the novel work undertaken in this thesis (the understanding and improvement in identification of concurrently presented earcons), as well as the its limitations. The following chapter demonstrates how the lower level auditory features of this chapter can, and have, been used to create "auditory displays" which can effectively communicate information through the medium of sound.

Chapter 3

Auditory Display

3.1 What is an Auditory Display?

In the previous chapter what sound is and how it is interpreted by the human auditory system was outlined. In this chapter, how such knowledge can be exploited to create auditory displays will be discussed.

Unfortunately, there is no explicit definition of an auditory display. However, for the purposes of this thesis, and based on the papers that have been published as part of the International Conference on Auditory Display (ICAD) [65], an auditory display can be considered to be the use of sound to communicate information about the state of an application or computing device to a user. Such displays have also been called auditory interfaces [54]. In the definition used here auditory displays comprise only of the use of sound to communicate from the computing device to the user. This thesis defines auditory interfaces to be a superclass of auditory displays which use sound (mainly through speech) to communicate from the user to the device (e.g. voice controlled car navigation systems [112]).

In this chapter advantages and disadvantages of using auditory displays both in isolation, and in addition to visual displays will be outlined. The major techniques by which data can be encoded into sound will be discussed, followed by the specific advantages of using spatialised sound (see Section 2.4) as part of an auditory display.

3.2 Advantages and Disadvantages of Auditory Display

3.2.1 Advantages of Auditory Display

There are several reasons why the inclusion of sound to communicate data in a computer interface is beneficial both with and without a visual interface. These reasons are outlined in the following sections.

Eyes Free Displays

One situation where there are advantages in using an auditory display, is as part of an interface for eyes busy situations, which require eyes free displays. In such situations the user is engaged in another important task whilst interacting with the computing device. Operating theatres, factory floors [68], operating controls on a car dashboard whilst driving the car [140], or in aircraft cockpits where pilots may not have sufficient visual ability to both monitor all of their instruments as well as concentrating on flying the plane [96], are all examples of situations where the computing task is important but cannot demand the user's full attention. In many respects designing interfaces for such situations poses similar problems to those of providing interfaces for the visually impaired (see next section), and hence auditory displays have similar advantages when used in such situations. As Newell [96] points out *"a particular person's capabilities are specified to a substantial degree by the environment within which he or she has to operate.*".

Whilst there are several situations in which auditory displays have been applied to eyes busy situations, one of the major areas of current research is the application of auditory displays to mobile computing devices such as PDA's (Personal Digital Assistants) and mobile telephones. When using such devices the user will be involved in activities which require a high demand of visual attention, such as walking, avoiding pedestrians, street furniture etc., which will be much more important than the computing task, and hence leave a limited amount of visual attention that can be used on the computing task [62]. These demands can be reduced by an auditory display which can allow for information presentation to be moved from the visual to the aural domain, both reducing the demands on the users' visual attention, as well as increasing the overall display space available on the device. Several systems have incorporated auditory displays into the designs for mobile devices. Sawhney and Schmandt [124], developed an audio based diary notification system (see Section 3.4). Mynatt *et al.* [94, 93], developed a status information system which used audio, as well as mobile tracking technologies to provide contextual information about other users. For example, if a user walked past an empty office they would be provided with an auditory cue to indicate how long that person had been out of their office. Using auditory cues in this way instead of visual cues allows the user to avoid interrupting their main walking task (by slowing down, due to needing to monitor the environment). Unfortunately there has been little empirical evaluation on such devices to determine the actual reduction on the user's visual attention they provide.

One study which has empirically investigated the use of sound on mobile devices was undertaken by Brewster [20]. He investigated how the addition of sound to a simple number entry task on a PDA keypad improved its usability. Simple sounds were used to indicate whether a button had been successfully pressed or not, and participants had to key in numbers whilst walking a predefined route. Brewster identified that the addition of sounds to this task allowed for significantly more numbers to be successfully entered by participants. Additionally, the size of the buttons on the visual display could be reduced when sounds were presented without a significant drop in performance. This indicates that in cases where sound is used in addition to a visual display it may allow more efficient use of that visual display to be achieved, which would be especially important when the visual display resources are limited. Brewster also found that when sounds were not present, the task was significantly more annoying than when sounds were present.

Orientation in Information

One feature of audio is its omnipresent quality. That is, unlike in a visual display were we can avoid looking at the display, we cannot "shut our ears" to avoid listening to an auditory display [68]. This makes audio advantageous as we may hear information before seeing it and thus use it to guide our eyes, or be able to "mark" interesting data in a large information space with an auditory source, which can be easily referred to whilst looking at some other area of the space [69]. The ability to encode the spatial location of a sound (see Section 2.5), allows a user's understanding of space to be exploited to help orientation in both real and virtual environments. Holland, Morse and Gedenryd [62], developed an audio GPS (global positioning system) to provide orientation and location information in real world environments. In many environments where GPS is used it may clearly be unsafe to look at a visual display, and hence audio is better suited. They found that the auditory display was useful at providing course bearing information, allowing the user to advance consistently in the correct direction without needing to avert their eyes from the terrain to look at a visual display.

Interfaces for Visually Impaired Users

The major way with which computing devices communicate with us is visually, through a visual display unit (VDU). This presents problems for blind and visually impaired people of which there are approximately 11.4 million currently resident in the United States (US) [121], and 2 million in the United Kingdom (UK) [120]. Auditory displays present a potential way for such people to interact with computers. Screen reader systems such as JAWS [49], can read information that is presented on the screen to the user, however such systems work with predominantly textual data. Mercator [91, 43] developed by Mynatt and colleagues, mapped the structure of a visual windowing interface, e.g. windows, menus etc., to a hierarchical structure, with windows at the top of the hierarchy, moving down to menus for those windows, entries in the menus etc, at lower levels in the hierarchy. Whilst speech was used, other forms of audio could also be used to present information about the current point in the hierarchy to the users. Other research has investigated how specific components of graphical displays can be made accessible to the visually impaired such as line graphs [76] (described further in Section 3.3.1), or with mathematic equations, where even small misinterpretations in the presentation of information can have a serious impact on the meaning of the equation [42].

Alerting and Alarms

Possibly the most common use of auditory displays is to alert users to problems or undesirable situations quickly. Because the ears, as already mentioned are omni-directional, the use of an auditory display to alert a user to problems is in many cases more suitable than indicating problems visually. When presenting problems visually the user might not be able to constantly monitor a visual display to identify problems, when performing surgery or operating plant machinery for example. This is a particular issue if the alarm indicates a critical issue which requires immediate attention, such as a heart rate monitor detecting that the patient's heart has stopped [45]. Guidelines for the design of warnings for use in intensive care units have been developed by Patterson, Edworthy and Shailer [108]. Amongst their recommendations were that sounds should repeat and get louder if the equipment was not attended to, as well as having an increasing gap between repetitions of the alarm sound increasing from 1-2 seconds up to 4 seconds to allow time for users to process the alarm messages.

3.2.2 Disadvantages with Auditory Displays

Whilst auditory displays have many advantages, they also suffer from various disadvantages that need to be considered when using them in practice. Some of these disadvantages taken from Kramer [68] are outlined below.

Low Spatial Fidelity

Whilst in a graphical computer display it is possible to present multiple concurrent applications close together and allow users to rapidly switch their attention between them, doing so in an auditory display is much more difficult. As already discussed in Section 2.4, whilst it is possible to discriminate that two sounds originate in different spatial locations when those sounds are separated by one degree in the azimuthal plane [88], this assumes that "free field" listening, not incorporating a synthetic auditory environment using a generalised head related transfer function (GHRTF) as might be used in an auditory display (see Section 3.4). When sounds are concurrently presented, the degree of spatial separation required increases. The degree of separation required is also dependent on the similarity of the sounds used [10]. This disadvantage can best be overcome by accepting the limitations of the human auditory system when designing auditory displays. However, when using concurrent sounds as part of an auditory display it is unclear where these limitations are, as little work or evaluation has been undertaken on auditory displays which use concurrent audio presentation. Work is therefore required to determine where these limitations are when sounds which would be used in an auditory display are concurrently presented.

Audio is Annoying

One of the claims made against the use of auditory displays are that they are annoying. Since as already explained, we cannot "listen away" to an auditory display in the same way as we can look away from a visual display, the user may become annoyed by the use of sound. However, Buxton [32] points out that we exhibit little influence in the ambient sounds of our working environment, such as air conditioning units, computers, telephones ringing etc. He further notes that sounds can either provide us with information and therefore help us in our task or can impede us in our task and can be categorised as noise. Therefore in order to reduce the annoyance of auditory displays sounds should be designed and used where they contain more information and less noise. This requires that auditory displays are designed effectively and that guidelines exist for auditory display designers to do so. This, as will be shown in Chapter 4, is a particular problem with the concurrent presentation of sounds, with little empirical evaluation of displays with such presentation being undertaken, and as such a lack of guidelines for designing concurrently present sounds effectively.

Another related issue to the annoyance of auditory displays is the volume, or loudness (see Section 2.3.2), of the sounds used. Imagine trying to work with a noisy pneumatic drill outside your window, communicating no other information than it exists. Patterson [106, 107] investigated problems with auditory warnings in aircraft cockpit environments. He found that a better safe than sorry approach was applied with warning sounds, with sounds being so loud that pilots would try to turn off any warning sound before dealing with the situation that the sound was warning about. Patterson recommended that the loudness of sounds should be reduced which would also reduce the annoyance of the sounds. Reducing the loudness of sounds works well in controlled environments like aircraft cockpits, where the loudness of ambient sound can be accurately predicted. However in other environments such as when using a mobile device, it may be difficult to predict ambient noise and as such determine the volume of an auditory display to avoid masking (see Section 2.3.2). In order to control for this, the volume of audio should be placed under the control of a user and the volumes of different sounds used in the auditory display should be kept as equal as possible since higher frequency sounds are louder at given intensities than lower frequency sounds at the same intensity (see Figure 2.2 from Section 2.3.2) [152].

Difficulties in Presenting Absolute Data

Unless speech is used in an auditory display (see Section 3.3.2), it is difficult to communicate large numbers of absolute values due to variations in the ability of different listeners' auditory systems. For example, as explained in Section 2.3.2 less than 1% of the population have perfect pitch; meaning that they can identify the pitch of a tone precisely without reference to any comparison tone [88]. The ability to localise sounds also degrades dependent on where the sound is relative to the user (see Section 2.4).

These interpretation issues reduce the accuracy with which data can be communicated through sound. This problem of auditory display is best solved by accepting the limitations of the human auditory system and taking these limitations into account when designing auditory displays [19].

Transience of Information

Because sound is defined as a waveform of varying pressure (see Section 2.2), and as such is constantly moving, sound "disappears" immediately after it has been presented [68]. This is an issue for long auditory messages such as speech (see Section 3.3.2) since human short term memory only allows 7 ± 2 chunks of information to be retained for any length of time [86]. Again this issue can best be dealt with by taking it into account when designing auditory displays to ensure that unreasonable demands are not placed on users' short term memory. Mitsopoulos and Edwards [87] have proposed that increasing the speed of playback of auditory encoded information may be effective in overcoming this problem. They proposed that Auditory Scene Analysis (ASA) could be employed in the design of user interface widgets such as checkboxes. The speed of presentation being used to allow auditory streams to form which would provide an overview of the state of the widgets, with lower speeds of presentation being used to provide more detailed information. Mitsopoulos and Edwards found this an effective way to present and overview of complex user interface widgets (such as menus) to users. Another way of overcoming the transience of information, as will be discussed in Chapter 4, is to concurrently present sounds to users, such that the need to remember large numbers of sounds is reduced, and comparisons between multiple data presented through sound can be better supported.

Lack of Orthogonality

Due to the subjective nature of sound, changing one attribute of sound may also cause the perception of other attributes of a sound to change. As shown in Section 2.3.2, loudness is dependent on the frequency of a sound. Hence by changing the frequency, the loudness of the sound will also change in sympathy. Changing the timbre will cause changes in the pitch of a sound since they are both derived in part from the frequency. This issue further contributes to the problems of presenting absolute data mentioned. Again, as with the problems of absolute data representation previously mentioned, lack of orthogonality must be accepted and taken into account in the design of auditory displays [19].

3.3 Mapping Data to Sound

In the previous section the advantages and disadvantages of auditory displays were outlined. However for an auditory display to be effective in communicating data, there must be some mapping between the data to be communicated and the sound which is presented to the user. In the same way that icons, pictures and text [128] can be used to encode information in visual displays, auditory displays have their own techniques for encoding data, which have their own advantages and disadvantages. As with visual displays, there are properties of the relationships between data and their representations that can be used to evaluate the relative advantages and disadvantages of different mapping techniques.

Linguistic mappings use a set of rules called a grammar [1] to define how symbols can be constructed, and also how they should be interpreted. Such mappings are powerful and allow for complex messages to be constructed and interpreted. Speech is an example of such a mapping system used in an auditory display. Sign mappings use a cause and effect relationship between data and representation [77]. For example

[1] Note the term grammar as used here should not be confused with the definition from Section 3.3.4 and used in the remainder of this thesis.

smoke is a sign of fire, and more smoke indicates a bigger fire. For example Mansur's [76] sound graphs used increases and decreases in pitch to indicate the rise or fall of a data variable. The remaining three mappings (Iconic, Metaphorical and Symbolic) [52] use progressively less obvious relationships between data and their representations to communicate information. With an iconic mapping, the sound directly reflects the properties of the world; a tapping noise for example, can be used to indicate the selection of a folder [52]. As such the meaning of the sound is obvious in the context of use. Metaphorical mappings use an "is like" relationship between data and sound, allowing designers to exploit users' real world knowledge of scenarios when there is no iconic mapping available. For example, Gaver's SonicFinder application (see Section 3.3.3) used a metaphorical mapping between the file copy operation and the sound of water being poured into a container. Whilst metaphoric mappings can be useful, it is important the metaphor is familiar to the user, otherwise its usefulness will be lost. Finally symbolic mappings do not rely on any pre-existing knowledge of mappings between data and sound, these mappings are to all extents arbitrary, requiring users to learn the mappings between data and sound explicitly. In the following sections the main techniques used to present information in sound as part of an auditory display will be discussed, however it should be remembered that these techniques only represent a subset of possibilities afforded by the previous discussion. It may be that other mappings, which have as yet not been considered in auditory displays may help to alleviate some of the problems as discussed in the next chapter of concurrently presented sounds in auditory display.

In auditory displays there are four main ways in which data can be encoded in sound (Sonification, Speech, Auditory Icons and Earcons). In the remainder of this section these four techniques, their advantages and disadvantages and examples of their usage will be discussed.

3.3.1 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"* [125], and can be considered from the previous discussion to be a sign system. There have been several systems which have successfully used sonification, one of the earliest being the Geiger counter developed by Hans Geiger in the early 20^{th} century [70]. The Geiger counter periodically "clicks" to inform the user as to the level of radiation in the environment. The closer together the clicks are, the greater the level of radiation. Tzelgov *et al.* [138] found that searching for radioactive sources using such an auditory display was significantly faster than when a visual display was used.

One of the major research domains for sonification however, has been the presentation of graphs to both sighted and visually impaired users. Mansur [76] pioneered the technique of using pitch to represent the *y* axis of a line graph and allowing the user to navigate the *x* axis with keys on the keyboard, such that as the user navigated along the *x* axis, the pitch representing the current value on the *y* axis would be played.

The use of such line graphs has been heavily studied [27, 144], and it has been shown that users are able to identify both key features of the graphs, as well as draw reasonable approximations of their shapes [28]. The use of auditory graphs has also been shown to be effective in the monitoring of stock market prices on mobile devices, where the continually changing prices of stock could not be continuously presented on the mobile device display [22].

Because of the temporal nature of audio, sonification allows users to get a rapid overview of data such that they can perceive trends in a way not possible with other mapping techniques such as speech (see Section 3.3.2). However, it is difficult for users to get precise values from sonification systems. Most sonification systems map data to the pitch of a sound since this is one of the few one-dimensional parameters of sound. As described in Section 2.3.2 very few people have absolute pitch, hence most of the population are only able to identify if a pitch is greater or lower than a comparison, and as shown by Neuhoff, Kramer and Wayand [95], listeners may make different judgments on the size of the difference between two pitches dependent on the first pitch being lower or higher than the second.

3.3.2 Speech

Whilst sonification is good for showing trends in large datasets, it is less useful for communicating absolute values, and therefore 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. Synthesized speech is very flexible as it can be created "on the fly" from a given text [133]. Concatenated speech from human speakers is less flexible in that all of the speech must pre-exist as audio recordings to be used, however it is generally of higher quality [133]. Speech has been used in many contexts, from screen readers for blind and visually impaired users such as JAWS [49], to telephone enquiry systems, airline cockpits and subway announcement systems [97]. 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 is mapped to sound will already be understood (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.* [111] note 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 is also problematic for the presentation of continuous data such as would visually be presented in a graph [153], as due to the speed of presentation available and the transience of an auditory display (see Section 3.2.2) it would be difficult to gain an overview of the data such as would be available with sonification (see Section 3.3.1), due to the time taken to present auditory messages and the demands on short term memory that would be required.

3.3.3 Auditory Icons

Auditory icons defined by Gaver [54] 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. Auditory icons have been successfully used in several systems. Gaver [52] developed an auditory interface, called Sonic Finder, to augment the graphical user interface (GUI) for the Apple Macintosh computer, where he used various everyday sounds to express user interaction in the interface. For example, when a folder was selected the sound of an object being struck was played, as the folder was dragged a scraping sound was used. When copying files the sound of water being poured was used, with the pitch of the sound indicating the amount of data copied. Whilst Gaver produced no empirical evaluation of Sonic Finder, informal demonstrations yielded positive comments from users.

Gaver, Smith and O'Shea [55] have also investigated how auditory icons perform when they are concurrently presented. The ARKola system simulated a cola production and bottling plant that was controlled by two users in physically separate locations. Operators were able to control the rate of several pieces of equipment such as the bottle capping machine, the bottle filling machine etc., as well as being able to fix broken equipment. Producing bottles of capped cola generated sales, whereas buying raw ingredients such as bottles, cola nuts etc., cost money. If supplies were allowed to run out or machines were run too fast (causing, for example, the capping machine being unable to cope with the rate of filled bottles arriving and causing bottles to be lost) profits fell. Machines breaking down also caused profit to fall, and raw materials to be lost if the machines were not quickly fixed. Users were instructed to make as much profit as possible, as efficiently as possible. Gaver, Smith and O'Shea implemented a set of auditory icons to represent activity and errors inside the plant. For example, the capping machine made the sound of clanking bottles to indicate that it was operational. The rhythm, or repetition rate of the sounds, indicated the speed of the machine. If a machine stopped or had broken down, the sound representing it would be stopped. Other sounds were used to indicate errors where materials were being lost, e.g. running a machine too fast would cause bottles to be lost, indicated by a breaking glass sound. Because of the number of processes involved in the factory, several sounds (up to fourteen) could be concurrently presented. Gaver took care to ensure that concurrent sounds were of equal loudness to avoid masking, which as stated in Section 2.3.2 could be a problems when sounds are concurrently presented.

ARKola indicates that concurrently presenting auditory icons to users can be effective. No formal evaluation of ARKola was performed, certainly as regards the identification of concurrently presented auditory icons, but there seem to be no adverse comments from users regarding the concurrency of audio presentation. This may seems surprising given the discussion on auditory scene analysis from Section 2.6, however as will be shown in the next section this is not as surprising as it may seem.

Designing Auditory Icons

Whilst there have been several examples of the use of auditory icons [52, 55, 2, 43], there is comparatively little known about how they should be designed, or what sounds should and should not be used to represent them. Some experiments have however been undertaken by Mynatt [91], who looked at how auditory display designers should make decisions on which sounds to use and what those sounds should represent as auditory icons. She proposed that there were four factors which affected the usability of auditory icons. These included the users' ability to identify the sounds used, how well the sound mapped onto the data that it was to represent, and the quality and other physical parameters of the sound [92]. In order for designers to deal with such factors she proposed a methodology to follow when producing auditory icons. This methodology recommended performing tests to have subjects identify sounds that might be used as auditory icons, testing to determine if the mappings chosen for those auditory icons were intuitive, determining the cohesiveness of the set of auditory icons and their mappings, and to identify any unwarranted interactions between auditory icons if they are to be presented simultaneously. On evaluating her methodology Mynatt observed that the ability of users to correctly identify 64 common sounds was poor, with only 13 of the sounds reaching 80% accuracy of identification. However, many of the participants were able to identify some aspects of the sound, e.g. the sound of a metal mailbox being closed was identified as the closing of a filing cabinet drawer. Mynatt also found that care needed to be taken to avoid similar sounds which could be confused with each other, for example the sound of a typewriter and a keyboard may be confused. As Mynatt states [91], *"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."*. Following such guidelines means that when concurrently presented, auditory icons will be quite dissimilar (according to Section 2.6), and thereby less likely to interfere with each other and be placed in the same auditory stream (see Section 2.6).

Advantages and Disadvantages of Auditory Icons

Auditory icons as implied from their definition, are everyday sounds the meaning of which should be intuitive and obvious in the context of their use. In terms of general mapping techniques they fall into the iconic and metaphorical categories. Therefore it is unnecessary to learn the mapping explicitly, in the same way as with sonification (see Section 3.3.1). Auditory icons also seem to be easily used in concurrent auditory displays such as with ARKola [55]. This can be understood in terms of Auditory Scene Analysis (ASA) (see Section 2.6), as the sounds used were very different from each other, and as Mynatt's design methodology states, similar sounds should not be used as auditory icons as they may be confused. That auditory icons should behave in such a way is not surprising since they are everyday sounds, and it is precisely these kinds of sounds our auditory system has been designed to separate [54].

Unfortunately, auditory icons do have drawbacks. Firstly, they do require an intuitive mapping (either iconic or a good metaphor) between the sound and the datum that the sound is to represent. In some cases this may not be possible [19]. Abstract operations such as renaming a file, or changing the font size of a document, are likely to prove difficult to find an appropriate sound for. Auditory icons are also difficult to parameterise to communicate more than one bit of information since sound synthesis techniques do not support modifying sounds along real world dimensions such as material, force of impact, etc. [98]. As Gaver [53] notes, standard synthesis techniques *"do not make it easy to change a sound from indicating a large wooden object, for instance, to one specifying a small metal one."*. In addition, when a sound is changed in such a way it may not be clear what that modification relates too, thereby forcing users to learn the mapping between sound and data. For example, what is the intuitive difference between a metal hitting sound and a wooden hitting sound when used in a computer interface?

3.3.4 Earcons

Earcons, as defined by Blattner *et al.* [13], are *"non-verbal audio messages used in the user-computer interface to provide information to the user about some computer object, operation, or interaction"*. Brewster [16] further refines this definition and defines earcons as *"abstract, synthetic tones that can be used in structured combinations to create auditory messages"*. 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 musical like sound and data. This gives earcons the advantage of being able to represent any event, object, or interaction in a human computer interface, the disadvantage being that the association between the sound and what it represents (unlike auditory icons) must, at least initially, be explicitly learned. Earcons being in the symbolic, rather than iconic category of mapping. There are four types of earcon (one-element, compound, hierarchical and transformational) [13] which are described below.

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 [13]. 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. Such earcons are in many ways analogous to the "new text message arrival" sound on mobile telephones. Whilst there may be different sounds to indicate messages from different types of people, work, home etc., there is no structured relationship between the sounds, each is unique and its meaning must be individually learned. For large datasets, or in cases where more than one parameter of the data must be communicated, the number of sounds, and mappings of data to those sounds that must be learned, could become extremely large. The following three types of earcon attempt to provide solutions to such situations.

Compound Earcons

Compound earcons are formed by concatenating one-element, or 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 in combinations to form earcons for the “open file” and “save file” operations [24].

When using compound earcons it is important to consider earcon length as the messages can easily become too long to be usable. Brewster [16], who has performed extensive work on the usability of both compound and hierarchical earcons, found that even short compound earcons such as those mentioned earlier could take up to 2.6 seconds to be played, which may be too long in many situations. Brewster, Wright and Edwards [26] showed that the use of compound parallel earcons, where each earcon part is played concurrently in opposite ears, could be effective in reducing the time taken to play compound earcons without reducing their intelligibility (see Section 4.2.2).

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 un-pitched rhythm might represent an error, the next level will alter the pitch of that rhythm to represent the type of error etc. This is summarised in Figure 3.1, taken from Blattner *et al.* [13].

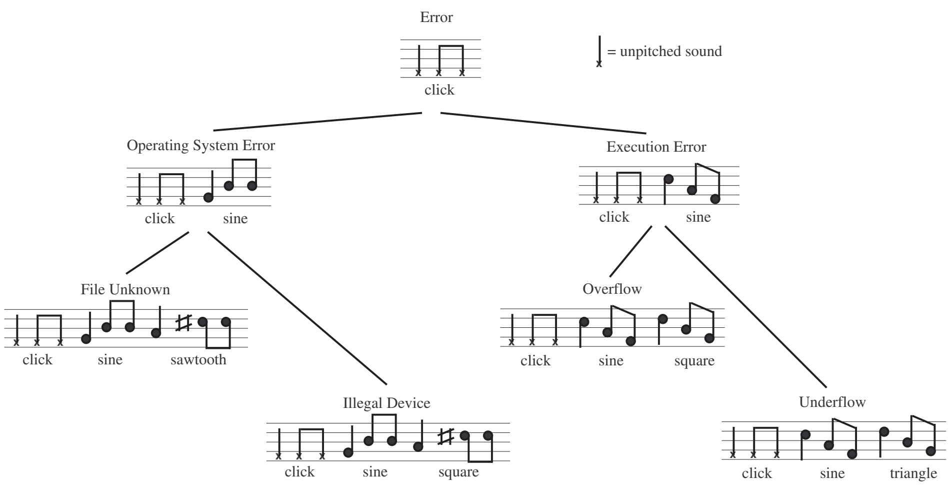

Figure 3.1: An overview of the “grammar” used to construct a set of hierarchical earcons representing computer error messages. Taken from Blattner et al. [13].

Work by Leplâtre and Brewster [72, 73] has shown that hierarchical earcons can be used to significantly improve interaction in mobile telephone based menus. Such menus are commonly difficult to interact with due to the lack of context over where a user is in the menu structure at a particular time [73]. Leplâtre and Brewster implemented a set of hierarchical earcons to deal with this problem, and evaluated mobile telephone tasks on a simulated mobile telephone, comparing a visual only menu condition and a visual plus auditory menu condition, which used earcons as well as the visual interface to provide the users with context in the menu. They found that the number of keystrokes required to successfully complete a task was significantly reduced in the visual plus auditory menu condition over the visual only menu condition. Unfortunately, due to technical constraints they did not evaluate the visual plus auditory menu condition on a real mobile device to determine the impact of the earcons in a more ecologically valid context. However, their work does indicate, as discussed in Section 3.2.1, that sound can be a useful addition to devices where visual display space is limited.

Transformational Earcons

Transformational earcons are similar to hierarchical earcons in that they are constructed around a "grammar", where there exists a consistent set of structured symbolic mappings from individual data parameters (such as file type) to individual sound parameters (such as timbre). Specific values of data parameters (e.g. a paint file) are then mapped to specific values of the corresponding auditory attribute (e.g. a piano timbre). See Section 8.3.5 for an example grammar of transformational earcons. Due to the consistency in the mappings used, a relatively large set of complex auditory messages can be represented by a relatively small set of rules making learning of those messages easier for users to undertake. This is advantageous over having to learn each individual mapping between data and sound explicitly, such as with compound earcons (see Section 3.3.4); it is only necessary to learn the rules by which earcons are constructed and the auditory parameters used, in order to understand the earcons, unlike one-element earcons where each data item has its own sound, without any structured relationship between all of the sounds and the data they represent. In the remainder of this thesis, wherever the term grammar is used it shall refer to such a system and should not be confused with more complex linguistic definitions such as used in speech. Transformational earcons have been less studied, however they share many similarities with hierarchical earcons, and as Blattner *et al.* [13] note, their principles may be used to shorten hierarchical earcons for "expert users". How this is done is shown in Figure 3.2, taken from Blattner *et al.* [13]. Because earcons are constructed around a grammar, and it is that which provides their ease of learning, earcons constructed from the same grammar may be quite similar to each other according to Section 2.6 (sharing the same timbre, pitch, rhythm etc.), and as such when concurrently presented may interfere with each other in a way that auditory icons appear not to. Recall from Section 3.3.3, that auditory icons used in the same auditory display should be different from each other, and as such when concurrently presented are more likely to stream correctly (see Section 2.6).

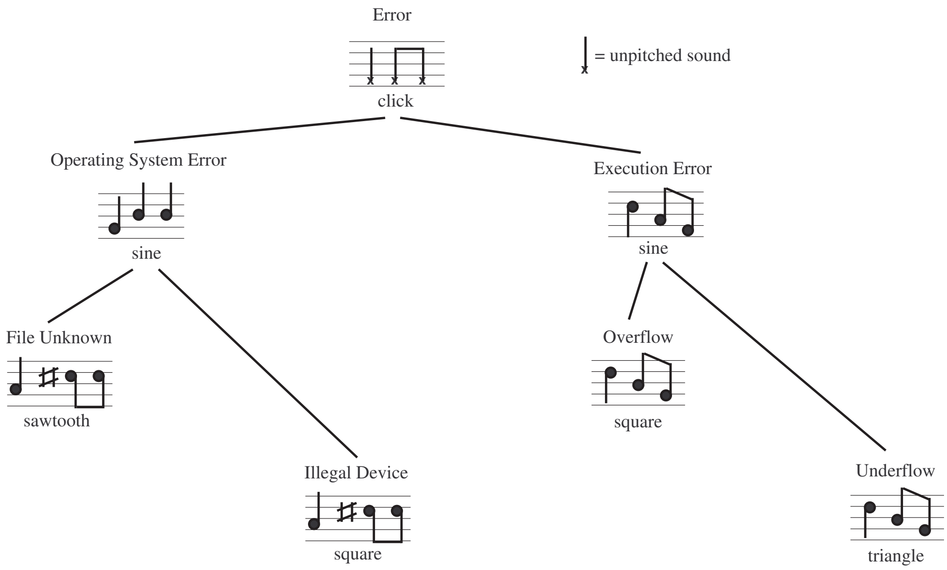

Figure 3.2: A shortening of hierarchical earcons from Figure 3.1, showing the similarities between the transformational and hierarchical earcon types. Taken from Blattner et al. [13].

Designing Earcons

In the previous sections the various different kinds of earcons that can be used to convey data in an auditory display were outlined. However, how can earcons be designed to accurately and unambiguously convey the information that is encoded in them to the user? For example, how different do two timbres need to be in order to be identifiable as representing different attributes, or how much should two pitches differ in order not to be confused? Brewster [16] 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 since, compound earcons are generally formed from one-element earcons and, as already stated in Section 3.3.4, transformational earcons are similar to hierarchical earcons. Brewster's work lead to a set of guidelines [25] for how auditory attributes should be used to develop earcons which are more easily understood by users. These guidelines are outlined in Table 3.1.

Advantages and Disadvantages of Earcons

Earcons have been shown to be useful in communicating the same data as auditory icons [23, 21, 17] as part of an auditory display. However, where as auditory icons have iconic or metaphorical mappings which rely on intuitive links between data and everyday sounds, earcons use symbolic mappings which require explicit learning.

Brewster [19], has proposed that auditory icons and earcons are at the extreme ends of a continuum from representational mappings to abstract mappings. This is consistent with the discussion of metaphorical,

Timbre	Musical instrument timbres that are subjectively easy to tell apart should be used. For example use MIDI patches brass and organ instead of brass 1 and brass 2. It is important that care is taken so that the timbres chosen can play at the registers chosen, since not all instruments can play all pitches.
Rhythm	Rhythms being used should be as different as possible. Using different numbers of notes can be used to do this effectively. However care must be taken to ensure that earcons are short enough to keep up with interactions in a computer interface. In order 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	The use of complex intra-earcon pitch structures can be effective in differentiating earcons if used with another attribute such as rhythm.
Register	If absolute identification of register is to be made then it should be used with care, with gross differences between the registers used.

Table 3.1: Table describing the main guidelines of earcon design from Brewster, Wright and Edwards (BWE) [25].

iconic and symbolic mappings from Section 3.3. Figure 3.3 shows how earcons and auditory icons can be considered in terms of these mapping techniques.

Given the clear relationship between auditory icons and earcons the advantages of earcons are really the disadvantages of auditory icons. Earcons are easy to create and parameterise, it is relatively easy to use tools to create and manipulate earcons along musical dimensions such as pitch and timbre, unlike auditory icons which require to be manipulated along the dimensions of natural sounds [53]. Since the relationship between the data to be communicated and the sound is abstract as opposed to representational, earcons can represent any data that may be required for them to communicate, as opposed to auditory icons which require an intuitive link between the data and the sound used to represent that data.

The disadvantages of earcons are also the advantages of auditory icons. Because the mapping between earcons and the data they represent is abstract, it must be learned. This requires the user to invest time to learn the mapping, something which is not required with auditory icons, where the relationship between the data and the sound is intuitive. However as indicated by Figure 3.3 it may be possible to use more metaphorical earcons, for example using a melody which decreases in loudness to indicate file deletion [52]. Such earcons have not yet been explicitly researched so it remains unclear what their effectiveness would be. Also, as discussed in Section 3.3.3, auditory icons can easily be used when more than one is concurrently presented. As will be shown in Chapter 4, this is problematic for earcons. Because of this complementary relationship between earcons and auditory icons it is likely that designers may wish to use both as part of an auditory display.

3.4 Use of 3D Sound in Auditory Display

In Section 3.3 the main four techniques (auditory icons, speech, sonification and earcons) which can be used to map data to sound in an auditory display were introduced. In this section spatialised audio, which has recently become available to auditory display designers, will be discussed. Spatialised or 3D sound has

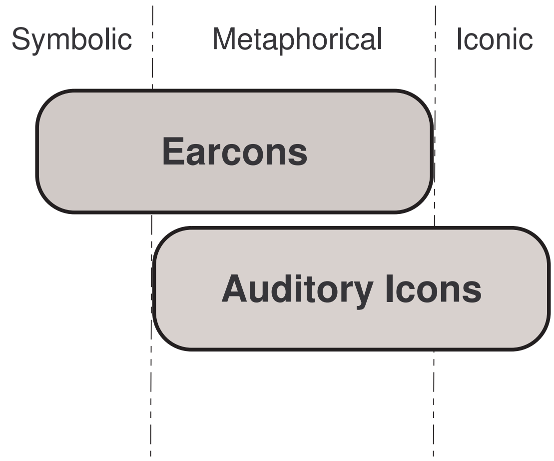

Figure 3.3: A diagrammatic representation of the mapping techniques that can be used between data and sound for both auditory icons and earcons.

already been outlined in Section 2.5; in this section the use of such technologies to create auditory displays will be explored.

There are two main ways in which the use of spatial audio can provide advantages in auditory displays. Firstly, the ability to position sounds in space can be used to encode a further parameter of data. Hence instead of using just melody, timbre, pitch etc., to encode information, the position of the sound can also be used [146]. This has advantages since the spatial location of objects is remembered even if there is no explicit determination to do so, such information tends to be learned automatically [75].

Walker *et al.* [143] compared the recall of appointments in a simple diary when that diary was presented either visually, or as synthesised speech with the spatial position that the speech was presented at encoding the time of the appointment. In the spatial audio condition, a 2D audio plane (only azimuth was used, all sounds had the same elevation) located laterally with the user's ears was mapped to a clock face metaphor, twelve o'clock being in front of the user, six o'clock behind the user and nine and three o'clock to the left and right of the user. In the visual condition, each day's appointments were presented as text which the user could scroll through using a standard scroll bar. Participants were presented with a day's appointments (one each at twelve, three, six and nine o'clock). Participants were asked to provide both relative (*"Did A occur before or after B?"*) and absolute (*"What occurred at X time?"*), information about the appointments. In the spatial audio condition each day's appointments were presented only once, in the visual condition participants were given eight seconds to scroll through the display (eight seconds being about the same time as the audio took to be presented).

Walker *et al.* found that the audio condition significantly improved the recall of both absolute and relative appointments, and many of the participants commented that the time of appointments came "for free" in the audio condition due to the spatial encoding of appointment time. Whilst the use of spatial location of an auditory source to encode data about it can be useful, in Walker *et al's.* system it is unclear how performance would be affected by the incorporation of more fine grained appointments, due to the low fidelity of spatial position encoding at certain positions around the head (see Section 2.4.2). Additionally, the amount of time taken to present the diary (eight seconds) may be too long and frustrating to users, especially if a user wants to be reminded what they are doing at 9PM and doesn't want to listen to all of the previous appointments first. If entries in the diary were concurrently presented this might reduce such problems.

The other advantage of spatial audio in auditory displays is related, and afforded by the first advantage. As has been shown in Section 2.6, an important part of Auditory Scene Analysis [14], is that two sounds coming from different places are more easily segregated by the auditory system than two sounds which originate from the same spatial location. This means that when using a spatial auditory display, it is possible to simultaneously present auditory information to the user and therefore increase the bandwidth of the auditory channel [146]. Whilst some systems, such as ARKola [55], have used concurrent sounds in a non-spatialised environment, such systems use auditory icons, which by design are dissimilar [91] according to the taxonomy of Section 2.6. Spatialised presentation has been shown to assist identification of concurrently presented sounds which may be much more similar than auditory icons.

Cohen and Ludwig [34] proposed that spatialisation could be used to effectively present both an audio implementation of a graphical user interface and to present a conference calling system. They noted that spatialisation could help to separate concurrently presented sounds and proposed a number of gestures that could be used to grab and move auditory sources in space to increase the separation between concurrent sounds, which should improve the identification of the sounds. Unfortunately no evaluation of these systems was carried out to determine their effectiveness.

Brewster, Wright and Edwards [26], who identified that compound earcons (see Section 3.3.4), which can take a long time to be played to a user, could be split such that each component earcon from which the compound earcon was composed could be presented concurrently to opposite ears. This reduced the time taken for the earcons to be presented, and allowed them to more effectively keep pace with events in the computer interface. These parallel earcons were evaluated by Brewster, Wright and Edwards [26] who found that although information was being presented to the user at a faster rate, identification of earcons was not significantly impaired when compared to sequentially presented compound earcons. Although Brewster, Wright and Edwards used the extremities of spatialisation, with one sound being presented to each ear, they did show that concurrent presentation of auditory information using spatial separation can be effective. Unfortunately, they did not investigate if parallel earcons significantly increased the demands on the users who were identifying them, which may be the case, due to an increase in the presentation rate of the data.

Finally of course, it is possible to combine both advantages of spatial audio, such that the spatial location of a sound encodes some data attribute, and there is more than one sound concurrently presented. Kobayashi and Schmandt [67] developed a browsing tool for speech based audio called Dynamic Soundscape. As with the work by Walker and Brewster [142], azimuth was used to encode a data parameter, in this case the current time counter of a speech based audio recording. As the recording was played it moved around the listener's head 1.2^o every second. The sound files used by the system were broken down into topics, as might be the case in news reports, which assisted the user in mapping a portion of the recording to a specific location (see Figure 3.4).

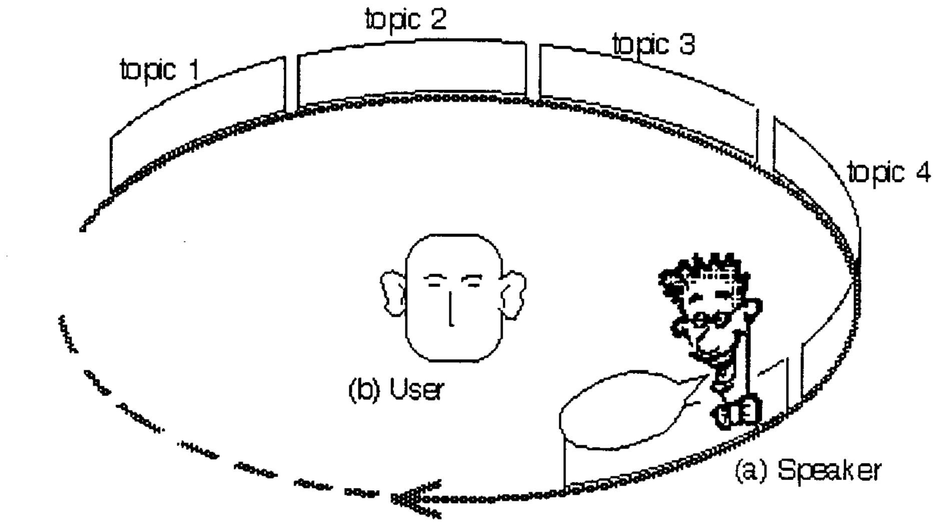

Figure 3.4: An illustration of Dynamic Soundscape showing how audio was mapped to time. Taken from Kobayashi and Schmandt [67].

If a listener found a topic that they were particularly interested in, or later wished to review information contained in a prior topic, they could use a knob or touchpad to select that topic, whereupon a new "speaker" would be spawned at that location (if one did not already exist) and would start to "read" the recording from that location. If an additional speaker already existed, this speaker could be selected, or "grabbed", by the user, and moved to the appropriate location. The original speaker would continue to carry on albeit at a reduced loudness (see Figure 3.5), allowing the user to monitor that speaker for any new relevant information by making use of the cocktail party effect discussed in Section 2.6.

Whilst Kobayashi and Schmandt performed an informal evaluation of their system in terms of the best way to select speakers, and how fast the audio should move around the user, they performed no evaluation studies of the usefulness of their system in terms of its suitability for those who need to browse large collections of speech based audio, such as those who produce transcripts from taped conversations etc.

Whilst the previous systems have dealt with the use of only one type of audio being concurrently presented at a time (e.g. speech, earcons etc.), it is possible to successfully combine different types of sounds into a spatialised auditory display as shown by Sawhney and Schmandt's Nomadic Radio system [123, 124]. Nomadic radio was a wearable computing system which acted as a personal notification system, informing the user of upcoming appointments, incoming email or text messages, as well as news items. As well as an auditory display, users interacted with the device through speech. Nomadic radio, as with the work de-

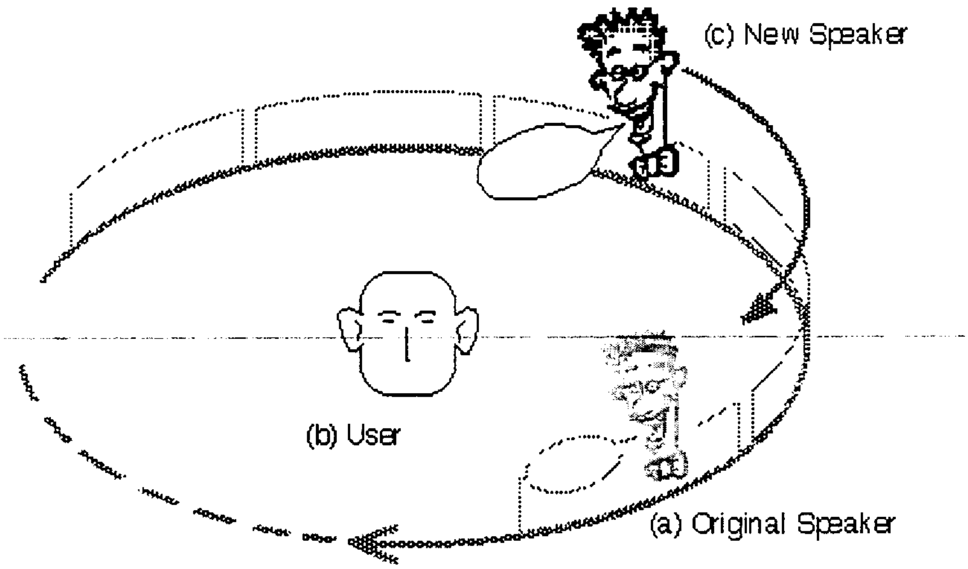

Figure 3.5: An illustration of Dynamic Soundscape showing a new speaker, the original speaker is represented as grey to indicate that its volume has been reduced. Taken from Kobayashi and Schmandt [67].

scribed previously on Walker *et al's* spatialised diary [141], used a clock metaphor encoded by azimuth to denote the time of arrival of messages as well as the position to play upcoming appointments from. When the system was activated there was the constant sound of running water at a low volume, when mail or other messages arrived they were represented as faster flowing water, the length of time the water flowed faster, the greater the amount of data being downloaded (and hence the length of the message). When messages arrived, the system used rules which it had been both explicitly told by the user as well as those learned from past experience, to determine how to notify the user of the incoming message. Depending on the priority of an arriving email message, the system could categorise the message and play an auditory cue (assigned by the user) dependent on the message's category. If the message was rated as important, or the user was not involved in an important task, nomadic radio could provide a spoken summary of the first 100 characters of the message [123]. Similar techniques existed for other types of messages, a telephone ringing for example, was used for currently low importance voice messages, and a short excerpt of a radio station's signature tune "jingle" was used to indicate a news summary. If a message was determined to be important enough by the user they could elect to play it in full.

Unfortunately, as with most systems which present concurrent data via a spatialised auditory display, Nomadic Radio was not fully evaluated. 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. This makes it difficult for designers to design systems which exploit such advantages of spatialised sound presentation. It is unclear therefore how many sounds can be presented, or how far apart those sounds need to be. How similar can those sounds be and can they be redesigned to improve identification. Without access to such information, using concurrent spatialised sound presentation may simply lead to an increase in the annoyance of the display (see Section 3.2.2). Lack of guidance in exploiting the concurrency of spatialised environments is unfortunately, as will be shown in the next chapter, a problem with all forms of concurrent auditory display.

3.5 Conclusions

This chapter has shown how research on sound introduced in Chapter 2 can be used to construct information displays that use sound to effectively communicate information to a user. Such auditory displays can be used to create effective user interfaces for mobile environments and for situations where users are visually impaired in some way, either temporarily such as with driving [140] or more permanently such as with blindness [41]. They can also be used to orient users in information and provide warning alerts and alarms [62, 108] . However, such auditory displays suffer several problems, some, such as presenting absolute data can be overcome by accepting such limitations in the design of audio. Others such as transience of data, may be overcome by the concurrent presentation of sound, allowing a reduction of the demands on users to remember information. Another more recent feature is synthetic spatialised sound presentation, which allows for the location of a sound to be used to map a data parameter, as well as increase the amount of information to a user, which may assist in the effective presentation of concurrent sound. Whilst systems which use spatialisation to increase the data presented to users have generally been shown to be useful, there has been very little empirical evaluation carried out on them. This makes it difficult to extract guidelines for future designers to build such displays and avoid problems with concurrently presented sounds, as discussed in Section 2.6. As will be shown in the next chapter, there is a great lack of design guidelines for concurrent auditory displays in general, which is likely to be a bigger problem for sounds which are similar to each other, such as earcons, even if they are spatially presented. If design guidelines for such displays are not identified, displays which use such concurrent presentation are likely to be annoying and of little use to users.

Chapter 4

Issues Involving Concurrent Audio Presentation

4.1 Introduction

The previous chapter investigated how sound could be used in human computer interfaces to effectively communicate information to users. In this chapter a smaller aspect of such auditory displays will be investigated: concurrent presentation of multiple auditory sources. This has already been mentioned in Chapter 3 through systems such as the personal messaging centre, Nomadic Radio [124], and the ARKola [52] bottling plant simulation. For the purposes of this thesis however, it is necessary to understand more about why the presentation of multiple audio sources may be useful, as well as some of the associated problems.

4.2 Why use Concurrent Audio Presentation?

4.2.1 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 [104], 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 [55] bottling plant simulator, and Nomadic Radio [123, 124], 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.* [143] (described in Section 3.4), 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 spatial proximity and articulation features of Auditory Scene Analysis (see Section 2.6.2), 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 described in Section 3.2.1 are claimed to be a key application area for auditory displays. Mobile devices suffer from both small screen displays [142], and that users may not be able to constantly attend to that visual display in the same way as with a desktop computer [62]. 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.

4.2.2 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.

Brewster, Wright and Edwards [26], as discussed in Section 3.4, 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 [116]) 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 be much more likely to require a shortening to their presentation time.

4.2.3 Browsing Multiple Data

One of the drawbacks of auditory displays is their temporal nature (see Section 3.2.2). This can make it difficult to make comparisons between multiple data in an auditory display, for example determining relationships between two sonified graphs [27]. 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 [13] 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, 134] (see Section 2.6) has long been an interesting problem for psychologists seeking to explain how the human auditory system works.

Brown *et al.* [27] investigated if users could identify key features of two concurrently presented sound graphs [76]. 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 (see Section 2.6). 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 [46] 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 [1]. 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 [14]. Unfortunately, Fernström and Bannon did not perform a user evaluation of their system. Sonic Browser does however 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 [46]. Similar advantages are apparent in the concurrent presentation of speech based audio, such as Kobayashi and Schmandt's Dynamic Soundscape [67] as described in Section 3.4, which used 3D sound presentation to allow users to simultaneously browse and monitor multiple parts of the same audio recording.

Hankinson and Edwards [59] 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. This work is an example of using the interactions that can occur with ASA, that this thesis seeks to avoid. Because Hankinson and Edwards used one-element earcons (see Section 3.3.4), they could introduce large modifications between the auditory attributes used for each earcon to avoid streaming problems, which as will be shown later, are a problem with hierarchical and transformational earcons.

4.3 Problems with Concurrent Audio Presentation

As described in the previous section and through the examples of Section 3.4 , there are clear advantages in using concurrent audio presentation as part of an auditory display. However, there are also problems which must be considered.

The major issue is that sounds which are concurrently presented may interfere with each other in undesirable ways. As Norman [100] notes on the interference of multiple warning alarms, *"they often conflict, and the resulting cacophony is distracting enough to hamper performance"*. As shown in Section 2.3.2, sounds which are concurrently presented, where one sound is a harmonic of another will interfere, such that the sound which is the harmonic will not be heard. In addition, one sound may be presented at a particular intensity to mask, or hide, another sound which is of a similar frequency and of lower intensity. Such simple interference can be easily controlled, however as discussed in Section 2.6 sounds may interfere with each other in more complex ways. These interferences can be difficult to predict and control.

Whilst there have been several studies which have used concurrently presented audio in some form or another [55, 124, 27, 26, 67, 4], there is very little guidance for designers of auditory displays who wish to exploit some of the advantages of concurrent audio presentation, to design sounds which will not interfere with each other. This may be a particular problem for the more powerful types of Earcon (transformational and hierarchical) (see Section 3.3.4), since any two earcons formed from the same grammar are likely to share certain attributes, such as timbre or rhythm, which would increase the likelihood that they would be placed in the same auditory stream, causing the interpretation of those earcons to become problematic. It is not possible to make arbitrary changes to earcons to increase differences between them and thus avoid such interferences, as there is a mapping between a data item and the sound which represents it (see Section 3.3.4). So if, for example, the Save operation is mapped to a melody played with a piano timbre, and two Save operations were played at the same time, it would not be possible to arbitrarily change the timbre of one earcon, as this would destroy the grammar which defines the mapping between data and sound, making it impossible to interpret the data encoded in an earcon. It may be possible to modify the design and presentation of the earcons to make them less likely to interfere, and as such increase their identification when concurrently presented. However, such modifications are likely to be constrained due to the need to preserve the earcon grammar. Since there are no guidelines for designers to use when designing displays which use concurrent earcons, it is unclear what modifications to earcon design and presentation will reduce undesirable streaming, or the degree of such modifications that will improve earcon streaming without destroying the grammar, both between earcons and the data they represent, which makes them powerful communicating sounds.

As a way of highlighting the problem of concurrent audio presentation, the next section describes an investigation into the use of concurrent earcons as part of a system designed to overcome the restricted visual displays of mobile computing devices. This study acted as a motivational exercise for this thesis, and indicates that when concurrently presented, earcons will interfere, causing their meaning to be lost or confused.

4.4 Dolphin

4.4.1 Overview

Dolphin was an attempt to overcome the visual display limitations of mobile devices (see Section 3.2.1) by augmenting the visual display with a spatialised auditory display (see Section 3.4). The technique used borrows many of the ideas from focus and context visualisation research [50, 51, 132]. In focus and context visualisation, the data which are most important (e.g. the procedure currently being debugged etc.) are given a large amount of the available display resource. The rest of the data is reduced in size in some way and is thus presented in less detail, allowing the user to retain a sense of orientation in large information spaces. As described in Section 3.2.1 providing a user with a sense of orientation in a large information space is one of the advantages of auditory displays.

In Dolphin, the visual display of the mobile device was used to represent the focus, and a spatialised auditory display was used to represent the context. Dolphin was used to display maps of theme parks (since this is an environment that users may be unfamiliar with, yet wish to navigate quickly and easily), using icons to represent theme park rides visually, and transformational earcons to represent rides in the auditory display. It was hoped that Dolphin would be the first step of a generic map based navigation system that would be initially applied to theme parks, hence this particular data mapping. However as the mapping between data and sound is abstract, the earcons could equally well represent any data values (see Chapter 8 for an example). The earcons were spatialised using a Generalised Head Related Transfer Function (GHRTF) (see Section 2.5) as found on the PURE Digital Sonic Fury soundcard [115]. Unfortunately no device was available to allow for dynamic respatialisation of the earcons based on listeners' head movements (see Section 2.5). Due to the problems of GHRTFs as outlined in Section 2.5.2, elevation was not used as a parameter to encode information.

In Dolphin the focus essentially "floated" over the context and users saw the focus on the PDA screen. The data which were to the right and forward of the focus were "played" in the auditory display to the right and forward of the user. The data which were to the left and rear of the focus were played to the left and rear of the user (see Figure 4.1).

Users navigated through the map via scrollbars on the visual display. The act of moving a part of the display from the focus to the context actually involved moving map items from the visual to the audio modality. When this occurred the visual representation of the map item (icon) was replaced with a spatialised audio representation (earcon). For example, scrolling to the right caused the left part of the focus to move from the visual display to the auditory display (and hence move from the focus to the context). In essence, a "lens" (the visual display) moved over a large information space. The data that the visual display

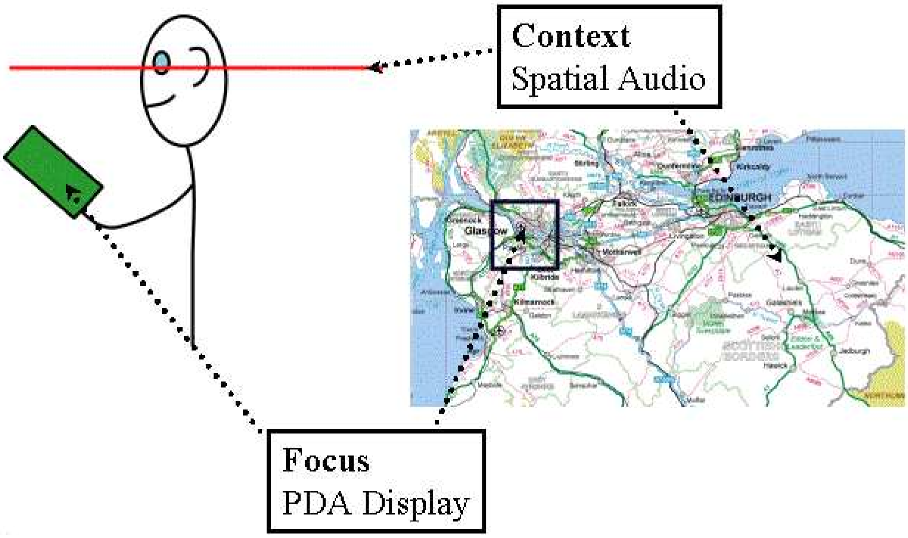

Figure 4.1: An overview of the Dolphin system showing how the focus and context were mapped to different modalities.

is over are represented visually; the rest of the information space is represented in audio.

Audio representations of map items remain the same relative distance from each other as when displayed on the visual display. This means there was limited spatial separation between concurrently presented sources, which may reduce the ability of spatial location to promote separate streaming between concurrently presented earcons (see Section 2.6), but does highlight the point that if a parameter is mapped to the spatial position of a sound as part of an auditory display, and sounds are concurrently presented, it may not be possible to ensure that sounds are spatially separated to ensure they can be separately streamed.

4.4.2 Cluttered Audio Spaces

One issue with Dolphin was that there was a much greater amount of information to be presented in the auditory display (context) than on the visual display (focus). For example, if a particular theme park map contained twenty seven rides, and therefore twenty seven items which would be represented either visually or aurally; the visual display would only be able to display around four rides at any time, leaving twenty three rides to be presented in audio. Twenty three concurrently presented earcons would clearly cause the user to become overloaded, and it was clear during pilot testing, that some way to reduce the amount of concurrently presented audio whilst still retaining the ability of the audio to present contextual information was important.

A technique called "priority zones" was developed to provide a framework for the rule-based reduction of the number of concurrently presented Earcons. Priority zones borrow many of the ideas of the Degree of Interest (DOI) function of Furnas's original fisheye concept [51]. The idea is that less important things, that are far away, should be given less display resource than closer, more important things. Far away, but

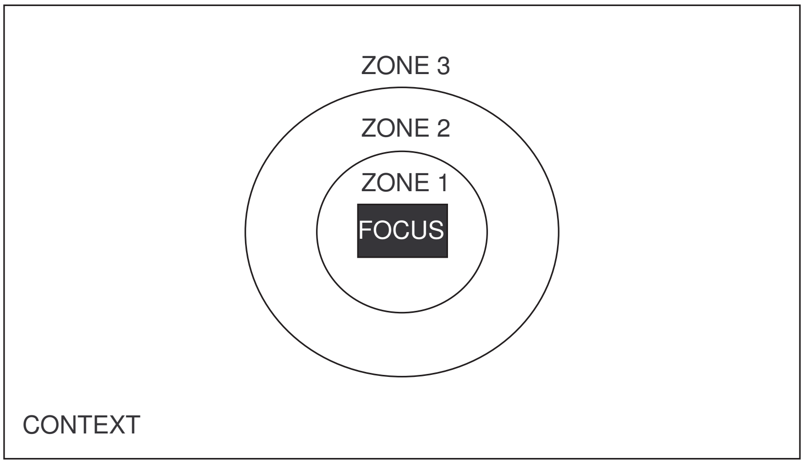

Figure 4.2: An example of priority zones used in Dolphin.

very important things, should have more resource than very unimportant but close things. It is simple in the visual domain, to determine what is meant by using “less resource” to display information; the size of the visual icon is reduced. In the audio domain determining what “less resource” means is more difficult. The technique employed in Sawhney and Schmandt’s Nomadic Radio personal notification system [124] was considered. There, more important messages were played using more detailed audio means. For example, for low priority messages, auditory icons were used, whereas for high importance messages, speech was used (see Section 3.4 for a more detailed discussion of Nomadic Radio). This approach was rejected since sounds would dynamically change their representations at different distances from the user which may become confusing. Reducing the volume at which an Earcon was presented was also considered. This is a direct analogy with the reduction of a visual stimuli size. However, the volume of a sound is an important cue to its distance (see Section 2.3.2), particularly when the object the sound is generated from is not familiar. Reducing the volume is therefore likely to confuse the user over the distances of objects. In addition, due to the similar properties of each earcon, such as similar registers, and therefore frequency, reducing the volume of one earcon may perceptually mask it when presented concurrently with a similar, louder earcon (see Section 2.3.2). Because of these issues, a more extreme solution to the problem of audio overload, which was to completely switch off audio that is not required, was employed.

In Dolphin, each of the rides (represented by an Earcon in the auditory display) is given a priority number between 1 and 3 which specifies its “importance”. The lower the number, the less important the ride. Numbers were allocated based on the highest value of the cost and intensity attribute of the ride (see Section 4.5.1 for a description of the attributes of each ride). Therefore a low cost, low intensity ride would be allocated a priority number of 1, whereas a low cost, high intensity ride would be allocated a priority of 3. Extending out from the focus, and fixed relative to it, in concentric circles are the priority zones (see Figure 4.2).

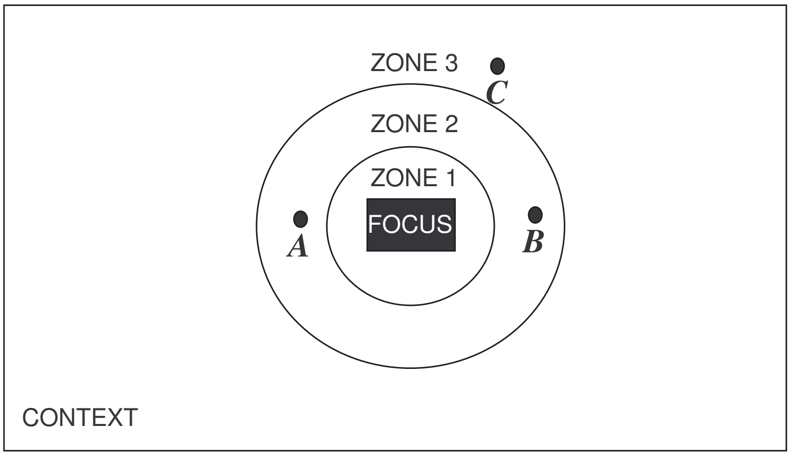

Figure 4.3: Example of an audio space with three Earcons, labelled A, B and C.

For a sound (representing a ride) to be played, it must lie in a priority zone with a number less than or equal to its priority number. This means that sounds are switched on and off dynamically as they move between zones. In doing so those audio sources that are unlikely to be important based on the user's current map location can be removed. For example, Figure 4.3 represents the 2D planar audio space for a particular map. The focus (which is represented visually on a mobile device display) is at the centre.

This particular map contains three Earcons, A, B and C. Earcon A represents a low intensity, low cost ride, Earcon B a medium intensity, low cost ride, and Earcon C a low intensity, high cost ride. According to the previously outlined system for allocating priority numbers, Earcon A will have a priority number of 1, Earcon B will have a priority number of 2 and Earcon C will have the priority number of 3. Therefore in this map Earcons B and C will be audible to the user since they are lying in a priority zone with a number less than or equal to their own. Earcon A lies in priority zone 2, and since it has a priority number of 1, will not be played. Figure 4.4 shows the same map after the user has moved the focus position by scrolling the visual display. As the priority zones are fixed relative to the focus they also move. Here, Earcon A will be played as it has moved from priority zone 2 to priority zone 1. However Earcon B has moved from priority zone 2 to priority zone 3 and will stop being played. Earcon C has not switched zones so will continue to be played.

4.5 Evaluation of Dolphin

4.5.1 Procedure

To determine if Dolphin would be a useful technique for displaying large information spaces on restrictive displays, an evaluation which compared participants' abilities to create shortest length (minimum) tours

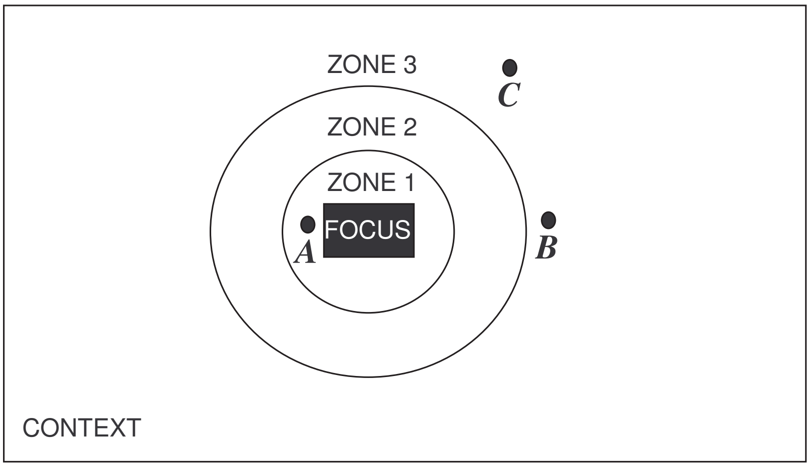

Figure 4.4: Example of the audio space from Figure 4.3, shown after the user has scrolled the visual display, and thus moved the focus.

around given types of theme park rides was carried out. To provide a comparison in performance, Dolphin was compared to a standard scrolling display. This can be considered to be Dolphin, albeit with all audio switched off. Such a comparison is useful since scrolling displays are a standard way to present large information spaces on small displays.

Sixteen participants undertook the experiment, which was of a within groups design. All participants were undergraduate students at Glasgow University, and ranged in age from 18-24, each was paid £5 on completion. There were two conditions in the experiment, the *standard scrolling* condition where participants navigated the map using scrollbars at the right and bottom of a 6x6 cm dialogue box, which simulated a Personal Digital Assistant's (PDA's) visual display (see Figure 4.5 for a screenshot of this display), and the *Dolphin* condition where participants navigated theme park maps using the Dolphin system as previously discussed. The visual interface of which was the same as that used for the standard scrolling condition (see Figure 4.5). Participants were randomly assigned to one of two groups to determine the order in which they undertook the conditions. Each condition consisted of two parts, a training phase and a testing phase. The order in which conditions were undertaken by each group is summarised in Table 4.1.

	First Training Session	**First Testing Session**	**Second Training Session**	**Second Testing Session**
Group 1	Standard Scrolling Condition	Standard Scrolling Condition	Dolphin Condition	Dolphin Condition
Group 2	Dolphin Condition	Dolphin Condition	Standard Scrolling Condition	Standard Scrolling Condition

Table 4.1: Table showing the procedure for the two groups of participants undertaking the experiment.

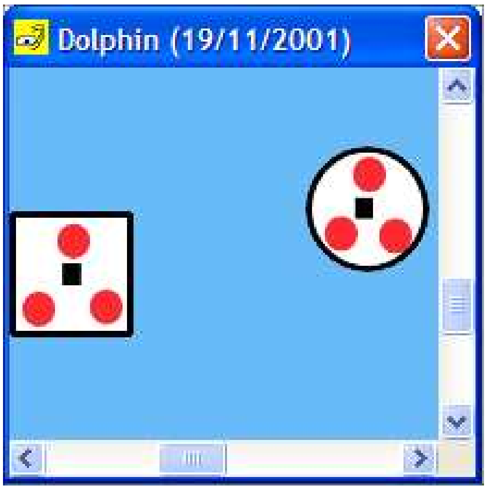

Figure 4.5: Example of the display from the standard scrolling condition, which is also an example of the visual interface (focus) from Dolphin.

Attribute	Possible Values	Description
Ride Type	Rollercoaster, Water Ride, Static Ride	Categorises theme park rides by their properties.
Ride Intensity	Low Intensity, Medium Intensity, High Intensity	How intense the ride is. Large, fast rollercoasters would be categorised as high intensity, whereas a miniature railway designed to transport customers around the park would be of low intensity.
Cost	Low Cost, Medium Cost, High Cost	How much it costs to be admitted to the ride.

Table 4.2: Table showing the attributes and their values encoded in the earcons and icons used in the experiment.

Earcons and Icons

As stated in Section 4.4.1, when a theme park ride was presented on the visual display it was represented with an icon, and when in the auditory display by an Earcon. Each Earcon and icon encoded the same three attributes of a theme park ride. These are shown with descriptions and possible values in Table 4.2.

Encoding Earcons

The attributes from Table 4.2 were encoded into Earcons according to the guidelines of Brewster, Wright and Edwards [25] (see Section 3.3.4), and were recorded from the output of a Roland Super JV-1080 MIDI (Musical Instrument Digital Interface) synthesiser. Although the guidelines of Brewster, Wright and Edwards note the differences between the various auditory parameters required to make earcons useful, they provide little guidance on which data attributes should be mapped to which auditory parameters. Norman [99], notes a difference between additive and substitutive dimensions, i.e. those dimensions where there is some concept of linear ordering, such as price; and substitutive dimensions, where there is only choice amongst many, such as with ride type. In the mapping of theme park rides to earcons, wherever possible, substitutive data parameters were mapped to substitutive auditory parameters (such as timbre), and additive data parameters were mapped to additive auditory parameters (such as pitch, loudness etc.).

Ride type was mapped to timbre. Three distinct timbres were used, with a grand piano (General MIDI patch No. 000), used to represent Rollercoasters, a violin (General MIDI patch No. 040) used to represent Water Rides, and a trumpet (General MIDI patch No. 056) used to represent Static Rides.

Ride Intensity was mapped to Melody, which is a combination of a rhythm and a pitch structure for that rhythm [116], as Brewster, Wright and Edward's [25] guidelines note that such combinations can be useful in differentiating earcons. The Melodies used for high, medium and low intensity rides are shown in Figure 4.6.

Figure 4.6: Melodies used to represent high, medium and low intensity theme park rides.

The cost of the ride was mapped to the register that the melody was played in. Although Brewster, Wright and Edward's guidelines [25] generally advise against register, the gross differences between the registers that guidelines recommend have been used, additionally the notes for each melody in each register have been staggered so that they sound slightly inharmonic (dissonant) when played with each other in order to help the earcons be distinguishable, and avoid lower level psychoacoustical perceptual interference (see Section 2.3.2). 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 highest register and high cost rides were played in the highest register. The registers used were the octave of C4 for low cost (approximately 261Hz), the octave of C5 for medium cost (approximately 523Hz) and the octave of C6 (approximately 1046Hz) for high cost (see Section 2.3.2).

Encoding Icons

In order to encode data in the visual icons, abstract icons as opposed to representational icons were used [13] as it was found to be difficult to encode all ride parameters in representational icons, and therefore to provide a fair comparison between Dolphin and the standard scrolling condition. As with the design of the earcons, Norman's distinction [99] between substitutive and additive dimensions was employed, with substitutive data parameters being encoded in substitutive attributes of the icons, and additive data parameters being encoded in additive attributes of the icons. The ride type was mapped to the shape of the icon, with a square, circle and triangle used to represent rollercoasters, static rides and water rides respectively. The intensity of the ride was mapped to the shade of dots on the icon, with dark, medium and light shades representing high, medium and low intensity rides. The cost of the ride was encoded using a

Figure 4.7: Examples of icons used to represent theme park rides.

different number of dots within the icon. High, medium and low cost rides were represented with three, two and one dots. Figure 4.7 provides examples of the icons used.

Training Phase

Because of the abstract mapping between the Earcons and icons, and the data they represent, it was important that participants were trained on them before undertaking the experiment. Training involved participants firstly being presented with a sheet describing the parameters of a theme park ride and how these were encoded into the icons (see Appendix B). Participants were then given 10 minutes of self guided training on the icons via a Web page, which displayed all of the icons that would be used in the experiment, together with descriptions of those icons (see Section A of the accompanying CD-ROM for the Web page used (Appendix M)). After 10 minutes participants were asked to independently identify three icons without any form of assistance. If they were unable to do so, the sheet describing the icons was returned along with the Web page, and a further five minutes training was allowed. After that time the test was carried out again. If participants were still unable to identify the three icons, they took no further part in the experiment. No participants however failed training.

In the Dolphin condition, after training on the icons had been successfully completed, participants were trained on the earcons used in a similar manner. Participants were provided with a sheet describing the grammar of the earcons (see Appendix B) as well as a Web page containing the earcons, their descriptions, and the complementary icon (see Section A of the accompanying CD-ROM for the Web page used (Appendix M)). Again, 10 minutes was allowed for participants to learn the earcons. After 10 minutes participants were asked to identify three individually presented earcons. If they were unable to do so, as with the icon training, a further five minutes was allowed before retesting took place. If participants were still unable to identify the earcons, they took no further part in the experiment. No participants however failed training.

This training technique for earcons is similar to that used by Brewster [16], who found that allowing participants to train by listening to the earcons used provided a significant improvement in task performance [18].

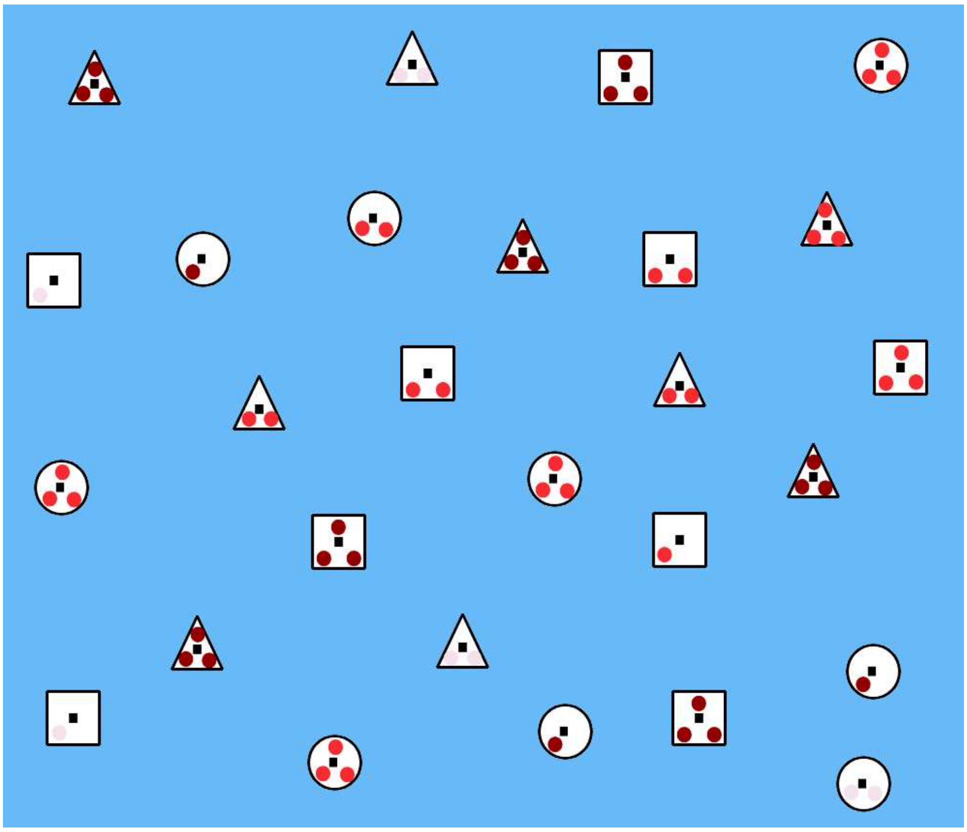

Figure 4.8: An example theme park map used in the experiment. Note the size of the map has been substantially reduced from the size used in the experiment.

Testing Phase

The testing phase of each condition comprised of participants making shortest path length tours around specific theme park rides in four simulated theme park maps. Each map was 992 x 850 pixels and contained 27 rides (around the size and number of rides that were found in example theme park maps). The same 27 rides were used for all maps, however the locations of the rides in each map were randomly assigned. An example map is shown in Figure 4.8.

Participants were provided with a question in a dialogue box which asked them to create a minimum tour around all of the rides on the map which contained the two ride attributes given. In real world situations it may not always be the case that all data parameters encoded in an earcon will always be required, and a subset of parameters may be all that is required, thereby asking about specific combinations of parameters is useful. See Figure 4.9 for an example of the dialogue used as well as an example question.

The main hypotheses of the experiment were:

H1 Participants will take less time to complete tours in the Dolphin condition than in the standard scrolling condition.

H2 Participants will create shorter routes in the Dolphin condition than in the standard scrolling condition.

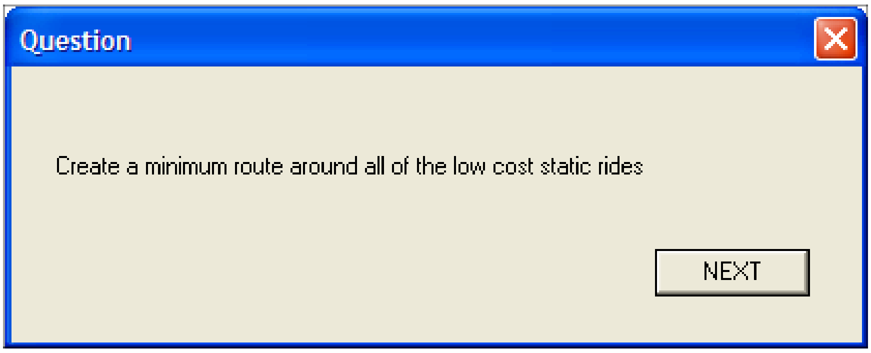

Figure 4.9: An example question as used in the experiment.

H3 There will be fewer occasions in the Dolphin condition where one or more rides will either be missed from a tour, or extra rides will be erroneously added to a tour.

The independent variable (IV) was the system used to present the maps, either Dolphin or the standard scrolling view. The dependent variables (DVs) were the time taken to complete each map, the length of the tour created by participants, and the number of occasions on which one or more rides were either missed, or erroneously added to the tour.

In addition, it was hypothesised that the Dolphin condition would reduce the subjective workload of participants required to carry out the task. Workload is important to determine since it is possible that whilst there may not be a significant variation between the objectively measurable data for each condition, if one condition subjectively increases workload a user may consider themselves to be overloaded and shed work, and as such lower their performance criteria or become psychologically distressed [60]. This may lead to a user refusing to use a system which incorporates one condition over the other, even if their is a significant objective improvement in task performance. A set of modified NASA Task Load Index (TLX) scales [60] were used to collect data to test this hypothesis.

NASA TLX scales allow for participants to rate on a set of unmarked scales, the subjective workload that they experienced during an experiment. A table of the attributes used to define workload in the modified TLX scales used in this experiment is shown in Table 4.3.

The scales used in this experiment have been modified to include the annoyance experienced attribute, which was originally included in the work of Brewster [16]. As noted by Buxton [32] (see section 3.2.2) auditory displays which are uninformative are likely to be annoying to a user. Therefore, measuring the annoyance of an auditory display is of importance when considering its usability. The overall preference attribute as devised by Brewster [16] has also been incorporated as a metric to provide an overall subjective rating of which condition made the experimental task easier. Before the testing phase proper, participants were given a trial run of the experiment with two maps not used in the testing phase, in order to familiarise themselves with the task they would be performing.

Attribute	End Points	Description
Mental Demand	*Low/High*	How much mental, visual and auditory activity was required? (e.g. thinking, deciding, calculating, looking, listening, scanning, searching)
Physical Demand	*Low/High*	How much physical activity was required?(e.g. pushing, pulling, turning, controlling)
Time Pressure	*Low/High*	How much time pressure did you feel because of the rate at which things occurred? (e.g. slow, leisurely, rapid, frantic)
Effort Expended	*Low/High*	How hard did you work (mentally and physically) to accomplish your level of performance?
Performance Level Achieved	*Poor/Good*	How successful do you think you were in doing the task set by the experimenter? How satisfied were you with your performance?
Frustration Experienced	*Low/High*	How much frustration did you experience? (e.g. were you relaxed, content, stressed, irritated, discouraged)
Annoyance Experienced	*Low/High*	How annoying did you find the graphics and/or the sounds in the condition?
Overall Preference	*Low/High*	Rate your preference for the two conditions. Which one made the task the easiest? The one with sounds or the one without.

Table 4.3: Table showing the attributes and descriptions of NASA TLX workload questionnaires used in the experiment. Modified from Hart and Staveland [60].

4.5.2 Results

For hypothesis one (H1), the time taken by each participant to create a tour for each map was collected. Each participant's overall score was determined to be the mean time taken for all those tours which incorporated all of the required rides and did not miss rides, or have additional incorrect rides added to a tour. The mean time taken, across all sixteen participants is graphically presented in Figure 4.10. Raw data are presented in Appendix C. A within groups t-test on this data failed to show significance ($t(15) = 1.54$, $p = 0.144$).

For hypothesis two (H2), the tour for each correctly completed map (i.e. those which did not miss or have additional rides which were not asked for), was converted into a percentage of the optimal shortest path length for that map. The average of each participant's percentage optimal path length was then taken. The average of these data across all participants is shown in Figure 4.11. Raw data are presented in Appendix C. A within groups t-test on these data failed to show significance ($t(15) = 0.03$, $p = 0.979$).

For hypothesis three (H3), the average number of tours where a ride was erroneously added or missed from a tour is graphically presented in Figure 4.12. Raw data are presented in Appendix C. A within groups t-test failed to show significance ($t(15) = 0.58$, $p = 0.572$).

As stated in the previous section, to measure participants' subjective workload, modified NASA TLX questionnaires were used. The results of these questionnaires is shown in Figure 4.13. Raw data are presented in Appendix C. To determine if the difference in workload was significant, each participant's scores for each attribute (excluding overall preference and annoyance experienced since these do not form

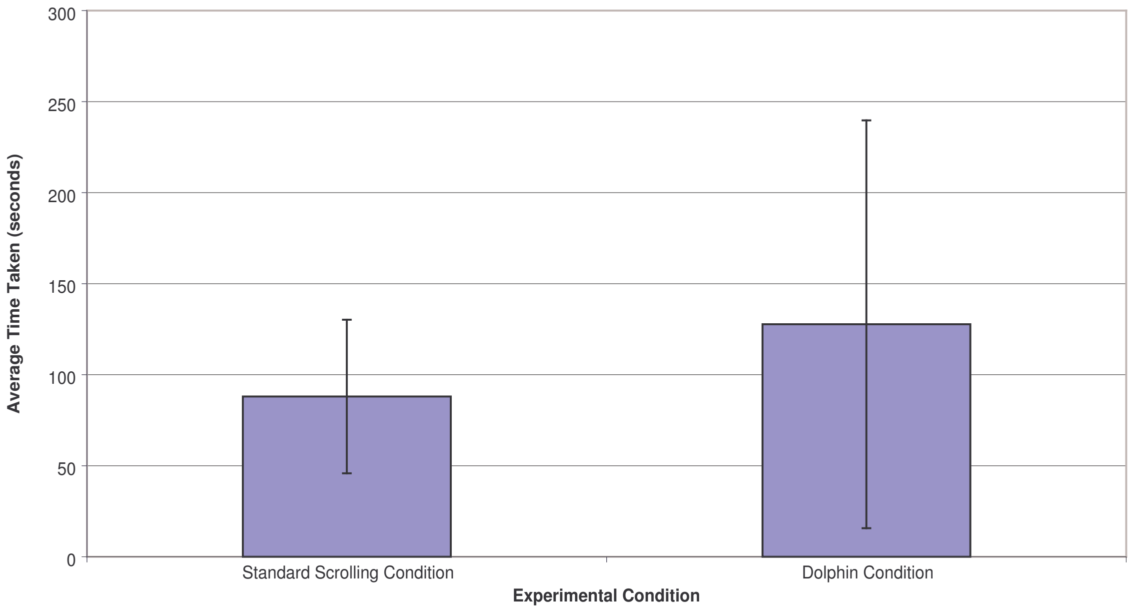

Figure 4.10: Graph showing the average time taken to complete a tour in the standard scrolling and Dolphin conditions, showing standard deviations.

part of the standard TLX set of attributes) were added together and a within groups t-test was performed. This t-test showed that workload was rated significantly higher in the Dolphin condition than in the standard scrolling condition (t(15) = 2.06, p = 0.050). To determine which attributes caused this increase in workload eight within groups t-tests were carried out, one for each modified NASA TLX attribute.

The t-tests showed that participants recorded significantly greater mental demand in the Dolphin condition than in the standard scrolling condition (t(15) = 2.10, p = 0.050). t-tests also showed significantly greater frustration recorded for the Dolphin condition than the standard scrolling condition (t(15) = 3.43, p = 0.004). Significantly greater annoyance was also recorded in the Dolphin condition than in the standard scrolling condition (t(15) = 4.76, $p < 0.001$). In addition participants rated overall preference of the standard scrolling condition to be significantly greater than that of the Dolphin condition (t(15) = 2.83, p = 0.013). The t-tests for physical demand (t(15) = 0.34, p = 0.7390), time pressure (t(15) = 1.63, p = 0.124), effort expended (t(15) = 2.02, p = 0.061) and performance level achieved (t(15) = 0.79, p = 0.444), were not found to be significant.

4.5.3 Discussion

The results show that there were no significant differences between the Dolphin and standard scrolling conditions for any of the objectively measurable data. There does however, appear to be a trend where the Dolphin condition has poorer performance than the standard scrolling condition. Some understanding of why this occurred may come from the modified NASA TLX questionnaire results. Participants reported that

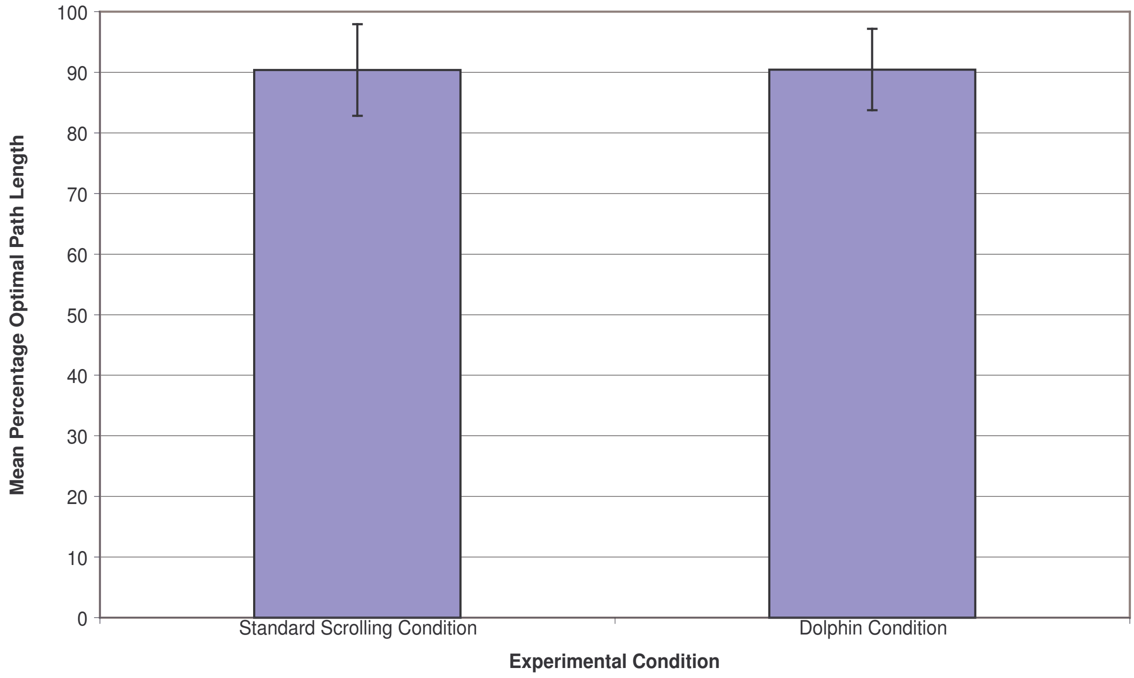

Figure 4.11: Graph showing the average percentage optimal path length tours for the standard scrolling and Dolphin conditions. Shown with standard deviations.

mental demand, frustration and annoyance were significantly greater in the Dolphin condition than in the standard scrolling condition. This indicates that not only did the auditory display used not assist participants in their given task, it actually caused a detrimental effect on perceived performance.

Understanding why the audio should cause such problems cannot be shown from the results of this experiment. However, informal discussions with participants after the experiment revealed that the sounds had merged together and interfered with each other causing the identification of earcon attributes to become difficult. Therefore in order to use concurrent earcons it is necessary to understand the extent of interferences between concurrently presented sounds and ways in which such interferences can be reduced. This issue will be investigated in the remainder of this thesis.

4.5.4 Dolphin Conclusions

It seems clear from the results of this experiment that to successfully employ the advantages of concurrent audio presentation as discussed in Section 4.2, it is essential to understand how concurrently presented sounds interfere with each other. Unfortunately, as already stated there is little guidance on how the techniques used as part of auditory displays (see Section 3.3) can be modified to reduce unwanted interference when concurrent presentation is used. Such guidance is likely to be of great importance to Earcons, since Earcons which are formed from the same grammar (see Section 3.3.4) are likely to share components such as timbre or melody. For example in the Dolphin condition, all rollercoasters shared the same timbre, and

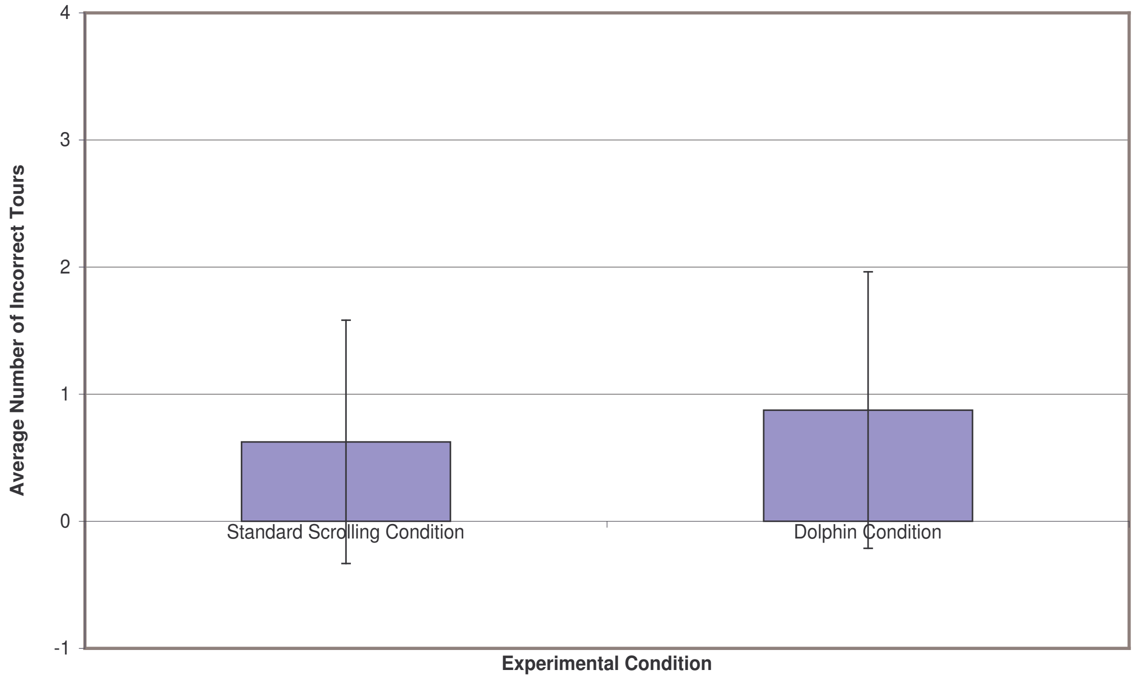

Figure 4.12: Graph showing the average number of tours either missing rides or incorporating incorrect rides into a tour in the standard scrolling and Dolphin conditions, showing standard deviations.

all high intensity rides shared the same melody. According to auditory scene analysis (ASA) (see Section 2.6), sounds which share components in such a way are likely to be merged together. Such issues are not as prevalent in other auditory display mapping techniques such as auditory icons where guidelines state that using similar sounds should be avoided [91]. For example, the ARKola system [55], where up to fourteen sounds could be concurrently presented, whilst only being informally evaluated did not appear to show significant problems due to the density of sounds presented. Indeed Gaver, Smith and O'Shea noted that *"Sounds served as shared reference points for partners, allowing to to refer directly to events they couldn't see"* [55]. In the Dolphin system it was considered that the use of spatialised sound would help to remove this issue, however this appears not to have been the case. It is therefore likely that spatially separating earcons does not guarantee that they will be perceived as distinct, and further work to determine the extent of interference and action that can be taken in combating it, needs to be undertaken.

4.6 Related Work

In this section the limited amount of work which has been undertaken investigating the problems of concurrently presented audio as part of an auditory display is reviewed and discussed. Several features from such work may be able to be applied to earcon design to improve the identification of concurrently presented earcons.

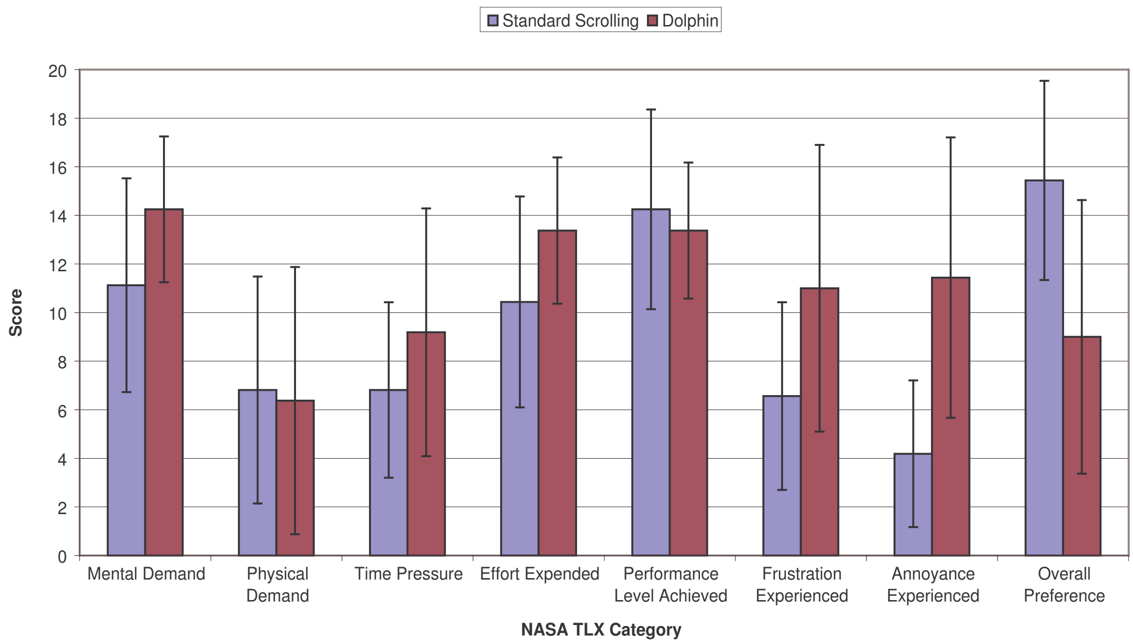

Figure 4.13: Graph showing the mean values for NASA TLX workload data for the standard scrolling and Dolphin conditions. Shown with standard deviations.

4.6.1 Gerth

Gerth [57] carried out several experiments which investigated how discrimination of concurrently presented audio was affected by both the density of sound (number of concurrent sounds presented), and differences between the timbre of concurrently presented sounds.

One experiment involved two complex sounds which were consecutively presented to participants with a short intervening delay. The first complex consisted of between one and four distinct temporal patterns (rhythms) which were concurrently presented using the same timbre (either piano, violin, "sewing machine" or "metal hits"). The second complex tone was the same as the first except that in 50% of cases, a rhythm was either removed or added to the complex. Participants had to determine if the two consecutively presented complexes were the same or different. Gerth identified that as sound source density increased (as the number of concurrently presented sounds increased), the number of errors in accuracy increased from around 2% to 20%.

Gerth's second experiment was the same as that previously discussed however, instead of all sounds in a complex having the same timbre, each was presented with a different timbre. Hence, instead of all rhythms being presented with a piano timbre, one would be presented with a violin, another with a piano, etc. Gerth found, as with his previous experiment, that as sound density increased the accuracy of responses dropped, from around 100% for one and two concurrent rhythms per complex, to around 80% for four concurrent rhythms per complex. Gerth claimed that using different timbres for each rhythm increased the accuracy of participant responses, however he failed to provide clear evidence to validate this assertion; the accuracy of responses seems to be the same irrespective of timbre differences. It may be that Gerth's task was so simple that a ceiling effect was observed in his results. If timbre differences between concurrent sound sources do however improve their discrimination, and evidence from auditory scene analysis research does exist to support this [129] (see Section 2.6), incorporating timbre differences between concurrently presented earcons may help their identification.

For Gerth's final experiment he changed his methodology, from identifying if two complex sounds were the same or different, to identifying exactly what the change between the two complexes was. In this experiment, participants had to identify what temporal component and timbre had either been deleted, added or substituted between the first and second sequentially presented complexes. In order to successfully undertake this task, participants were first trained to identify the sixteen possible sounds that could be included in a complex (four timbres combined with the four temporal patterns). If participants could not reach 80% accuracy in identifying those sounds, their results were not used. In order for participants to be able to identify the temporal patterns used, Gerth used the names "regular", "fast", "slow" and "chopped" to represent each pattern. In some respects this is like the identification required to decode earcons, since the mapping between data and sound is abstract and must be learned. However the descriptions Gerth used were likely representative descriptions of the temporal patterns and may not be entirely the same as those used for earcons such as in Dolphin (see Section 4.5). Gerth found that, as with his previous experiments, performance fell as the density of sounds was increased. The accuracy of participant responses in identifying the change between the two complex sounds was found to be lower than the simple identification of sameness or difference in the two sounds from the first two experiments, however accuracy stayed around the 60% correct level.

Gerth's work indicates several issues that may be relevant to concurrent earcon research, notably that increasing the number of earcons which are concurrently presented is likely to lead to a decrease in the number which can be correctly identified. He showed that this trend still occurs even when modifications to the sound (in the form of a unique timbre for each temporal pattern) are introduced.

However, Gerth's research incorporates maximum possible differences between concurrently presented sounds, with each having a different rhythm and timbre. It may not always be possible to guarantee this with earcons since there is a concrete mapping between sounds and the data they represent which would be broken if maximum differences between sounds were forcibly incorporated.

4.6.2 Multi-Talker Speech

Brungart, Ericson and Simpson [29] have investigated how the intelligibility of multi-talker speech environments (where more than one talker is speaking at a time) can be improved.

In their experiments, they looked at the identification of coordinate response measure (CRM) speech intelligibility tests [90]. 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",

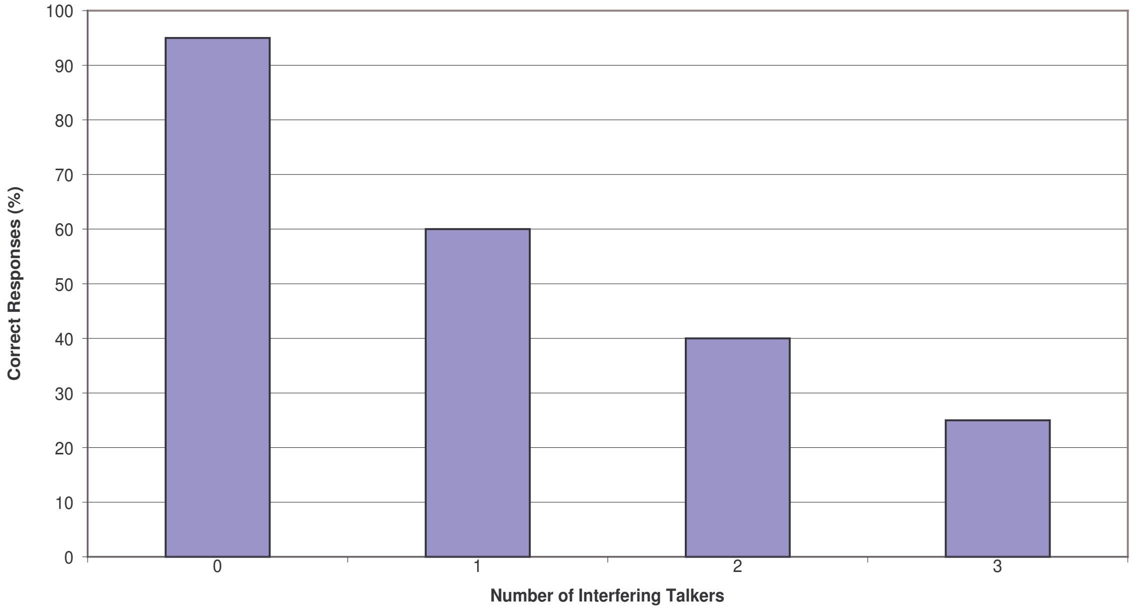

Figure 4.14: Graph showing the percentage of correct performances on a coordinate response measure (CRM) task with 0, 1, 2 and 3 competing talkers. Taken from Brungart, Ericson and Simpson [29].

"Ringo", "Eagle", "Arrow", "Hopper", "Tiger" or "Laker", color is one of red, green, blue or white and number is between one and eight. Participants had to listen for a particular call sign and record the colour and number associated with it.

Brungart, Ericson and Simpson investigated how several modifications to the presentation of the various concurrent speeches affected task performance amongst participants. In investigating how the number of competing talkers (density in terms of Gerth [57], see previous section) influenced performance, they found a similar effect to that of Gerth. As the number of concurrent talkers was increased from zero to three, the percentage of correct responses fell from around 95% to around 20% (See Figure 4.14).

Gerth found that increasing the density of his complex sounds, increased error rates from 2% (98% correct) for complexes with one or two components, to 20% errors (80% correct) for complexes with three to four component sounds. The trend identified by Brungart, Ericson and Simpson is therefore much steeper than that identified by Gerth. This raises the question of which trend more accurately represents the identification of concurrently presented earcons when the number concurrently presented is varied. On one hand, Earcons, as stated in Section 4.6.1, are much more like the temporal patterns used by Gerth; although the task that Gerth used is much simpler than would be required for Earcons, where individual identification of parameters would be required, rather than differences between two consecutively presented sequences of sounds. However, Earcons which are formed from the same grammar are likely to share components (See Section 3.3.4), as is the case in the CRM tasks of Brungart, Ericson and Simpson, who also used the same talker to speak all of the texts; Gerth ensured maximum differences between each component in his complex sounds. Therefore although it is likely that increasing the number of concurrently presented earcons will cause their identification to fall, it is difficult to predict if that fall will be steep, like Brungart, Ericson and Simpson, or more shallow such as with Gerth.

Further experiments by Brungart, Ericson and Simpson where the target talker (the voice speaking the phrase containing the participant's call sign) was of a different sex than the other concurrently presented phrases, found that this increased task performance to 80% when there was one other competing talker, and task performance was increased to 60% when there were three competing talkers; although task performance still fell as the number of talkers was increased. Changing the sex of the speaker, causes a change in the fundamental frequency of the speaker's voice, and thus changes the timbre of the voice, although as already discussed in Section 2.6 changes in other parameters of sound such as pitch may also occur. Gerth also claimed that the introduction of timbre differences between concurrently presented sounds increased participants' ability to detect concurrently presented stimuli, and such differences may improve the identification of concurrently presented earcons.

Brungart, Ericson and Simpson also identified that by placing each concurrent talker in a different spatial location in azimuth (See Figure 2.4) (at -45^o, 0^o, 45^o and 90^o relative to the listener), performance could also be improved, although again, as the number of talkers was increased, performance fell.

Unfortunately Brungart, Ericson and Simpson did not perform statistical analyses on their data to determine if the improvements in performance were significant. Their work does point however to several sound manipulation techniques which could be applied to concurrently presented earcons, in order to improve earcon identification.

4.6.3 The Computational Auditory Scene Synthesizer

An alternative approach to solving the problems of concurrently presented sounds in an auditory display from that attempted by Gerth [57] and Brungart, Ericson and Simpson [29], is to control all of the sounds that might be presented in an computer interface from a central point.

Papp [104], 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 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 which incorporated some auditory scene analysis 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 etc.).

Whilst 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. Firstly, as explained in Section 2.6, it is difficult to predict how different auditory scene analysis factors influence each other. This is even harder for complex auditory sources since most ASA research is done on simple sinusoidal tones [14]. 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 4.3), however it can also be difficult for other auditory information types, such as auditory icons (see Section 3.3.3), where the modification of a sound may cause identification problems for the user [53]. 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 since 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 [128].

4.7 Research Questions

The discussion of Dolphin, and the work on auditory displays from both Chapter 3 and Section 4.2, have shown that whilst auditory displays can make effective use of concurrent sound presentation, great care is required to ensure that the sounds used in such a display do not interfere with each other, causing the display to be ineffective. Unfortunately, little guidance is available to auditory display designers to make informed decisions about the amount of audio, or modification to that audio, which would be beneficial in reducing unwanted interactions between sounds. This is especially the case for earcons, where Dolphin is only one of two example auditory displays to evaluate the usability of concurrently presented earcons (the other being Brewster, Wright and Edwards's parallel earcons [26], as described in Section 3.4). Whilst auditory scene analysis (see Section 2.6) and the work described in the previous section indicates that it is possible to modify earcons to influence their streaming and thus reduce unwanted interactions between them, it is unclear what modifications will be useful. As discussed in Section 4.3, due to the grammar earcons are formed from there is a limitation to the extent of any modifications since if the mapping between data and sound is broken, it will be irrelevant if all concurrently presented earcons stream separately. If earcons are to be successfully concurrently presented in an auditory display, work must be undertaken into both determining the extent of the problems of concurrently presented earcons as well as modifications to their design which can be used to reduce such problems. By doing so, guidance for future auditory display designers on the use of concurrently presented earcons can be determined. Guidance, which is currently lacking in the literature.

The remainder of this thesis investigates how earcons perform when concurrently presented and what steps may be taken by auditory display designers to combat the issues of undesirable interactions between concurrently presented earcons, such as occurred in Dolphin. These investigations will lead to a set of guidelines which can be used by future designers of auditory displays which use concurrent audio, and specifically concurrent earcons, to design more effective and usable displays. To tackle this problem four research questions are derived, the answers to which will be explored through the following chapters:

RQ1 What is the effect on earcon identification of varying the number of earcons which are concurrently presented?

RQ2 How can concurrently presented earcons be more robustly designed and presented in order to increase the number which can be identified?

RQ3 What is the impact of presenting concurrent earcons in a spatialised auditory environment?

RQ4 How much do modifications to the design and presentation of concurrently presented earcons affect performance in "real world" tasks?

These research questions provide a basis for a thorough investigation of the field of concurrent earcon presentation, as well as allowing a relationship to prior concurrent audio research (as outlined in the previous sections) to be established.

Research Question 1 attempts to identify if there is a trend in earcon identification when the number of earcons which are concurrently presented is varied. Work by both Gerth [57] and Brungart, Ericson and Simpson [29] has identified such a trend, however as stated in Section 4.6.2, there are differences between their work in exactly how identification performance decreases when the number of concurrent sources is increased. This research question will allow the trend of number of earcons versus their identification to be identified.

The research from the previous section, as well as that described in Section 2.6, has shown that the modification of sounds can affect the way they are streamed by the auditory system. Research Question 2 will investigate ways in which earcons can be better designed and presented so that such interferences are reduced. Gerth found an increase in identification when different timbres were used for different rhythms, and Brungart, Ericson and Simpson identified a similar finding when they used different sex speakers for different texts. In doing so however, the rules that earcons are constructed from (the earcon grammar) must be retained to ensure the relationships between earcons are not lost and therefore one of their strengths is destroyed. This may mean that the extent of any modifications on earcon design and presentation are constrained, making it important that any improvements gained from a specific modification are determined in isolation to other modifications. This is especially important, as such modifications may place constraints on earcon designers such as having multiple similar timbres, or needing to constrain the structure of melodies (see Section 6.2.3), which without individually investigating each earcon modification may be needlessly placed on earcon design.

One of the features of auditory scene analysis that is likely to be beneficial at improving the number of concurrently presented earcons which can be attended to, is the use of spatialisation, or 3D sound presentation, to separate concurrently presented earcons. As Bregman notes *"it would be a good bet that sounds* *emanating from the same spatial location have been created by the same source"* [14]. Also as discussed in Section 3.4, many systems which have successfully used concurrent audio presentation have also used spatial presentation. Therefore Research Question 3 will investigate how concurrent earcon identification is affected by spatial separation. However, when considering the impact of spatialisation in auditory displays both practical and technical limitations must be considered. As discussed in Section 2.4, the localisation of sound depends on differences in sound reaching the left and right ears. This requires either properly calibrated speakers, or the user to wear headphones, which may not always be possible or convenient. In addition, although spatialisation was used in the Dolphin system (see Section 4.4) undesirable interactions between the earcons still occurred. It is therefore likely that it may not always be possible to separate concurrent earcons enough to cause them to stream separately. Investigating spatialisation in isolation to other features is therefore useful.

Finally, although the Dolphin system identified that concurrently presented earcons are likely to be a problem, it provided little guidance on exactly what features and aspects of the earcons caused problems. Therefore Research Questions 1-3 will investigate concurrent earcons using a more abstract experimental methodology, in order to be able to determine in exactly which way any modifications to the design or presentation of earcons affects their identification. This creates the problem of understanding how the useful modifications to earcons which have been identified, perform in more realistic environments and applications. Research Question 4 will attempt to determine this.

4.8 Conclusions

This chapter has explored the use of auditory displays where more than one sound is presented simultaneously. These concurrent auditory displays can increase both the rate and amount of data which can be presented to users through audio, as well as allowing users to more easily make comparisons between multiple data. Whilst several examples of auditory displays which use concurrent sound presentation exist, comparatively little guidance on the number of sounds which should be concurrently presented, or how those sounds should be designed for such situations, exists. In addition, concurrently presented sounds may interfere with each other in undesirable ways, and any parameters of data encoded in the sound may be impossible to determine. This is especially problematic with earcons since they are formed from a common grammar, and are likely to be very similar (sharing the same timbre, rhythm etc.), which according to auditory scene analysis will make it more likely that concurrently presented earcons will interfere with each other.

Whilst some research [57, 29] has investigated ways to improve the identification of concurrently presented sounds as part of an auditory display, very little research [26] has yet been undertaken to identify how to improve the identification of concurrently presented earcons in such environments. The Dolphin system, which incorporated concurrent earcon presentation was not successful, with a significant increase in users' perceived frustration and annoyance over a standard scrolling display. Participants claimed that the earcons interfered with each other and thus became unusable.

This thesis proposes an investigation into concurrently presented earcons to identify how well they are identified, as well as how their identification can be improved in concurrent auditory displays. Four research question are proposed as a method to investigate concurrently presented earcons. These four questions are explored in the remaining chapters of this thesis.

Chapter 5

Number of Earcons versus Identification

5.1 Introduction

In the previous chapter several questions derived from the main research statement posed in this thesis were described. In this chapter an experiment is described which attempts to answer one of these questions (RQ1), *"What is the effect on earcon identification of varying the number of earcons which are concurrently presented?"*.

There are several reasons why understanding how concurrent earcon identification is affected by the number of earcons which are concurrently presented is important. Firstly, although the work on Dolphin (see Section 4.4) has indicated that the identification of concurrent earcons is difficult, there is no empirical evidence to support this claim.

Whilst there has been some work on the identification of different types of concurrently presented auditory stimuli, such as Brungart *et al's* [29, 31] work on concurrent speech identification and Gerth's [57] work on the identification of concurrently presented rhythms, both of which are described in Section 4.6, there has been only one study, by Brewster, Wright and Edwards [25], which has empirically investigated the identification of concurrently presented compound earcons (see Section 7.1.1). That work only considered two concurrent earcons and incorporated large differences in both the spatial location of the earcons and the registers in which each was presented. No work which investigates how the more powerful hierarchical and transformational earcon types are identified in concurrent presentation situations has been undertaken. As described in Section 4.3, these earcon types are believed to be the most problematic when concurrently presented. Due to earcons sharing the same grammar and as such being similar according to the work of Section 2.6, they are likely to stream in such a way that makes the identification of their attributes difficult. It would be useful therefore to determine how the identification of concurrently presented earcons relates to these previously undertaken studies in order to help determine future research directions to improve their identification.

Finally, one of the aims of this thesis is to provide a set of guidelines which can be used to guide the designers of future auditory displays which use concurrent earcons, to make informed choices to improve the identifiability of such earcons that they may use in a specific application. Guidelines which are currently lacking in the literature. It may be that reducing the number of earcons concurrently presented from say 4 to 3, which may cause an increase in earcon identification for example, from 0.5 to 3, would be an acceptable solution for an auditory display designer when using concurrent presentation of sounds. It is also the case that although a reduction in performance was identified by increasing the number of concurrently presented sounds by both Gerth [57] and Brungart, Ericson and Simpson [29], the degree to which performance dropped was different in each case (see Section 4.6). Understanding of how varying the number of earcons concurrently presented affects their identification is therefore important to contextualise the answers of the other research questions of Section 4.7, and therefore the remainder of the work in this thesis.

5.2 Methodology

5.2.1 Issues With The Experimental Design

One of the issues with the evaluation of Dolphin (see Section 4.4), was that whilst neither condition was significantly better than the other, there was no indication as to why this occurred, other than the anecdotal evidence about the problems with concurrent identification of earcons from experimental participants. In the Dolphin experiment, data on earcon identification could not be recorded with enough fidelity to identify why and in what way earcons interfered with each other. If an investigation into the presentation of concurrent earcons is to yield the maximum understanding of how earcons work concurrently, it is essential to focus more tightly on the aspect of interest, namely the identification of concurrently presented earcons, at the expense of the more ecological aspects of any work such as existed in Dolphin. This raises some limitations to the applicability of this experiment and the experiments in the following two chapters. It will not be possible to claim, simply by showing an increase in participants' ability to identify concurrently presented earcons, that any presentation modifications, or redesign of the earcons will lead to performance improvements in real interfaces which use concurrent earcons. However, since there is so little understanding on the identification of concurrently presented earcons, and due to the gap between the psychological investigations of auditory scene analysis and how it would be implemented in auditory displays, investigating this topic through further experiments like Dolphin may yield little information to understand how modifications to the design and presentation of concurrently presented earcons affect their identification. The ecological validity and applicability of the work contained in this chapter, and Chapters 6 and 7, is explored further in Chapter 8.

With regard to these issues, an experimental technique similar to that of Gerth [57] was chosen to investigate how varying the number of concurrently presented earcons affected their identification. Whilst Gerth however, was only interested if participants could identify differences between two consecutive sequences of concurrently presented sounds, earcons, as already stated in Section 3.3.4, need to be accurately identified.

It is not sufficient to simply present two consecutive sequences of concurrent earcons, and ask whether

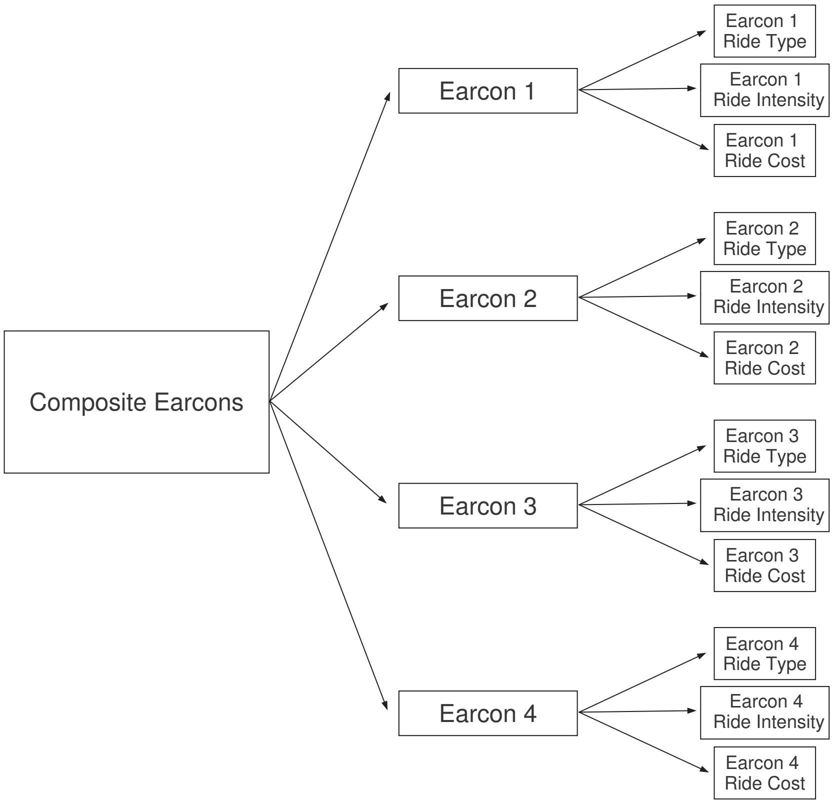

Figure 5.1: A diagram showing the stages involved in the identification of concurrent earcons. Participants must first recognise and separate both perceptual and cognitive processes, the earcons contained within the composite sound. Participants must then use cognitive processes to recall the mapping between earcon sound attributes and the labels associated with those sound attributes.

the second sequence contained more or less earcons than the first. In the experiment described in this chapter, participants were asked to listen to sequences of concurrently presented earcons and identify the attributes of each earcon presented. The steps involved by participants in this identification are outlined in Figure 5.1.

5.2.2 Procedure

Sixty four participants undertook the experiment, which was of a between groups design. All participants were aged 18-24, and comprised of a mix of both males and females. All were paid £5 on completion. The earcons used were the same as those from the Dolphin investigation (see Section 4.4). There is no reason to consider that the earcons from Dolphin were worse than any other set of earcons, and using the same set of earcons may assist in further understanding the problems with Dolphin. As with Dolphin, each earcon encoded three parameters of a theme park ride: the ride type (either rollercoaster, 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 dependent variables (DV's) were the number of earcons, ride types, ride intensities and ride costs successfully identified. Understanding how the identification of earcon attributes varies with the number presented is important, since as discussed in Section 4.4, 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 such cases actual performance may be closer to the identification of individual earcon attributes than total earcons. Additionally, the number of attributes encoded in each earcon are around the maximum that can be realistically incorporated, and to date not other work has used more than three attributes in each earcon. Therefore, determining the individual attribute performance may also indicate what would happen to identification performance if the number of data attributes encoded in each earcon was reduced.

In addition it was hypothesised 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 [60] were used to collect data to test this hypothesis (see Section 4.5.1).

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, whilst not incorporating concurrent earcon presentation, since only one earcon was presented at a time, allowed a general comparison of earcon identification to be made back to the work of Brewster [16], and a determination of the quality of the earcons used in both this experiment and Dolphin (see Section 4.4) 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.

Issues with the Experiment

The experiment reported in this chapter was originally part of a larger between groups experiment which also incorporated the experiment from Chapter 6. Originally, the only intended investigation into varying the number of concurrently presented earcons was comparing three concurrently presented earcons with four concurrently presented earcons. It was felt that this could be achieved best by including the three earcon condition as one of the conditions of the experiment described in Chapter 6. However, after the three earcon condition had been performed it became clear that further studies of varying the number of concurrently presented earcons would be necessary. Therefore the larger experiment was split, with one experiment (reported in this chapter) investigating the effect of varying the number of concurrently presented earcons, and another (reported in Chapter 6) investigating how earcons can be more robustly designed and presented to increase earcon identification. However, because the three earcon condition had already been performed, there is some linkage between the two experiments. Therefore although this is a between groups experiment, participants who performed a condition of this experiment also performed a condition from the experiment described in Chapter 6. The way in which conditions from this and the experiment from Chapter 6 were paired together is shown in Appendix D. Note that for each pair of conditions the order in which participants performed them was counterbalanced in order to remove any order effects in the data. Since no participant performed two conditions of the experiment described in this chapter, between groups statistical analyses are relatively easy to perform, and there should be no anomalies in the results which could be caused by the interleaving of the experimental conditions of the two experiments.

Training Phase

The training phase was similar to that of the Dolphin evaluation 4.4 and involved participants reading a sheet which described the grammar of the earcons used (see Appendix E), followed by ten minutes of self guided training via a Web page interface (see Section B of the accompanying CD-ROM for the Web page (Appendix M)), 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 five minutes of training before retaking the test. If they were still unable to successfully identify the earcons, they took no further part in the experiment. Overall, one participant failed to complete the training phase. The purpose of this training methodology was to quickly allow the participants to learn the earcons. Brewster [18] compared a number of different training methods and identified that providing participants with a description of the earcons, and allowing them to listen to the the earcons was an effective training method. He did note however, that it may not be the most common way that participants are trained in real world applications, and a simple description of the earcons may be more likely. The work presented in this thesis however, investigates the identification of concurrent earcons and it would be unfortunate if the results of these experiments were confounded due to participants being unable to identify the earcons individually, hence a “best case” training methodology was chosen.

Testing Phase

The testing phase involved participants listening to twenty 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 diotically presented (monaural presentation to both ears) to participants seven times in rapid succession (see Figure 5.2 for an example of presentation from the three earcon condition). The number of earcons in each set was dependent on the experimental condition, with four, three, two and one earcons used in the four, three, two and one earcon conditions. Seven repetitions being, for the four earcon condition, approximately four times greater than the number of repetitions used by Brewster, Wright and Edwards [24] when investigating single earcon identification. Earcons were repeated seven

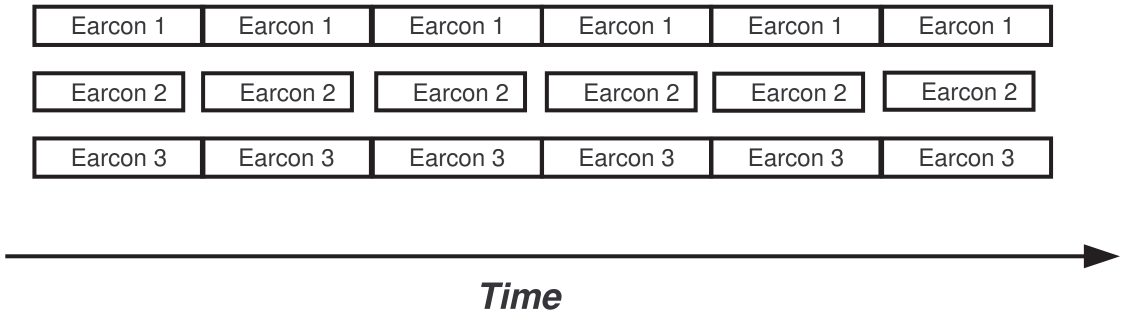

Figure 5.2: An example of the presentation of concurrent earcons. Showing how each set of earcons were concurrently presented seven times. Taken from the three earcon condition.

times as Bregman [14] notes that it is difficult to listen in detail to multiple streams at the same time, hence participants would need to re-listen to the earcons to retrieve all of the information contained within them. However, it may be the case that the number of earcon repetitions has an effect on earcon identification, and if more or fewer repetitions were used, performance may increase or decrease. If participants did not have to identify all of the earcons then this would reduce the need for so many repetitions and may overcome some of the issues that exist with multiple repetitions. However, this would mean that the experiment would have to ask participants to identify differences between two sets of sequentially presented earcon sets. This, as already discussed in Section 5.2.1 is an unrealistic approximation of earcon identification.

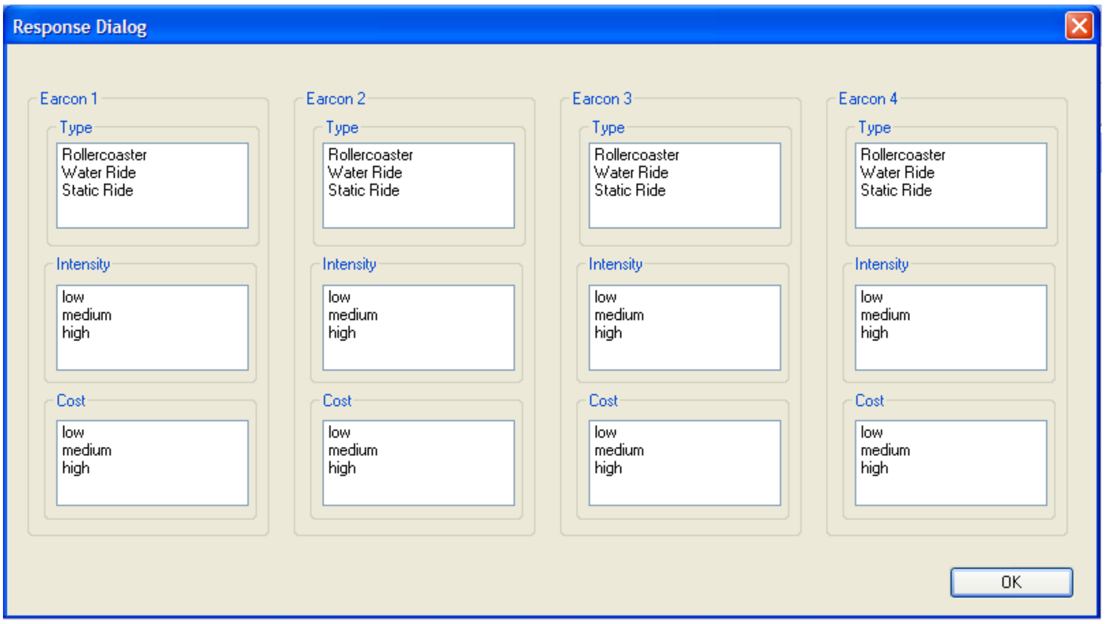

Figure 5.3: A screenshot of the dialog used by participants to fill in earcons identified. Taken from the four earcon condition.

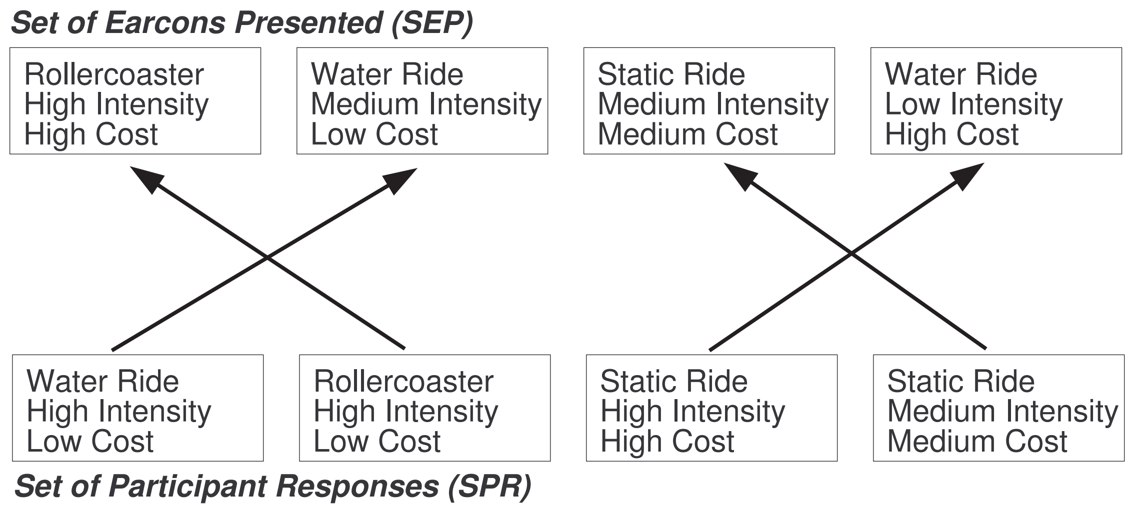

Figure 5.4: An example from the four earcon condition of how the set of participant responses (SPR) were mapped to the set of earcons presented (SEP) in order to determine the number of correctly identified earcons, ride types, ride intensities and ride costs.

Participants had to identify the attributes of each earcon in a set and record what those attributes were in a dialogue box. See Figure 5.3 for an example of the dialog box as used in the four earcon condition. The experiment was of a forced choice design. Participants were presented with the response dialog box as the earcons started to be played and were able to fill in the dialog box as the earcons were being presented, or wait until the earcon presentations had ended before filling in responses. Between successive sets of earcons, a mandatory rest break for participants of at least four seconds was introduced, since some research suggests that this is the period of time required by the auditory streaming mechanism to fully reset itself [14], 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 familiarise 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) concurrently 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, i.e. they had been correctly identified.

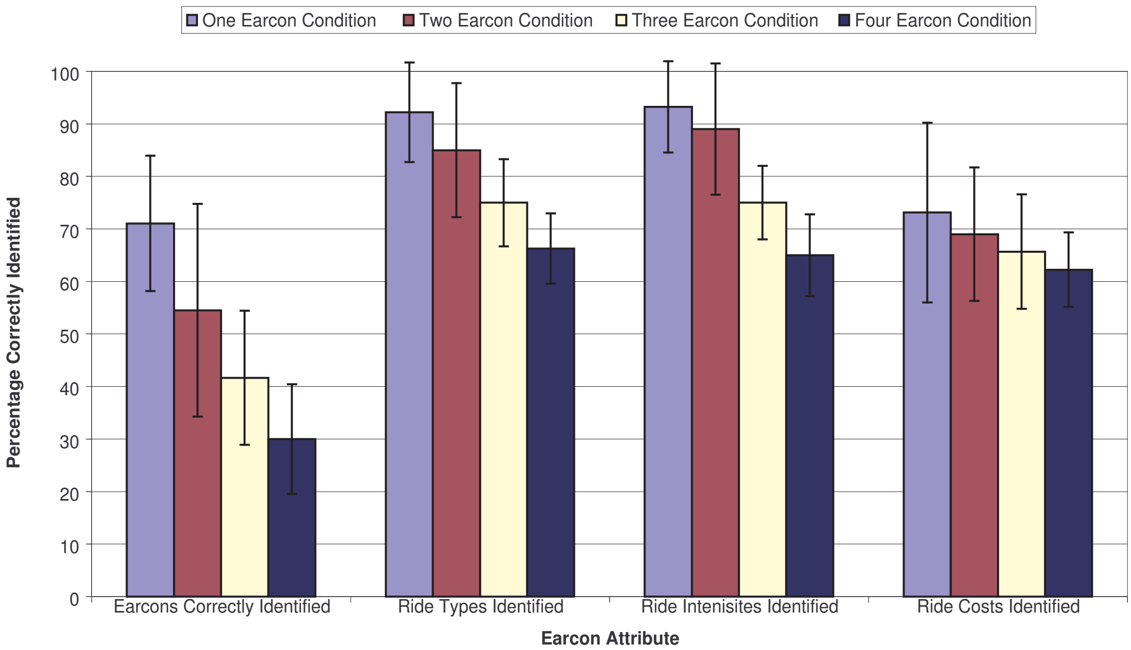

Figure 5.5: Graph showing the average proportion of earcons, ride types, ride intensities and ride costs correctly identified for the one, two, three and four earcon conditions.

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 which 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 5.4.

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) * 100 = 66%. The average proportion of correctly identified earcons, ride types, ride intensities and ride costs across all participants is presented graphically in Figure 5.5. Raw data are presented in Appendix F.

To determine if any of the differences shown in Figure 5.5 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

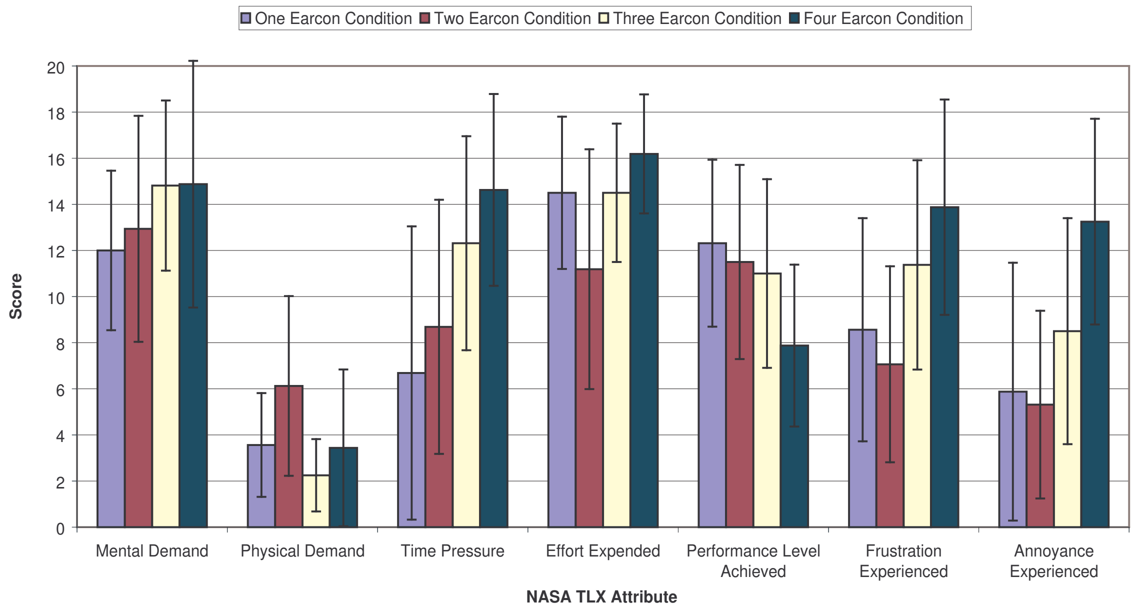

Figure 5.6: Graph showing the mean values for NASA TLX workload data for the one, two, three and four earcon conditions.

be significant (F(3,60) = 23.28, $p < 0.001$). *Post hoc* Tukey Honesty Significance Difference (HSD) tests [64] showed that earcons were significantly better identified in the one earcon condition than in the two, three and four earcon conditions ($p < 0.05$). 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 ($p < 0.05$), and ride types in the one, two and three earcon conditions were significantly better identified than those in the four earcon condition ($p < 0.05$). 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 one and two earcon conditions than in the three and four earcon conditions ($p < 0.05$), and ride intensities in the one, two and three earcon conditions were significantly better identified than those in the four earcon condition ($p < 0.05$). 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 4.5.1) to determine variations in subjective workload. A graphical summary of these data is presented in Figure 5.6. Raw data are presented in Appendix F. 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 analysis of variance (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 ANOVAs, 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 earcon condition ($p < 0.05$). Effort was also rated significantly higher in the one earcon condition than in the two earcon condition. For time pressure, the ANOVA again showed significance ($F(2,45) = 9.28$, $p < 0.001$). *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 ($p < 0.05$). 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. 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 ($p < 0.05$).

5.4 Discussion

The results from the previous section have confirmed the conclusion from the Dolphin experiment from Section 4.4, that trying to identify concurrently presented earcons is difficult, with on average only 30% of earcons correctly identified in the four earcon condition; four earcons being the average number of earcons concurrently presented in Dolphin. The results from the one earcon condition showed around 70% accuracy in earcon identification. Although using a different experimental procedure, Brewster [16] 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 5.7 which shows best fit trend lines between the number of concurrently presented earcons, and the percentage of earcons and earcon attributes

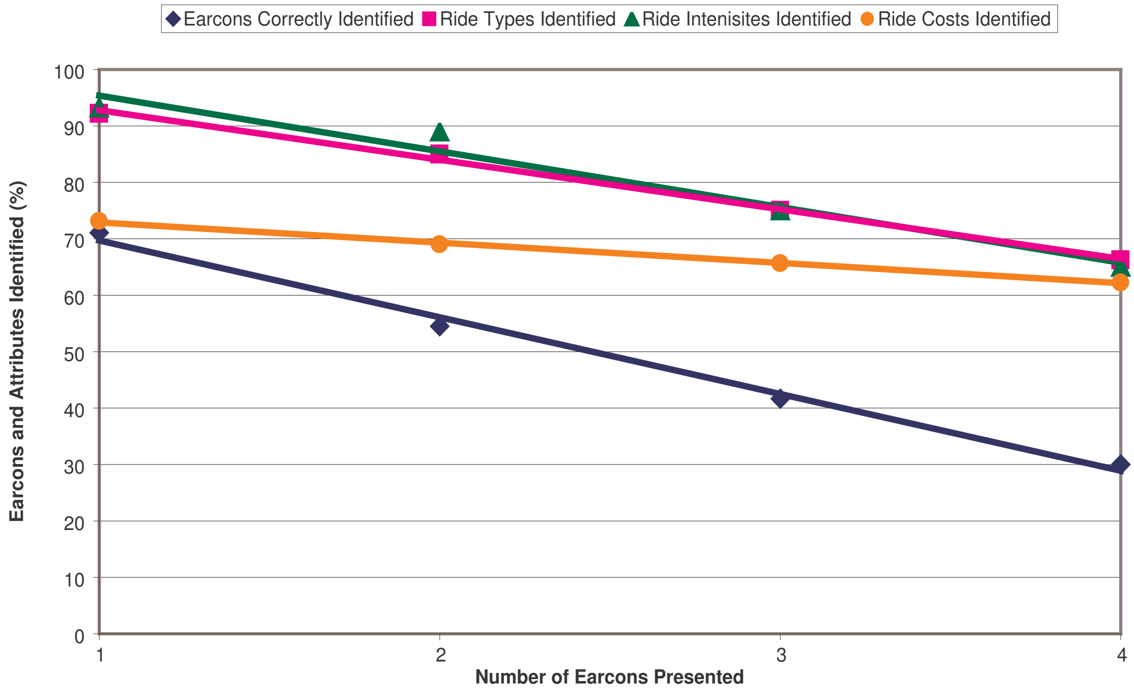

Figure 5.7: Graph showing best fit trend lines for the one, two, three and four earcon conditions, showing how the proportion of earcons and earcon attributes correctly identified is affected by the number of concurrently presented earcons.

correctly identified. The trend for the percentage of correctly identified earcons agrees both with the work of Brungart, Ericson and Simpson [29], and Gerth [57], described in Section 4.6. However, it shows the same relatively steep drops in performance as Brungart, Ericson and Simpson who investigated the effect on the identification of a target CRM (Coordinate Response Measure) phrase when the number of concurrent talkers were increased. They found for zero, one, two and three interfering talkers, similar levels of identification to the proportion of earcons identified in the one, two, three and four earcon conditions. This indicates that modifications undertaken by Brungart, Ericson and Simpson to improve identification in concurrent talker environments may be useful in improving the identification of concurrently presented earcons. In addition, Figure 5.7 shows that whilst 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 slightly inharmonic register intervals between the three registers chosen to represent ride cost (see Section 4.5.1). Indicating such intervals may reduce the likelihood of concurrent earcons interfering with each other, and they will be better identified.

Whilst 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.

As discussed in Section 4.4, 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 performance. Additionally, the earcons used encoded three separate data attributes, which is around the maximum number which 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 therefore the results of this experiment 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.

5.4.1 Guidelines for Concurrent Earcon Presentation

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

- Increasing the number of concurrently presented earcons significantly reduces the proportion of the earcons which 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.

- 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.

5.5 Conclusions

In conclusion, this chapter has provided answers to one of the research questions posed in Section 4.7 namely, RQ1: *"What is the effect on earcon identification of varying the number of earcons which are concurrently presented?"*. Increasing the number of concurrently presented earcons from one to two reduces the proportion which can be successfully identified from 70% (identification levels similar to work on earcons undertaken by Brewster [16]), to around 50%, with further increases in the number of concurrently presented earcons reducing the proportion of earcons successfully identified even more. Whilst concurrent earcon identification is poor, it should be recalled from Section 4.4 that it is not always the case that all parameters of an earcon are required, and the identification of individual earcon parameters faired much better, with identification for the timbre and melody encoded attributes falling from 90% to 70% as the number of earcons concurrently presented was increased from one to four. Additionally, three data attributes were encoded in each earcon (the maximum realistic number possible); this task can therefore be seen as a worst case scenario, with a reduction in the complexity of each earcon possibly increasing the proportion of earcons identified. However, it is clear that trying to identify concurrently presented earcons is a difficult task. It is important therefore to understand how the design and presentation of earcons can be modified to reduce the impact on their identification when the number of earcons concurrently presented is increased.

Chapter 6

Engineering More Robust Earcons

6.1 Introduction

As shown in the previous chapter, varying the number of earcons which are concurrently presented to a user has a significant effect on the proportion which can be successfully identified. With an increase in the number of concurrently presented earcons causing a significant reduction in the proportion of earcons that can be successfully identified.

Although reducing the number of concurrently presented earcons significantly increases the proportion which can be identified, there are still specific interactions between them that are likely to prove problematic. Take as an example problem two earcons as used in Dolphin (see Section 4.4) representing a high intensity, high cost rollercoaster, and a medium intensity, high cost rollercoaster. In other words, two earcons which because of the grammar they are derived from, share the same timbre, and will be played in the same register; they will differ only in melody (see Figure 6.1 (A) and (B)). If these two earcons were concurrently presented, Auditory Scene Analysis (ASA) (see Section 2.6) indicates that the only information available to perceptually segregate the earcons would be the listener's *a priori* knowledge of the melody of the two earcons. In addition, the shared timbre and register of the earcons would weight the auditory system to group the two earcons in the same auditory stream (see Figure 6.1 (C)). This makes the user's task of identifying the earcons difficult, if not impossible.

The only way to overcome such problems when concurrently presenting earcons is through changes to the design and presentation of those earcons. One way of doing so is to vary one of the other earcon attributes as well as the melody, such as presenting each earcon with a distinct timbre. The guidelines of

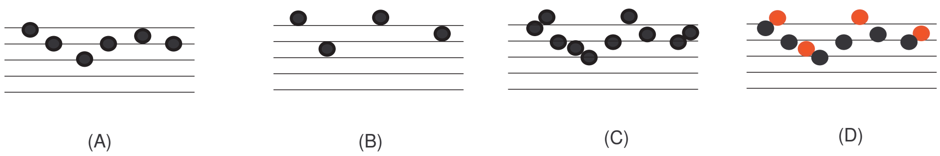

Figure 6.1: Two melodies representing, (A) high intensity, and (B) medium intensity, high cost theme park rides. (C) The same two earcons concurrently presented with the same, and (D) with different timbres (denoted by note colour). Note how the addition of timbre differences makes the earcons easier to separate.

Brewster, Wright and Edwards [25] (BWE) acknowledge this is an effective way of discriminating between earcons. Compare the concurrent presentation of the earcons from Figure 6.1 (C) where they are presented with the same timbre, and as shown in Figure 6.1 (D) where they are presented with different timbres (denoted by a different colour). Including this difference in timbre helps to make the two earcons stand out more, and will encourage them to be placed in separate auditory streams.

Modifying the timbre in addition to the melody would provide more information for the auditory system to separate the earcons (see Section 2.6). However, the strength of the transformational and hierarchical earcon types, as already discussed in Section 3.3.4, is due to the grammar that a particular set of earcons is derived from. If arbitrary changes are made to earcons in order for them to be more easily discriminated, the grammar that connects them is likely to be destroyed. Therefore, in making modifications to the design and presentation of earcons for concurrent presentation situations, a tension exists between modifying the earcons enough so that even similar earcons can be uniquely identified, but not modifying them so much that the grammar that binds them together (and thus provides the mapping between data and sound), is destroyed.

This chapter therefore, investigates the issue of separating concurrently presented earcons, and provides answers to one of the research questions in Section 4.7 (RQ2), *"How can concurrently presented earcons be more robustly designed and presented in order to increase the number which can be identified?"*.

6.1.1 Earcons and Auditory Scene Analysis

The experiment described in this chapter applies ASA research (see Section 2.6) to the design and presentation of concurrently presented earcons in order to improve their identification. There are several reasons why auditory scene analysis is an appropriate technique to apply to concurrent earcon identification. Earcons are modified and designed according to musical parameters such as timbre, rhythm, melody etc. (see Section 3.3.4) and much auditory scene analysis research can easily be considered in terms of these parameters. For example, Singh's [129] and Singh and Hirsh's work [131] on the influence of harmonic structure in auditory scene analysis (as mentioned in Section 2.6.2) can be considered in terms of timbre (see Section 2.3.2), a key earcon design parameter [24]. The work of van Noorden [139] (see Section 2.6.2), which investigated the effect of frequency separation and repetition rate on auditory streaming can be considered in terms of melody, and the register of earcon presentation. Frequency, in the context of van Noorden, can be considered to be a component of pitch (see Section 2.3.2), and hence as the register component from the guidelines of BWE (see Section 3.3.4). Understanding the relationship between low level attributes of sound as investigated by ASA research, and higher level auditory parameters used to design earcons (as discussed in Chapter 2) makes it straightforward to consider ASA features which may improve the identification of earcons. In addition, the use of auditory scene analysis was accurate in predicting that the identification of concurrent earcons would be problematic (see Section 4.3 for a discussion, and Chapter 5 for evidence to support this), therefore auditory scene analysis should also be able to offer useful modifications to concurrently presented earcons to make them more easily identifiable.

Whilst the application of auditory scene analysis to the identification of concurrent earcons should be productive, there are several issues with the design of most auditory scene analysis studies which need to be borne in mind when applying such work to the identification of concurrently presented earcons. Although a large body of work which identifies the effect of changing individual auditory parameters on the perception of a sinusoidal tonal sequence exists, there is little research which investigates the relative influence of multiple changes to auditory attributes on the perception of a sound source (see Section 2.6). Therefore all modifications to the design and presentation of concurrent earcons which prove to be effective should be combined and evaluated to ensure they do not conflict, and thus reduce earcon identification performance. Auditory scene analysis research tends to involve a participant listening to a composite sound, whilst some auditory parameter is gradually changed until the perception of the sound a participant is listening to changes by either being perceived as one, or two, distinct auditory streams [14]. Identifying earcons is more complex, requiring the user to identify the auditory parameters of each earcon and recall the mapping of data to those parameters. Therefore, although experimental work may show that certain features of a sound will influence it to be perceived as either one or two streams, it may not be the case that the application of those features to earcon design will have a significant impact on concurrent earcon identification. Additionally, any modification to the design of earcons is likely to constrain the earcon grammar in some way, for example reducing the number of timbres which can be used to represent different values of a data parameter. Therefore it is important to compare each auditory scene analysis modification to a set of standard earcons, compliant with the earcon design guidelines of BWE (see Section 3.3.4), to identify if that modification actually does improve concurrent earcon identification, and therefore ensuring unnecessary constraints are not placed onto earcon designers.

Because BWE's earcon design guidelines have been experimentally evaluated [16], it is known how well earcons following these guidelines are identified when presented in isolation to each other. Due to the already mentioned lack of clear design guidelines for presenting concurrent audio as part of an auditory display, and the lack of clarity on the specific impact of individual auditory scene analysis features using BWE's guidelines provides a solid basis to build upon and is a logical starting point to investigate modifications to improve the identification of concurrently presented earcons. This approach may also allow for existing earcon sets to be "retrofitted", at minimal additional design effort, to allow for there presentation of those earcons in concurrent presentation situations.

6.1.2 Applying Auditory Scene Analysis to Earcons

Since earcons are extremely complex in the context of ASA research, and there is little guidance on how different ASA modifications to a sound interact, the approach of the experiment described in this chapter is to apply ASA research to modify the design and presentation of earcons in a way that is likely to directly increase the identification of specific earcon attributes.

The remainder of this section will discuss in detail two aspects of ASA and how they will be applied to redesign earcon parameters (timbre and melody modifications), and two features which will alter the presentation of the earcons (timing and training).

Modifying Earcon Timbre

One feature that is important in ASA is timbre [14]. Whilst as discussed in Section 2.3.2, there is no universal definition of timbre, known elements have been found to be important in auditory scene analysis. Singh [129, 130], Cusack and Roberts [36], and Singh and Hirsh [131] found that alternately presenting two complex tones (A and B), each of which contained a different number of harmonics (see Section 2.3.2), a known element of timbre, required less separation between the fundamental frequency of the A and B tones to be heard as two separate streams (one composed of A tones and one composed of B tones), than when each tone contained the same number of harmonics. Brungart, Ericson and Simpson [29], found that in a CRM (Coordinate Response Measure) identification task, having different texts spoken by different sex speakers, effectively changing the timbre of the speaker's voice, produced a 20% improvement in the identification of a target spoken phrase.

Timbre therefore exerts a strong influence in how an auditory scene is perceived and how audio is grouped into streams. This thesis claims that in cases where concurrent earcons are presented which are of the same ride type, thereby sharing the same timbre, the influence the shared timbre exerts on the grouping of the earcons is so strong that it greatly reduces the effect of other differences between the earcons which would be used to influence their streaming, such as register and melody.

If each concurrently presented earcon was presented with a different timbre, the influence that timbre exerts would work at separating rather than merging the earcons. Unfortunately, since timbre forms part of the earcon grammar, arbitrary changes which are likely to have the maximum improvement in earcon discrimination are also likely to destroy the grammar, and therefore make earcon identification difficult, if not impossible. It may be possible however to use slightly different timbres for each earcon, which would affect the influence timbre exerts in the auditory streaming process without adversely affecting the interpretation of the earcons. Rigas [118], as described in Section 2.3.2 grouped instruments into subjectively similar groups. Therefore instead of mapping a particular ride type to a single timbre, it could be mapped to a group of similar timbres, any one of which could represent that ride type. In such a way, if more than one earcon of a particular ride type was presented concurrently, each would have a different timbre from the same timbre group. Whilst the earcon design guidelines of Brewster, Wright and Edwards [25] recommend against using very similar timbres, it is not important that participants can explicitly differentiate between the different timbres used for each ride type; only that they can differentiate between the three groups of timbres used for different ride types. The different timbres for each ride type should weaken the effect of timbre in the grouping of concurrently presented earcons, and increase the importance placed on other differences between concurrently presented earcons (such as melody and register) in the grouping process.

One issue with such a solution however is that a sufficient number of instrument groups must be available (one for each distinct data value encoded by earcon timbre), and each instrument group must contain a sufficient number of instruments (one for each earcon that is to be concurrently presented which encodes the same data value). It may be difficult to identify enough distinct instrument groups to represent all data parameters that a user may wish.

Modifying Earcon Melodies

One ASA feature, which can be categorised under the common fate principle (see Section 2.6.2), is the robustness of tones which glide in frequency to be grouped together. That is, sequences of tones which consistently rise, fall or stay the same in frequency over time (see Figure 6.2 for an example). Bregman [14] notes that when listeners were presented with a frequency glide, which was occluded in places by broadband noise, thereby masking (see Section 2.3.2) part of the frequency glide, listeners associated the frequency glide as continuing through the noise burst, as opposed to hearing two separate frequency glides on either side of the noise. Van Noorden [139], as discussed in Section 2.6.2, found that if two sequences of tones were concurrently presented they could be forced to be perceived as being separate by increasing the frequency difference between the tones of each sequence and reducing the time between consecutive tones, thereby increasing the tempo, or playing speed of the sequence [116]. Frequency glides conform to the research of van Noorden with small increases in frequency (pitch in the context of this thesis (see Section 2.3.2)) between consecutive notes fusing those notes into a coherent stream.

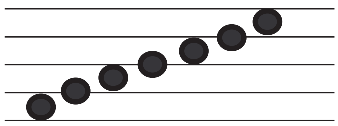

Figure 6.2: Example of an upward frequency glide.

One problem with this approach is that if two glides are played concurrently, they may cross over each other (see Figure 6.3 (A)). Studies which have investigated this have shown that at the point of intersection of the two glides, grouping will occur by frequency proximity, rather than by common fate [136, 58, 39]. The example glides in Figure 6.3 (A), would therefore be perceived as two glides which bounce apart (Figure 6.3 (B)), rather than two glides which cross (Figure 6.3 (C)). Halpern [58] identified that if the timbre of one of the glides was different, the glides would be heard as shown in Figure 6.3 (C). Such a result was also identified by Tougas and Bregman [136]. However, as timbre is loosely defined (see Section 2.3.2), 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 3.3.4). 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

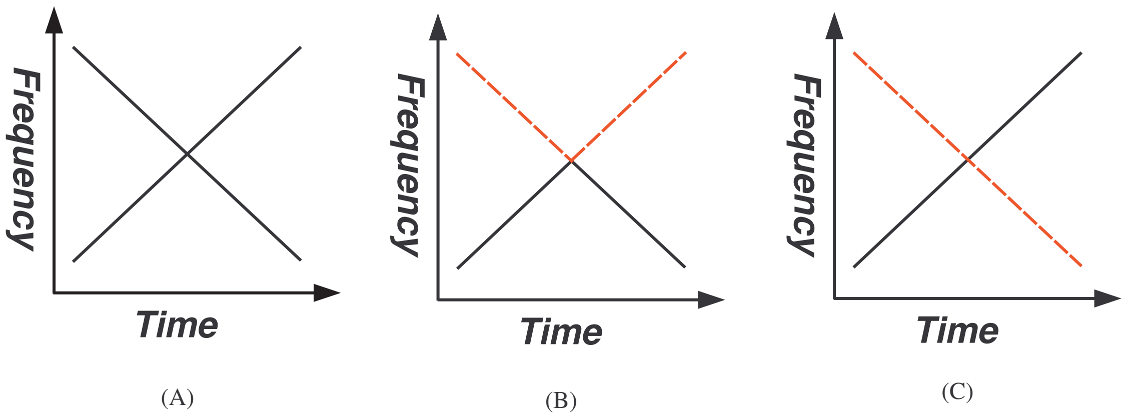

Figure 6.3: An example of two crossing frequency glides (A) and how they will be perceived by the auditory system if they are presented with the same timbre (B) and with different timbres (C).

timbre, which might be enough for frequency glides to stream separately. This issue is further discussed in Section 6.2.3.

Onset Asynchronicity of Earcons

Whilst modifications to the design of earcons may improve their identification, it may also be the case that modifying the presentation of concurrent earcons will increase their identification.

Auditory Scene Analysis research shows that sounds which start at different times tend to be grouped separately. Research which has investigated temporal differences in the presentation of sound has shown that such differences exert a strong influence in how those sounds are grouped. Bregman and Pinker [15] found that when a tone formed of two separate, harmonically related tones (a complex tone), was presented asynchronously with a third tone, the greater the onset to onset difference that was introduced between the two tones which comprised the complex tone, the more likely it was that the three tones would be perceived as a melodic sequence rather than an alternating complex and simple tone. Therefore as the time between the presentation of the harmonic tones increased, the influence of the harmonic relationship in the grouping process was reduced as the influence of the temporal distance between both of the tones increased. Deutsch [38] found that when melodies, the components of which were alternately presented in opposite ears, with a corresponding noise in the other ear, that introducing onset-to-onset differences between the melodic component and the noise of 15ms significantly improved the identification of the melody. Darwin and Ciocca [37] showed that mistiming an harmonic component in a complex tone by 300ms removed the contribution of the mistimed component to the perceived pitch of the complex tone (see Section 2.3.2). Therefore, if a short gap is placed between the starts of concurrent earcons, will this allow them to be better identified?

Increased Earcon Training

As described in Section 3.3.4, because the relationship between an earcon and the data it represents is abstract, training participants to correctly identify earcons is important if they are to be effectively used in a human computer interface [18]. In a similar way, there is evidence which suggests that a lot of auditory scene analysis situations are influenced by the listener's prior knowledge of the stimuli [14]. For example, work by Dowling [40] found that familiarising participants with an arbitrary tone sequence improved the detection of that sequence when it was concurrently presented with other tones. Given that learning individual earcons is important, and learning of melodies improves their detection in a composite sound, will familiarisation with concurrently presented earcons improve their identification?

6.1.3 Why Not Spatial Audio?

One feature of concurrent earcon presentation, that will be absent from this chapter, is the use of a distinct spatial location for each concurrently presented earcon. The advantages that the incorporation of spatial audio presentation brings to auditory displays have already been discussed in Section 3.4. Spatialised audio allows for an increase in the amount of data that can be concurrently presented, as well as a new auditory parameter (the spatial location of sound) to encode a data attribute [124, 67, 141].

This thesis argues that whilst spatial audio can be used to improve concurrent earcon identification (see next chapter), there are several reasons why investigating non-spatial concurrent earcon design and presentation modifications in isolation to spatial presentation is advantageous. Firstly, there are several situations in which good quality spatial audio may not be practical. As explained in Section 2.4, the perception of spatial location is largely dependent on the differences in auditory cues reaching the left and right ears. This means that to use spatial audio, a user would need to either wear headphones, or have a set of properly calibrated speakers. There are several application areas where this may be difficult. If concurrent spatialised earcons are used in a mobile telephone interface [73], a user may not wish to wear headphones just to perform simple and short interactions, thereby rendering any advantages of spatial audio useless. Additionally, in cases where the spatial location of information is used to encode a data parameter, such as was the case in Dolphin [79, 80] (see Section 4.4), it may not be possible to move audio sources far enough apart to allow them to be distinct, this may be a particular problem due to the similarity of earcons. Recall from Section 2.4.2 that to separate concurrently presented sounds in space is as much a function of their similarity as their angular separation. Hence, an understanding of how earcons perform when concurrently presented in a non-spatialised environment, and how identification can be improved in such environments is essential.

6.2 Methodology

Sixty four participants between the ages of 18-24 undertook this experiment, which was of a between groups design. As already stated in Section 5.2.1, this experiment was originally intertwined with the experiment from Chapter 5, but was later split, with each condition in this experiment being coupled with a condition from the experiment described in Chapter 5. The pairing of conditions from this experiment and the experiment as discussed in Chapter 5 is provided in Appendix D. There were five conditions (the *original earcon set condition*, the *melody altered earcon set condition*, the *multi-timbre 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 was paid £5 on completion. As with the experiment from Chapter 5, each condition consisted of a training phase and a testing phase, which are described in Section 6.2.1 and Section 6.2.2. The earcons used in this experiment were the same as those used in Dolphin (see Section 4.4), however various modifications, dependent on the experimental condition, to the design and presentation of the earcons were carried out. These modifications are described later in this section. Each earcon encoded three parameters of a theme park ride: the ride type (either rollercoaster, 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 earcons and their attributes which had been modified according to ASA principles would be significantly better identified than earcons which only complied with the guidelines of BWE. The independent variable (IV) was the modification undertaken to the earcons, and the dependent variables were the number of earcons, ride types, ride intensities and ride costs correctly identified. Additionally it was hypothesised that modifications to the earcons would significantly reduce the subjective workload of participants. Modified NASA TLX questionnaires were used evaluate this hypothesis.

6.2.1 Training Phase

As with the experiment described in Chapter 5, the training phase provided participants with an opportunity to learn the earcons that they would be asked to identify in the testing phase.

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 the earcons derived from the grammar (see Section C of the accompanying CD-ROM for the Web pages used for each condition (Appendix M)). Participants were allowed to listen to the earcons individually for ten minutes. After ten minutes participants were asked to 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 five minutes. After this time the test was retaken. If participants were still unable to correctly identify the three earcons they took no further part in the experiment. One participant failed to complete the training phase.

6.2.2 Testing Phase

The testing phase of this experiment was similar to the testing phase used for the experiment in Chapter 5. It comprised of participants listening to twenty sets of diotically, 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 twenty sets of concurrent earcons were used for all conditions, and were randomly presented to avoid any 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 5.3 for an example of the dialogue box used). As with the previous experiment, a mandatory four second break was introduced between successive sets of earcons to allow the Auditory Scene Analysis mechanism to fully reset [14].

Before carrying out the testing phase proper, participants carried out a shortened version involving two sets of earcons not used in the experiment, in order to familiarise themselves with the experimental condition.

6.2.3 Conditions

This section outlines the conditions used in the experiment, each of which incorporates an addition (informed by ASA) to either the design or presentation of earcons which are compliant with the earcon design guidelines of BWE [25].

Original Earcon Set Condition

The earcons used in this condition incorporated no modifications to either their design or presentation. The earcons were the same as those described in Section 3.3.4, and the presentation was the same as used in the experiment described in the previous chapter. As such this condition is the same as the four earcon condition of the previous experiment, and, instead of re-running that condition, the data from the four earcon condition (see Section 5.2.2) were used in this experiment.

Multi-Timbre Earcon Set Condition

As discussed in Section 6.1.2, presenting each earcon with a different timbre may improve their identification. Whilst as discussed in Section 4.3, it would not be possible to use grossly different timbres for each presented earcon as this would destroy the grammar of the earcons (e.g. presenting two concurrent roller-coasters, one with a trumpet and another with a piano), using a set of similar timbres for each ride type which may be used interchangeably to represent that ride type attribute could be employed. This would ensure that no two earcons were concurrently presented with the same timbre, whilst not destroying the "grammar" that interconnects the earcons and provides their advantages (see Section 3.3.4).

Since timbre is not fully understood, determining what is and is not a similar timbre is difficult. The work of Rigas [118], which identified groups of MIDI instruments which were subjectively determined to be similar (see Section 2.3.2), was used to provide guidance in selecting similar timbres. Three groups of timbres were chosen, and three similar timbres from each group (there were no cases in the 20 stimuli sets of 4 earcons that all four earcons had the same ride type) were chosen to represent each ride type. Table 6.1 shows the instruments and their general MIDI patch numbers used to represent rollercoasters, static rides and water rides in the multi-timbre earcon set condition. Whilst BWE earcon design guidelines recommend against MIDI patches so similar to each other as those in Table 6.1, it is not important that participants can differentiate between the three timbres used for each ride type, only that they can differentiate between the three groups of timbres used for different ride types. The different timbres for each ride type should however weaken the effect of timbre in the grouping of concurrently presented earcons, and increase the importance placed on other differences between concurrently presented earcons in the grouping process, such as melody and register.

This condition does however place limitations on the timbres that can be used to encode data parameters. It is necessary to be able to determine enough timbre groups for each data parameter which would be encoded as timbre, and for those timbre groups to contain enough timbres such that there is one distinct timbre for each earcon (encoding the same parameter) which may be concurrently presented. Therefore if three earcons with the same timbre mapped data parameter are concurrently presented, there would need to be three separate timbres from the same timbre group available so that each earcon was played with a different timbre.

Because Rigas [118] used MIDI instruments in his categorisation experiment, and MIDI does not define exactly the timbres that are used by a midi synthesiser or how they are generated, there is a variation in the timbres of different synthesisers. With the synthesiser used in this work (Roland super JV-1080), in order to have three groups of instruments as distinct as possible, a Marimba, Shamisen (a plucked Japanese instrument) and Kalimba (a “thumb” piano) were used to represent water rides. On the synthesiser used these were subjectively distinct from the other groups and were subjectively similar to each other, whereas other groups that might be used according to the guidelines of Rigas, seemed subjectively similar when played on the synthesiser used in this work.

Ride Type	**Instrument**	**General MIDI Patch No.**
Rollercoaster	Acoustic Grand Piano	000
Rollercoaster	Electric Acoustic Piano	001
Rollercoaster	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

Table 6.1: Table showing the timbres used to represent ride types in the multi-timbre earcon set condition.

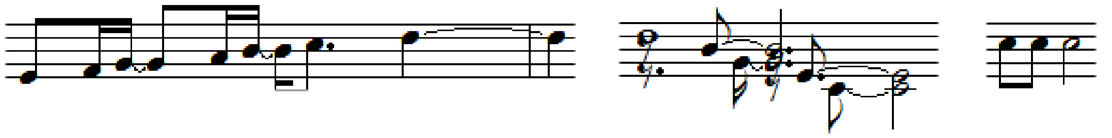

Figure 6.4: Examples of the melodies used for high, medium and low intensity theme park rides as used in the melody altered earcon condition.

Melody Altered Earcon Set Condition

As previously discussed, sequences of tones which consistently rise, fall or stay the same in pitch are more robustly segregated by the auditory scene analysis system (see Section 2.6). In this condition, the melodies of the earcons from Section 4.5.1, which were used in the original earcon set condition, were altered to use frequency glides in the hope that this would increase their robustness and ability to be identified when concurrently presented. The revised melodies for high, medium and low intensity theme park rides are shown in Figure 6.4.

As also previously discussed, if two glides cross each other, they are likely to group according to pitch proximity rather than by common fate. However, if each glide is presented with a different timbre, then this will assist the desired grouping of the glides. It is unclear however, how different two timbres would need to be to create such a situation. It may be the case that the changes in timbre between concurrent earcons provided by the multi-timbre earcon set condition would be sufficient. However, the melody altered earcon set condition places severe limitations on the number of possible melodies which can be incorporated into an earcon grammar. Further limitations on the grammar are imposed by the multi-timbre earcon set condition. It would be unwise if both conditions were combined before each had been shown to be individually useful, in case one was not effectual in increasing the identification of earcons, and guidelines from that condition unnecessarily placed constraints on the design of earcons for concurrent presentation. In addition, most earcon stimuli sets used in the experiment which had a different melody were also accompanied by changes in another parameter. Therefore if this cue is effective, an improvement in concurrent earcon identification may be found without explicit modifications to other earcon parameters.

Staggered Onset Condition

As previously discussed in Section 2.6.2, staggering the onsets of two tones can have a significant influence in the way they will be grouped by the auditory system. This condition investigated if staggering the onsets of concurrently presented earcons has a similar effect in improving the identification of those earcons. In this condition although the four earcons in each set were still presented concurrently, a 300ms onset-to-onset delay was introduced between the presentations of each earcon. An example of the presentation method used is shown in Figure 6.5. The order of presentation of the earcons in each earcon set was randomised between participants to avoid any order effects affecting the results, due to one earcon always being presented before another. In terms of Figure 6.5, the earcon allocated as earcon 1, earcon 2 etc. was randomly varied between participants.

Extended Training Condition

As previously discussed, familiarity with identifying concurrent earcons would be expected to increase their identification. In this condition a second testing phase was introduced between the training and testing phases, which presented participants with the application shown in Figure 6.6. Participants were allowed to listen to two sets of four concurrently presented earcons, which were not used in the testing phase. This allowed participants to switch on and off individual earcons in order to understand the kinds of interactions that might occur when multiple earcons are concurrently presented. The list at the bottom of the application (see Figure 6.6), shows all possible permutations of the four earcons. Clicking a list item changed the order of the earcons in the top part of the application to the selected permutation. Training was self guided and participants were instructed to listen to the changes that occurred when a new earcon was added or removed from the mix. Participants had to select all of the possible permutations in the list before finishing the training. On average this training phase took twenty minutes to complete.

6.3 Results

6.3.1 Identified Earcons and Attributes

As with the previous experiment, participants' responses to the earcons presented were collected, and the number of correctly identified earcons, ride types, ride intensities and ride costs 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 multi-timbre earcon set and staggered onset earcon conditions over the original earcon set condition. To determine the overall improvement in earcon identification, the features

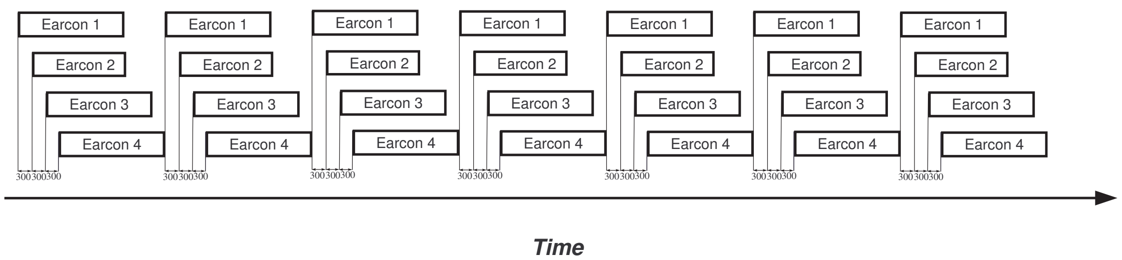

Figure 6.5: An example of concurrent earcon presentation from the staggered onset earcon condition. The four earcons are still concurrently presented seven times, however a 300ms onset-to-onset delay between the start of each earcon is introduced.

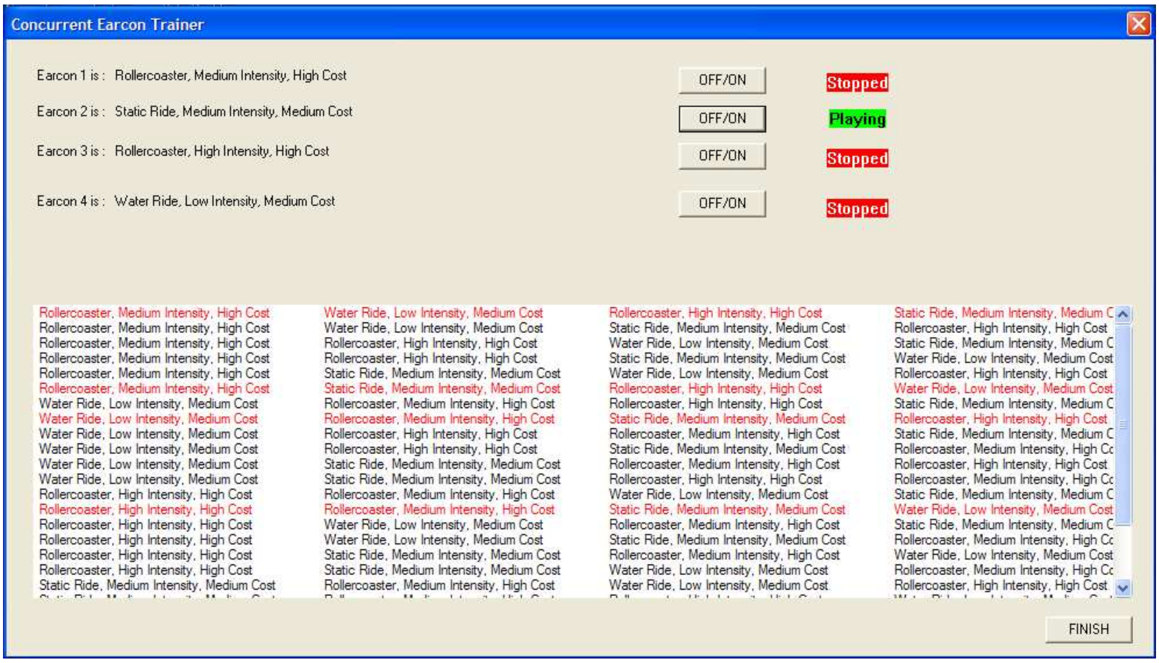

Figure 6.6: A screenshot of the concurrent earcon training tool, designed to give participants an opportunity to understand how earcons interfere with each other when concurrently presented.

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 6.7. Raw data are presented in Appendix H.

To determine if any of the differences in the data presented in Figure 6.7 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 number of earcons identified in the staggered onset and final conditions were significantly greater than the original earcon set condition ($p < 0.05$). 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 multi-timbre earcon set, the staggered onset and the final conditions had a significantly greater number identified than in the original earcon set condition ($p < 0.05$). 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 ride intensity in the staggered onset and multi-timbre earcon set conditions were significantly better identified than in the melody altered earcon set condition ($p < 0.05$). The ANOVA for ride costs identified was not significant ($F(5,90) = 0.31$, $p = 0.907$).

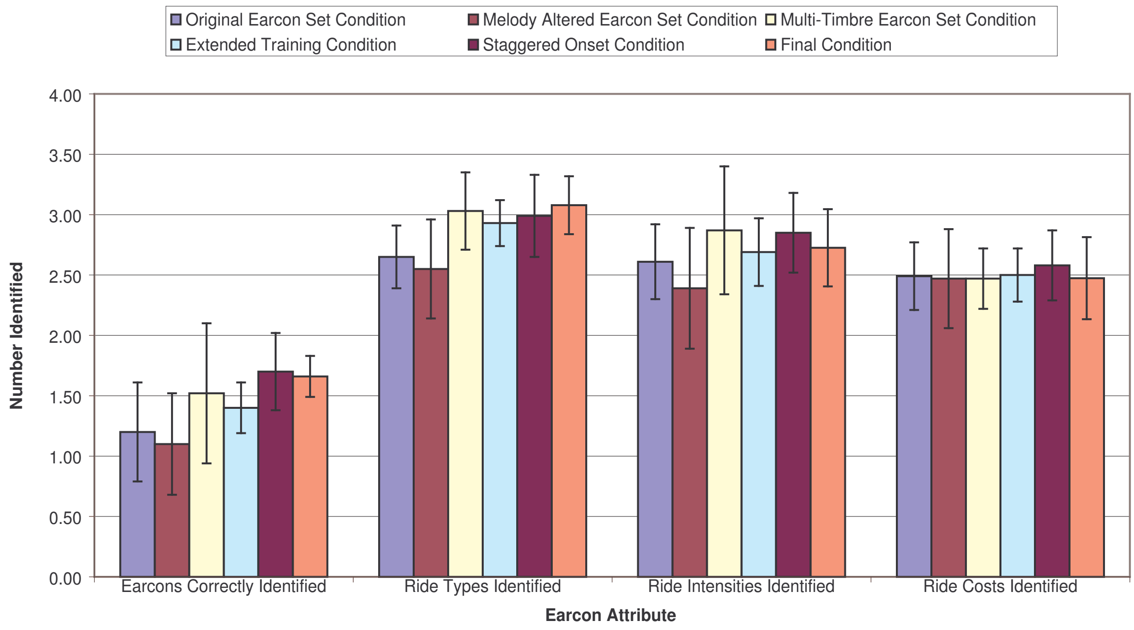

Figure 6.7: Graph showing the average number of earcons, number of correctly identified ride types, ride intensities, ride costs and their standard deviations for the original earcon set, melody altered earcon set, multi-timbre earcon set, extended training, staggered onset and final experimental conditions.

6.3.2 Workload Data

As with the experiment of Chapter 5, NASA TLX workload ratings were recorded. The average score for each rating in each condition is presented graphically in Figure 6.8. Raw data are presented in Appendix H. 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 analysis of variance (ANOVA) on these data failed to show significance ($F(5,90) = 0.69$, $p = 0.629$).

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 ($p < 0.05$).

6.4 Discussion

The results show that both the multi-timbre 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 which were correctly identified in the staggered onset condition was also significantly greater than in the original earcon set condition.

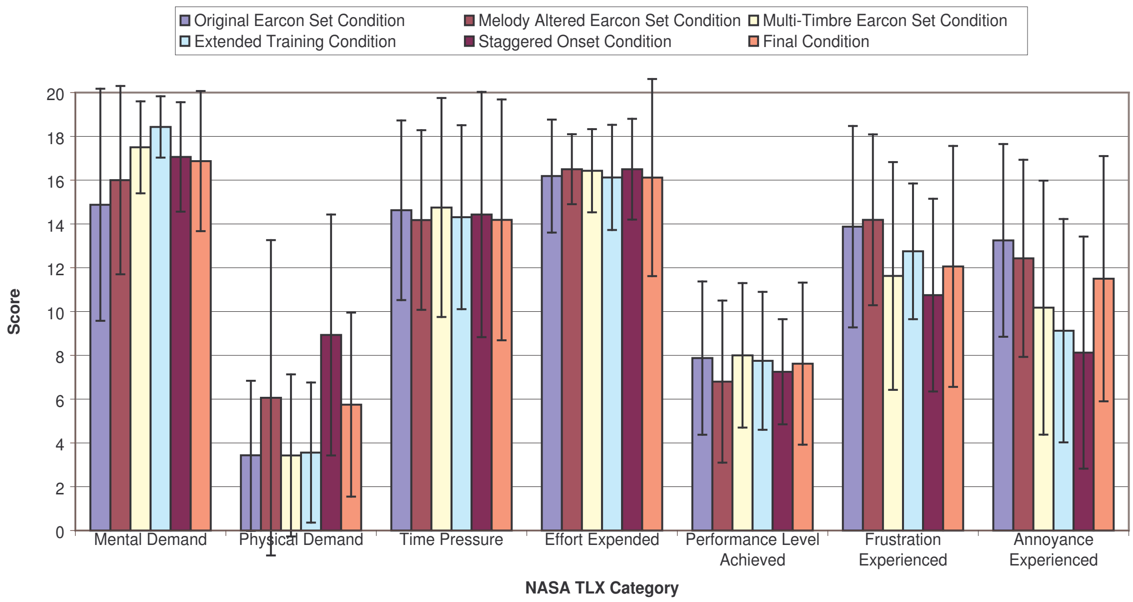

Figure 6.8: Graph showing the mean values for NASA TLX workload data with standard deviations for the original earcon set, melody altered earcon set, multi-timbre earcon set, extended training, staggered onset and final experimental conditions. Shown with standard deviations.

The final condition which incorporated the staggered onset condition and multi-timbre 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 multi-timbre earcon set conditions. Given that both the staggered onset and multi-timbre earcon set conditions significantly improved the identification of ride type, they may both have solved the same earcon interaction problems (see Section 6.1).

Whilst 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. This result again echoes the claim of Buxton from Section 3.2.2, that by improving the effectiveness of auditory displays they will be less annoying to users. 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 appears to outperform the melody altered earcon set condition in the number of earcons, ride types, ride intensities and ride costs correctly identified. As was discussed in Section 6.2.3 describing the melody altered earcon set condition, the identification of crossing 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 which had different ride intensities would also vary in timbre, thereby assisting the correct interpretation of the melody [58]; or both earcons would be played in different registers, and as such would be inharmonicaly separated so that the melodies would not actually cross in pitch. 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 multi-timbre 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 may therefore be required.

The issues with the melody altered earcon set condition provide supporting evidence to the claim made in Section 6.1.1, that because certain modifications to sounds have been shown to influence how sounds are grouped together to form auditory streams, it can not be assumed that incorporating such features into the design of concurrently presented earcons will lead to significant improvements in the identification of those earcons. This is especially important to consider given that certain ASA features, such as those incorporated into the melody altered earcon set condition, constrain the design of the earcon grammar and as such, the number of different data values an earcon is capable of encoding. Therefore identifying that a specific ASA feature improves the identification of concurrent earcons must be experimentally determined, and cannot be assumed.

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 multi-timbre earcon set conditions 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 (see Figure 6.6). On average participants spent 13 minutes on the first training session, and 8 minutes 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.

Although significant improvements in earcon identification were found in some conditions, these improvements were not large and overall identification performance was low which raises questions over the practical usefulness of the modification. However, from the results of the previous experiment described in Chapter 5, it was considered that four concurrently presented earcons of maximum complexity was a "worst case scenario". 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, have shown that whilst increasing the number of concurrently presented sounds still causes a reduction in performance, that reduction is much less severe [29]. Hence if modifications to the design and presentation of earcons were carried out, the gradient in the graph from Figure 5.7 would be expected to be much flatter.

Although modifications to the design and presentation of Earcons were found to have significantly improved their identification, in order to avoid destroying the grammar of the earcons and therefore making it impossible to retrieve the data encoded within them, the degree of such modifications were constrained. The degree of such constraints 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 modification to concurrently presented earcons were all applied to a pre-existing set derived from the guidelines of BWE (see Section 3.3.4), and used only those musical elements that the guidelines discuss. For example, the use of dynamics (the variation in the degree of loudness of a composition [116]) was not used as part of the earcon design. Additionally some of the aspects of BWE's guidelines were grouped, such as using melody to represent ride intensities rather than just using rhythm, since Brewster's work shows that combinations of parameters in such a way allows for earcons to be more robustly identified. Whilst following BWE's guidelines as a starting point to investigate the issues of concurrently presented earcons is logical (since they are the only evaluated guidelines for earcon construction), it may be that more practically useful results would be found by a radical redesign of the earcons rather than the "augmentation" of those earcons with specific modifications. Such a redesign would remove the reliance on the underlying guidelines of BWE, however it would be necessary not only to determine how well those earcons were identified concurrently but also how they could be parameterised to encode information and how easily the mappings between earcons and the data they represent can be learned, issues which the work of Brewster [16] has already investigated and provided solutions for.

Guidelines for Concurrent Earcon Presentation

From the results of the experiment undertaken in this chapter, the following guidelines for the concurrent presentation of non-spatialised earcons can be derived. These guidelines will help designers who need to concurrently present earcons do so more effectively, allowing the earcons to be more accurately identified.

- When timbre is used to encode a data parameter, each concurrently presented earcon should have a different timbre. The guidelines of Brewster, Wright and Edwards [25], 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 [118] can be used to determine distinct instrument groupings.

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

6.5 Conclusions

This chapter has provided answers to one of the research question posed in Section 4.7, namely RQ2: *"How can concurrently presented earcons be more robustly designed and presented in order to increase the number that can be identified?"*. The experimental work from this chapter has shown that by presenting each earcon with a unique timbre as well as having a 300ms onset-to-onset gap between the starts of concurrently presented earcons, provides a significant increase in the number of earcons and their attributes which can be successfully identified. Increasing the training given to participants, and modifying the melodies of the earcons did not lead to significant improvements in earcon identification. Additionally, modifying the timbres of concurrently presented earcons only lead to an increase in the identification of the timbre encoded earcon attribute. It did not lead, as suggested in Section 6.1.2, to a significant reduction in the influence of timbre in the grouping process, leading to a significant increase in the number of earcons which could be identified. Due to the need to avoid destroying the grammar, the degree of modifications which could be applied to concurrent earcon presentation was limited. Incorporating larger modifications between the earcons would improve streaming, but would put the earcon grammar at risk. The results showed that although there were significant improvements in identification those improvements were small and of questionable "real world" practicality. Therefore, it may be more effective to reduce the number of data parameters encoded in each earcon, "freeing up" an auditory parameter which can incorporate large arbitrary changes without destroying the earcon grammar. However such an approach would lead to a reduction in the data which could be encoded in each earcon. The guidelines derived from this work however, should allow designers who require concurrent earcon presentation to make more informed decisions when designing auditory displays.

Chapter 7

The Impact of Spatialisation on Concurrent Earcon Identification

7.1 Introduction

In the previous two chapters, work investigating how the identification of concurrently presented earcons was affected by varying the number of earcons concurrently presented, as well as modifying the design and presentation of those earcons to incorporate principles derived from auditory scene analysis (ASA) research was discussed. This work however does not include one feature of audio presentation which should provide a significant improvement in concurrent earcon identification: spatial presentation, where each concurrent earcon is presented in a unique spatial location. The work described in this chapter comprises two experiments designed to contribute towards answering RQ2: *"How can concurrently presented earcons be more robustly designed and presented in order to increase the number which can be identified?"*, and RQ3: *"What is the impact of presenting concurrent earcons in a spatialised auditory environment?"*, from Section 4.7.

One experiment compares the performance of non-spatialised earcons formed from the guidelines described in Chapter 6, with spatialised earcons created solely from the guidelines of Brewster, Wright and Edwards (BWE) [25]. The second experiment investigates if spatially presented earcons formed from the guidelines for concurrent earcon design and presentation from Chapter 6, are better identified than spatialised earcons formed solely from the guidelines of BWE [25].

7.1.1 Previous Spatial Research

Whilst there has been some research which has investigated systems that use spatial audio as a mean of presenting information (see Section 3.4), and further research into the basics of how spatialisation works as part of ASA, little empirical research has been undertaken to determine if the spatial presentation of concurrently presented audio is superior to non-spatial presentation in auditory displays, especially with earcons.

The only pre-existing study investigating the impact of the spatial presentation of earcons on their identification, was performed by Brewster, Wright and Edwards [26]. They were concerned that the length of time compound earcons (see Section 3.3.4) took to present was excessive, and could mean that they would fail to keep up with interactions in a human computer interface. As a way to solve these problems parallel earcons were proposed. These consisted of compound earcons which were split, with the parts of each compound earcon being presented to opposite ears, thereby halving the time taken to play each earcon. Whilst Brewster, Wright and Edwards used spatial position to separate the earcons, they did not do so in the way described in Section 2.4, or in the way which this thesis determines spatialisation. Brewster, Wright and Edwards used stereo panning, which incorporates only the IID described in Section 2.4.1; this thesis regards spatialisation of sound to involve the convolution of a sound via an HRTF (see Section 2.5). BWE also modified the earcons that would be presented to the left and right ears such that there was at least an octave difference [116] between them, thus encouraging each component earcon to be placed in a separate auditory stream. This was possible since the compound earcon parts presented to the left and right ears were formed from different grammars and as such were not similar to each other in the same way that two earcons derived from the same grammar would be. Those in the left ear represented objects such as “file”, “folder”, “application” etc. Those in the right ear represented actions for the objects such as “open”, “print” etc. Brewster, Wright and Edwards found that there was no significant difference in identification between the use of standard compound and parallel earcons. However, since they had to modify their one-element earcons to ensure the spatially separated parallel earcons would be separately streamed, separating sounds spatially may, as was also indicated by Dolphin (see Section 4.4) and the discussion of the concurrent minimum auditory angle from Section 2.4.2, not be sufficient to discriminate concurrently presented earcons when it is the only cue available. Therefore it is important to determine both the impact of spatial separation and the impact of the non-spatial presentation guidelines identified in previous chapters.

7.2 Spatial versus Non-Spatial Earcons

7.2.1 Motivation

Previous research described in Chapter 6, has shown that the identification of concurrently presented earcons based on the guidelines of Brewster, Wright and Edwards [25] can be significantly improved by presenting each earcon with a unique timbre as well as staggering the start of each earcon by at least 300ms. It is believed that presenting each earcon in a different spatial position by the use of a 3D audio environment will further improve the identification of concurrently presented earcons. The experiment in this section therefore compares the identification of earcons designed using the guidelines of BWE [25], but with each presented in a unique spatial location, with earcons which also incorporate the guidelines for concurrent

	First Training Session	First Testing Session	Second Training Session	Second Testing Session
Group 1	Original Spatial Condition	Original Spatial Condition	Revised Non-Spatial Condition	Revised Non-Spatial Condition
Group 2	Revised Non-Spatial Condition	Revised Non-Spatial Condition	Original Spatial Condition	Original Spatial Condition

Table 7.1: Table showing the procedure for the two groups of participants undertaking the spatial versus non-spatial experiment.

earcon presentation from Chapter 6, but with all earcons presented in a common spatial location. This is important to determine since if the concurrent earcon guidelines are superior to spatial presentation, this would mean that devices would not need to support the computational requirements of spatialisation and other issues involved in spatial audio as described in Section 6.1.3 to effectively present concurrent earcons. On the other hand, if spatialisation is superior to the concurrent earcon guidelines, this may remove the requirement for designers to have multiple similar timbres for each presented earcon. As described in section 6.2.3, selecting such timbres can be difficult, as well as limiting the number of potential data values that can be mapped to the timbre attribute of an earcon (see Section 6.1.2).

7.2.2 Methodology

Sixteen participants aged 18-24, none of whom had taken part in the previous experiments, undertook the experiment described in this section, which was of a within groups design and involved two conditions, the *spatial* condition and the *non-spatial* condition, both of which are explained below. Written consent was obtained from all participants prior to the experiment, and all were paid £5 on completion. Participants were randomly assigned to one of two groups to determine the order in which they undertook the conditions. Each group contained the same number of participants. The experiment followed a similar procedure to those described in Chapters 5 and 6, with each condition having a training and testing phase. The order of conditions for each group is shown in Table 7.1.

The main hypothesis of the experiment was that earcons in the spatial condition, which did not incorporate the concurrent earcon guidelines from Chapter 6 but were spatially separated, would be better identified than those from the non-spatial condition which incorporated the concurrent earcon guidelines from Chapter 6 but were spatially collocated. The independent variable (IV) was the design and presentation of the earcons, and the dependent variables (DV's) were the number of earcons, ride types, ride intensities and ride costs successfully identified.

Training Phase

The training phase provided participants with an opportunity to learn the earcons appropriate to the condition they were to perform in the testing phase, and was the same as the previous experiments. This involved participants reading a sheet describing the grammar of the earcons used (see Appendix I), followed by ten minutes of self guided training on the earcons via a Web page interface (see Section D of the accompanying CD-ROM (Appendix M)). After this time participants were asked to identify three independently presented earcons without any form of assistance or feedback. If they were unable to do so, a further five minutes of training was provided after which the test was retaken. If participants were still unable to correctly identify the three test earcons, they took no further part in the experiment. When participants had successfully completed this phase they carried onto the testing phase. No participants however failed this phase.

Testing Phase

The testing phase comprised, as in the previous experiment from Chapter 6, of participants listening to twenty sets of four concurrently presented earcons which repeated seven times and trying to identify the attributes of each earcon. Variations in the presentation of earcons between the two conditions are described in the following sections. Participants recorded their selection in a computer based dialogue box as shown in Figure 5.3, and were given no feedback as to the correctness of their responses. The same twenty stimuli sets were used for both conditions, but were randomly presented to avoid any order effect. After each testing phase each participant completed a modified NASA TLX questionnaire to measure subjective workload.

Spatial Condition

The spatial condition used the earcons described in Section 4.5.1. Each earcon was presented in a different spatial location on a lateral plane collocated with the listener's ears. Due to the lengthy procedure and specialised equipment required to create individualised HRTFs (see Section 2.5.2), and since many of the potential applications of this work (as described in Chapter 1) would make it currently infeasible to produce individualised HRTF's for each user, the earcons have been spatialised using a generalised HRTF (GHRTF). GHRTF's as described in Section 2.5.2, are particularly poor at encoding elevation, hence the earcons were only separated on the horizontal plane. In terms of Figure 2.4, the sounds varied spatial location in azimuth only, they did not vary in elevation. The GHRTF used was that found on the PURE Digital Sonic Fury PC sound card [1].

Each earcon was placed approximately 20cm from the listener's head, one at each of the four main points of the compass. The placement of the earcons is summarised in Figure 7.1. To overcome some of the problems of GHRTFs, a Polhemus Fastrak 6 degree of freedom tracking device [114] was used to dynamically respatialise the sounds relative to the participant as they moved and rotated their head, hence incorporating the small head movements that can be used to localise and discriminate sound sources (see Section 2.5.2).

During pilot testing, participants responded that the earcons were presented too quickly. Initially, there was no time between consecutive repetitions of the earcons, since Bregman [14] notes that rapidly repeating a stimulus allows streaming to build up over time. However, in response to pilot participants' comments,

[1]This card is marketed as the Turtle Beach Santa Cruz in the USA

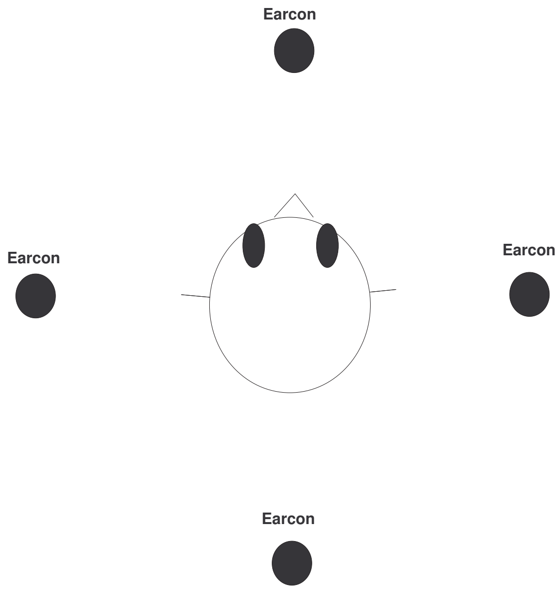

Figure 7.1: An illustration showing how the earcons used in the spatial condition were located relative to the listener's head.

a 1.5 second break was inserted between consecutive presentations of the earcons. This time, as recommended by Patterson, Edworthy and Shailer [108] in guidelines for the use of auditory warnings in hospital equipment, was designed to give listeners thinking time between consecutive stimuli presentations. This issue is further discussed in Section 7.2.3.

Non-Spatial Condition

As previously described in Section 7.2.1, the non-spatial condition comprised the earcons from Section 4.5.1 whilst also incorporating the concurrent earcon guidelines from Chapter 6. As such it is the same as the final condition described in Chapter 6. However, as with the spatial condition, a 1.5 second break was incorporated between consecutive presentations of the earcons.

7.2.3 Results

Identified Earcons and Attributes

As with the experiments described in Chapters 5 and 6, the numbers of correctly identified ride types, ride intensities and ride costs were collected, and from this the number of correctly identified earcons was determined using the same method as described in Section 5.3.1. The average number of correctly identified earcons and earcon attributes across all participants for each condition is shown in Figure 8.8. Raw data are presented in Appendix J.

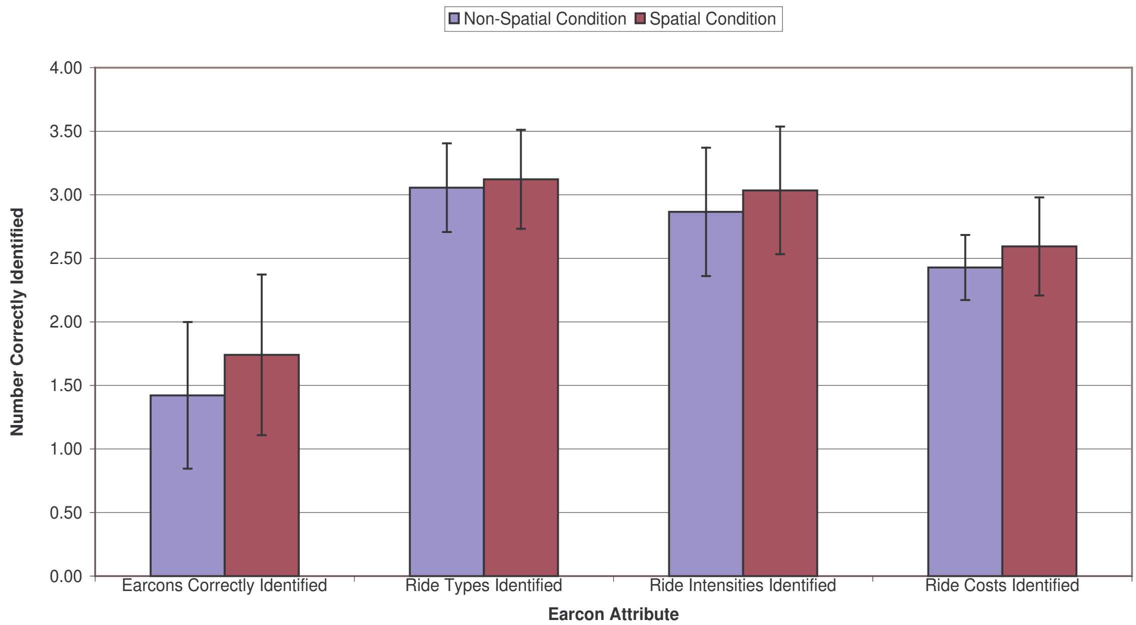

Figure 7.2: Graph showing the average number of earcons, number of correctly identified ride types, ride intensities and ride costs for the non-spatial and original spatial experimental conditions. Shown with standard deviations

To determine if any of the differences shown in Figure 7.2 were significant, four within groups t-tests were carried out, one on the number of earcons correctly identified and one each on the number of correctly identified ride types, ride intensities and ride costs.

The t-test on the number of correctly identified earcons showed that participants identified significantly more earcons in the spatial condition than in the non-spatial condition ($t(15) = 2.61$, $p = 0.020$). The t-test for the number of correctly identified ride costs showed that participants identified significantly more ride costs in the spatial condition than in the non-spatial condition ($t(15)= 2.37$, $p = 0.031$). The t-test on the number of correctly identified ride types ($t(15) = 0.55$, $p = 0.591$) and ride intensities ($t(15) = 1.42$, $p = 0.176$) were not significant.

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 4.5.1). A summary of these data across all participants is presented in Figure 7.3. Raw data are presented in Appendix J. To determine if workload was significantly different, overall workload for each participant was calculated by summing the individual scores for each attribute (excluding overall preference and annoyance as these do not form part of the standard TLX set of scales). A within groups t-test on these results failed to show significance ($t(15) = 1.60$, $p = 0.130$). Individual t-tests for annoyance experienced ($t(15) = 0.30$, $p = 0.765$) and overall preference ($t(15) = 0.59$, $p = 0.564$) failed to show significance.

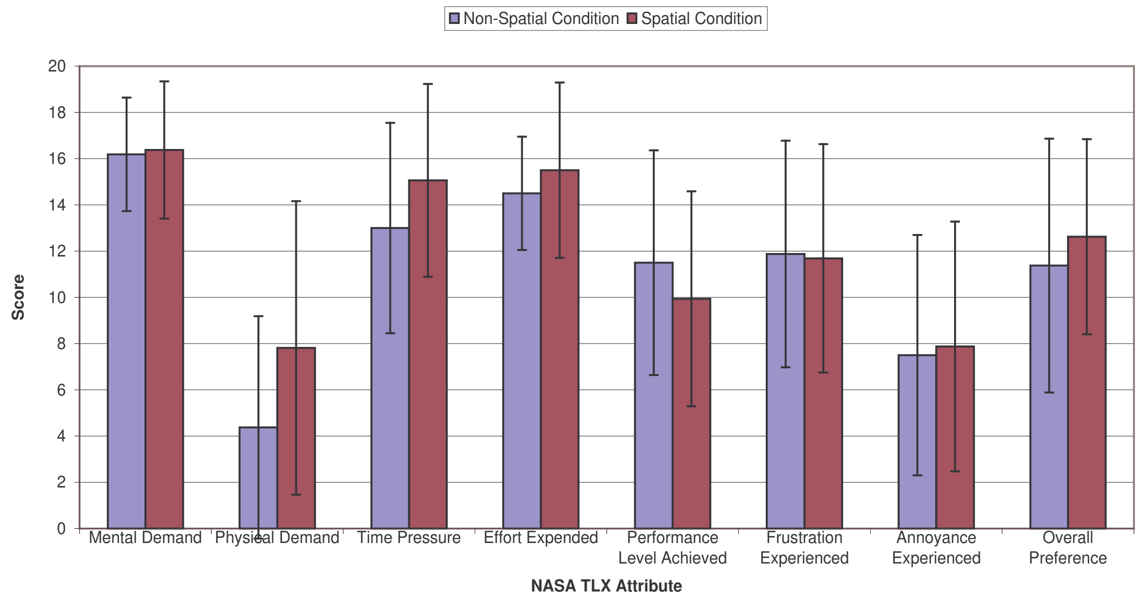

Figure 7.3: Graph showing the mean values for NASA TLX workload data for the non-spatial and original spatial experimental conditions. Shown with standard deviations.

7.2.4 Discussion

The results show that having a unique spatial location for each concurrently presented earcon can significantly improve identification, over concurrently presented earcons which are spatially collocated. This holds even when the earcons which are spatially collocated have been modified to make them more easily identifiable than the earcons which are at spatially unique locations. The results also show that the identification of the ride cost attribute is significantly improved in the spatial condition. This may be due to cases in the non-spatial condition where two earcons which differ only in ride cost (which is encoded as the register the earcon is played in) are concurrently presented, and the timbre and melody of the earcons then dominate in the grouping process, streaming the two sounds together and thus forming a composite sound from the belongingness principle from Section 2.6.2. As stated in Section 4.5.1, registers were chosen to avoid harmonic intervals between the earcons, however this may not have provided a strong enough grouping effect to dominate the timbre and melody similarity of such earcons. The results of the modified NASA TLX questionnaires failed to show any significant interactions between the conditions, which indicates that spatialisation did not increase subjective workload on task performance. However as with the work described in Chapter 6, the overall increase in identification performance is low and again, spatialisation would not appear to be practically useful on its own.

As described in Section 7.2.2, the experiment incorporated a 1.5 second break between consecutive presentations of each set of earcons, as recommended in guidelines by Patterson, Edworthy and Shailer [108] on hospital warning alarms. They claimed that these breaks allowed users to process the alarm and decide on a response. In the previous experiments (Chapters 5 and 6) there was no significant gap between consecutive earcon presentations, as Bregman [14] explains that constant repetition allows the effects of auditory streaming to accumulate. This poses the question *"What is more important when trying to understand concurrently presented earcons, a 1.5 second gap between consecutive presentations, or no gap between consecutive presentations?"*. Whilst it was possible to evaluate this question by comparing the results of the non-spatial condition with the results of the final condition from Chapter 6, between groups statistical analyses for both the objective and subjective data failed to show significance. However, such results do not provide any conclusive answers and it may be useful for future work to specifically investigate the impact of gap time between concurrent earcon presentation.

In conclusion, spatially separating the locations of concurrently presented earcons allows the listener to correctly group earcon attributes as whole earcons more easily than when all earcons are located in the same spatial location. The ASA modifications applied to earcon design as described in Chapter 6, tended to increase the number of earcon attributes that were correctly identified rather than the number of whole earcons correctly identified. The experimental results described in this section tend to show an increase in the number of correctly identified earcons but not an increase in the number of correctly identified attributes of the earcons. Therefore it seems that spatial concurrent earcon presentation, improves the ability of participants to group the right attributes with each other to form correct earcons, rather than improving the identification of individual earcon attributes. Hence, if the modifications to the earcons and their presentation as used in the non-spatial condition were combined with the spatial placement used in the spatial condition, this may result in further improvements to the identification of concurrently presented earcons. This is important to determine given the overall low identification of the earcons by participants. Such modifications may also allow earcons to be placed closer together if data were mapped to spatial position due to the differences between the earcons (see Section 2.4.2).

7.3 Spatial versus Revised Spatial Earcons

7.3.1 Motivation

The previous section has shown the importance of spatialisation when presenting concurrent earcons. The presentation of earcons, based only on the guidelines of BWE [25], which were presented in a spatial audio environment, significantly outperformed non-spatially presented earcons which had been revised based on the guidelines for concurrent earcon presentation from Chapter 6.

This result indicates a possible divergence of guidance for designers to improve the identification of concurrently presented earcons. One possibility is that if spatial audio presentation is available, earcons should be presented in spatially distinct locations and be designed strictly to the guidelines of BWE [25]. If spatial audio presentation is unavailable however, earcons should be designed with the guidelines of concurrent earcon presentation described in Section 6.4. Alternately, the identification of concurrent earcons may be further improved by the incorporation of the guidelines of concurrent earcon identification described in Section 6.4 as well as the spatial presentation of the earcons. As shown by Best, van Schaik and Carlile [10] (see Section 3.2.2), spatial location is not a totally dominating factor in auditory scene analysis, and in applications which seek to encode some parameter of data as the position of an earcon (see Section 3.4), sufficient spatial separation of earcons to avoid interference with each other may not be possible. As Bregman [14] (pp. 302) notes *"the human auditory system does not give an overriding importance to the spatial cues for belongingness but weighs these cues against all the others. When these cues all agree, the outcome is a clear perceptual organisation, but when they do not, we can have a number of outcomes*.". Thereby having other differences between concurrently presented earcons, such as those provided by the guidelines of Section 6.4, should work with spatialisation to desirably stream the earcons. Given the results from the previous experiment and those from Chapter 6 this is important and would allow users to be better able to identify information encoded in earcons and as such may increase the practical usefulness of concurrently presented earcons. It is important therefore to determine how well concurrently presented earcons which incorporate the guidelines from Section 6.4 are identified when each earcon is presented in a spatially distinct location.

The experiment described in this section investigates the impact on identification of concurrently presented spatialised earcons based only on the guidelines of BWE [25], compared to concurrently presented spatialised earcons which also incorporate the concurrent earcon guidelines from Chapter 6.

7.3.2 Methodology

The procedure and methodology of this experiment is largely the same as that of the experiment described in Section 7.2. Again, sixteen participants between the ages 18-24 undertook the experiment, none of whom had taken part in the previous experiment. The experiment was of a within groups design, with participants randomly assigned to one of two groups. There were two conditions, the *"spatial(2)"* condition and the *"revised spatial"* condition. The conditions are explained below. The order in which participants undertook the conditions was counterbalanced to avoid any order effects. Each condition consisted of a training and testing phase. The order of the conditions for each group of participants is summarised in Table 7.2. Written consent was obtained prior to the experiment from all participants, and all were paid £5 on completion.

The main hypothesis of the experiment was that earcons and earcon attributes in the revised spatial condition, which incorporated the concurrent earcon guidelines from Chapter 6 and were spatially separated, would be better identified those from the spatial(2) condition which were only spatially separated and incorporated no other modifications to their design or presentation. The independent variable (IV) was the design and presentation of the earcons, and the dependent variables (DV's) were the number of earcons, ride types, ride intensities and ride costs successfully identified, and participant responses for modified NASA TLX questionnaires.

	First Training Session	First Testing Session	Second Training Session	Second Testing Session
Group 1	Spatial(2) Condition	Spatial(2) Condition	Revised Spatial Condition	Revised Spatial Condition
Group 2	Revised Spatial Condition	Revised Spatial Condition	Spatial(2) Condition	Spatial(2) Condition

Table 7.2: Table showing the procedure for the two groups of participants undertaking the spatial versus revised spatial experiment

Spatial(2) Condition

The spatial(2) condition, is, as its name suggests, the same as the spatial condition described in Section 7.2.2. The name has been changed to avoid confusion with the results from the spatial condition.

Revised Spatial Condition

The revised spatial condition was the same as the spatial(2) condition, but used the earcons from the final condition of the experiment in Chapter 6, which incorporated the concurrent earcon guidelines from Section 6.4.

Because of the 300ms onset-to-onset time difference between the start of each earcon there would have existed a predictable spatial pattern in the order of earcon presentation. If the spatial positions of the first, second, third and fourth earcons to be presented were always the same. This could have lead to a learning effect which would have been undesirable and could confound the results. To overcome this issue, the spatial positions of the first, second, third and fourth earcons, were randomly alternated for each trial in the condition. However, the order of presentation was held constant during each trial.

7.3.3 Results

Identified Earcons and Attributes

As with the previous spatial experiment (Section 7.2), the numbers of correctly identified ride types, ride intensities and ride costs were collected, and from these the number of correctly identified earcons was determined using the method described in Section 5.3.1. The average values for these data are summarised graphically in Figure 7.4. Raw data are presented in Appendix J.

To determine if any of the differences shown in Figure 7.4 were statistically significant, four within groups t-tests were carried out, one on the number of earcons correctly identified, and one each on the number of correctly identified ride types, ride intensities and ride costs.

The t-test on the number of correctly identified ride types showed that participants identified significantly more ride types in the revised spatial condition than in the spatial(2) condition ($t(15) = 3.11$, $p = 0.007$). The t-tests for the number of earcons ($t(15) = 1.73$, $p = 0.104$), ride intensities ($t(15) = 0.81$, $p = 0.429$), and ride costs identified ($t(15) = 0.95$, $p = 0.356$), failed to show significance.

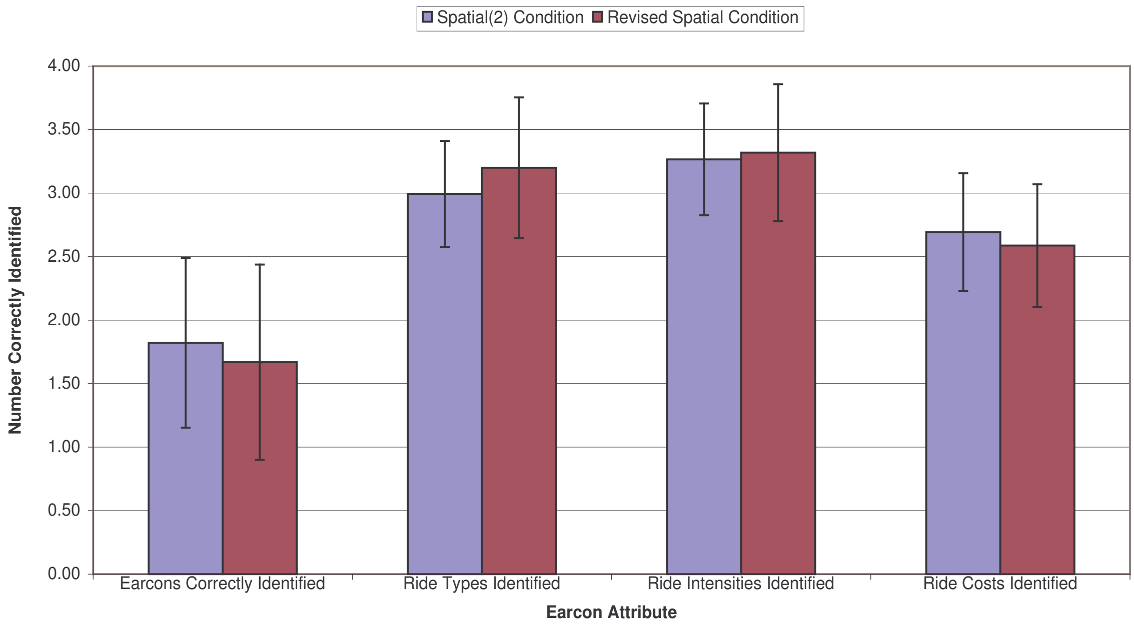

Figure 7.4: Graph showing the average number of earcons, number of correctly identified ride types, ride intensities and ride costs for the second original spatial and revised spatial experimental conditions. Shown with standard deviations.

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 4.5.1). The averages of these data across all participants is presented in Figure 7.5. Raw data are presented in Appendix J. To determine if workload differed significantly between conditions, overall workload for each participant was calculated by summing the individual scores for each attribute (excluding overall preference and annoyance as these do not form part of the standard TLX set of scales). A within groups t-test on these results showed significance ($t(15) = 3.32$, $p = 0.005$). To determine which individual workload factors caused this increase in workload, eight within groups t-tests were carried out, one for each modified NASA TLX attribute.

Participants reported that performance level achieved ($t(15) = 3.36$, $p = 0.004$) was significantly higher in the revised spatial condition than in the spatial(2) condition. Participants also reported significantly lower physical demand ($t(15) = 2.75$, $p = 0.015$) in the revised spatial condition than in the spatial(2) condition. Participants also reported significantly lower time pressure ($t(15) = 2.52$, $p = 0.024$) in the revised spatial condition than in the spatial(2) condition. The t-tests for mental demand ($t(15) = 1.50$, $p = 0.154$), effort expended ($t(15) = 1.25$, $p = 0.231$), frustration experienced ($t(15) = 1.18$, $p = 0.258$) and annoyance experienced ($t(15) = 1.52$, $p = 0.150$), failed to show significance. Participants did not express a significant overall preference for either condition ($t(15) = 1.63$, $p = 0.123$).

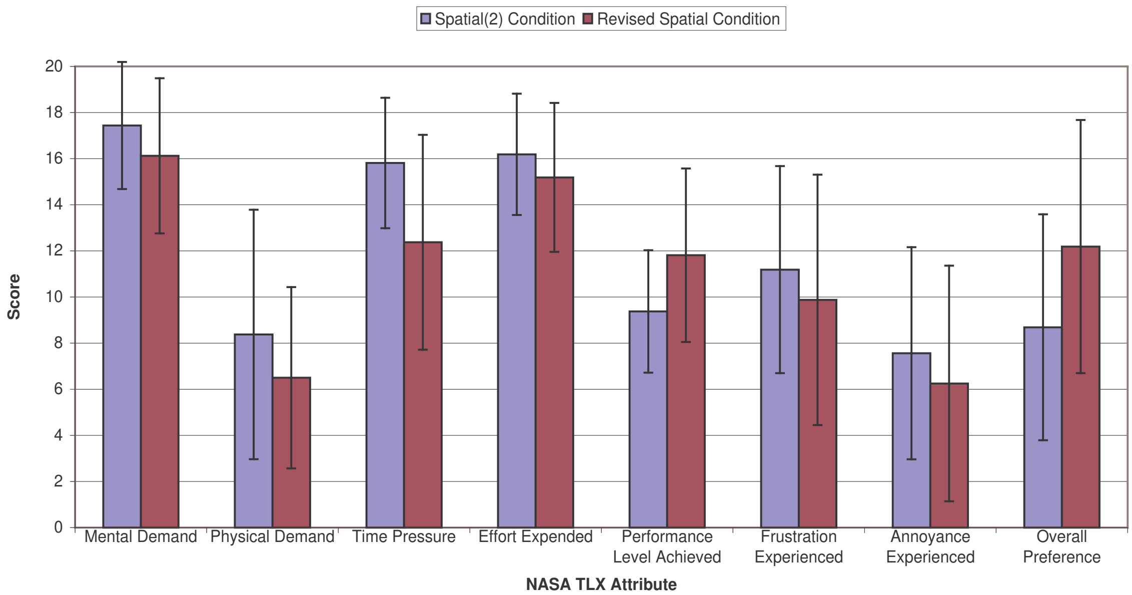

Figure 7.5: Graph showing the mean values for NASA TLX workload data for the second original spatial and revised spatial experimental conditions. Shown with standard deviations.

7.3.4 Discussion

The results from Section 7.3.3 show that the identification of earcons in a spatialised audio environment can be improved by the application of the concurrent earcon guidelines from Chapter 6. The identification of ride type, encoded as timbre, was significantly improved in the revised spatial condition. The subjective physical demand of participants was significantly lower in the revised spatial condition, this may indicate that less head movement (using the headtracking device) was required for participants to perceptually separate the earcons, which may be useful in reducing fatigue if a display incorporating spatial earcons is to be used for a prolonged period. However, further studies are required to identify if this is the case. Participants also reported significantly lower time pressure and higher perceived performance in the revised spatial condition, which may make any interface which uses such a technique a more enjoyable experience for users. However, overall results remain low and may not practicaly improve performance in real world tasks. This issue is investigated in Chapter 8.

Although the number of ride types that were correctly identified was significantly greater in the revised spatial condition than in the spatial(2) condition, the actual number of earcons identified was not significantly different in either condition. In many ways this is predictable, given the previous work on non-spatial earcon modification that the revised spatial condition incorporated (see Chapter 6), where modifications to the design and presentation of earcons tended to increase the number of earcon parameters that were successfully identified, rather than the total number of earcons identified. As discussed in Section 3.2.2 however, spatialisation is not a totally dominating parameter in ASA, hence having more cues available to separate different earcons from each other is advantageous in reducing the reliance solely on spatial location to separate earcons, and may allow the separation between earcons to be reduced, which would be advantageous if spatial location was used to map a data parameter.

Concurrent Earcon Guidelines

From these experiments, the following guidelines which extend those previously discovered and described in Chapters 5 and 6 have been identified. These guidelines can be used for future designers of auditory displays which use concurrent earcon presentation to communicate information.

- The use of spatialised presentation with headtracking significantly improves the identification of concurrently presented earcons. Therefore spatialised presentation should be employed whenever practically possible in cases where earcons will be concurrently presented.
- Whilst spatial audio significantly increases the number of concurrent earcons that can be identified, it does not always guarantee that earcons will be desirably streamed, hence the maximum amount of angular (in azimuth) separation between concurrently presented earcons should be used. The guidelines of Chapters 5 and 6 should also be incorporated when using spatial presentation to improve earcon identification.

7.4 Conclusions

The two experiments described in this chapter have significantly contributed to answering two of the research questions posed in this thesis (see Section 4.7), namely, RQ2: *"What is the impact of presenting earcons in a spatialised auditory environment?"*, and RQ3: *"How can earcons be more robustly designed and presented to increase the number which can be concurrently identified?"*.

The first experiment investigated the effect on concurrent earcon identification when earcons were designed solely from the guidelines of BWE [25], but were presented in spatially distinct locations (the spatial condition). This method was compared to the non-spatial presentation of concurrent earcons which were based on the guidelines of BWE [25], but also included the concurrent earcon guidelines described in Section 6.4, which had already been shown to significantly improve the identification of non-spatial, concurrently presented earcons (the non-spatial condition). Participants were found to have identified significantly more earcons and ride costs in the spatial condition than in the non-spatial condition. These results showed that the separation of earcons in space is an effective way to concurrently present them. Indeed it is a superior method than if those earcons were non-spatially presented but included presentation and design modification which have already been shown to improve earcon identification in concurrent situations.

The second experiment answers the obvious question arising from the first experiment. Can the use of earcons which incorporate the non-spatial concurrent earcon guidelines from Section 6.4 be used in a spatial audio environment to further improve the identification of concurrently presented earcons? This is an important question to answer as in some applications it may not be possible to sufficiently separate earcons in space to allow them to be uniquely identified; such as when a data parameter is mapped to spatial location [141, 124] (see Section 3.4). Best, van Schaik and Carlile [10] have shown that interference may still occur with up to 60^o separation between concurrently presented audio sources. The experiment therefore compared identification of earcons based only on the guidelines of BWE [25], which were spatially presented (the spatial(2) condition), to earcons which were also spatially presented but also incorporated the non-spatial earcon guidelines of Section 6.4 (the revised spatial condition). The results showed participants identified significantly more ride types in the revised spatial condition, and that participants' subjective physical demand and time pressure were significantly lower in the revised spatial condition than the spatial(2) condition. The lower physical demand indicates that less physical movement of participants' heads was required when more cues for the separation of sounds existed, as was the case in the revised spatial condition.

Both of the experiments in this chapter investigated only the maximum possible azimuthal separations between concurrently presented earcons. As discussed in Section 3.4, one of the advantages of spatial presentation of sound is that spatial location can be used to map a data parameter. As already discussed, when mapping spatial location to a data parameter it may not be possible to arbitrarily ensure such maximum differences between the spatial location of sounds. Therefore the work of this chapter has only investigated the "best case scenario" for spatial sound presentation, whereas the experiments of Chapters 5 and 6 have investigated the worst case of spatial separation (where all sounds are spatially collocated). As discrimination of concurrently presented spatialised sounds is a function of both the similarity of those sounds as well as their azimuthal separation (see Section 2.4.2), it would be useful if future work investigated this relationship, so that the "middle ground" could also be mapped out. Additionally, although both experiments have investigated four concurrently presented earcons, reducing this number (as suggested by the guidelines in Section 5.4), in cases where spatial location is not mapped to a particular data parameter, would allow greater separation between concurrently presented earcons, and, by the discussion of Section 2.4.2, should increase the number of earcons which can be identified. This is important since although significant improvements in earcon identification were found by both spatialisation and other modifications to the design and presentation of the earcons, the magnitude of these improvements was not large and it is questionable if such modifications would be of practical use in "real world" interfaces, allowing users to exploit the advantages of concurrent auditory display discussed in Section 4.2. However, the experiments which have been described in both this chapter, and Chapters 5 and 6, have elicited a great deal of knowledge and information as to the limitations of concurrent earcon presentation, as well as improving the identification of concurrently presented earcons. This knowledge should assist future designers of auditory displays which use concurrent earcons, to produce more effective and informative displays which, in the terms of Buxton [32], provide more information and less noise.

Chapter 8

An Ecological Investigation into Concurrent Earcons

8.1 Introduction

Chapters 5, 6 and 7 have investigated how well the complex types of earcons which are formed from a structured grammar are identified when they are concurrently presented. Investigations which have varied the number of earcons concurrently presented, modified the design and presentation of earcons based on auditory scene analysis research, as well as having each earcon in a different spatial location have been carried out. The results of this work have lead to a set of guidelines for concurrent earcon presentation which are outlined below:

Guideline 1: Increasing the number of concurrently presented earcons significantly reduces the proportion of the earcons which 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 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, Wright and Edwards [25], 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 [118] 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 300ms onset-to-onset gap between the starts of concurrently presented earcons will make the earcons more identifiable to users.

Guideline 5: The use of spatialised presentation with headtracking significantly improves the identification of concurrently presented earcons. Therefore spatialised presentation should be employed whenever practically possible in cases where earcons will be concurrently presented.

Guideline 6: Whilst spatial audio significantly increases the number of concurrent earcons that can be identified, it does not always guarantee that earcons will be desirably streamed, hence the maximum amount of angular (in azimuth) separation between concurrently presented earcons should be used. Guidelines 1-5 should also be incorporated when using spatial presentation to improve earcon identification.

One of the issues of the previously undertaken work, and thus the guidelines derived from it, is that it uses an abstract methodology. In other words, although the previously undertaken work allows the issues in the identification of concurrently presented earcons to be understood, the tasks undertaken by participants in Chapters 5, 6 and 7, do not represent "real" tasks that may be performed using concurrent earcons in a human computer interface (a justification of the experimental design used in Chapters 5, 6 and 7 is given in Section 5.2.1). Whilst identifying concurrently presented earcons would be part of the task users would perform, it would be only part of the task; users would use their identification of the earcons to make a decision in some larger task. If the guidelines previously mentioned concerning concurrent earcon presentation are to be claimed to be useful in real world interfaces, their impact on performance in such tasks must be determined. This is important given the overall low identification scores from the previously undertaken work which call into question the practical usefulness of the guidelines in a real world context.

In this chapter an experiment which investigates the research question (RQ4), *"How much do modifications to the design and presentation of concurrently presented earcons affect performance in "real world" tasks?"*, from Section 4.7, is described and its implications are discussed.

8.2 Methodology

8.2.1 Outline of Evaluation Domain

One of the advantages of concurrently presented audio in an auditory display (as discussed in Section 4.2.1) is an increase in communications bandwidth. Such an increase in bandwidth may be of specific advantage in auditory displays for mobile devices which have a small visual display, where information could be moved from the visual domain to the auditory domain and relieve demands on the user's visual system [62], which would be required for more important activities. Such reasons mean mobile tasks are ideally suited to evaluating concurrently presented earcons in more ecologically valid situations.

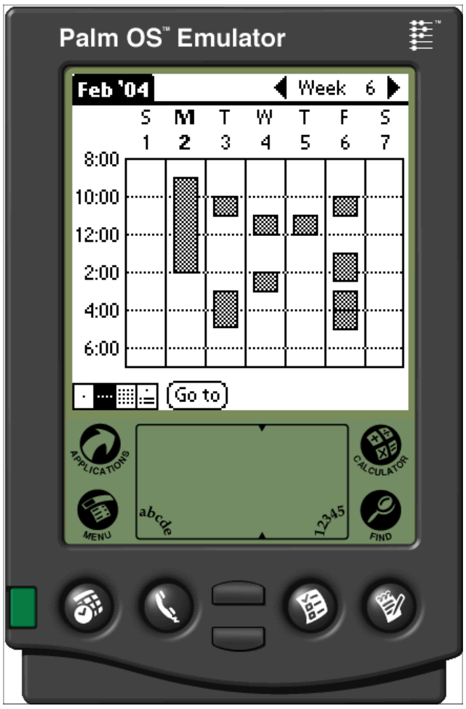

Figure 8.1: An example of the week at a glance view from the standard Palm date book application.

In choosing a task with which to evaluate concurrently presented earcons, it is important that the task is well understood. Dolphin (see Section 4.4), whilst highlighting the problems with concurrent earcons, is itself a relatively new idea. Because of this, even improving the identification of the earcons may not be enough to improve task performance since Dolphin may have other, currently unknown issues, which could confound any results. Due to such issues, the task of browsing a diary was chosen for the evaluation.

Diaries are ubiquitous mobile applications, and several studies of their usage have been undertaken [66, 102, 11]. Although the majority of these studies have been carried out on desktop systems, they provide useful information on the kinds of tasks that users perform and diary features which are useful. Blandford [11] identified that one of the most important features of a diary was to support "prospective remembering", and Payne [109] noted in another study that users regularly looked ahead a week or so at a time, and used such information to assist with preparation for appointments (e.g. to help decide what to wear the next day (pp. 92)). Such claims are also described by a much earlier study by Kincaid [66] who noted that "week at a glance" and "day at a glance" features were amongst the most popular diary features for users. Payne [109] however noted that, *"Most electronic calendars severely compromise the browsability of calendars. Often entries are accessed for a single day at a time (making it impossible to simultaneously review entries for other days)"*. Such issues are compounded on mobile devices due to their small visual displays which restricts the information which can be presented. For example, Figure 8.1 shows a "week at a glance" type view from the standard Palm PDA [103] diary, which can only show when appointments are taking place, and not provide information about those appointments without switching the view, losing the “at a glance” information available.

Studies on diary systems have also shown that users will categorise appointments in some way. Blandford [11] notes that in paper based diaries, users would use different coloured pens or underlining to denote the priority of appointments, with other “symbols” being used to denote particular details of those appointments.

8.2.2 Outline of Experimental Task

In the experiment described in this chapter, participants had to browse audio diaries, each of which consisted of five days, with four appointments on each day. Attributes of each diary entry (the type of appointment, the importance of appointment, and the time of day of the appointment) were encoded as earcons and used to represent appointments. Although the attributes chosen are to some degree arbitrary, due to the abstract mapping between data attributes and earcons, other attributes of the data could easily be mapped to the earcons by users, thereby allowing them to customise the diary to suit themselves. Lack of user customisation is a claim frequently made against electronic diaries [11]. The earcons for each day are concurrently presented, thus increasing the rate of data presentation and reducing the overall time to play one day’s appointments (see Section 4.2.1). This should allow participants to quickly move between different days and build up a “picture” of that week’s diary. The procedure of the experiment is further discussed in the following section.

8.3 Procedure

8.3.1 Introduction

Sixteen participants between 18-24 undertook the experiment described in this section which was of a within groups design. Written consent was obtained from all participants, and all were paid £5 on completion. Participants were randomly assigned to one of four groups to determine the order in which they undertook the experiment. There were two conditions, the *non-spatial* condition and the *spatial* condition. Each condition comprised of a training and a testing phase. Both conditions are described later in this section. Both conditions involved browsing several diaries and answering questions about them. There were two sets of diaries (A and B). To avoid any learning effects, participants used a different set of diaries for each condition. The order of presentation for each diary set was counterbalanced between conditions to avoid order effects. Prior to carrying out any condition participants undertook a baseline measuring phase, this is described later. The details of each group are shown in Table 8.1.

		Training Session 1	**Testing Session 1**	**Training Session 2**	**Testing Session 2**
Group 1	Baseline Measuring Phase	Non-Spatial Condition	Non-Spatial Condition Diary Set A	Spatial Condition	Spatial Condition Diary Set B
Group 2	Baseline Measuring Phase	Spatial Condition	Spatial Condition Diary Set A	Non-Spatial Condition	Non-Spatial Condition Diary Set B
Group 3	Baseline Measuring Phase	Non-Spatial Condition	Non-Spatial Condition Diary Set B	Spatial Condition	Spatial Condition Diary Set A
Group 4	Baseline Measuring Phase	Spatial Condition	Spatial Condition Diary Set B	Non-Spatial Condition	Non-Spatial Condition Diary Set A

Table 8.1: Table showing the order in which each group of participants undertook the experiment.

8.3.2 Baseline Measuring Phase

One of the advantages of using audio in a mobile device, is that it removes some of the demands on users' visual attention (see Section 3.2.1). To determine if there was an effect on the user's visual demands which could be attributed to one of the experimental conditions, a metric was included to measure how participants' walking speed varied between conditions. Checking a diary whilst walking is a reasonably common occurrence, where it is the walking task which is of most importance, as users must avoid obstacles and continue to walk in the correct direction towards their destination [62], thereby requiring them to continually avert their eyes from the device.

Whilst it is possible to directly measure the distance walked by a participant during an experiment and the time taken, making summaries across all participants would cause a widely spread distribution of scores. Clark-Carter, Heyes and Howarth [33] have noted that each person has a preferred walking speed which is most efficient for them, which can be affected by factors such as weight, age and leg length. To normalise such issues between participants, they proposed percentage preferred walking speed (PPWS). Here, the speed participants walk during the condition is converted into a percentage of some pre-measured walking speed. For example, if a participant walked at 2mph in the experimental condition, and had previously walked at 4mph in a pre-testing measurement, the PPWS of that participant would be 50%. PPWS was originally developed as a way of evaluating the quality of mobile guides for the visually impaired [33, 111], but has recently been used for mobile interfaces in general [113]. When used with mobile devices PPWS also provides a rough global guide of the impact of a device or application on a user, since PPWS is measured in relation to optimal walking speed.

The baseline measuring phase therefore involved measuring each participant's normal walking speed. Participants were instructed to walk a predetermined route around three cones ten times, whilst wearing (but not using) the device used in the testing phase (see Section 8.3.4). See Figure 8.2 for the route used. Participants were instructed to walk at what they felt was a comfortable pace. The time taken for participants

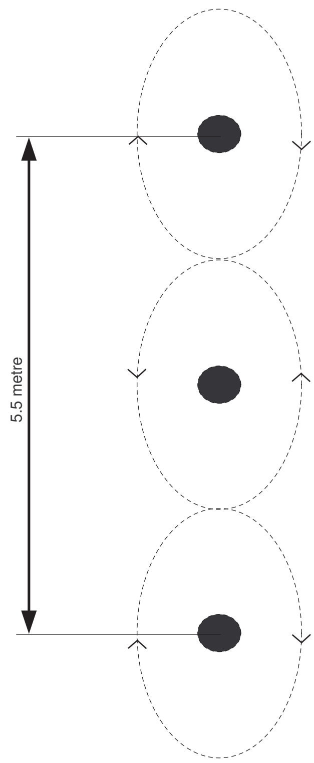

Figure 8.2: An example of the route walked by participants whilst carrying out the experiment.

to complete each lap was recorded using a computer program. The average time taken to walk the middle six laps was used as that participant's baseline walking speed.

8.3.3 Training Phase

A training phase similar to that from the experiments described in Chapters 4, 5, 6 and 7 was used. It involved participants reading a sheet which described the grammar of the earcons used (see Appendix K), followed by ten minutes of self-guided training via a Web page (see Section E of the accompanying CD-ROM for the Web page (Appendix M)), 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 five minutes of training, before retaking the test. If they were still unable to successfully identify the earcons, they took no further part in the experiment. No participants however failed training.

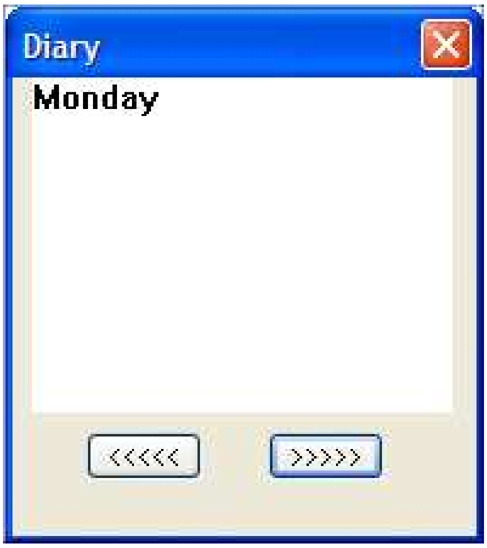

Figure 8.3: An example of the visual display of the experiment. Participants could switch between days by using the "<<<<<" and ">>>>>" buttons.

8.3.4 Testing Phase

In the testing phase participants browsed twelve diaries. Each diary consisted of five days (Monday - Friday), and each day consisted of four appointments. Each appointment was represented as an earcon. All four earcons were concurrently presented. There were differences to the design and presentation of the earcons depending on the condition the participant was performing. These modifications are described in Sections 8.3.6 and 8.3.7. Only one day's appointments were presented at a time, which were continuously repeated. Participants could switch to the next, or previous day by using buttons at the bottom of a 6 x 6 cm dialogue box which represented the visual display of a mobile computing device. The name of the current day was also presented on this display. See Figure 8.3 for an example of this dialogue.

Whilst browsing the diary participants attempted to answer a question about it. Each question asked about only two of the earcon attributes. As discussed in Section 4.4, it is not always necessary to know all three earcon attributes. Since there are many different types of questions which can be asked about data in a diary, two types of question were used. One type of question (Type S) asked how many times certain kinds of appointments occurred in a week. This type of question required a single, numerical answer. A dialogue illustrating such a question is shown in Figure 8.4. The other type of question required one or more answers which were not numerical (Type M). A dialogue which illustrates such a question is shown in Figure 8.5. When participants had decided on an answer for the question, they clicked the "NEXT QUESTION" button on the dialogue (see Figure 8.5) to be presented with the next diary.

An equal number of questions of each type were used, and there were an even number of questions which asked about each possible pair of data parameters. All questions were randomly presented to avoid order effects.

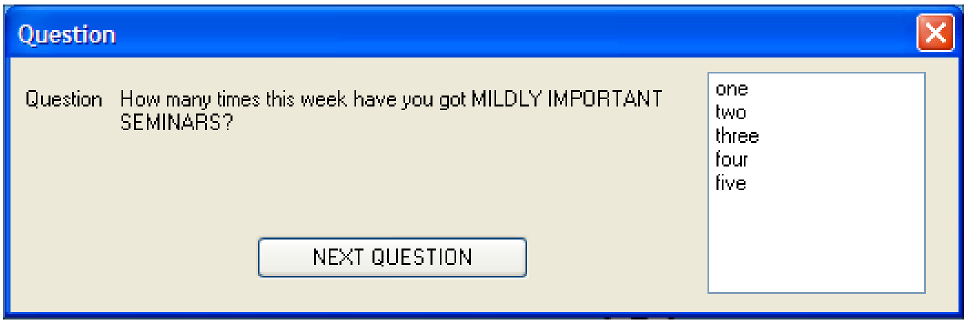

Figure 8.4: An example of a question dialogue from the experiment, where only one answer was required.

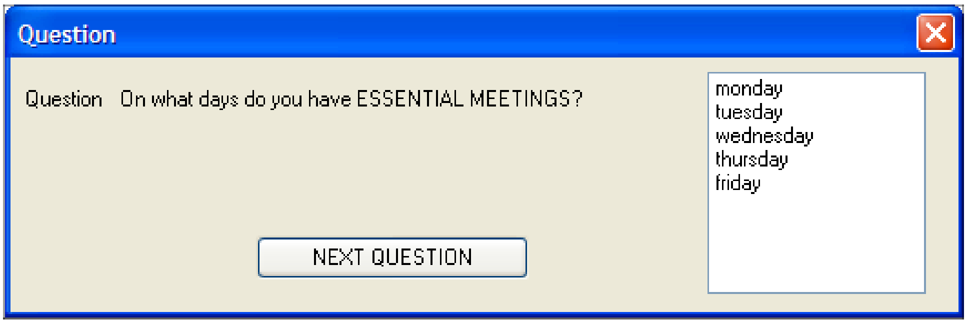

Figure 8.5: An example of a question dialogue from the experiment, where more than one was required.

Participants carried out the testing phase on a Xybernaut Mobile Assistant V [151] wearable computer, via a touch sensitive display (see Figure 8.6). Sound was presented through a set of Sennheiser HD-25 headphones. Whilst carrying out this task participants walked the route as described in Section 8.3.2.

There were three main hypotheses in the experiment:

H1 Participants will have a greater PPWS in the spatial condition than in the non-spatial condition.

H2 Participants will answer more type S questions correctly in the spatial condition than in the non-spatial condition.

H3 Participants will answer more type M questions correctly in the spatial condition than in the non-spatial condition.

The independent variable (IV) was the design and presentation of the earcons for each condition (see Sections 8.3.6 and 8.3.7). The dependent variables (DV's) were, the average time for each participant to walk a lap, the number of correctly answered type S questions and the number of correctly answered type M questions. Additionally, it was hypothesised that participants' workload would be significantly reduced in the spatial condition. To determine this, as with the previous experiments of Chapters 4, 5, 6 and 7, modified NASA TLX questionnaires (see Section 4.5.1) were filled in by each participant after they had completed each condition.

Figure 8.6: An example of how the Xybernaut device was worn by participants and (insert) the device itself.

Before the testing phase proper, each participant carried out a reduced version involving two diaries, not used in the testing phase proper, to familiarise themselves with the experimental procedure.

8.3.5 Earcons Used

The earcons used in this experiment were the same as those used for the previous experiments; however they were remapped to represent appointment types that may be found in a diary. Three parameters of appointments were chosen to be encoded in the earcons: the type of appointment (either meeting, lecture or seminar), the importance of the appointment (either mildly important, very important or essential), and the time of the appointment (either early morning, mid-morning or late afternoon).

Appointment type was mapped to timbre. Three distinct timbres were used: a grand piano (General MIDI patch No. 000), was used to represent meetings, a violin (General MIDI patch No. 040) was used to represent lectures, and a trumpet (General MIDI patch No. 056) was used to represent seminars.

Ride Intensity was mapped to melody, which is a combination of a rhythm and a pitch structure for that

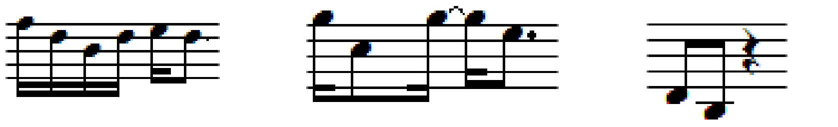

Figure 8.7: Melodies used to represent mildly important, very important and essential appointments.

rhythm. The melodies used for mildly important, very important and essential appointments are shown in Figure 8.7.

Finally the time of the appointment was mapped to the register that the melody was played in. The registers used were the octave of C4 for early morning, the octave of C5 for mid morning and the octave of C6 for late afternoon. Each melody was staggered in each register to avoid the same melodies having musically harmonic intervals as suggested by Guideline 2 (see Section 8.1).

8.3.6 Non-Spatial Condition

In the non-spatial condition, the earcons representing each day's appointments were dioticly presented (monaural presentation to both ears) to participants continuously in rapid succession. The earcons used were those described in Section 8.3.5, and were continuously concurrently presented until a new day was selected.

8.3.7 Spatial Condition

In the spatial condition, the concurrent earcon guidelines from Chapter 5, 6 and 7 were implemented on the earcons from Section 8.3.5. Three timbres were chosen to represent each appointment type. See Table 8.2 for the timbres used. Earcons were fixed in spatially distinct locations egocentrically [34] relative to a participant's head (see Figure 7.1 for the placement of earcons). The presentation of the earcons in this condition can be regarded as largely similar to that used in the revised spatial condition from Section 7.3.

One difference with the revised spatial condition from Section 7.3 however, is that the use of dynamic respatialisation, using a position and orientation sensor mounted on the headphones (see Section 7.2.2), was not incorporated. Primarily this was due to the device used in previous experiments being unsuitable for use in a mobile context. Another device which was capable of detecting movement in a reduced number of dimensions was considered, however it was not available in a reasonable amount of time to be incorporated.

Additionally, the Xybernaut device used in the experiment did not incorporate a hardware based HRTF, as the experiments from Chapters 4 and 7 did. In this experiment a lower quality spatialisation system which is incorporated into Microsoft DirectX [78] was used. The spatialisation used in this experiment can

Appointment Type	Instrument	General MIDI Patch No.
Meeting	Acoustic Grand Piano	000
Meeting	Electric Acoustic Piano	001
Meeting	Electric Grand Piano	002
Lecture	Tuba	058
Lecture	French Horn	060
Lecture	Synth Brass 1	062
Seminar	Marimba	012
Seminar	Shamisen	106
Seminar	Kalimba	108

Table 8.2: Table showing the timbres used to represent appointment types in the spatial condition.

therefore be considered of lower quality than that previously used. However, the spatialisation system can be considered as more realistic for current generation mobile devices, which do not currently incorporate hardware based HRTFs, or have enough processing power to use full software based HRTFs.

8.4 Results

8.4.1 Correctly Answered Questions

Each correctly answered type S question (those requiring only one answer) was awarded a mark. Each participant could have therefore received up to 6 marks for type S questions. The average mark across all participants for each condition is shown in Figure 8.8. Raw data are presented in Appendix L. The difference in performance between the two conditions was not found to be significant ($t(15) = 0.50$, $p = 0.627$).

A mark was allocated for each type M question (those requiring one or more answers). For each correctly identified part of the answer, a proportion of that mark was awarded. For example if the correct answer was Monday, Tuesday and Wednesday, but a participant only gave Monday and Tuesday, 2/3 of a mark would be awarded. Again the maximum possible score was 6. The average mark across all participants is shown in Figure 8.8. Raw data are presented in Appendix L. The difference in performance between the conditions was not found to be significant ($t(15) = 0.49$, $p = 0.629$).

8.4.2 PPWS

The walking speed of each participant was measured in the same way as described in Section 8.3.2; the time taken for each lap was recorded, the first and last two lap times were discarded, and the remaining laps were averaged. This average was then converted into a percentage of the lap time recorded in the baseline measuring phase (see Section 8.3.2), to yield that participant's PPWS. An average of all participants' PPWS for each condition is shown in Figure 8.9. Raw data are presented in Appendix L. The difference in performance was not found to be significant ($t(15) = 1.46$, $p = 0.164$).

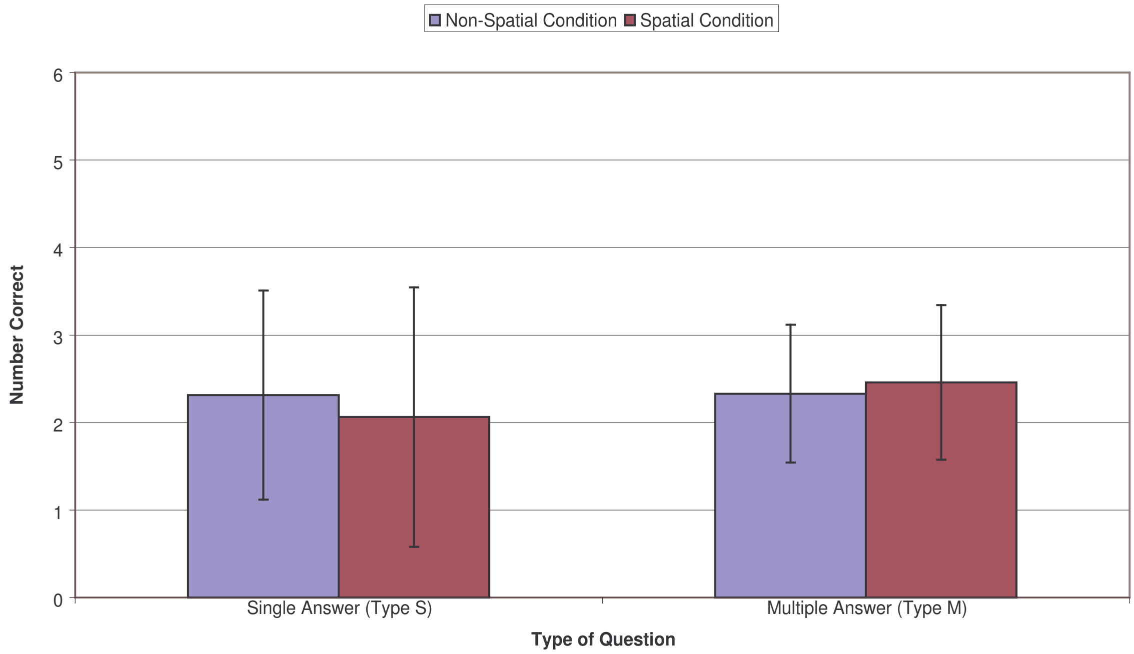

Figure 8.8: Graph showing the mean number of type S and type M questions correctly identified in the experiment. Shown with standard deviations.

8.4.3 Modified NASA TLX Workload Ratings

As mentioned in Section 8.3.4, modified NASA TLX questionnaires were used to assess participants' subjective workload. A graphical summary of these results is shown in Figure 8.10. Raw data are presented in Appendix L. To determine if workload differed significantly between the conditions, overall workload for each participant was calculated by summing the individual scores for each attribute (excluding overall preference and annoyance as these do not form part of the standard TLX set of scales). A within groups t-test on these results showed significance ($t(15) = 2.06$, $p = 0.045$). To determine which attributes caused this difference in workload, eight within groups t-tests were performed, one for each TLX attribute. These showed that participants had recorded significantly lower mental demand ($t(15) = 3.08$, $p = 0.008$) and annoyance ($t(15) = 3.07$, $p = 0.008$) in the spatial condition. Participants also recorded significantly higher overall preference ($t(15) = 4.07$, $p = 0.001$) in the spatial condition. The tests for physical demand ($t(15) = 0.64$, $p = 0.530$), time pressure ($t(15) = 1.80$, $p = 0.092$), effort expended ($t(15) = 1.16$, $p = 0.263$), performance level achieved ($t(15) = 0.93$, $p = 0.368$) and frustration experienced ($t(15) = 0.46$, $p = 0.650$) were not significant.

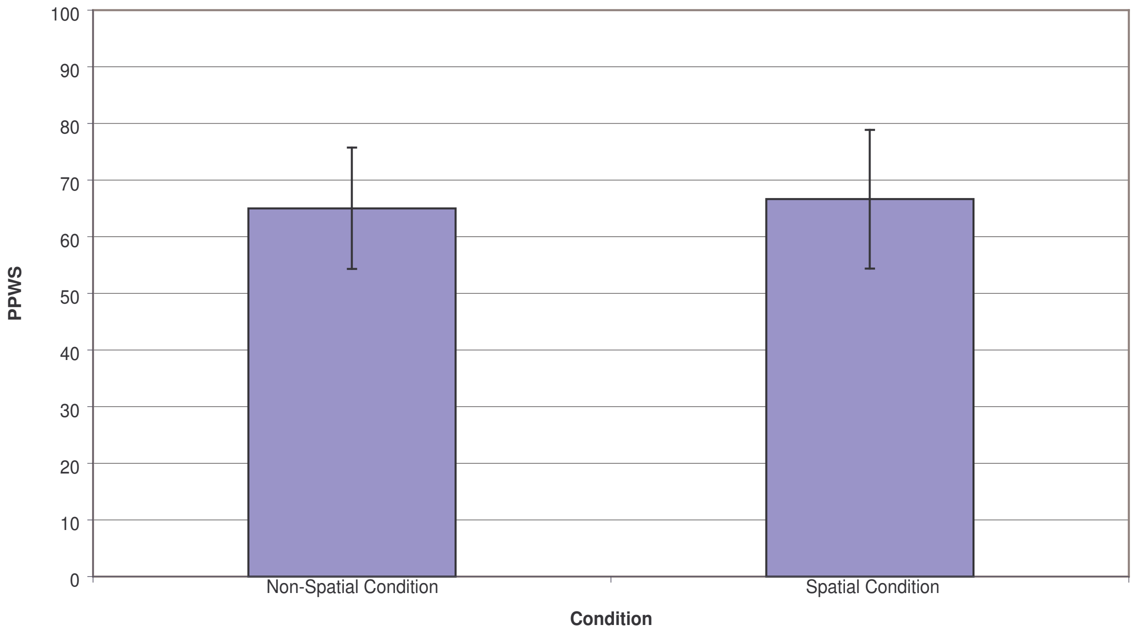

Figure 8.9: Graph showing participants' mean percentage preferred walking speed (PPWS) recorded for each condition. Shown with standard deviations.

8.5 Discussion

The results from the previous section showed that there were no significant results for any of the objectively measured data parameters. Neither accuracy of response on either type of question, or PPWS, was significantly affected by modifications to the design and presentation of the earcons used. However, significant differences were identified in the results of the modified NASA TLX questionnaires, with mental demand and annoyance being significantly reduced in the spatial condition. Overall preference was also significantly higher in the spatial condition.

Therefore although the modifications to the design and presentation of the earcons did not significantly improve task performance, they did have a significant impact on participants' perceptions of each condition. The significant reduction in annoyance indicates that participants found the spatial condition to be less distracting which may make them more likely to enjoy using the system. A reduction in mental demand may be important in mobile devices since there are likely to be other, more important mental demands on participants, such as navigating their environments, though these subjective improvements are not reflected in the objectively measured data. However, although significant results were found, as with the previous experiments overall task performance was low, indicating that participants still found it extremely difficult to separate the concurrently presented earcons, which means that identifying concurrently presented earcons is still difficult and further research will be required before concurrent earcons can exploit the advantages of concurrent presentation as discussed in Section 4.2.

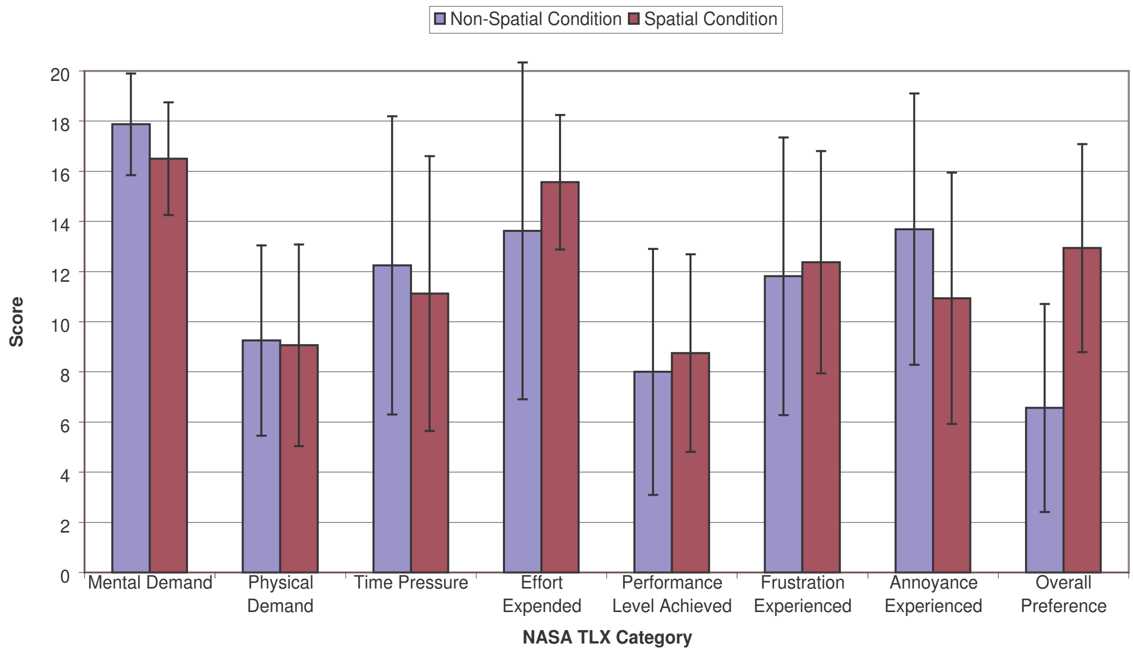

Figure 8.10: Graph showing mean results from the modified NASA TLX questionnaires. Shown with standard deviations.

It is unclear from the results of the experiment why the improvements in earcon design and presentation did not significantly improve task performance, whilst significantly affecting participants' attitudes to the task. There are however several possibilities.

Firstly, there were technical variations in the equipment used for this experiment and the previous experiments described in this thesis. Notably, due to the mobile element of this experiment, a general hardware based HRTF, as used in the experiments from Chapter 7, could not be used. Due to the computational requirements of HRTFs and the performance of the Xybernaut device used in these experiments, it was also not possible to use a software based HRTF. Therefore the standard 3D algorithm of Microsoft's DirectX, which is to use standard stereo reproduction with some "muffling" (attenuation of higher frequency components) of sounds placed to the rear of the user was used [7]. In addition, as already explained in Section 8.3.7, a headtracking device was unavailable for use, so it was not possible for participants to use small head movements to better separate the earcons. Due to the poor sound spatialisation system available, this may have seriously reduced the advantages of spatialisation. To determine the actual effect the spatialisation hardware had on performance, the previous experiments from Chapter 7 should be undertaken on the hardware used for this experiment which would allow for hardware dependent variations in performance to be undertaken. In addition to the restrictions on the quality of the spatialisation system used in the experiment, the guidelines from Chapter 5 were not incorporated into the spatial condition. These guidelines were not incorporated, as varying the number of concurrently presented earcons has such a large effect on their identification. It was felt, that reducing the number of earcons which were concurrently presented would not allow for a fair comparison between the two conditions. The impact of the other guidelines on real world earcon identification would not be able to be determined since varying the number of earcons presented would have such a large effect on the results. The identification of four concurrently presented earcons, from the results of Chapters 6 and 7, is known to be low, and by reducing this number a significant effect may have been observed. Additionally other studies [48] have indicated that the modifications to earcon identification from Chapters 6 and 7 may flatten the trend shown in Figure 5.7, thereby perhaps allowing more earcons to be presented but without a significant drop in identification. In order to explore such issues, further experiments should be carried out which do vary the number of concurrently presented earcons between the conditions, whilst incorporating the guidelines from Chapters 6 and 7. Initially the experiments should be of the same methodology as used in Chapters 5, 6 and 7, in an attempt to identify if there is a flattening of the trend of Figure 5.7, as indicated there might be by both Brungart, Ericson and Simpson [31] and Folds and Gerth [48]. Further experiments involving the methodology of this chapter should then be carried out to re-evaluate the full set of guidelines in a more ecologically valid context.

Alternately, as indicated in Section 8.1 the lack of improvement in task performance may be down to the modifications to the design and presentation of the earcons not providing a large enough improvement in earcon identification to significantly impact on task performance. Whilst in many ways this may be caused by the previous possible reason, it may be the case that even with better spatialisation there would be no significant improvement in task performance. The tasks which participants were asked to perform on each set of four earcons was different to previous experiments. In the previous experiments, participants had to identify each of the four earcons which were presented. In the experiment described in this chapter, each participant must make a decision if an earcon with particular attributes exists in the mix of four earcons. It is likely, due to the scoring system used in this experiment, that an incorrect decision will have a greater impact on the results than in previous experiments. This is a likely situation with other “real world” tasks as well. Therefore, a greater improvement in earcon identification would be necessary to identify a significant improvement in performance above the subjective measures already identified.

In order to identify it the issues outlined above contributed to the lack of significance in the results, further investigations involving concurrently presented earcons are required.

8.6 Conclusions

The experiment described in this chapter attempted to answer the research question *“How much do modifications to the design and presentation of concurrently presented earcons affect performance in ‘real world’ tasks?”* (RQ4), as outlined in Section 4.7. The previous chapters have investigated the identification of concurrently presented earcons using a more abstract experimental design, this chapter sought to ground the results of previous work in a more ecological context.

The experiment compared participants’ abilities to answer questions on mobile audio diaries, where each diary entry was represented as an earcon. Two sets of earcons were compared, one non-spatially presented set based on the guidelines of Brewster, Wright and Edwards [25], and another set which were spatially presented and incorporated the guidelines of concurrent earcon presentation from Chapters 5, 6 and 7 (albeit without the use of a headtracking device to dynamically respatialise the sounds). Analysis of several objective measures found that neither condition was significantly different to the other. This runs contrary to the research presented in Chapters 6 and 7 which showed that concurrently presented earcons were significantly better identified when the incorporated the guidelines from Chapter 6 and 7. However, it was determined that participants' subjective workload was significantly reduced in the spatial condition, with significant reductions in mental demand and annoyance being observed. This indicates that the modifications to earcon design and presentation do have some positive impact on task performance albeit not to the extent that was expected. Whilst significant results were found, overall performance by participants, as was also found by the experiments of Chapters 6 and 7, was low, indicating that further research into the concurrent identification of earcons is required if concurrent earcons are to be practically useful in an auditory display.

There are several possible reasons why the experiment failed to produce the results expected. Firstly, the system used to spatialise the earcons was of considerably lower quality than that previously used. This may have made it difficult to separate the earcons. Additionally, whilst the earcon identification task used in this experiment can be regarded as being easier than in previous experiments, the consequences on overall task performance of an incorrect determination are much greater. This is likely to be a standard feature of any "real world" task and as such indicates that a greater increase in earcon identification using the experimental methodology of Chapter 6 would be needed to increase performance on a specific task like diary browsing. This also indicates that if the experimental methodology of this chapter had been employed in Chapters 5, 6 and 7, our understanding of concurrent earcon identification would be significantly worse than it currently is due to a likely increase in non-significant results.

It is not possible, given the results of this experiment, to provide conclusive answers to the research question *"How much do modifications to the design and presentation of concurrently presented earcons affect performance in "real world" tasks?"*, from Section 4.7. Potential reasons for the lack of significant results were given in the previous paragraph, however there is no evidence available to determine their validity. However, overall user performance was low and it is clear that further research into both the effect on ecological task performance of concurrently presented earcons, as well as research to further improve the identification of concurrently presented earcons should be undertaken.

Chapter 9

Conclusions

The final chapter of this thesis provides a summary of the work undertaken, as well as its limitations. In addition possible future directions in the study of concurrently presented earcons are discussed.

9.1 Summary of the Thesis

9.1.1 Literature Review and Motivation

This thesis started by introducing an initial motivation for this work, that concurrently presenting multiple sounds as part of a human computer interface can be advantageous, allowing for example, an increase in the bandwidth of data presentation. However when concurrently presenting sounds they may interfere with each other making it difficult to determine the data which are encoded in each sound. Chapter 2 introduced a basic overview of sound and how it is perceived by the human auditory system, as well as providing some background information to several features used later in the thesis, such as spatialised sound presentation and Auditory Scene Analysis. The chapter concluded with a short discussion on why sounds interfere with each other when concurrently presented, and provided a taxonomy of the dimensions by which such interference can be considered. Chapter 3 followed on from the basic introduction of sound by introducing the topic of auditory displays, systems which use audio to communicate information to users. The advantages and disadvantages of using such systems were outlined, and the four main means of encoding data in sound (sonification, speech, auditory icons and earcons) were described. Earcons as discussed in Section 3.3.4, were structured sounds which could be easily parameterised. The most powerful types of earcon (hierarchical and transformational) were composed of a grammar, therefore allowing them to be more easily learned on account of their internal structure. It was also discussed that the grammar made earcons from the same set similar, with those earcons sharing the same timbre, melody or register. Thereby, earcons used in a particular auditory display will sound quite similar to each other. The chapter concluded with a discussion of spatialised sound presentation as part of an auditory display and several example systems (many of which used concurrent sound presentation) were described.

Chapter 4 elaborated on the work of the previous two chapters by discussing concurrent sound presentation in detail. With the use of examples and counter examples, the advantages of using concurrent audio were described. The disadvantages of concurrent audio were then described, the most notable of which being that sounds which are similar to each other (according to the taxonomy described in Chapter 2) will interfere, making it difficult for a user to extract data from the sounds. Although examples of the use of concurrent audio in various systems exist, very few of these systems have been evaluated to determine problems with such concurrent presentation of audio. In addition, there are no guidelines available for designers to use when constructing auditory displays which use concurrent audio to do so effectively. It was argued that this would be a specific problem for concurrently presented earcons formed from the same grammar, as these would have sounds which are similar according to the taxonomy of Chapter 2. In addition, although modifying the earcons to make them more distinguishable would be possible, it is important not to change the earcons so much that they lose their relationship with each other and the data they represent, and thus the "grammar" which allows their easy identification, is destroyed. No guidelines to allow designers to redesign earcons for concurrent presentation existed. A study which compared user performance in a map navigation task, where one condition involved a solely visual display and the other incorporated concurrent earcons to represent map items which were not displayed on the visual display, found that the use of earcons did not significantly improve participants' speed or efficiency in the map navigation task, additionally participants' subjective workload was significantly increased in the condition which used the earcons. Some participants described that the earcons "fused" together such that their individual attributes were difficult to distinguish. This was argued as an example of the problems of concurrent earcon presentation. The chapter then discussed the limited amount of previous work which had been undertaken to investigate concurrent sound presentation in auditory displays. It was determined from this that identifying concurrently presented earcons was a difficult problem, and if earcons were to be used in such a way, an investigation into how well they were identified when concurrently presented, and how they could be better designed to increase identification needed to be carried out. The thesis then described and justified four research questions, the answers to which would be investigated in the following chapters.

9.1.2 Original Thesis Work

Chapter 5 investigated how the identification of concurrently presented earcons was affected by the number concurrently presented. Sets of one, two, three and four earcons were concurrently presented, and participants had to identify all of the data attributes encoded in each earcon. Results showed that as the number of earcons concurrently presented increased above one, the proportion of those earcons which could be identified was reduced, with correct earcon identification falling from 70% correct for one presented earcon, to 30% for four concurrently presented earcons. This was confirmed by subjective workload assessments by participants which showed that as the number of concurrently presented earcons was increased above one subjective workload increased. It was argued that the results of this experiment validated the conclusion of Chapter 4 that identifying concurrently presented earcons was difficult.

Once the problems of concurrent audio had been identified, Chapter 6 considered how the identification of earcons could be improved by modifications to the design and presentation of those earcons. Sets of four earcons were concurrently presented to participants, with each set being presented multiple times with different modifications based on Auditory Scene Analysis (ASA) research (see Section 2.6), to either the design or presentation of the earcons. Those modifications which significantly improved the identification of earcons or their attributes (incorporating a 300ms onset-to-onset gap between concurrently presented earcon, and presenting each earcon with a different timbre) were combined to measure the overall improvement in earcon identification. A discussion of the results and guidelines for the design and presentation of concurrent earcons in non-spatial environments was presented.

Chapter 7 extended the work of the previous two by considering the identification of concurrent earcons when those earcons were presented in different spatial locations around a user's head (A discussion on the reasons why spatial presentation was investigated separately from non-spatial presentation was provided in Chapter 6), two studies investigating concurrent spatial presentation of concurrent earcons were discussed. The first experiment compared the identification of non-spatially presented earcons which incorporated the guidelines identified in Chapter 6, with spatially presented earcons. *The results showed that the identification of earcons and their attributes was significantly (though not largely) better when spatial presentation was used.* The second experiment compared the identification of two sets of spatially presented earcons. One condition presented earcons spatially, the other condition also spatially presented earcons but incorporated the guidelines of concurrent earcon presentation from Chapter 6. That experiment showing that again significant improvements in earcon attribute identification occurred when spatialisation and the earcon guidelines of Chapter 6 were combined. Although, as with the previous experiment the magnitude of the differences in identification were not great enough to ensure that the improvements in identification would be practically useful. These results were discussed and guidelines for concurrent earcon presentation identified. Guidelines from these experiments were identified, and these extended those from Chapters 5 and 6.

Because the experimental work from Chapters 5-7 was of an abstract nature, Chapter 8 sought to ground the guidelines for earcon presentation in a more ecologically valid scenario. A discussion of mobile based diaries and some of their problems was undertaken. An audio based diary was created to identify the net improvement in task performance of the earcons described in Chapters 5-7. Participants had to walk a route whilst carrying out browsing tasks on the diary, answering summary questions about entries in the diary. Each diary contained four appointments per day, each of which was presented as an earcon. A set of earcons which incorporated the guidelines from Chapters 5-7 were compared to a set which did not. The results of this experiment failed to show any significant differences in performance between the two conditions. However, overall subjective workload was significantly reduced in the condition which incorporated the guidelines from Chapters 5-7. The possible reasons for a lack of significance between the two conditions was discussed and possible future directions were outlined.

9.2 Contributions of the Thesis

This thesis has presented the first investigation which has identified the extent to which concurrently presented earcons interfere with each other, and the impact on earcon identification of such interference. This thesis has also carried out the first investigation to empirically evaluate the effectiveness of modifications to the design and presentation of earcons to avoid earcons interfering with each other. Additionally, this thesis has identified and considered that the more powerful earcon types (transformational and hierarchical, see Section 3.3.4) cannot be arbitrarily modified to make them more easily identifiable when concurrently presented, as the relationships between the earcons that make them powerful communicating sounds, their grammar, will be destroyed, thus making it difficult to identify the data encoded in the earcons. Previous studies [104] have failed to take this important aspect of concurrent earcon presentation into account. The novel contributions of this thesis will be outlined in further detail based around the four research questions, introduced in Chapter 1, and further justified in Section 4.7.

RQ1: What is the effect on earcon identification of varying the number of earcons which are concurrently presented?

One issue identified from the literature review was that varying the number of concurrently presented auditory sources affected their identification. Work by Brungart, Ericson and Simpson [29], on the identification of concurrently presented speech, and work on concurrently presented melodies by Gerth [57], found a linear relationship between the number of auditory sources presented, and performance on an identification task. However, their work disagreed as to the magnitude of the relationship between number of sources and identification performance, with Gerth proposing that there was a much lower degradation in performance as the number of auditory sources increased. RQ1 sought to identify if such a linear relationship existed between the number of concurrently presented earcons and, by identifying the gradient of any such relationship, how difficult it was to identify concurrently presented earcons.

Chapter 5 investigated this question via an experiment where sets of one, two, three and four earcons were concurrently presented, and participants had to identify all three data parameters encoded in each earcon. The results showed that there was a linear relationship between the number of concurrently presented earcons, and the proportion of those earcons which could be identified. This trend agrees most closely with the work of Brungart, Ericson and Simpson [29], with performance for earcon identification dropping from 70% correct for one “concurrently” presented earcon, to 30% for four concurrently presented earcons (see Figure 5.7). Identification of earcon attributes showed the same trend but with much higher levels of identification. The identification of timbre and melody reduced from 90% for one earcon to around 65% for four. Identification of register was lower, and consistent with earlier work with individual earcons by Brewster, Wright and Edwards [16]. More surprisingly, the identification of register was similar irrespective of the number of earcons concurrently presented. This was possibly due to prior modifications to the earcons to avoid lower level interference (see Section 2.3.2), indicating that modifications to the design and presentation of the earcons may improve identification. Additionally, if the number of data parameters encoded in each earcon was reduced (the earcons used in this work encoding the maximum number of parameters, and therefore being a "worst case" scenario), the individual attribute results indicate that overall earcon identification may be higher. Participants' subjective workload, measured by a set of modified NASA TLX questionnaires [60], agreed with the identification performance results showing significant increases in workload as the number of concurrently presented earcons was increased.

RQ2: How can concurrently presented earcons be more robustly designed and presented in order to increase the number which can be identified?

The work on RQ1 identified that significant reductions in performance occur when more than one earcon is concurrently presented. However, several studies on concurrent auditory sources other than earcons have shown that improvements in identification can occur if the sounds which are concurrently presented are modified to increase the differences between them [29, 31, 57]. Such a conclusion is also validated by the discussion of Auditory Scene Analysis (ASA) from Section 2.6.

Chapter 6 investigated how modifications to the design and presentation of concurrent earcons based on ASA could improve earcon identification. Several ASA principles which were considered to offer the best possibility of improving earcon identification were applied to a set of earcons. Each modification to the earcons was evaluated separately, as the modifications, if successful, would constrain the design of the earcons in some way. In making these changes however, it was important to ensure that concurrently presented earcons are not changed so much that the "grammar" that linked them together was destroyed and their identification became harder; so all modifications were designed such that they would work within the grammar of the existing set of earcons used.

An experiment using the same methodology as that of Chapter 5 was carried out. This experiment used four concurrently presented earcons to allow easy comparison back to previous work, as well as avoiding any possibility of a ceiling effect in the data, and therefore represents a "worst case scenario" in earcon identification. Modifying the earcons, such that each concurrently presented earcon with the same timbre encoded attribute was presented with a different timbre from the same instrument group was found to significantly improve identification of the timbre of the earcon. In addition, staggering the onset of each earcon by 300ms was found to significantly improve earcon identification as well as earcon timbre identification. Combining both features together again showed significant improvement in the identification of earcons over those with no modifications; however improvements were not great enough to be significantly better than any one individual modification. This was possibly due to the constraints on the extent of modifications which could be applied in order to preserve the earcon grammar. If fewer attributes could be encoded in each earcon, this would leave major auditory attributes which may be arbitrarily modified to promote particular streaming, albeit at a loss of the communicating power of the earcons. Whilst the previously described modifications to the earcons showed a significant improvement, this improvement was not large, with only an average increase in identification of 0.5 earcons. Such an improvement would, as indicated by the work of Chapter 8, not be enough to allow for a practical improvement in user task performance. However, if fewer earcons were concurrently presented and such an improvement could be maintained, 0.5 earcons would be a much greater improvement in identification. Other studies [48, 57, 29], indicate that although the trend between the number of auditory sources presented and identification (as shown in Figure 5.7) still exists with modifications to the auditory sources, the gradient can be significantly flattened. Although the impact of such a modification on earcon would need to be determined.

RQ3: What is the impact of presenting concurrent earcons in a spatialised auditory environment?

One aspect of concurrent earcon identification which can also be considered as contributing towards an answer for RQ2, is the impact on identification of presenting each earcon in a distinct position around the user's head. There are several reasons for investigating such a presentation modification in isolation from those investigated under RQ2. Most notably, as discussed in Section 2.5, spatial sound perception uses differences in the auditory signals reaching the left and right ears to determine sound source direction, and thus to separate concurrent sounds in space. An Auditory Display which used such a technique would require a user to wear headphones, or have a set of properly calibrated speakers with the user located at a specific point [126]. There are several scenarios where such a situation may be impractical or impossible. In addition, from the literature review in Section 3.4, it was identified that when spatial sound presentation is used as part of an auditory display, spatial position of a sound source may be mapped to some other data parameter, such as time [141, 124]. In such situations there is a structured mapping between sound location and data, such a mapping (as with the mapping between data attributes and sound parameters in earcons) cannot be arbitrarily changed to promote better streaming as the mapping will be destroyed. Therefore it cannot be assumed that simply by spatialising sounds, those sounds will be far enough apart that they will not interfere (see Section 2.4.2). Taking these issues into account shows the validity of investigating the spatialisation of concurrent earcons as a separate feature to those investigated in Chapter 6.

The work described in Chapter 7 examined two features of spatialised earcon presentation. The first was to compare earcon identification performance of non-spatially presented earcons which incorporated the earcon design guidelines which had been identified from the work on RQ2, with spatially presented earcons which had no other design or presentation modifications. In other words, if spatial presentation is available to designers, is earcon identification superior to non-spatial presentation which incorporates previously identified design and presentation guidelines? An experiment similar to those used for the work on RQ1 and RQ2 was undertaken. The results showed that earcon identification performance was indeed superior to non-spatial presentation; additionally, the identification of earcon register was significantly improved by spatial earcon presentation. The obvious question to follow from these results is will the application of the concurrent earcon guidelines discovered from the work of RQ2 to spatial earcon presentation further increase the identification of those earcons? This is important due to the issues of sufficient spatial separation previously discussed (see Section 2.4.2), which may mean that inspite of the superiority of spatial presentation for concurrent earcon identification, it may still be the case that earcons cannot be sufficiently separated, requiring the implementation of the non-spatial earcon presentation guidelines. As Bregman [14] (pp. 302) notes " *the human auditory system does not give an overriding importance to the spatial cues for belongingness but weighs these cues against all the others. When these cues all agree, the outcome is a clear perceptual organisation, but when they do not, we can have a number of outcomes.*". An experiment of the same methodology as that previously undertaken was carried out. This experiment compared the identification of concurrent earcons which were spatially presented, and concurrent earcons which were spatially presented and incorporated the guidelines identified in the work of RQ2. The results failed to show a significant difference between the number of earcons identified, but did show a significant improvement in the identification of the timbre encoded earcon parameter in the spatial condition which incorporated the concurrent earcon presentation guidelines. Whilst these results show that the non-spatial earcon presentation guidelines can be used in spatial presentation situations, and provide significant improvements in identification, over spatially presented earcons which do not incorporate the non-spatial earcon presentation guidelines, the magnitude of those improvements is, as with the work on RQ2 low. The relationships that exist between concurrently presented earcons are so strong that the modifications and degree of those modifications that can be applied to desirably stream the earcons do not have a large impact and such may be of limited practical benefit.

RQ4: How much do modifications to the design and presentation of concurrently presented earcons affect performance in "real world" tasks?

The work undertaken in answering the previous research questions elicited a great deal of knowledge on the identification of concurrently presented earcons, as well as a set of guidelines which could be used by designers of auditory displays to make more informed choices when incorporating concurrent earcons into their displays. However, due to the nature of the experiments undertaken in this work, it could not be assumed that the improvements in earcon identification found would translate into improvements in real human computer interface tasks. Such performance is important to determine given the significant yet low improvements in earcon identification from the previous experiments. Chapter 8 investigated this issue by comparing performance between earcons which included the guidelines for concurrent earcon identification, identified from the work with RQ2 and RQ3 with earcons which did not. Again, to ensure their was no ceiling effect in the data, four earcons were concurrently presented in each condition. The experiment attempted to exploit some of the key advantages argued for the use of both auditory displays (see Section 3.2.1), and concurrent sound presentation (see Section 4.2). Participants had to perform a diary browsing task and answer questions about the entries in the diary. Diary entries were encoded as earcons. Participants had to walk a predefined route whilst carrying out the task to simulate checking the diary "on the move".

It was hypothesised that participants would answer more questions correctly, walk closer to their optimal walking speed, and would record significantly lower subjective workload (measured via modified NASA TLX questionnaires), when the earcons were modified to incorporate the guidelines for concurrent earcon presentation previously identified. The results, however, did not show a significant variation between the conditions for any objectively measurable performance metric. Participants did record significantly lower subjective workload when the earcon design guidelines were incorporated. An improvement in subjective workload is important and should make a system using concurrent earcon presentation more acceptable to users [60]. However, it is only possible to consider this result as a partial answering of RQ4, and more work needs to be undertaken to identify the impact of concurrent earcon presentation on "real world" task performance. This issue is further discussed in Section 9.3.

As previously discussed, the work of this thesis represents the most comprehensive and thorough investigation into the identification of concurrently presented earcons undertaken. Whilst other studies have used concurrently presented earcons [25], or have indicated that problems exist in their concurrent presentation [104], no study has previously empirically investigated their identification when their number is varied, or how they can be modified to improve their identification. The studies of this thesis have found that the identification of concurrent earcons as part of an auditory display is difficult and that there is a limited amount (at least given the approach of this thesis) which can be done to practically reduce the problems of concurrent earcon identification. Whilst it is possible to apply modifications to improve concurrent earcon identification, it must be concluded given the results of Chapters 6 and 7 and the experiment from Chapter 8 that the degree of modifications that can be applied allows for limited practical benefit. Any practical benefit is likely to be found, as indicated in Section 6.4 and by other studies [29], when the number of earcons concurrently presented is significantly reduced (see Section 9.3).

The results of this thesis indicate that although concurrently presenting earcons can bring significant advantages (see Section 4.2), because of the inherent relationships between earcons from the same grammar it is difficult to leverage these advantages. Based on the results of the work contained within this thesis, designers who wish to employ concurrent earcons as part of an auditory display should consider their use carefully and reduce the number to be presented as much as possible. Additionally the use of simpler earcons such as compound or one-element (see Section 3.3.4) may be more practical since there is no requirement, as with the transformational and hierarchical types, to share attributes and as such greater differences may be introduced between concurrent earcons. Other possibilities to improve concurrent earcon identification performance are discussed in Section 9.4. Additionally specific guidelines (given below) for the improvement of concurrent earcon identification derived from the work of this thesis have been identified and should be used whenever concurrent earcons are employed.

Guideline 1: Increasing the number of concurrently presented earcons significantly reduces the proportion of the earcons which 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 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, Wright and Edwards [25], 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 [118] 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 300ms onset-to-onset gap between the starts of concurrently presented earcons will make the earcons more identifiable to users.

Guideline 5: The use of spatialised presentation with headtracking significantly improves the identification of concurrently presented earcons. Therefore spatialised presentation should be employed whenever practically possible in cases where earcons will be concurrently presented.

Guideline 6: Whilst spatial audio significantly increases the number of concurrent earcons that can be identified, it does not always guarantee that earcons will be desirably streamed, hence the maximum amount of angular (in azimuth) separation between concurrently presented earcons should be used. Guidelines 1-5 should also be incorporated when using spatial presentation to improve earcon identification.

The work presented in this thesis can therefore be considered as defending the thesis statement described in Chapter 1 that *"Identifying concurrently presented earcons where those earcons are constructed from a grammar is difficult. Whilst modifications can be undertaken to significantly improve earcon identification, these modifications are constrained due to the need to preserve the mapping between data and sound"*.

9.3 Limitations of the Thesis

Whilst this thesis has contributed a great deal of knowledge on the identification of concurrently presented earcons, there are a few limitations to its work which must be outlined.

Firstly, this thesis has been unable to show that the earcon guidelines identified through work in Chapters 5 to 7 improve task performance in real world usage. Whilst the earcon design guidelines do significantly reduce subjective workload in real world tasks, and as such are likely to be useful if implemented when concurrent earcons are presented, without showing an improvement in task performance their impact is limited. As already discussed in Section 8.5 there are several possible reasons why no improvement in task performance was found in the experiment of Chapter 8. Although, it is not possible with the data available to make definitive conclusions as to the lack of improvement in task performance, it is likely that the overall poor identification of earcons in Chapters 6 and 7 have a strong impact. Further work which applies the concurrent earcon guidelines in different types of environments and situations, as well as a reimplementation of the experiments from Chapter 7 using the spatialisation system from the experiment described in Chapter 8 would be useful to further understand and overcome this issue.

Another limitation of this thesis is the emphasis on consistency between the experiments of Chapters 5, 6 and 7. Namely that sets of four concurrently presented earcons were used for the experiments in Chapter 6 and 7. This was due to two factors, firstly it allowed comparisons to other work on the identification of concurrent sounds which tended to use four concurrent sources [29, 57] to be made, additionally other researchers claim that it is reasonable to concurrently present four sounds in real interfaces [12, 105, 48]. It was also unclear what the impact of the earcon modifications of Chapter 6 and the spatialisation of earcons from Chapter 7 would be. It was important that no ceiling effect in the results was obtained. In other words, any modification to the design and presentation of the earcons should not make their identification so easy that there are an insufficient number of earcons presented to measure the actual improvement in identification. Therefore the work of this thesis has investigated the boundary of concurrent earcon identification. Whilst it can be reasonably assumed based on other work (see Section 4.6), that reducing the number of concurrently presented earcons and applying the concurrent earcon guidelines would improve identification, it is not clear what the numerical improvement would be. Would a linear relationship still exist between the identification of one, two and three concurrently presented earcons? Further work applying the concurrent earcon guidelines to sets of one, two and three concurrently presented earcons would yield useful results which would allow designers to make even better informed decisions when using concurrent earcon presentation.

9.4 Future Work

In the previous sections, the main contributions of the thesis and the limitations of those contributions were outlined. This section discusses ways in which the work of this thesis may be expanded upon to further improve our understanding of concurrent earcon presentation and identification, and its relationship to other auditory displays.

9.4.1 Spatial Separation versus Identification

A final issue with the work of this thesis concerns the identification of earcons in a spatialised environment. Whilst the thesis has shown improvements in earcon identification when those earcons are spatially separated, it has not investigated the relationship between the degree of spatial separation and earcon identification. As discussed in Section 2.4.2, the ability to distinguish two sounds concurrently presented in space, is a function of both the azimuthal distance between those sounds in space and the similarity of those sounds. This is consistent with the discussion of auditory scene analysis from Section 2.6. Best, van Schaik and Carlile [10], as discussed in Section 3.2.2, found that to separate two broadband sounds in space could require 60^o separation in azimuth before they could confidently be perceived as separate. In the experimental work of this thesis, maximum azimuthal differences between concurrently presented earcons were used to ensure separation in space. However, if a designer wished to present concurrent earcons, and map some data parameter to spatial position (one of the advantages of spatial sound presentation as described in Section 3.4), it would be useful to know how close two earcons could be situated before they began to interfere with each other, and the extent of such interferences. This thesis has only considered the two extremities, where earcons are all collocated, and where they are maximally separated in azimuth. Mapping the “middle group” would give designers who wished to exploit such advantages a great deal more knowledge to make informed decisions.

9.4.2 Reducing Earcon Parameters

One feature from the results of the experimental work of this thesis, is that even when modifications to the design and presentation of earcons yield significant results, the magnitude of such increases is small. In part this may be due to investigating the “worst case scenario” of earcon identification, as discussed in Section 9.3, but is also likely due to the constraints placed on any modifications, due to the need to protect the earcon grammar and the mapping between sound and data, where arbitrarily large modifications between concurrently presented earcons cannot be introduced. As indicated by the results of the experiment in Chapter 5, reducing the number of data parameters encoded in each earcon may increase the identification of those earcons. Additionally, this would leave auditory parameters that did not form part of the earcon grammar. As such, arbitrarily large modifications may be introduced between concurrently presented earcons, along that parameter, which would be likely to further influence the auditory streaming process. However, incorporating such a feature would reduce the amount of information that could be encoded in each earcon. Additionally, the earcon designer would be faced with choices concerning which auditory parameters to use as part of the earcon grammar and auditory parameters to increase the differences between concurrently presented earcons. However, given the limitations of the modification applied to the work of this thesis, such a study may be useful.

9.4.3 Identification in Mixed Concurrent Auditory Environments

It is highly unlikely that most interfaces which would use concurrent earcon presentation would only use earcons. As already discussed in Section 3.3 there are several ways of encoding data in sound, and it is likely that a designer would use a variety of methods including earcons, dependent on the use to which the data would be put. It would be important for such designers to understand how concurrently presented earcons interacted with other auditory display techniques such as sonification or speech. However such a study is extremely complex, and would take many experiments to determine solutions, given the scope of the solution space that would be available. However it would allow the relationship between the work of this thesis and other previously undertaken studies to be further explored.

9.5 Final Remarks

In conclusion, whilst the presentation of concurrent sound as part of an auditory display has several advantages, there are a number of issues that must be considered. One issue being that sounds which are concurrently presented may interfere with each other, making it difficult to determine the information encoded within them. This thesis has identified that a lack of guidelines for the design and presentation of concurrent sounds as part of an auditory display exist, making it difficult for designers to exploit the advantages of concurrent sound presentation. The work of this thesis has identified that the presentation of concurrent earcons makes it difficult for those earcons to be identified by users. Work has been undertaken showing the extent of the problem and how identification of concurrent earcons is affected by the number of earcons concurrently presented. This thesis has sought to reduce the problems of concurrent earcon identification, by bringing together guidelines for the design of effective earcons [26] and research into auditory scene analysis [14]. Empirically evaluated modifications to the design and presentation of earcons have been developed, that significantly improve their identification when concurrently presented, without destroying the relationships between individual earcons that make them powerful communicating sounds. However in avoiding destroying the grammar, the degree to which these modifications impact earcon identification is limited, leading to little practical increase in earcon identification. These modifications to the design and presentation of concurrent earcons have also been applied and evaluated in a “real world” interface. This evaluation showed that incorporating the design and presentation modifications significantly reduced participants’ subjective workload, although no significant improvement in task performance was found. The work of this thesis has yielded a set of empirically evaluated guidelines for the presentation of concurrent earcons, which can be used by future designers of auditory displays to make informed design decisions, and as such significantly improve our understanding of concurrently presented earcons.

Appendix A

Glossary

Attack: The time between a sound starting and that sound reaching a steady amplitude [116].

Attenuation: A reduction in the amplitude of certain frequencies in a sound [152].

Dynamic Respatialisation: The use of a position and orientation sensing device to adjust the position of auditory sources relative to a listener as the listener moves their head.

Free Field: Listening to spatialised sounds generated from monaural sound sources. I.e. instead of synthesising sounds coming from a location, the sound is actually generated from that location.

Head Shadowing: The inability of higher frequency sounds to diffract around the head causing the head to cast an auditory "shadow" across the ears.

Human Auditory System: The human system used to detect and derive meaning from a sound.

Kemar Dummy: A plastic dummy of the head and shoulders of a human being, used in the recording and creation of head related transfer functions (HRTFs).

Melody: A rhythmic structure where successive notes may have different pitches.

Rhythm: The relative durations of the notes and gaps between notes in a particular sequence of notes.

Scale: A sequence of pitch intervals referenced to a particular pitch. This thesis uses a chromatic scale which consists of eight distinct pitch intervals, each of which raises the perceived pitch of a sound by an even amount.

Spectra: The frequencies and relative amounts of those frequencies contained within a sound.

Tempo: The speed of a composition [116].

Tone: An auditory waveform composed of sinusoids.

Appendix B

Dolphin Evaluation - Training Sheets

This appendix contains copies of the training sheets from the Dolphin evaluation described in Section 4.4.

Appendix C

Dolphin Evaluation - Raw Data

This appendix presents the raw results from the experiment described in Chapter 4.

C.1 Average Time to Create a Tour

Participant	Dolphin Condition (sec)	Standard Scrolling Condition (sec)
1	217	174
2	124	59
3	54	66
4	142	145
5	73	73
6	488	96
7	59	74
8	206	56
9	119	109
10	117	97
11	126	102
12	63	89
13	43	52
14	0	0
15	142	145
16	73	73

C.2 Percentage Optimal Path Length

Participant	Dolphin Condition	Standard Scrolling Condition
1	95	99
2	85	87
3	84	86
4	88	87
5	81	93
6	88	91
7	90	100
8	100	96
9	92	92
10	97	95
11	84	91
12	100	95
13	100	91
14	95	67
15	87	87
16	81	89

C.3 Tours with Missing or Added Rides

Participant	Dolphin Condition	Standard Scrolling Condition
1	3	0
2	0	0
3	1	1
4	0	0
5	0	2
6	0	1
7	0	3
8	3	1
9	2	0
10	2	1
11	1	0
12	1	1
13	3	0
14	3	3
15	0	0
16	0	2

C.4 Modified NASA TLX Workload Ratings

Participant	Modified NASA TLX Attribute	Standard Scrolling Condition	Dolphin Condition
1	Annoyance Experienced	3	19
2	Annoyance Experienced	7	15
3	Annoyance Experienced	2	14
4	Annoyance Experienced	9	17
5	Annoyance Experienced	6	16
6	Annoyance Experienced	1	18
7	Annoyance Experienced	4	13
8	Annoyance Experienced	1	3
9	Annoyance Experienced	1	9
10	Annoyance Experienced	1	18
11	Annoyance Experienced	2	2
12	Annoyance Experienced	10	14
13	Annoyance Experienced	7	6
14	Annoyance Experienced	6	7
15	Annoyance Experienced	5	5
16	Annoyance Experienced	2	7
1	Effort Expended	5	11
2	Effort Expended	5	18
3	Effort Expended	6	14
4	Effort Expended	12	14
5	Effort Expended	8	18
6	Effort Expended	20	7
7	Effort Expended	11	14
8	Effort Expended	11	11
9	Effort Expended	5	13
10	Effort Expended	16	18

Participant	Modified NASA TLX Attribute	Standard Scrolling Condition	Dolphin Condition
11	Effort Expended	12	14
12	Effort Expended	12	14
13	Effort Expended	6	10
14	Effort Expended	12	13
15	Effort Expended	14	14
16	Effort Expended	12	11
1	Frustration Experienced	3	16
2	Frustration Experienced	9	13
3	Frustration Experienced	6	20
4	Frustration Experienced	7	17
5	Frustration Experienced	9	17
6	Frustration Experienced	4	2
7	Frustration Experienced	17	16
8	Frustration Experienced	1	3
9	Frustration Experienced	1	9
10	Frustration Experienced	10	19
11	Frustration Experienced	7	7
12	Frustration Experienced	6	8
13	Frustration Experienced	7	7
14	Frustration Experienced	6	5
15	Frustration Experienced	8	11
16	Frustration Experienced	4	6
1	Mental Demand	9	11
2	Mental Demand	6	13
3	Mental Demand	2	16
4	Mental Demand	11	13
5	Mental Demand	11	17
6	Mental Demand	18	7
7	Mental Demand	13	18
8	Mental Demand	13	11
9	Mental Demand	3	17
10	Mental Demand	16	18
11	Mental Demand	11	15
12	Mental Demand	13	17
13	Mental Demand	10	13
14	Mental Demand	13	15
15	Mental Demand	15	14
16	Mental Demand	14	13
1	Overall Preference	19	9
2	Overall Preference	18	7
3	Overall Preference	20	4
4	Overall Preference	20	1
5	Overall Preference	17	5
6	Overall Preference	11	11
7	Overall Preference	8	13
8	Overall Preference	20	1
9	Overall Preference	19	7
10	Overall Preference	19	1
11	Overall Preference	15	16
12	Overall Preference	13	10
13	Overall Preference	10	10
14	Overall Preference	11	17
15	Overall Preference	15	15
16	Overall Preference	12	17
1	Performance Level Achieved	11	11

Participant	Modified NASA TLX Attribute	Standard Scrolling Condition	Dolphin Condition
2	Performance Level Achieved	16	11
3	Performance Level Achieved	20	10
4	Performance Level Achieved	16	13
5	Performance Level Achieved	16	12
6	Performance Level Achieved	20	19
7	Performance Level Achieved	9	12
8	Performance Level Achieved	20	19
9	Performance Level Achieved	19	11
10	Performance Level Achieved	13	15
11	Performance Level Achieved	10	12
12	Performance Level Achieved	15	12
13	Performance Level Achieved	10	14
14	Performance Level Achieved	13	17
15	Performance Level Achieved	12	12
16	Performance Level Achieved	8	14
1	Physical Demand	1	1
2	Physical Demand	8	10
3	Physical Demand	10	19
4	Physical Demand	5	6
5	Physical Demand	5	5
6	Physical Demand	11	2
7	Physical Demand	10	7
8	Physical Demand	1	1
9	Physical Demand	3	9
10	Physical Demand	14	18
11	Physical Demand	2	2
12	Physical Demand	3	3
13	Physical Demand	3	2
14	Physical Demand	15	4
15	Physical Demand	12	5
16	Physical Demand	6	8
1	Time Pressure	7	11
2	Time Pressure	4	14
3	Time Pressure	6	14
4	Time Pressure	10	9
5	Time Pressure	5	18
6	Time Pressure	1	1
7	Time Pressure	9	5
8	Time Pressure	7	5
9	Time Pressure	1	9
10	Time Pressure	4	8
11	Time Pressure	10	11
12	Time Pressure	11	11
13	Time Pressure	2	2
14	Time Pressure	9	16
15	Time Pressure	11	11
16	Time Pressure	12	2

Appendix D

Condition Grouping For Non-Spatial Experiments

As discussed in Section 5.2.2, the conditions from the experiments described in Chapters 5 and 6 were interleaved. With each participant performing one condition of each experiment in a counterbalanced order. The table below shows how the conditions from each experiment were paired.

Condition from experiment in Chapter 5		**Condition from experiment in Chapter 6**
Four Earcon Condition (aka Original Earcon Set Condition from Chapter 6)	*was paired with*	Melody Altered Earcon Set Condition
Three Earcon Condition	*was paired with*	Multi-timbre Earcon Set Condition
N/A		Extended Training Condition
Two Earcon Condition	*was paired with*	Staggered Onset Condition
One Earcon Condition	*was paired with*	Final Condition

Appendix E

Number of Earcons versus Identification - Training Sheets

This appendix contains copies of the training sheets from the experiment described in Chapter 5.

Appendix F

Number of Earcons versus Identification - Raw Data

This appendix presents the raw results from the experiment described in Chapter 5.

F.1 One Earcon Condition

F.1.1 Earcons and Earcon Attributes Identified

Participant	Attribute	Value
1	Average Percentage Rides Identified	70.00
2	Average Percentage Rides Identified	70.00
3	Average Percentage Rides Identified	80.00
4	Average Percentage Rides Identified	75.00
5	Average Percentage Rides Identified	45.00
6	Average Percentage Rides Identified	75.00
7	Average Percentage Rides Identified	80.00
8	Average Percentage Rides Identified	50.00
9	Average Percentage Rides Identified	70.00
10	Average Percentage Rides Identified	85.00
11	Average Percentage Rides Identified	50.00
12	Average Percentage Rides Identified	80.00
13	Average Percentage Rides Identified	70.00
14	Average Percentage Rides Identified	90.00
15	Average Percentage Rides Identified	80.00
16	Average Percentage Rides Identified	67.00
1	Average Percentage Ride Types Identified	100.00
2	Average Percentage Ride Types Identified	90.00
3	Average Percentage Ride Types Identified	95.00
4	Average Percentage Ride Types Identified	90.00
5	Average Percentage Ride Types Identified	95.00
6	Average Percentage Ride Types Identified	95.00
7	Average Percentage Ride Types Identified	95.00
8	Average Percentage Ride Types Identified	80.00
9	Average Percentage Ride Types Identified	100.00
10	Average Percentage Ride Types Identified	100.00
11	Average Percentage Ride Types Identified	75.00
12	Average Percentage Ride Types Identified	100.00
13	Average Percentage Ride Types Identified	70.00
14	Average Percentage Ride Types Identified	100.00
15	Average Percentage Ride Types Identified	100.00
16	Average Percentage Ride Types Identified	90.00
1	Average Percentage Ride Intensities Identified	100.00
2	Average Percentage Ride Intensities Identified	95.00
3	Average Percentage Ride Intensities Identified	100.00

Participant	Attribute	Value
4	Average Percentage Ride Intensities Identified	100.00
5	Average Percentage Ride Intensities Identified	90.00
6	Average Percentage Ride Intensities Identified	100.00
7	Average Percentage Ride Intensities Identified	100.00
8	Average Percentage Ride Intensities Identified	90.00
9	Average Percentage Ride Intensities Identified	90.00
10	Average Percentage Ride Intensities Identified	95.00
11	Average Percentage Ride Intensities Identified	95.00
12	Average Percentage Ride Intensities Identified	92.00
13	Average Percentage Ride Intensities Identified	65.00
14	Average Percentage Ride Intensities Identified	90.00
15	Average Percentage Ride Intensities Identified	100.00
16	Average Percentage Ride Intensities Identified	90.00
1	Average Percentage Ride Costs Identified	70.00
2	Average Percentage Ride Costs Identified	75.00
3	Average Percentage Ride Costs Identified	85.00
4	Average Percentage Ride Costs Identified	85.00
5	Average Percentage Ride Costs Identified	30.00
6	Average Percentage Ride Costs Identified	80.00
7	Average Percentage Ride Costs Identified	80.00
8	Average Percentage Ride Costs Identified	70.00
9	Average Percentage Ride Costs Identified	55.00
10	Average Percentage Ride Costs Identified	90.00
11	Average Percentage Ride Costs Identified	45.00
12	Average Percentage Ride Costs Identified	90.00
13	Average Percentage Ride Costs Identified	65.00
14	Average Percentage Ride Costs Identified	90.00
15	Average Percentage Ride Costs Identified	80.00
16	Average Percentage Ride Costs Identified	80.00

F.1.2 Modified NASA TLX Workload Ratings

Participant	Attribute	Value
1	Mental Demand	10
2	Mental Demand	10
3	Mental Demand	11
4	Mental Demand	9
5	Mental Demand	20
6	Mental Demand	6
7	Mental Demand	16
8	Mental Demand	15
9	Mental Demand	17
10	Mental Demand	10
11	Mental Demand	11
12	Mental Demand	11
13	Mental Demand	12
14	Mental Demand	13
15	Mental Demand	11
16	Mental Demand	10
1	Physical Demand	2
2	Physical Demand	7
3	Physical Demand	4
4	Physical Demand	1

Participant	Attribute	Value
5	Physical Demand	6
6	Physical Demand	6
7	Physical Demand	4
8	Physical Demand	1
9	Physical Demand	7
10	Physical Demand	2
11	Physical Demand	6
12	Physical Demand	3
13	Physical Demand	4
14	Physical Demand	2
15	Physical Demand	1
16	Physical Demand	1
1	Time Pressure	3
2	Time Pressure	3
3	Time Pressure	7
4	Time Pressure	1
5	Time Pressure	20
6	Time Pressure	1
7	Time Pressure	10
8	Time Pressure	1
9	Time Pressure	14
10	Time Pressure	1
11	Time Pressure	15
12	Time Pressure	1
13	Time Pressure	12
14	Time Pressure	13
15	Time Pressure	1
16	Time Pressure	4
1	Effort Expended	16
2	Effort Expended	17
3	Effort Expended	12
4	Effort Expended	9
5	Effort Expended	16
6	Effort Expended	10
7	Effort Expended	13
8	Effort Expended	7
9	Effort Expended	17
10	Effort Expended	17
11	Effort Expended	16
12	Effort Expended	15
13	Effort Expended	17
14	Effort Expended	18
15	Effort Expended	16
16	Effort Expended	16
1	Performance Level Achieved	12
2	Performance Level Achieved	16
3	Performance Level Achieved	10
4	Performance Level Achieved	18
5	Performance Level Achieved	3
6	Performance Level Achieved	11
7	Performance Level Achieved	11
8	Performance Level Achieved	13
9	Performance Level Achieved	11
10	Performance Level Achieved	18
11	Performance Level Achieved	16

Participant	Attribute	Value
12	Performance Level Achieved	10
13	Performance Level Achieved	12
14	Performance Level Achieved	11
15	Performance Level Achieved	12
16	Performance Level Achieved	13
1	Frustration Experienced	5
2	Frustration Experienced	10
3	Frustration Experienced	6
4	Frustration Experienced	1
5	Frustration Experienced	19
6	Frustration Experienced	4
7	Frustration Experienced	10
8	Frustration Experienced	13
9	Frustration Experienced	11
10	Frustration Experienced	1
11	Frustration Experienced	11
12	Frustration Experienced	12
13	Frustration Experienced	10
14	Frustration Experienced	10
15	Frustration Experienced	11
16	Frustration Experienced	3
1	Annoyance Experienced	2
2	Annoyance Experienced	11
3	Annoyance Experienced	6
4	Annoyance Experienced	1
5	Annoyance Experienced	19
6	Annoyance Experienced	1
7	Annoyance Experienced	2
8	Annoyance Experienced	13
9	Annoyance Experienced	14
10	Annoyance Experienced	1
11	Annoyance Experienced	8
12	Annoyance Experienced	5
13	Annoyance Experienced	3
14	Annoyance Experienced	5
15	Annoyance Experienced	2
16	Annoyance Experienced	1

F.2 Two Earcon Condition

F.2.1 Earcons and Earcon Attributes Identified

Participant	Attribute	Value
1	Average Percentage Rides Identified	30.00
2	Average Percentage Rides Identified	55.00
3	Average Percentage Rides Identified	67.50
4	Average Percentage Rides Identified	47.50
5	Average Percentage Rides Identified	32.50
6	Average Percentage Rides Identified	85.00
7	Average Percentage Rides Identified	30.00
8	Average Percentage Rides Identified	80.00
9	Average Percentage Rides Identified	67.50

Participant	Attribute	Value
10	Average Percentage Rides Identified	55.00
11	Average Percentage Rides Identified	32.50
12	Average Percentage Rides Identified	57.50
13	Average Percentage Rides Identified	40.00
14	Average Percentage Rides Identified	32.50
15	Average Percentage Rides Identified	82.50
16	Average Percentage Rides Identified	77.50
1	Average Percentage Ride Types Identified	67.50
2	Average Percentage Ride Types Identified	90.00
3	Average Percentage Ride Types Identified	95.00
4	Average Percentage Ride Types Identified	95.00
5	Average Percentage Ride Types Identified	92.50
6	Average Percentage Ride Types Identified	100.00
7	Average Percentage Ride Types Identified	55.00
8	Average Percentage Ride Types Identified	95.00
9	Average Percentage Ride Types Identified	87.50
10	Average Percentage Ride Types Identified	82.50
11	Average Percentage Ride Types Identified	75.00
12	Average Percentage Ride Types Identified	80.00
13	Average Percentage Ride Types Identified	80.00
14	Average Percentage Ride Types Identified	75.00
15	Average Percentage Ride Types Identified	97.50
16	Average Percentage Ride Types Identified	100.00
1	Average Percentage Ride Intensities Identified	65.00
2	Average Percentage Ride Intensities Identified	97.50
3	Average Percentage Ride Intensities Identified	87.50
4	Average Percentage Ride Intensities Identified	97.50
5	Average Percentage Ride Intensities Identified	62.50
6	Average Percentage Ride Intensities Identified	100.00
7	Average Percentage Ride Intensities Identified	85.00
8	Average Percentage Ride Intensities Identified	100.00
9	Average Percentage Ride Intensities Identified	100.00
10	Average Percentage Ride Intensities Identified	97.50
11	Average Percentage Ride Intensities Identified	75.00
12	Average Percentage Ride Intensities Identified	95.00
13	Average Percentage Ride Intensities Identified	80.00
14	Average Percentage Ride Intensities Identified	95.00
15	Average Percentage Ride Intensities Identified	95.00
16	Average Percentage Ride Intensities Identified	97.50
1	Average Percentage Ride Costs Identified	60.00
2	Average Percentage Ride Costs Identified	65.00
3	Average Percentage Ride Costs Identified	72.50
4	Average Percentage Ride Costs Identified	52.50
5	Average Percentage Ride Costs Identified	62.50
6	Average Percentage Ride Costs Identified	85.00
7	Average Percentage Ride Costs Identified	55.00
8	Average Percentage Ride Costs Identified	85.00
9	Average Percentage Ride Costs Identified	80.00
10	Average Percentage Ride Costs Identified	72.50
11	Average Percentage Ride Costs Identified	65.00
12	Average Percentage Ride Costs Identified	70.00
13	Average Percentage Ride Costs Identified	70.00
14	Average Percentage Ride Costs Identified	45.00
15	Average Percentage Ride Costs Identified	90.00
16	Average Percentage Ride Costs Identified	80.00

F.2.2 Modified NASA TLX Workload Ratings

Participant	Attribute	Value
1	Mental Demand	17
2	Mental Demand	8
3	Mental Demand	17
4	Mental Demand	18
5	Mental Demand	14
6	Mental Demand	14
7	Mental Demand	18
8	Mental Demand	7
9	Mental Demand	7
10	Mental Demand	11
11	Mental Demand	9
12	Mental Demand	18
13	Mental Demand	19
14	Mental Demand	10
15	Mental Demand	16
16	Mental Demand	4
1	Physical Demand	7
2	Physical Demand	8
3	Physical Demand	12
4	Physical Demand	2
5	Physical Demand	3
6	Physical Demand	8
7	Physical Demand	10
8	Physical Demand	3
9	Physical Demand	3
10	Physical Demand	5
11	Physical Demand	13
12	Physical Demand	4
13	Physical Demand	12
14	Physical Demand	3
15	Physical Demand	3
16	Physical Demand	2
1	Time Pressure	17
2	Time Pressure	6
3	Time Pressure	16
4	Time Pressure	2
5	Time Pressure	7
6	Time Pressure	5
7	Time Pressure	11
8	Time Pressure	10
9	Time Pressure	1
10	Time Pressure	5
11	Time Pressure	18
12	Time Pressure	13
13	Time Pressure	6
14	Time Pressure	4
15	Time Pressure	14
16	Time Pressure	4
1	Effort Expended	16
2	Effort Expended	8
3	Effort Expended	18

Participant	Attribute	Value
4	Effort Expended	18
5	Effort Expended	9
6	Effort Expended	11
7	Effort Expended	19
8	Effort Expended	4
9	Effort Expended	4
10	Effort Expended	6
11	Effort Expended	12
12	Effort Expended	16
13	Effort Expended	10
14	Effort Expended	6
15	Effort Expended	15
16	Effort Expended	7
1	Performance Level Achieved	5
2	Performance Level Achieved	12
3	Performance Level Achieved	7
4	Performance Level Achieved	18
5	Performance Level Achieved	12
6	Performance Level Achieved	16
7	Performance Level Achieved	9
8	Performance Level Achieved	13
9	Performance Level Achieved	14
10	Performance Level Achieved	4
11	Performance Level Achieved	8
12	Performance Level Achieved	10
13	Performance Level Achieved	13
14	Performance Level Achieved	10
15	Performance Level Achieved	16
16	Performance Level Achieved	17
1	Frustration Experienced	10
2	Frustration Experienced	6
3	Frustration Experienced	17
4	Frustration Experienced	3
5	Frustration Experienced	10
6	Frustration Experienced	6
7	Frustration Experienced	8
8	Frustration Experienced	3
9	Frustration Experienced	3
10	Frustration Experienced	2
11	Frustration Experienced	8
12	Frustration Experienced	11
13	Frustration Experienced	2
14	Frustration Experienced	8
15	Frustration Experienced	12
16	Frustration Experienced	4
1	Annoyance Experienced	4
2	Annoyance Experienced	6
3	Annoyance Experienced	14
4	Annoyance Experienced	3
5	Annoyance Experienced	2
6	Annoyance Experienced	6
7	Annoyance Experienced	2
8	Annoyance Experienced	1
9	Annoyance Experienced	1
10	Annoyance Experienced	1

Participant	Attribute	Value
11	Annoyance Experienced	12
12	Annoyance Experienced	4
13	Annoyance Experienced	8
14	Annoyance Experienced	10
15	Annoyance Experienced	8
16	Annoyance Experienced	3

F.3 Three Earcon Condition

F.3.1 Earcons and Earcon Attributes Identified

Participant	Attribute	Value
1	Average Percentage Rides Identified	58.33
2	Average Percentage Rides Identified	25.00
3	Average Percentage Rides Identified	38.33
4	Average Percentage Rides Identified	50.00
5	Average Percentage Rides Identified	45.00
6	Average Percentage Rides Identified	33.33
7	Average Percentage Rides Identified	50.00
8	Average Percentage Rides Identified	28.33
9	Average Percentage Rides Identified	26.67
10	Average Percentage Rides Identified	30.00
11	Average Percentage Rides Identified	50.00
12	Average Percentage Rides Identified	60.00
13	Average Percentage Rides Identified	33.33
14	Average Percentage Rides Identified	26.67
15	Average Percentage Rides Identified	53.33
16	Average Percentage Rides Identified	58.33
1	Average Percentage Ride Types Identified	81.67
2	Average Percentage Ride Types Identified	75.00
3	Average Percentage Ride Types Identified	61.67
4	Average Percentage Ride Types Identified	86.67
5	Average Percentage Ride Types Identified	73.33
6	Average Percentage Ride Types Identified	65.00
7	Average Percentage Ride Types Identified	73.33
8	Average Percentage Ride Types Identified	76.67
9	Average Percentage Ride Types Identified	63.33
10	Average Percentage Ride Types Identified	75.00
11	Average Percentage Ride Types Identified	90.00
12	Average Percentage Ride Types Identified	75.00
13	Average Percentage Ride Types Identified	71.67
14	Average Percentage Ride Types Identified	70.00
15	Average Percentage Ride Types Identified	86.67
16	Average Percentage Ride Types Identified	81.67
1	Average Percentage Ride Intensities Identified	80.00
2	Average Percentage Ride Intensities Identified	73.33
3	Average Percentage Ride Intensities Identified	75.00
4	Average Percentage Ride Intensities Identified	86.67
5	Average Percentage Ride Intensities Identified	75.00
6	Average Percentage Ride Intensities Identified	75.00
7	Average Percentage Ride Intensities Identified	80.00
8	Average Percentage Ride Intensities Identified	75.00

Participant	Attribute	Value
9	Average Percentage Ride Intensities Identified	75.00
10	Average Percentage Ride Intensities Identified	61.67
11	Average Percentage Ride Intensities Identified	78.33
12	Average Percentage Ride Intensities Identified	86.67
13	Average Percentage Ride Intensities Identified	63.33
14	Average Percentage Ride Intensities Identified	66.67
15	Average Percentage Ride Intensities Identified	73.33
16	Average Percentage Ride Intensities Identified	80.00
1	Average Percentage Ride Costs Identified	75.00
2	Average Percentage Ride Costs Identified	45.00
3	Average Percentage Ride Costs Identified	66.67
4	Average Percentage Ride Costs Identified	60.00
5	Average Percentage Ride Costs Identified	75.00
6	Average Percentage Ride Costs Identified	65.00
7	Average Percentage Ride Costs Identified	71.67
8	Average Percentage Ride Costs Identified	50.00
9	Average Percentage Ride Costs Identified	60.00
10	Average Percentage Ride Costs Identified	50.00
11	Average Percentage Ride Costs Identified	70.00
12	Average Percentage Ride Costs Identified	83.33
13	Average Percentage Ride Costs Identified	61.67
14	Average Percentage Ride Costs Identified	61.67
15	Average Percentage Ride Costs Identified	78.33
16	Average Percentage Ride Costs Identified	75.00

F.3.2 Modified NASA TLX Workload Ratings

Participant	Attribute	Value
1	Mental Demand	18
2	Mental Demand	11
3	Mental Demand	17
4	Mental Demand	8
5	Mental Demand	18
6	Mental Demand	15
7	Mental Demand	16
8	Mental Demand	7
9	Mental Demand	16
10	Mental Demand	12
11	Mental Demand	17
12	Mental Demand	12
13	Mental Demand	17
14	Mental Demand	19
15	Mental Demand	18
16	Mental Demand	16
1	Physical Demand	1
2	Physical Demand	4
3	Physical Demand	5
4	Physical Demand	1
5	Physical Demand	1
6	Physical Demand	1
7	Physical Demand	3
8	Physical Demand	3
9	Physical Demand	6

Participant	Attribute	Value
10	Physical Demand	2
11	Physical Demand	1
12	Physical Demand	2
13	Physical Demand	1
14	Physical Demand	2
15	Physical Demand	1
16	Physical Demand	2
1	Time Pressure	16
2	Time Pressure	8
3	Time Pressure	17
4	Time Pressure	11
5	Time Pressure	17
6	Time Pressure	15
7	Time Pressure	11
8	Time Pressure	7
9	Time Pressure	4
10	Time Pressure	17
11	Time Pressure	4
12	Time Pressure	15
13	Time Pressure	14
14	Time Pressure	18
15	Time Pressure	13
16	Time Pressure	10
1	Effort Expended	14
2	Effort Expended	9
3	Effort Expended	17
4	Effort Expended	10
5	Effort Expended	17
6	Effort Expended	16
7	Effort Expended	15
8	Effort Expended	9
9	Effort Expended	14
10	Effort Expended	17
11	Effort Expended	14
12	Effort Expended	12
13	Effort Expended	17
14	Effort Expended	18
15	Effort Expended	18
16	Effort Expended	15
1	Performance Level Achieved	7
2	Performance Level Achieved	16
3	Performance Level Achieved	14
4	Performance Level Achieved	12
5	Performance Level Achieved	7
6	Performance Level Achieved	13
7	Performance Level Achieved	10
8	Performance Level Achieved	17
9	Performance Level Achieved	9
10	Performance Level Achieved	12
11	Performance Level Achieved	12
12	Performance Level Achieved	11
13	Performance Level Achieved	3
14	Performance Level Achieved	4
15	Performance Level Achieved	15
16	Performance Level Achieved	14

Participant	Attribute	Value
1	Frustration Experienced	15
2	Frustration Experienced	9
3	Frustration Experienced	18
4	Frustration Experienced	9
5	Frustration Experienced	15
6	Frustration Experienced	9
7	Frustration Experienced	8
8	Frustration Experienced	7
9	Frustration Experienced	15
10	Frustration Experienced	7
11	Frustration Experienced	1
12	Frustration Experienced	15
13	Frustration Experienced	17
14	Frustration Experienced	14
15	Frustration Experienced	11
16	Frustration Experienced	12
1	Annoyance Experienced	14
2	Annoyance Experienced	13
3	Annoyance Experienced	12
4	Annoyance Experienced	8
5	Annoyance Experienced	9
6	Annoyance Experienced	10
7	Annoyance Experienced	14
8	Annoyance Experienced	3
9	Annoyance Experienced	13
10	Annoyance Experienced	1
11	Annoyance Experienced	1
12	Annoyance Experienced	15
13	Annoyance Experienced	3
14	Annoyance Experienced	11
15	Annoyance Experienced	6
16	Annoyance Experienced	3

F.4 Four Earcon Condition

F.4.1 Earcons and Earcon Attributes Identified

Participant	Attribute	Value
1	Average Percentage Rides Identified	17.50
2	Average Percentage Rides Identified	38.75
3	Average Percentage Rides Identified	23.75
4	Average Percentage Rides Identified	45.00
5	Average Percentage Rides Identified	18.75
6	Average Percentage Rides Identified	38.75
7	Average Percentage Rides Identified	25.00
8	Average Percentage Rides Identified	41.25
9	Average Percentage Rides Identified	22.50
10	Average Percentage Rides Identified	20.00
11	Average Percentage Rides Identified	23.75
12	Average Percentage Rides Identified	38.75
13	Average Percentage Rides Identified	23.75
14	Average Percentage Rides Identified	45.00

Participant	Attribute	Value
15	Average Percentage Rides Identified	18.75
16	Average Percentage Rides Identified	41.25
1	Average Percentage Ride Types Identified	63.75
2	Average Percentage Ride Types Identified	66.25
3	Average Percentage Ride Types Identified	56.25
4	Average Percentage Ride Types Identified	70.00
5	Average Percentage Ride Types Identified	65.00
6	Average Percentage Ride Types Identified	66.25
7	Average Percentage Ride Types Identified	58.75
8	Average Percentage Ride Types Identified	78.75
9	Average Percentage Ride Types Identified	66.25
10	Average Percentage Ride Types Identified	61.25
11	Average Percentage Ride Types Identified	72.50
12	Average Percentage Ride Types Identified	67.50
13	Average Percentage Ride Types Identified	56.25
14	Average Percentage Ride Types Identified	70.00
15	Average Percentage Ride Types Identified	65.00
16	Average Percentage Ride Types Identified	78.75
1	Average Percentage Ride Intensities Identified	55.00
2	Average Percentage Ride Intensities Identified	70.00
3	Average Percentage Ride Intensities Identified	72.50
4	Average Percentage Ride Intensities Identified	75.00
5	Average Percentage Ride Intensities Identified	55.00
6	Average Percentage Ride Intensities Identified	70.00
7	Average Percentage Ride Intensities Identified	75.00
8	Average Percentage Ride Intensities Identified	60.00
9	Average Percentage Ride Intensities Identified	60.00
10	Average Percentage Ride Intensities Identified	61.25
11	Average Percentage Ride Intensities Identified	57.50
12	Average Percentage Ride Intensities Identified	70.00
13	Average Percentage Ride Intensities Identified	72.50
14	Average Percentage Ride Intensities Identified	75.00
15	Average Percentage Ride Intensities Identified	55.00
16	Average Percentage Ride Intensities Identified	60.00
1	Average Percentage Ride Costs Identified	60.00
2	Average Percentage Ride Costs Identified	62.50
3	Average Percentage Ride Costs Identified	53.75
4	Average Percentage Ride Costs Identified	68.75
5	Average Percentage Ride Costs Identified	62.50
6	Average Percentage Ride Costs Identified	62.50
7	Average Percentage Ride Costs Identified	70.00
8	Average Percentage Ride Costs Identified	68.75
9	Average Percentage Ride Costs Identified	57.50
10	Average Percentage Ride Costs Identified	55.00
11	Average Percentage Ride Costs Identified	48.75
12	Average Percentage Ride Costs Identified	73.75
13	Average Percentage Ride Costs Identified	53.75
14	Average Percentage Ride Costs Identified	68.75
15	Average Percentage Ride Costs Identified	62.50
16	Average Percentage Ride Costs Identified	68.75

F.4.2 Modified NASA TLX Workload Ratings

Participant	Attribute	Value
1	Mental Demand	20
2	Mental Demand	16
3	Mental Demand	13
4	Mental Demand	20
5	Mental Demand	3
6	Mental Demand	14
7	Mental Demand	12
8	Mental Demand	15
9	Mental Demand	19
10	Mental Demand	20
11	Mental Demand	15
12	Mental Demand	18
13	Mental Demand	16
14	Mental Demand	14
15	Mental Demand	3
16	Mental Demand	20
1	Physical Demand	1
2	Physical Demand	11
3	Physical Demand	6
4	Physical Demand	4
5	Physical Demand	1
6	Physical Demand	1
7	Physical Demand	4
8	Physical Demand	1
9	Physical Demand	1
10	Physical Demand	4
11	Physical Demand	5
12	Physical Demand	2
13	Physical Demand	11
14	Physical Demand	1
15	Physical Demand	1
16	Physical Demand	1
1	Time Pressure	13
2	Time Pressure	18
3	Time Pressure	13
4	Time Pressure	4
5	Time Pressure	13
6	Time Pressure	19
7	Time Pressure	9
8	Time Pressure	15
9	Time Pressure	15
10	Time Pressure	20
11	Time Pressure	14
12	Time Pressure	18
13	Time Pressure	18
14	Time Pressure	19
15	Time Pressure	13
16	Time Pressure	13
1	Effort Expended	20
2	Effort Expended	15
3	Effort Expended	11
4	Effort Expended	16
5	Effort Expended	17

Participant	Attribute	Value
6	Effort Expended	14
7	Effort Expended	17
8	Effort Expended	15
9	Effort Expended	17
10	Effort Expended	20
11	Effort Expended	13
12	Effort Expended	18
13	Effort Expended	15
14	Effort Expended	14
15	Effort Expended	17
16	Effort Expended	20
1	Performance Level Achieved	7
2	Performance Level Achieved	11
3	Performance Level Achieved	12
4	Performance Level Achieved	13
5	Performance Level Achieved	10
6	Performance Level Achieved	6
7	Performance Level Achieved	7
8	Performance Level Achieved	8
9	Performance Level Achieved	3
10	Performance Level Achieved	1
11	Performance Level Achieved	11
12	Performance Level Achieved	3
13	Performance Level Achieved	11
14	Performance Level Achieved	6
15	Performance Level Achieved	10
16	Performance Level Achieved	7
1	Frustration Experienced	11
2	Frustration Experienced	18
3	Frustration Experienced	11
4	Frustration Experienced	10
5	Frustration Experienced	15
6	Frustration Experienced	18
7	Frustration Experienced	18
8	Frustration Experienced	13
9	Frustration Experienced	3
10	Frustration Experienced	18
11	Frustration Experienced	7
12	Frustration Experienced	18
13	Frustration Experienced	18
14	Frustration Experienced	18
15	Frustration Experienced	15
16	Frustration Experienced	11
1	Annoyance Experienced	15
2	Annoyance Experienced	16
3	Annoyance Experienced	15
4	Annoyance Experienced	17
5	Annoyance Experienced	15
6	Annoyance Experienced	11
7	Annoyance Experienced	15
8	Annoyance Experienced	5
9	Annoyance Experienced	1
10	Annoyance Experienced	18
11	Annoyance Experienced	13
12	Annoyance Experienced	14

Participant	Attribute	Value
13	Annoyance Experienced	16
14	Annoyance Experienced	11
15	Annoyance Experienced	15
16	Annoyance Experienced	15

Appendix G

Engineering More Robust Earcons - Training Sheets

This appendix contains copies of the training sheets from the experiment described in Chapter 6.

Appendix H

Engineering More Robust Earcons - Raw Data

This appendix presents the raw results from the experiment described in Chapter 6.

H.1 Original Earcon Set Condition

For raw data for the number of earcons and ride attributes identified see section F.1.1. For raw results for the modified NASA TLX workload ratings, see section F.1.2.

H.2 Melody Altered Earcon Set Condition

H.2.1 Earcons and Earcon Attributes Identified

Participant	Attribute	Value
1	Average Number Rides Identified	0.75
2	Average Number Rides Identified	1.60
3	Average Number Rides Identified	1.35
4	Average Number Rides Identified	1.10
5	Average Number Rides Identified	0.05
6	Average Number Rides Identified	1.10
7	Average Number Rides Identified	1.40
8	Average Number Rides Identified	1.30
9	Average Number Rides Identified	1.45
10	Average Number Rides Identified	0.55
11	Average Number Rides Identified	0.65
12	Average Number Rides Identified	1.65
13	Average Number Rides Identified	1.35
14	Average Number Rides Identified	1.10
15	Average Number Rides Identified	0.90
16	Average Number Rides Identified	1.30
1	Average Number Ride Types Identified	2.30
2	Average Number Ride Types Identified	3.10
3	Average Number Ride Types Identified	2.90
4	Average Number Ride Types Identified	2.35
5	Average Number Ride Types Identified	1.20
6	Average Number Ride Types Identified	2.55
7	Average Number Ride Types Identified	2.70
8	Average Number Ride Types Identified	2.65
9	Average Number Ride Types Identified	2.65
10	Average Number Ride Types Identified	2.60
11	Average Number Ride Types Identified	2.55
12	Average Number Ride Types Identified	2.70
13	Average Number Ride Types Identified	2.90
14	Average Number Ride Types Identified	2.35

Participant	Attribute	Value
15	Average Number Ride Types Identified	2.75
16	Average Number Ride Types Identified	2.65
1	Average Number Ride Intensities Identified	1.90
2	Average Number Ride Intensities Identified	2.90
3	Average Number Ride Intensities Identified	2.80
4	Average Number Ride Intensities Identified	2.65
5	Average Number Ride Intensities Identified	1.00
6	Average Number Ride Intensities Identified	2.10
7	Average Number Ride Intensities Identified	2.75
8	Average Number Ride Intensities Identified	2.35
9	Average Number Ride Intensities Identified	2.95
10	Average Number Ride Intensities Identified	2.15
11	Average Number Ride Intensities Identified	2.10
12	Average Number Ride Intensities Identified	2.75
13	Average Number Ride Intensities Identified	2.80
14	Average Number Ride Intensities Identified	2.65
15	Average Number Ride Intensities Identified	2.15
16	Average Number Ride Intensities Identified	2.35
1	Average Number Ride Costs Identified	2.40
2	Average Number Ride Costs Identified	2.60
3	Average Number Ride Costs Identified	2.55
4	Average Number Ride Costs Identified	2.60
5	Average Number Ride Costs Identified	1.00
6	Average Number Ride Costs Identified	2.65
7	Average Number Ride Costs Identified	2.85
8	Average Number Ride Costs Identified	2.75
9	Average Number Ride Costs Identified	2.40
10	Average Number Ride Costs Identified	2.45
11	Average Number Ride Costs Identified	2.40
12	Average Number Ride Costs Identified	2.65
13	Average Number Ride Costs Identified	2.55
14	Average Number Ride Costs Identified	2.60
15	Average Number Ride Costs Identified	2.35
16	Average Number Ride Costs Identified	2.75

H.2.2 Modified NASA TLX Workload Ratings

Participant	Attribute	Value
1	Mental Demand	20
2	Mental Demand	18
3	Mental Demand	15
4	Mental Demand	20
5	Mental Demand	7
6	Mental Demand	15
7	Mental Demand	10
8	Mental Demand	17
9	Mental Demand	19
10	Mental Demand	20
11	Mental Demand	17
12	Mental Demand	19
13	Mental Demand	17
14	Mental Demand	15
15	Mental Demand	7

Participant	Attribute	Value
16	Mental Demand	20
1	Physical Demand	15
2	Physical Demand	17
3	Physical Demand	16
4	Physical Demand	20
5	Physical Demand	20
6	Physical Demand	18
7	Physical Demand	20
8	Physical Demand	15
9	Physical Demand	18
10	Physical Demand	20
11	Physical Demand	17
12	Physical Demand	19
13	Physical Demand	7
14	Physical Demand	15
15	Physical Demand	17
16	Physical Demand	16
1	Time Pressure	13
2	Time Pressure	18
3	Time Pressure	15
4	Time Pressure	4
5	Time Pressure	13
6	Time Pressure	20
7	Time Pressure	8
8	Time Pressure	15
9	Time Pressure	15
10	Time Pressure	16
11	Time Pressure	13
12	Time Pressure	18
13	Time Pressure	13
14	Time Pressure	20
15	Time Pressure	13
16	Time Pressure	13
1	Effort Expended	17
2	Effort Expended	17
3	Effort Expended	15
4	Effort Expended	18
5	Effort Expended	19
6	Effort Expended	15
7	Effort Expended	14
8	Effort Expended	17
9	Effort Expended	17
10	Effort Expended	20
11	Effort Expended	17
12	Effort Expended	18
13	Effort Expended	17
14	Effort Expended	15
15	Effort Expended	19
16	Effort Expended	17
1	Performance Level Achieved	3
2	Performance Level Achieved	8
3	Performance Level Achieved	11
4	Performance Level Achieved	3
5	Performance Level Achieved	9
6	Performance Level Achieved	7

Participant	Attribute	Value
7	Performance Level Achieved	7
8	Performance Level Achieved	5
9	Performance Level Achieved	8
10	Performance Level Achieved	1
11	Performance Level Achieved	13
12	Performance Level Achieved	2
13	Performance Level Achieved	13
14	Performance Level Achieved	7
15	Performance Level Achieved	9
16	Performance Level Achieved	3
1	Frustration Experienced	13
2	Frustration Experienced	18
3	Frustration Experienced	11
4	Frustration Experienced	10
5	Frustration Experienced	17
6	Frustration Experienced	18
7	Frustration Experienced	14
8	Frustration Experienced	15
9	Frustration Experienced	3
10	Frustration Experienced	16
11	Frustration Experienced	13
12	Frustration Experienced	18
13	Frustration Experienced	13
14	Frustration Experienced	18
15	Frustration Experienced	17
16	Frustration Experienced	13
1	Annoyance Experienced	13
2	Annoyance Experienced	16
3	Annoyance Experienced	14
4	Annoyance Experienced	7
5	Annoyance Experienced	17
6	Annoyance Experienced	11
7	Annoyance Experienced	13
8	Annoyance Experienced	7
9	Annoyance Experienced	1
10	Annoyance Experienced	10
11	Annoyance Experienced	17
12	Annoyance Experienced	15
13	Annoyance Experienced	17
14	Annoyance Experienced	11
15	Annoyance Experienced	17
16	Annoyance Experienced	13

H.3 Multi-Timbre Earcon Set Condition

H.3.1 Earcons and Earcon Attributes Identified

Participant	Attribute	Value
1	Average Number Rides Identified	2.00
2	Average Number Rides Identified	0.60
3	Average Number Rides Identified	1.60
4	Average Number Rides Identified	1.50

Participant	Attribute	Value
5	Average Number Rides Identified	2.00
6	Average Number Rides Identified	1.00
7	Average Number Rides Identified	1.60
8	Average Number Rides Identified	1.25
9	Average Number Rides Identified	1.40
10	Average Number Rides Identified	0.80
11	Average Number Rides Identified	2.15
12	Average Number Rides Identified	2.35
13	Average Number Rides Identified	1.00
14	Average Number Rides Identified	1.30
15	Average Number Rides Identified	1.75
16	Average Number Rides Identified	2.00
1	Average Number Ride Types Identified	3.15
2	Average Number Ride Types Identified	2.45
3	Average Number Ride Types Identified	3.10
4	Average Number Ride Types Identified	2.90
5	Average Number Ride Types Identified	3.40
6	Average Number Ride Types Identified	2.60
7	Average Number Ride Types Identified	2.80
8	Average Number Ride Types Identified	2.95
9	Average Number Ride Types Identified	3.55
10	Average Number Ride Types Identified	2.80
11	Average Number Ride Types Identified	3.40
12	Average Number Ride Types Identified	3.10
13	Average Number Ride Types Identified	2.75
14	Average Number Ride Types Identified	2.80
15	Average Number Ride Types Identified	3.35
16	Average Number Ride Types Identified	3.40
1	Average Number Ride Intensities Identified	3.35
2	Average Number Ride Intensities Identified	1.75
3	Average Number Ride Intensities Identified	3.00
4	Average Number Ride Intensities Identified	3.55
5	Average Number Ride Intensities Identified	2.90
6	Average Number Ride Intensities Identified	2.95
7	Average Number Ride Intensities Identified	2.85
8	Average Number Ride Intensities Identified	2.75
9	Average Number Ride Intensities Identified	2.30
10	Average Number Ride Intensities Identified	3.25
11	Average Number Ride Intensities Identified	3.30
12	Average Number Ride Intensities Identified	1.90
13	Average Number Ride Intensities Identified	2.50
14	Average Number Ride Intensities Identified	2.90
15	Average Number Ride Intensities Identified	3.55
16	Average Number Ride Intensities Identified	3.10
1	Average Number Ride Costs Identified	2.65
2	Average Number Ride Costs Identified	2.70
3	Average Number Ride Costs Identified	2.25
4	Average Number Ride Costs Identified	2.35
5	Average Number Ride Costs Identified	2.25
6	Average Number Ride Costs Identified	2.50
7	Average Number Ride Costs Identified	2.70
8	Average Number Ride Costs Identified	2.30
9	Average Number Ride Costs Identified	2.30
10	Average Number Ride Costs Identified	2.25
11	Average Number Ride Costs Identified	2.45

Participant	Attribute	Value
12	Average Number Ride Costs Identified	3.20
13	Average Number Ride Costs Identified	2.30
14	Average Number Ride Costs Identified	2.50
15	Average Number Ride Costs Identified	2.50
16	Average Number Ride Costs Identified	2.25

H.3.2 Modified NASA TLX Workload Ratings

Participant	Attribute	Value
1	Mental Demand	19
2	Mental Demand	17
3	Mental Demand	19
4	Mental Demand	19
5	Mental Demand	18
6	Mental Demand	19
7	Mental Demand	16
8	Mental Demand	13
9	Mental Demand	16
10	Mental Demand	20
11	Mental Demand	19
12	Mental Demand	13
13	Mental Demand	18
14	Mental Demand	19
15	Mental Demand	19
16	Mental Demand	16
1	Physical Demand	1
2	Physical Demand	6
3	Physical Demand	15
4	Physical Demand	1
5	Physical Demand	1
6	Physical Demand	1
7	Physical Demand	3
8	Physical Demand	1
9	Physical Demand	6
10	Physical Demand	2
11	Physical Demand	1
12	Physical Demand	8
13	Physical Demand	3
14	Physical Demand	2
15	Physical Demand	3
16	Physical Demand	1
1	Time Pressure	18
2	Time Pressure	16
3	Time Pressure	17
4	Time Pressure	20
5	Time Pressure	18
6	Time Pressure	17
7	Time Pressure	15
8	Time Pressure	9
9	Time Pressure	6
10	Time Pressure	19
11	Time Pressure	3
12	Time Pressure	20

Participant	Attribute	Value
13	Time Pressure	12
14	Time Pressure	19
15	Time Pressure	15
16	Time Pressure	12
1	Effort Expended	16
2	Effort Expended	13
3	Effort Expended	17
4	Effort Expended	17
5	Effort Expended	18
6	Effort Expended	18
7	Effort Expended	15
8	Effort Expended	15
9	Effort Expended	15
10	Effort Expended	20
11	Effort Expended	17
12	Effort Expended	13
13	Effort Expended	15
14	Effort Expended	18
15	Effort Expended	19
16	Effort Expended	17
1	Performance Level Achieved	9
2	Performance Level Achieved	9
3	Performance Level Achieved	5
4	Performance Level Achieved	8
5	Performance Level Achieved	7
6	Performance Level Achieved	10
7	Performance Level Achieved	8
8	Performance Level Achieved	11
9	Performance Level Achieved	11
10	Performance Level Achieved	8
11	Performance Level Achieved	14
12	Performance Level Achieved	8
13	Performance Level Achieved	3
14	Performance Level Achieved	3
15	Performance Level Achieved	12
16	Performance Level Achieved	2
1	Frustration Experienced	13
2	Frustration Experienced	5
3	Frustration Experienced	18
4	Frustration Experienced	18
5	Frustration Experienced	13
6	Frustration Experienced	15
7	Frustration Experienced	7
8	Frustration Experienced	9
9	Frustration Experienced	15
10	Frustration Experienced	10
11	Frustration Experienced	1
12	Frustration Experienced	11
13	Frustration Experienced	17
14	Frustration Experienced	4
15	Frustration Experienced	16
16	Frustration Experienced	14
1	Annoyance Experienced	12
2	Annoyance Experienced	6
3	Annoyance Experienced	18

Participant	Attribute	Value
4	Annoyance Experienced	18
5	Annoyance Experienced	6
6	Annoyance Experienced	14
7	Annoyance Experienced	14
8	Annoyance Experienced	3
9	Annoyance Experienced	9
10	Annoyance Experienced	1
11	Annoyance Experienced	1
12	Annoyance Experienced	15
13	Annoyance Experienced	5
14	Annoyance Experienced	11
15	Annoyance Experienced	15
16	Annoyance Experienced	15

H.4 Extended Training Condition

H.4.1 Earcons and Earcon Attributes Identified

Participant	Attribute	Value
1	Average Number Rides Identified	1.60
2	Average Number Rides Identified	1.45
3	Average Number Rides Identified	1.75
4	Average Number Rides Identified	1.75
5	Average Number Rides Identified	1.10
6	Average Number Rides Identified	1.40
7	Average Number Rides Identified	1.20
8	Average Number Rides Identified	1.45
9	Average Number Rides Identified	1.50
10	Average Number Rides Identified	1.45
11	Average Number Rides Identified	1.35
12	Average Number Rides Identified	1.60
13	Average Number Rides Identified	1.15
14	Average Number Rides Identified	1.15
15	Average Number Rides Identified	1.40
16	Average Number Rides Identified	1.15
1	Average Number Ride Types Identified	2.95
2	Average Number Ride Types Identified	3.30
3	Average Number Ride Types Identified	3.10
4	Average Number Ride Types Identified	2.90
5	Average Number Ride Types Identified	2.60
6	Average Number Ride Types Identified	2.95
7	Average Number Ride Types Identified	2.90
8	Average Number Ride Types Identified	3.00
9	Average Number Ride Types Identified	3.10
10	Average Number Ride Types Identified	3.00
11	Average Number Ride Types Identified	3.10
12	Average Number Ride Types Identified	3.05
13	Average Number Ride Types Identified	2.55
14	Average Number Ride Types Identified	2.95
15	Average Number Ride Types Identified	2.70
16	Average Number Ride Types Identified	2.80
1	Average Number Ride Intensities Identified	2.55

Participant	Attribute	Value
2	Average Number Ride Intensities Identified	2.70
3	Average Number Ride Intensities Identified	2.80
4	Average Number Ride Intensities Identified	2.95
5	Average Number Ride Intensities Identified	2.75
6	Average Number Ride Intensities Identified	3.05
7	Average Number Ride Intensities Identified	2.45
8	Average Number Ride Intensities Identified	2.85
9	Average Number Ride Intensities Identified	2.60
10	Average Number Ride Intensities Identified	3.25
11	Average Number Ride Intensities Identified	2.35
12	Average Number Ride Intensities Identified	2.90
13	Average Number Ride Intensities Identified	2.55
14	Average Number Ride Intensities Identified	2.05
15	Average Number Ride Intensities Identified	2.65
16	Average Number Ride Intensities Identified	2.55
1	Average Number Ride Costs Identified	2.45
2	Average Number Ride Costs Identified	2.10
3	Average Number Ride Costs Identified	2.65
4	Average Number Ride Costs Identified	2.90
5	Average Number Ride Costs Identified	2.40
6	Average Number Ride Costs Identified	2.50
7	Average Number Ride Costs Identified	2.30
8	Average Number Ride Costs Identified	2.45
9	Average Number Ride Costs Identified	2.60
10	Average Number Ride Costs Identified	2.15
11	Average Number Ride Costs Identified	2.45
12	Average Number Ride Costs Identified	2.60
13	Average Number Ride Costs Identified	2.75
14	Average Number Ride Costs Identified	2.85
15	Average Number Ride Costs Identified	2.60
16	Average Number Ride Costs Identified	2.30

H.4.2 Modified NASA TLX Workload Ratings

Participant	Attribute	Value
1	Mental Demand	18
2	Mental Demand	19
3	Mental Demand	15
4	Mental Demand	18
5	Mental Demand	20
6	Mental Demand	20
7	Mental Demand	18
8	Mental Demand	18
9	Mental Demand	20
10	Mental Demand	18
11	Mental Demand	20
12	Mental Demand	17
13	Mental Demand	17
14	Mental Demand	18
15	Mental Demand	19
16	Mental Demand	20
1	Physical Demand	1
2	Physical Demand	5

Participant	Attribute	Value
3	Physical Demand	7
4	Physical Demand	11
5	Physical Demand	1
6	Physical Demand	8
7	Physical Demand	4
8	Physical Demand	1
9	Physical Demand	1
10	Physical Demand	3
11	Physical Demand	2
12	Physical Demand	1
13	Physical Demand	1
14	Physical Demand	8
15	Physical Demand	2
16	Physical Demand	1
1	Time Pressure	18
2	Time Pressure	17
3	Time Pressure	19
4	Time Pressure	5
5	Time Pressure	14
6	Time Pressure	16
7	Time Pressure	19
8	Time Pressure	20
9	Time Pressure	17
10	Time Pressure	13
11	Time Pressure	12
12	Time Pressure	13
13	Time Pressure	13
14	Time Pressure	13
15	Time Pressure	14
16	Time Pressure	6
1	Effort Expended	16
2	Effort Expended	16
3	Effort Expended	13
4	Effort Expended	18
5	Effort Expended	20
6	Effort Expended	16
7	Effort Expended	18
8	Effort Expended	15
9	Effort Expended	11
10	Effort Expended	15
11	Effort Expended	19
12	Effort Expended	15
13	Effort Expended	15
14	Effort Expended	17
15	Effort Expended	14
16	Effort Expended	20
1	Performance Level Achieved	7
2	Performance Level Achieved	9
3	Performance Level Achieved	11
4	Performance Level Achieved	11
5	Performance Level Achieved	4
6	Performance Level Achieved	9
7	Performance Level Achieved	2
8	Performance Level Achieved	5
9	Performance Level Achieved	9

Participant	Attribute	Value
10	Performance Level Achieved	9
11	Performance Level Achieved	8
12	Performance Level Achieved	9
13	Performance Level Achieved	11
14	Performance Level Achieved	6
15	Performance Level Achieved	12
16	Performance Level Achieved	2
1	Frustration Experienced	14
2	Frustration Experienced	13
3	Frustration Experienced	18
4	Frustration Experienced	14
5	Frustration Experienced	11
6	Frustration Experienced	13
7	Frustration Experienced	17
8	Frustration Experienced	16
9	Frustration Experienced	12
10	Frustration Experienced	15
11	Frustration Experienced	7
12	Frustration Experienced	9
13	Frustration Experienced	11
14	Frustration Experienced	8
15	Frustration Experienced	15
16	Frustration Experienced	11
1	Annoyance Experienced	11
2	Annoyance Experienced	12
3	Annoyance Experienced	20
4	Annoyance Experienced	7
5	Annoyance Experienced	1
6	Annoyance Experienced	7
7	Annoyance Experienced	15
8	Annoyance Experienced	5
9	Annoyance Experienced	5
10	Annoyance Experienced	15
11	Annoyance Experienced	6
12	Annoyance Experienced	6
13	Annoyance Experienced	3
14	Annoyance Experienced	8
15	Annoyance Experienced	14
16	Annoyance Experienced	11

H.5 Staggered Onset Condition

H.5.1 Earcons and Earcon Attributes Identified

Participant	Attribute	Value
1	Average Number Rides Identified	1.05
2	Average Number Rides Identified	1.85
3	Average Number Rides Identified	2.05
4	Average Number Rides Identified	1.30
5	Average Number Rides Identified	1.15
6	Average Number Rides Identified	1.95
7	Average Number Rides Identified	1.80

Participant	Attribute	Value
8	Average Number Rides Identified	1.85
9	Average Number Rides Identified	1.85
10	Average Number Rides Identified	1.60
11	Average Number Rides Identified	1.85
12	Average Number Rides Identified	1.95
13	Average Number Rides Identified	1.65
14	Average Number Rides Identified	1.45
15	Average Number Rides Identified	2.25
16	Average Number Rides Identified	1.70
1	Average Number Ride Types Identified	2.95
2	Average Number Ride Types Identified	3.25
3	Average Number Ride Types Identified	3.25
4	Average Number Ride Types Identified	3.15
5	Average Number Ride Types Identified	2.80
6	Average Number Ride Types Identified	3.65
7	Average Number Ride Types Identified	2.35
8	Average Number Ride Types Identified	3.30
9	Average Number Ride Types Identified	3.20
10	Average Number Ride Types Identified	3.15
11	Average Number Ride Types Identified	2.65
12	Average Number Ride Types Identified	2.65
13	Average Number Ride Types Identified	2.60
14	Average Number Ride Types Identified	2.65
15	Average Number Ride Types Identified	3.25
16	Average Number Ride Types Identified	3.10
1	Average Number Ride Intensities Identified	2.30
2	Average Number Ride Intensities Identified	2.85
3	Average Number Ride Intensities Identified	3.40
4	Average Number Ride Intensities Identified	2.85
5	Average Number Ride Intensities Identified	2.65
6	Average Number Ride Intensities Identified	3.25
7	Average Number Ride Intensities Identified	2.65
8	Average Number Ride Intensities Identified	3.20
9	Average Number Ride Intensities Identified	3.30
10	Average Number Ride Intensities Identified	2.80
11	Average Number Ride Intensities Identified	2.25
12	Average Number Ride Intensities Identified	2.65
13	Average Number Ride Intensities Identified	3.00
14	Average Number Ride Intensities Identified	2.85
15	Average Number Ride Intensities Identified	3.10
16	Average Number Ride Intensities Identified	2.60
1	Average Number Ride Costs Identified	2.55
2	Average Number Ride Costs Identified	2.95
3	Average Number Ride Costs Identified	2.75
4	Average Number Ride Costs Identified	2.20
5	Average Number Ride Costs Identified	2.25
6	Average Number Ride Costs Identified	2.50
7	Average Number Ride Costs Identified	2.35
8	Average Number Ride Costs Identified	2.55
9	Average Number Ride Costs Identified	2.50
10	Average Number Ride Costs Identified	2.70
11	Average Number Ride Costs Identified	2.30
12	Average Number Ride Costs Identified	2.25
13	Average Number Ride Costs Identified	3.00
14	Average Number Ride Costs Identified	2.45

Participant	Attribute	Value
15	Average Number Ride Costs Identified	3.20
16	Average Number Ride Costs Identified	2.80

H.5.2 Modified NASA TLX Workload Ratings

Participant	Attribute	Value
1	Mental Demand	16
2	Mental Demand	12
3	Mental Demand	20
4	Mental Demand	19
5	Mental Demand	16
6	Mental Demand	17
7	Mental Demand	20
8	Mental Demand	17
9	Mental Demand	14
10	Mental Demand	15
11	Mental Demand	14
12	Mental Demand	20
13	Mental Demand	20
14	Mental Demand	17
15	Mental Demand	20
16	Mental Demand	16
1	Physical Demand	12
2	Physical Demand	16
3	Physical Demand	18
4	Physical Demand	2
5	Physical Demand	10
6	Physical Demand	10
7	Physical Demand	10
8	Physical Demand	6
9	Physical Demand	5
10	Physical Demand	10
11	Physical Demand	8
12	Physical Demand	5
13	Physical Demand	20
14	Physical Demand	1
15	Physical Demand	6
16	Physical Demand	4
1	Time Pressure	17
2	Time Pressure	10
3	Time Pressure	19
4	Time Pressure	2
5	Time Pressure	17
6	Time Pressure	6
7	Time Pressure	18
8	Time Pressure	12
9	Time Pressure	15
10	Time Pressure	12
11	Time Pressure	20
12	Time Pressure	20
13	Time Pressure	18
14	Time Pressure	7
15	Time Pressure	19

Participant	Attribute	Value
16	Time Pressure	19
1	Effort Expended	17
2	Effort Expended	14
3	Effort Expended	20
4	Effort Expended	19
5	Effort Expended	18
6	Effort Expended	17
7	Effort Expended	18
8	Effort Expended	16
9	Effort Expended	12
10	Effort Expended	13
11	Effort Expended	13
12	Effort Expended	17
13	Effort Expended	18
14	Effort Expended	16
15	Effort Expended	19
16	Effort Expended	17
1	Performance Level Achieved	10
2	Performance Level Achieved	6
3	Performance Level Achieved	9
4	Performance Level Achieved	10
5	Performance Level Achieved	6
6	Performance Level Achieved	10
7	Performance Level Achieved	4
8	Performance Level Achieved	6
9	Performance Level Achieved	10
10	Performance Level Achieved	8
11	Performance Level Achieved	7
12	Performance Level Achieved	4
13	Performance Level Achieved	5
14	Performance Level Achieved	3
15	Performance Level Achieved	8
16	Performance Level Achieved	10
1	Frustration Experienced	8
2	Frustration Experienced	12
3	Frustration Experienced	18
4	Frustration Experienced	11
5	Frustration Experienced	8
6	Frustration Experienced	11
7	Frustration Experienced	18
8	Frustration Experienced	10
9	Frustration Experienced	2
10	Frustration Experienced	8
11	Frustration Experienced	14
12	Frustration Experienced	14
13	Frustration Experienced	5
14	Frustration Experienced	6
15	Frustration Experienced	14
16	Frustration Experienced	13
1	Annoyance Experienced	4
2	Annoyance Experienced	4
3	Annoyance Experienced	17
4	Annoyance Experienced	4
5	Annoyance Experienced	7
6	Annoyance Experienced	9

Participant	Attribute	Value
7	Annoyance Experienced	14
8	Annoyance Experienced	4
9	Annoyance Experienced	1
10	Annoyance Experienced	1
11	Annoyance Experienced	8
12	Annoyance Experienced	4
13	Annoyance Experienced	16
14	Annoyance Experienced	10
15	Annoyance Experienced	13
16	Annoyance Experienced	14

H.6 Final Condition

H.6.1 Earcons and Earcon Attributes Identified

Participant	Attribute	Value
1	Average Number Rides Identified	1.55
2	Average Number Rides Identified	1.65
3	Average Number Rides Identified	1.50
4	Average Number Rides Identified	1.70
5	Average Number Rides Identified	1.60
6	Average Number Rides Identified	1.95
7	Average Number Rides Identified	1.55
8	Average Number Rides Identified	1.50
9	Average Number Rides Identified	1.45
10	Average Number Rides Identified	1.60
11	Average Number Rides Identified	1.40
12	Average Number Rides Identified	2.00
13	Average Number Rides Identified	1.89
14	Average Number Rides Identified	1.70
15	Average Number Rides Identified	1.82
16	Average Number Rides Identified	1.70
1	Average Number Ride Types Identified	3.25
2	Average Number Ride Types Identified	2.70
3	Average Number Ride Types Identified	2.60
4	Average Number Ride Types Identified	3.40
5	Average Number Ride Types Identified	3.15
6	Average Number Ride Types Identified	3.10
7	Average Number Ride Types Identified	3.05
8	Average Number Ride Types Identified	3.00
9	Average Number Ride Types Identified	3.00
10	Average Number Ride Types Identified	3.40
11	Average Number Ride Types Identified	2.85
12	Average Number Ride Types Identified	3.10
13	Average Number Ride Types Identified	3.10
14	Average Number Ride Types Identified	2.85
15	Average Number Ride Types Identified	3.20
16	Average Number Ride Types Identified	3.50
1	Average Number Ride Intensities Identified	2.70
2	Average Number Ride Intensities Identified	2.35
3	Average Number Ride Intensities Identified	2.25
4	Average Number Ride Intensities Identified	2.85

Participant	Attribute	Value
5	Average Number Ride Intensities Identified	2.55
6	Average Number Ride Intensities Identified	3.20
7	Average Number Ride Intensities Identified	2.75
8	Average Number Ride Intensities Identified	2.55
9	Average Number Ride Intensities Identified	2.70
10	Average Number Ride Intensities Identified	2.20
11	Average Number Ride Intensities Identified	2.60
12	Average Number Ride Intensities Identified	3.20
13	Average Number Ride Intensities Identified	3.20
14	Average Number Ride Intensities Identified	2.60
15	Average Number Ride Intensities Identified	2.98
16	Average Number Ride Intensities Identified	2.93
1	Average Number Ride Costs Identified	2.25
2	Average Number Ride Costs Identified	2.35
3	Average Number Ride Costs Identified	2.15
4	Average Number Ride Costs Identified	2.60
5	Average Number Ride Costs Identified	2.25
6	Average Number Ride Costs Identified	3.00
7	Average Number Ride Costs Identified	2.65
8	Average Number Ride Costs Identified	1.90
9	Average Number Ride Costs Identified	2.65
10	Average Number Ride Costs Identified	2.00
11	Average Number Ride Costs Identified	2.30
12	Average Number Ride Costs Identified	3.00
13	Average Number Ride Costs Identified	3.00
14	Average Number Ride Costs Identified	2.30
15	Average Number Ride Costs Identified	2.70
16	Average Number Ride Costs Identified	2.48

H.6.2 Modified NASA TLX Workload Ratings

Participant	Attribute	Value
1	Mental Demand	15
2	Mental Demand	17
3	Mental Demand	16
4	Mental Demand	20
5	Mental Demand	20
6	Mental Demand	18
7	Mental Demand	20
8	Mental Demand	15
9	Mental Demand	18
10	Mental Demand	20
11	Mental Demand	17
12	Mental Demand	19
13	Mental Demand	7
14	Mental Demand	15
15	Mental Demand	17
16	Mental Demand	16
1	Physical Demand	4
2	Physical Demand	12
3	Physical Demand	4
4	Physical Demand	1
5	Physical Demand	4

Participant	Attribute	Value
6	Physical Demand	12
7	Physical Demand	1
8	Physical Demand	1
9	Physical Demand	1
10	Physical Demand	4
11	Physical Demand	12
12	Physical Demand	11
13	Physical Demand	8
14	Physical Demand	5
15	Physical Demand	3
16	Physical Demand	9
1	Time Pressure	20
2	Time Pressure	3
3	Time Pressure	13
4	Time Pressure	11
5	Time Pressure	20
6	Time Pressure	20
7	Time Pressure	17
8	Time Pressure	17
9	Time Pressure	18
10	Time Pressure	19
11	Time Pressure	19
12	Time Pressure	5
13	Time Pressure	7
14	Time Pressure	11
15	Time Pressure	15
16	Time Pressure	12
1	Effort Expended	16
2	Effort Expended	16
3	Effort Expended	16
4	Effort Expended	20
5	Effort Expended	15
6	Effort Expended	4
7	Effort Expended	20
8	Effort Expended	19
9	Effort Expended	19
10	Effort Expended	19
11	Effort Expended	19
12	Effort Expended	19
13	Effort Expended	7
14	Effort Expended	15
15	Effort Expended	19
16	Effort Expended	15
1	Performance Level Achieved	4
2	Performance Level Achieved	8
3	Performance Level Achieved	5
4	Performance Level Achieved	12
5	Performance Level Achieved	1
6	Performance Level Achieved	4
7	Performance Level Achieved	9
8	Performance Level Achieved	11
9	Performance Level Achieved	12
10	Performance Level Achieved	6
11	Performance Level Achieved	11
12	Performance Level Achieved	11

Participant	Attribute	Value
13	Performance Level Achieved	2
14	Performance Level Achieved	5
15	Performance Level Achieved	11
16	Performance Level Achieved	10
1	Frustration Experienced	16
2	Frustration Experienced	17
3	Frustration Experienced	8
4	Frustration Experienced	1
5	Frustration Experienced	20
6	Frustration Experienced	16
7	Frustration Experienced	7
8	Frustration Experienced	19
9	Frustration Experienced	18
10	Frustration Experienced	3
11	Frustration Experienced	11
12	Frustration Experienced	11
13	Frustration Experienced	12
14	Frustration Experienced	13
15	Frustration Experienced	11
16	Frustration Experienced	10
1	Annoyance Experienced	5
2	Annoyance Experienced	17
3	Annoyance Experienced	8
4	Annoyance Experienced	1
5	Annoyance Experienced	20
6	Annoyance Experienced	16
7	Annoyance Experienced	7
8	Annoyance Experienced	19
9	Annoyance Experienced	18
10	Annoyance Experienced	3
11	Annoyance Experienced	11
12	Annoyance Experienced	11
13	Annoyance Experienced	12
14	Annoyance Experienced	11
15	Annoyance Experienced	13
16	Annoyance Experienced	12

Appendix I

Spatial Experiments - Training Sheets

This appendix contains copies of the training sheets from the experiments described in Chapter 7.

Appendix J

Spatial Experiments - Raw Data

This appendix presents the raw results from the experiments described in Chapter 7.

J.1 Spatial versus Non-Spatial Conditions

J.1.1 Earcons and Earcon Attributes Identified

Participant	Attribute	Non-Spatial Condition	Spatial Condition
1	Average Rides Identified	0.55	1.30
2	Average Rides Identified	2.50	2.65
3	Average Rides Identified	1.45	2.25
4	Average Rides Identified	0.85	1.10
5	Average Rides Identified	0.95	0.70
6	Average Rides Identified	0.95	0.70
7	Average Rides Identified	2.50	1.95
8	Average Rides Identified	1.50	2.60
9	Average Rides Identified	1.75	1.75
10	Average Rides Identified	1.40	2.10
11	Average Rides Identified	1.70	1.45
12	Average Rides Identified	0.95	1.95
13	Average Rides Identified	1.50	1.80
14	Average Rides Identified	0.80	1.15
15	Average Rides Identified	1.40	1.80
16	Average Rides Identified	2.00	2.60
1	Average Ride Types Identified	2.95	2.85
2	Average Ride Types Identified	3.65	3.65
3	Average Ride Types Identified	2.40	3.30
4	Average Ride Types Identified	2.60	2.95
5	Average Ride Types Identified	2.95	2.90
6	Average Ride Types Identified	3.15	2.65
7	Average Ride Types Identified	3.65	3.30
8	Average Ride Types Identified	3.05	3.35
9	Average Ride Types Identified	3.35	3.20
10	Average Ride Types Identified	2.95	3.70
11	Average Ride Types Identified	3.30	3.05
12	Average Ride Types Identified	2.85	3.40
13	Average Ride Types Identified	3.55	2.55
14	Average Ride Types Identified	2.40	2.60
15	Average Ride Types Identified	2.95	3.25
16	Average Ride Types Identified	3.15	3.25
1	Average Ride Intensities Identified	1.55	2.45
2	Average Ride Intensities Identified	3.35	3.50
3	Average Ride Intensities Identified	3.50	3.30
4	Average Ride Intensities Identified	2.75	2.40
5	Average Ride Intensities Identified	2.70	2.55
6	Average Ride Intensities Identified	2.50	2.10

Participant	Attribute	Non-Spatial Condition	Spatial Condition
7	Average Ride Intensities Identified	3.55	3.10
8	Average Ride Intensities Identified	2.50	3.30
9	Average Ride Intensities Identified	3.20	3.25
10	Average Ride Intensities Identified	2.90	3.30
11	Average Ride Intensities Identified	3.00	2.90
12	Average Ride Intensities Identified	2.55	3.70
13	Average Ride Intensities Identified	3.15	3.35
14	Average Ride Intensities Identified	2.40	2.40
15	Average Ride Intensities Identified	3.05	3.25
16	Average Ride Intensities Identified	3.20	3.70
1	Average Ride Costs Identified	2.25	2.55
2	Average Ride Costs Identified	2.80	3.15
3	Average Ride Costs Identified	2.70	2.85
4	Average Ride Costs Identified	2.00	2.50
5	Average Ride Costs Identified	2.10	1.85
6	Average Ride Costs Identified	2.20	2.00
7	Average Ride Costs Identified	2.70	2.55
8	Average Ride Costs Identified	2.40	3.00
9	Average Ride Costs Identified	2.70	2.65
10	Average Ride Costs Identified	2.55	2.80
11	Average Ride Costs Identified	2.20	2.05
12	Average Ride Costs Identified	2.30	2.30
13	Average Ride Costs Identified	2.20	2.75
14	Average Ride Costs Identified	2.50	2.55
15	Average Ride Costs Identified	2.50	2.85
16	Average Ride Costs Identified	2.75	3.10

J.1.2 Modified NASA TLX Workload Ratings

Participant	Attribute	Non-Spatial Condition	Spatial Condition
1	Annoyance Experienced	6	9
2	Annoyance Experienced	16	9
3	Annoyance Experienced	2	2
4	Annoyance Experienced	9	9
5	Annoyance Experienced	3	18
6	Annoyance Experienced	18	19
7	Annoyance Experienced	8	5
8	Annoyance Experienced	6	6
9	Annoyance Experienced	13	15
10	Annoyance Experienced	5	5
11	Annoyance Experienced	11	5
12	Annoyance Experienced	12	8
13	Annoyance Experienced	5	9
14	Annoyance Experienced	2	2
15	Annoyance Experienced	2	2
16	Annoyance Experienced	2	3
1	Effort Expended	15	19
2	Effort Expended	15	13
3	Effort Expended	16	20
4	Effort Expended	14	6
5	Effort Expended	13	18
6	Effort Expended	18	18
7	Effort Expended	10	17

Participant	Attribute	Non-Spatial Condition	Spatial Condition
8	Effort Expended	16	13
9	Effort Expended	14	15
10	Effort Expended	13	10
11	Effort Expended	18	18
12	Effort Expended	12	14
13	Effort Expended	15	19
14	Effort Expended	10	13
15	Effort Expended	17	18
16	Effort Expended	16	17
1	Frustration Experienced	10	6
2	Frustration Experienced	20	15
3	Frustration Experienced	3	3
4	Frustration Experienced	13	15
5	Frustration Experienced	11	19
6	Frustration Experienced	18	19
7	Frustration Experienced	8	13
8	Frustration Experienced	17	10
9	Frustration Experienced	13	13
10	Frustration Experienced	18	5
11	Frustration Experienced	11	16
12	Frustration Experienced	14	12
13	Frustration Experienced	12	14
14	Frustration Experienced	7	13
15	Frustration Experienced	4	5
16	Frustration Experienced	11	9
1	Mental Demand	15	18
2	Mental Demand	18	19
3	Mental Demand	18	20
4	Mental Demand	16	17
5	Mental Demand	14	18
6	Mental Demand	19	19
7	Mental Demand	14	18
8	Mental Demand	20	12
9	Mental Demand	12	13
10	Mental Demand	17	11
11	Mental Demand	13	18
12	Mental Demand	18	12
13	Mental Demand	17	19
14	Mental Demand	13	16
15	Mental Demand	16	14
16	Mental Demand	19	18
1	Overall Preference	17	10
2	Overall Preference	2	18
3	Overall Preference	17	14
4	Overall Preference	6	8
5	Overall Preference	18	10
6	Overall Preference	10	7
7	Overall Preference	17	13
8	Overall Preference	12	18
9	Overall Preference	14	11
10	Overall Preference	2	17
11	Overall Preference	15	8
12	Overall Preference	4	18
13	Overall Preference	15	10
14	Overall Preference	11	16

Participant	Attribute	Non-Spatial Condition	Spatial Condition
15	Overall Preference	14	7
16	Overall Preference	8	17
1	Performance Level Achieved	16	14
2	Performance Level Achieved	9	11
3	Performance Level Achieved	17	15
4	Performance Level Achieved	4	6
5	Performance Level Achieved	14	10
6	Performance Level Achieved	2	2
7	Performance Level Achieved	16	11
8	Performance Level Achieved	12	15
9	Performance Level Achieved	9	5
10	Performance Level Achieved	6	15
11	Performance Level Achieved	16	11
12	Performance Level Achieved	12	12
13	Performance Level Achieved	9	4
14	Performance Level Achieved	19	9
15	Performance Level Achieved	10	3
16	Performance Level Achieved	13	16
1	Physical Demand	4	11
2	Physical Demand	2	12
3	Physical Demand	1	12
4	Physical Demand	1	1
5	Physical Demand	1	1
6	Physical Demand	14	16
7	Physical Demand	6	19
8	Physical Demand	2	2
9	Physical Demand	10	9
10	Physical Demand	1	2
11	Physical Demand	13	5
12	Physical Demand	1	1
13	Physical Demand	1	3
14	Physical Demand	11	16
15	Physical Demand	1	2
16	Physical Demand	1	13
1	Time Pressure	10	19
2	Time Pressure	20	16
3	Time Pressure	4	5
4	Time Pressure	18	14
5	Time Pressure	11	20
6	Time Pressure	20	19
7	Time Pressure	9	19
8	Time Pressure	18	15
9	Time Pressure	12	16
10	Time Pressure	13	10
11	Time Pressure	13	18
12	Time Pressure	10	12
13	Time Pressure	14	17
14	Time Pressure	7	10
15	Time Pressure	14	18
16	Time Pressure	15	13

J.2 Spatial(2) versus Revised Spatial Conditions

J.2.1 Earcons and Earcon Attributes Identified

Participant	Attribute	Spatial(2) Condition	Revised Spatial Condition
1	Average Rides Identified	2.70	2.55
2	Average Rides Identified	2.95	3.05
3	Average Rides Identified	2.45	2.40
4	Average Rides Identified	1.50	1.15
5	Average Rides Identified	2.00	2.70
6	Average Rides Identified	1.75	1.40
7	Average Rides Identified	0.60	0.60
8	Average Rides Identified	2.70	2.60
9	Average Rides Identified	1.45	0.80
10	Average Rides Identified	1.65	1.10
11	Average Rides Identified	2.05	1.85
12	Average Rides Identified	0.85	1.25
13	Average Rides Identified	1.05	1.00
14	Average Rides Identified	1.75	1.30
15	Average Rides Identified	1.65	1.10
16	Average Rides Identified	2.05	1.85
1	Average Ride Types Identified	3.20	3.55
2	Average Ride Types Identified	3.60	3.80
3	Average Ride Types Identified	3.35	3.90
4	Average Ride Types Identified	2.60	2.60
5	Average Ride Types Identified	3.10	3.75
6	Average Ride Types Identified	2.75	3.00
7	Average Ride Types Identified	2.90	2.80
8	Average Ride Types Identified	3.60	3.90
9	Average Ride Types Identified	2.85	2.45
10	Average Ride Types Identified	2.95	3.20
11	Average Ride Types Identified	3.20	3.45
12	Average Ride Types Identified	2.20	2.40
13	Average Ride Types Identified	2.50	2.60
14	Average Ride Types Identified	2.60	3.10
15	Average Ride Types Identified	2.90	2.80
16	Average Ride Types Identified	3.60	3.90
1	Average Ride Intensities Identified	3.55	3.90
2	Average Ride Intensities Identified	3.70	3.70
3	Average Ride Intensities Identified	3.65	3.95
4	Average Ride Intensities Identified	3.30	3.20
5	Average Ride Intensities Identified	3.35	3.50
6	Average Ride Intensities Identified	3.10	3.20
7	Average Ride Intensities Identified	2.10	2.05
8	Average Ride Intensities Identified	3.50	3.85
9	Average Ride Intensities Identified	2.85	2.65
10	Average Ride Intensities Identified	2.95	2.75
11	Average Ride Intensities Identified	3.40	3.45
12	Average Ride Intensities Identified	3.20	3.20
13	Average Ride Intensities Identified	2.70	3.10
14	Average Ride Intensities Identified	3.55	2.95
15	Average Ride Intensities Identified	3.70	3.70
16	Average Ride Intensities Identified	3.65	3.95
1	Average Ride Costs Identified	3.45	2.90
2	Average Ride Costs Identified	3.20	3.30
3	Average Ride Costs Identified	3.10	2.45

Participant	Attribute	Spatial(2) Condition	Revised Spatial Condition
4	Average Ride Costs Identified	2.90	2.50
5	Average Ride Costs Identified	2.50	3.25
6	Average Ride Costs Identified	2.55	2.20
7	Average Ride Costs Identified	1.70	1.80
8	Average Ride Costs Identified	3.05	2.65
9	Average Ride Costs Identified	2.20	2.30
10	Average Ride Costs Identified	2.50	1.90
11	Average Ride Costs Identified	2.60	2.40
12	Average Ride Costs Identified	2.15	2.40
13	Average Ride Costs Identified	2.50	2.40
14	Average Ride Costs Identified	3.00	2.40
15	Average Ride Costs Identified	2.50	3.25
16	Average Ride Costs Identified	3.20	3.30

J.2.2 Modified NASA TLX Workload Ratings

Participant	Attribute	Spatial(2) Condition	Revised Spatial Condition
1	Annoyance Experienced	8	10
2	Annoyance Experienced	6	6
3	Annoyance Experienced	1	1
4	Annoyance Experienced	4	1
5	Annoyance Experienced	5	2
6	Annoyance Experienced	17	11
7	Annoyance Experienced	17	15
8	Annoyance Experienced	6	5
9	Annoyance Experienced	9	13
10	Annoyance Experienced	7	4
11	Annoyance Experienced	1	1
12	Annoyance Experienced	10	16
13	Annoyance Experienced	4	4
14	Annoyance Experienced	7	4
15	Annoyance Experienced	9	2
16	Annoyance Experienced	10	5
1	Effort Expended	17	18
2	Effort Expended	20	19
3	Effort Expended	15	15
4	Effort Expended	15	18
5	Effort Expended	18	18
6	Effort Expended	15	11
7	Effort Expended	19	19
8	Effort Expended	13	16
9	Effort Expended	14	14
10	Effort Expended	19	16
11	Effort Expended	16	17
12	Effort Expended	17	18
13	Effort Expended	10	11
14	Effort Expended	17	10
15	Effort Expended	19	11
16	Effort Expended	15	12
1	Frustration Experienced	16	15
2	Frustration Experienced	6	6
3	Frustration Experienced	1	1
4	Frustration Experienced	11	12

Participant	Attribute	Spatial(2) Condition	Revised Spatial Condition
5	Frustration Experienced	13	3
6	Frustration Experienced	14	10
7	Frustration Experienced	20	19
8	Frustration Experienced	14	5
9	Frustration Experienced	10	17
10	Frustration Experienced	10	4
11	Frustration Experienced	12	16
12	Frustration Experienced	14	15
13	Frustration Experienced	10	11
14	Frustration Experienced	5	7
15	Frustration Experienced	12	8
16	Frustration Experienced	11	9
1	Mental Demand	19	19
2	Mental Demand	18	16
3	Mental Demand	19	19
4	Mental Demand	18	19
5	Mental Demand	10	14
6	Mental Demand	15	10
7	Mental Demand	15	15
8	Mental Demand	19	19
9	Mental Demand	18	19
10	Mental Demand	19	19
11	Mental Demand	16	16
12	Mental Demand	20	19
13	Mental Demand	14	15
14	Mental Demand	19	8
15	Mental Demand	20	15
16	Mental Demand	20	16
1	Overall Preference	2	4
2	Overall Preference	6	17
3	Overall Preference	1	20
4	Overall Preference	10	12
5	Overall Preference	4	17
6	Overall Preference	8	12
7	Overall Preference	13	18
8	Overall Preference	7	15
9	Overall Preference	9	5
10	Overall Preference	9	5
11	Overall Preference	18	3
12	Overall Preference	18	12
13	Overall Preference	12	9
14	Overall Preference	10	13
15	Overall Preference	5	16
16	Overall Preference	7	17
1	Performance Level Achieved	12	13
2	Performance Level Achieved	10	13
3	Performance Level Achieved	9	13
4	Performance Level Achieved	12	14
5	Performance Level Achieved	12	17
6	Performance Level Achieved	11	13
7	Performance Level Achieved	12	12
8	Performance Level Achieved	10	15
9	Performance Level Achieved	8	6
10	Performance Level Achieved	3	8
11	Performance Level Achieved	8	5

Participant	Attribute	Spatial(2) Condition	Revised Spatial Condition
12	Performance Level Achieved	10	13
13	Performance Level Achieved	6	5
14	Performance Level Achieved	12	14
15	Performance Level Achieved	6	12
16	Performance Level Achieved	9	16
1	Physical Demand	11	7
2	Physical Demand	7	3
3	Physical Demand	2	1
4	Physical Demand	18	10
5	Physical Demand	18	14
6	Physical Demand	10	5
7	Physical Demand	11	11
8	Physical Demand	3	3
9	Physical Demand	9	9
10	Physical Demand	5	8
11	Physical Demand	7	5
12	Physical Demand	1	1
13	Physical Demand	3	3
14	Physical Demand	16	12
15	Physical Demand	5	5
16	Physical Demand	8	7
1	Time Pressure	14	17
2	Time Pressure	13	13
3	Time Pressure	19	10
4	Time Pressure	18	19
5	Time Pressure	19	14
6	Time Pressure	12	12
7	Time Pressure	19	19
8	Time Pressure	16	10
9	Time Pressure	18	18
10	Time Pressure	15	7
11	Time Pressure	13	16
12	Time Pressure	17	11
13	Time Pressure	11	11
14	Time Pressure	20	2
15	Time Pressure	14	10
16	Time Pressure	15	9

Appendix K

Diary Experiment - Training Sheets

This appendix contains copies of the training sheets from the experiment described in Chapter 8.

Appendix L

Diary Experiment - Raw Data

This appendix presents the raw results from the experiment described in Chapter 8.

L.1 Type S & M Question Scores

Participant	Condition	Type S Question Score	Type M Question Score
1	Non-Spatial Condition	4	2.8
2	Non-Spatial Condition	3	3.0
3	Non-Spatial Condition	2	2.3
4	Non-Spatial Condition	3	2.0
5	Non-Spatial Condition	2	4.0
6	Non-Spatial Condition	3	2.1
7	Non-Spatial Condition	5	2.1
8	Non-Spatial Condition	2	3.3
9	Non-Spatial Condition	2	0.8
10	Non-Spatial Condition	1	1.6
11	Non-Spatial Condition	2	3.3
12	Non-Spatial Condition	2	1.8
13	Non-Spatial Condition	0	1.5
14	Non-Spatial Condition	3	2.3
15	Non-Spatial Condition	2	2.1
16	Non-Spatial Condition	1	2.3
1	Spatial Condition	1	1.2
2	Spatial Condition	3	3.0
3	Spatial Condition	1	2.8
4	Spatial Condition	2	1.4
5	Spatial Condition	5	2.8
6	Spatial Condition	1	1.7
7	Spatial Condition	1	2.7
8	Spatial Condition	1	4.3
9	Spatial Condition	1	1.8
10	Spatial Condition	2	3.2
11	Spatial Condition	3	1.7
12	Spatial Condition	3	1.8
13	Spatial Condition	0	2.7
14	Spatial Condition	1	3.8
15	Spatial Condition	5	2.8
16	Spatial Condition	3	1.8

L.2 PPWS Scores

Participant	PPWS Non-Spatial Condition	PPWS Spatial Condition
1	65.56	74.33
2	70.55	71.88
3	70.14	69.70
4	71.43	65.63
5	67.31	74.47
6	77.34	80.80
7	78.34	75.93
8	47.42	47.33
9	44.81	43.90
10	57.61	59.22
11	57.61	53.54
12	68.05	76.16
13	49.25	50.19
14	64.86	65.53
15	75.36	74.88
16	74.81	82.77

L.3 Modified NASA TLX Workload Ratings

Participant	Attribute	Non-Spatial Condition	Spatial Condition
1	Annoyance Experienced	18	16
2	Annoyance Experienced	9	5
3	Annoyance Experienced	18	17
4	Annoyance Experienced	9	3
5	Annoyance Experienced	11	11
6	Annoyance Experienced	13	8
7	Annoyance Experienced	16	12
8	Annoyance Experienced	18	15
9	Annoyance Experienced	17	20
10	Annoyance Experienced	18	16
11	Annoyance Experienced	19	11
12	Annoyance Experienced	5	4
13	Annoyance Experienced	3	8
14	Annoyance Experienced	8	6
15	Annoyance Experienced	19	11
16	Annoyance Experienced	18	12
1	Effort Expended	19	18
2	Effort Expended	13	11
3	Effort Expended	12	13
4	Effort Expended	1	13
5	Effort Expended	1	15
6	Effort Expended	18	19
7	Effort Expended	19	18
8	Effort Expended	14	13
9	Effort Expended	2	20
10	Effort Expended	19	18
11	Effort Expended	18	14
12	Effort Expended	17	16
13	Effort Expended	10	13

Participant	Attribute	Non-Spatial Condition	Spatial Condition
14	Effort Expended	18	18
15	Effort Expended	17	14
16	Effort Expended	20	16
1	Frustration Experienced	19	15
2	Frustration Experienced	13	7
3	Frustration Experienced	18	17
4	Frustration Experienced	12	4
5	Frustration Experienced	11	11
6	Frustration Experienced	7	12
7	Frustration Experienced	15	15
8	Frustration Experienced	16	14
9	Frustration Experienced	14	20
10	Frustration Experienced	18	16
11	Frustration Experienced	5	14
12	Frustration Experienced	5	8
13	Frustration Experienced	2	5
14	Frustration Experienced	11	12
15	Frustration Experienced	5	13
16	Frustration Experienced	18	15
1	Mental Demand	20	19
2	Mental Demand	17	14
3	Mental Demand	16	16
4	Mental Demand	13	12
5	Mental Demand	17	19
6	Mental Demand	19	15
7	Mental Demand	19	18
8	Mental Demand	16	14
9	Mental Demand	20	20
10	Mental Demand	20	15
11	Mental Demand	19	17
12	Mental Demand	19	17
13	Mental Demand	15	15
14	Mental Demand	19	19
15	Mental Demand	19	16
16	Mental Demand	18	18
1	Overall Preference	4	14
2	Overall Preference	5	16
3	Overall Preference	1	2
4	Overall Preference	5	11
5	Overall Preference	12	14
6	Overall Preference	7	16
7	Overall Preference	10	16
8	Overall Preference	7	12
9	Overall Preference	11	11
10	Overall Preference	5	10
11	Overall Preference	1	17
12	Overall Preference	5	17
13	Overall Preference	14	10
14	Overall Preference	12	8
15	Overall Preference	1	17
16	Overall Preference	5	16
1	Performance Level Achieved	3	7
2	Performance Level Achieved	6	9
3	Performance Level Achieved	2	4
4	Performance Level Achieved	14	16

Participant	Attribute	Non-Spatial Condition	Spatial Condition
5	Performance Level Achieved	12	12
6	Performance Level Achieved	17	9
7	Performance Level Achieved	7	8
8	Performance Level Achieved	8	10
9	Performance Level Achieved	11	6
10	Performance Level Achieved	10	11
11	Performance Level Achieved	2	6
12	Performance Level Achieved	5	5
13	Performance Level Achieved	16	18
14	Performance Level Achieved	6	5
15	Performance Level Achieved	2	6
16	Performance Level Achieved	7	8
1	Physical Demand	13	12
2	Physical Demand	9	6
3	Physical Demand	4	4
4	Physical Demand	12	13
5	Physical Demand	5	4
6	Physical Demand	15	14
7	Physical Demand	6	7
8	Physical Demand	6	6
9	Physical Demand	11	10
10	Physical Demand	6	6
11	Physical Demand	10	10
12	Physical Demand	5	5
13	Physical Demand	16	17
14	Physical Demand	12	14
15	Physical Demand	11	10
16	Physical Demand	7	7
1	Time Pressure	19	15
2	Time Pressure	4	4
3	Time Pressure	3	3
4	Time Pressure	9	9
5	Time Pressure	5	5
6	Time Pressure	16	7
7	Time Pressure	13	14
8	Time Pressure	18	17
9	Time Pressure	1	1
10	Time Pressure	18	17
11	Time Pressure	16	16
12	Time Pressure	15	13
13	Time Pressure	12	12
14	Time Pressure	14	14
15	Time Pressure	16	17
16	Time Pressure	17	14

Appendix M

Companion CD-ROM

This appendix contains a CD-ROM containing the earcons used in the experiments from Chapters 4 to 8.

Bibliography

[1] C. Ahlberg and B. Shneiderman. Visual information seeking using the filmfinder. In *Proceedings of CHI 94*, volume 2, pages 433–434, Boston, Massachusetts, 1994. ACM Press.

[2] M. C. Albers. Auditory cues for browsing, surfing and navigating. In *Proceedings of ICAD 96*, pages 85–90, Palo Alto, California, 1996. ICAD.

[3] M. C. Albers and E. Bergman. The audible web: Auditory enchancements for mosaic. In *Proceedings of CHI 95*, volume 2, pages 318–319, Denver, Colorado, 1995. ACM Press.

[4] P. M. Aoki, M. Romaine, M. H. Szymanski, J. D. Thornton, D. Wilson, and A. Woodruff. The mad hatter's cocktail party: A social mobile audio space supporting multiple simultaneous conversations. In *Proceedings of CHI 2003*, volume 1, pages 425–432, Ft. Lauderdale, Florida, 2003. ACM Press.

[5] B. Arons. A review of the cocktail party effect. *Journal of the American Voice I/O Society*, 12(7):35–50, 1992.

[6] American Standards Association. Acoustical termonology si. Technical report, American Standards Association, 1960.

[7] B. Bargen and P. Donnelly. *Inside DirectX*. Microsoft Programming Series. Microsoft Press, Redmond, Washington, 1998.

[8] D. R. Begault. *3-D Sound for virtual reality and multimedia*. Academic Press, Boston, 1994.

[9] F. Bellotti, R. Berta, A. De Gloria, and M. Margarone. Using 3d sound to improve the effectiveness of the advanced driver assistance systems. *Personal and Ubiquitous Computing*, 6(3):155–163, 2002.

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

[11] A. E. Blandford and T. R. G. Green. Group and individual time management tools: What you get is not always what you need. *Personal and Ubiquitous Computing*, 5(4):213–230, 2001.

[12] M. M. Blattner, A. L. Papp, and E. P. Glinert. Sonic enhancement of two-dimensional graphic displays. In G. Kramer, editor, *Auditory Display: Sonification, Audification, and Auditory Interfaces*, pages 447–470. Addison-Wesley, Reading, Massachusetts, 1994.

[13] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg. Earcons and icons: Their structure and common design principles. *Human Computer Interaction*, 4(1):11–44, 1989.

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

[15] A. S. Bregman and S. Pinker. Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, 32(1):20–31, 1978.

[16] S. A. Brewster. *Providing a structured method for integrating non-speech audio into human-computer interfaces*. PhD thesis, Department of Computer Science, University of York, 1994.

[17] S. A. Brewster. Using earcons to improve the usability of tool palettes. In *Proceedings of CHI 98*, volume 2, pages 297–298, Los Angeles, California, 1998. ACM Press.

[18] S. A. Brewster. Using non-speech sounds to provide navigation cues. *ACM Transactions on CHI*, 5(2):224–259, 1998.

[19] S. A. Brewster. Non-speech auditory output. In J. Jako and A. Sears, editors, *The Human Computer Interaction Handbook*, pages 220–239. Lawrence Erlbaum Associates, 2002.

[20] S. A. Brewster. Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing*, 6(3):188–205, 2002.

[21] S. A. Brewster and M. G. Crease. Making menus musical. In *Proceedings of Interact 97*, volume 1, pages 389–396, Sydney, Australia, 1997. ACM Press.

[22] S. A. Brewster and R. Murray. Presenting dynamic information on mobile computers. *Personal Technologies*, 2(4):209–212, 2000.

[23] S. A. Brewster, P. C. Wright, A. J. Dix, and A. D. N. Edwards. The sonic enhancement of graphical buttons. In *Proceedings of Interact 95*, volume 1, pages 43–48, Lillehammer, Norway, 1995. Chapman Hall.

[24] S. A. Brewster, P. C. Wright, and A. D. N. Edwards. A detailed investigation into the effectiveness of earcons. In G. Kramer, editor, *Auditory Display: Sonification, Audification, and Auditory Interfaces*, pages 471–498. Addison-Wesley, Reading, Massachusetts, 1994.

[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*, volume 2, pages 155–159, Huddersfield, UK, 1995. Springer.

[26] 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*, 43(2):153–175, 1995.

[27] L. Brown, S. Brewster, R. Ramloll, W. Yu, and B. Riedel. Browsing modes for exploring sonified line graphs. In *Proceedings of BCS-HCI 2002*, volume 2, pages 6–9, London, UK, 2002. BCS.

[28] L. M. Brown and S. A. Brewster. Drawing by ear: Interpreting sonified line graphs. In *Proceedings of ICAD 2003*, pages 152–156, Boston, Massachusetts, 2003. ICAD.

[29] 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*, volume 1, pages 424–430, Kyoto, Japan, 2002. ICAD.

[30] D. S. Brungart and B. D. Simpson. The effects of spatial seperation in distance on the informational and energetic masking of a nearby speech signal. *Journal of the Acoustical Society of America*, 112(2):664–676, 2002.

[31] D. S. Brungart and B. D. Simpson. Optimizing the spatial configuration of a seven talker display. In *Proceedings of ICAD 2003*, pages 188–193, Boston, Massachusetts, 2003. ICAD.

[32] W. Buxton. Introduction to this special issue on nonspeech audio. *Human Computer Interaction*, 4(1):1–9, 1989.

[33] D. D. Clark-Carter, A. D. Heyes, and C. I. Howarth. The efficiency and walking speed of visually impaired people. *Ergonomics*, 29(6):779–789, 1986.

[34] M. Cohen and L. F. Ludwig. Multidimensional audio window management. *International Journal of Man-Machine Studies*, 34:319–336, 1991.

[35] Creative. http://www.creative.com, 2001.

[36] R. Cusack and B. Roberts. Effects of differences in timbre on sequential grouping. *Perception and Psychophysics*, 62(5):1112–1120, 2000.

[37] C. J. Darwin and V. Ciocca. Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component. *Journal of the Acoustical Society of America*, 91(1):2281–3390, 1992.

[38] D. Deutsch. Binaural integration of melodic patterns. *Perception and Psychophysics*, 25(5):339–405, 1979.

[39] D. Deutsch. Grouping mechanisms in music. In D. Deutsch, editor, *The Psychology of Music*, pages 299–348. Academic Press, San Diego, 2nd edition, 1999.

[40] W. J. Dowling. Rythmic groups and subjective chunks in memory for melodies. *Perception and Psychophysics*, 14(1):37–40, 1973.

[41] A. D. N. Edwards. Soundtrack: An auditory interface for blind users. *Human Computer Interaction*, 4(1):45–66, 1989.

[42] A. D. N. Edwards, I. J. Pitt, S. A. Brewster, and R. D. Stevens. Multiple modalities in adapted interfaces. In A. D. N. Edwards, editor, *Extra-Ordinary Human-Computer Interaction: Interfaces for Users with Disabilities*, pages 221–243. Cambridge University Press, Cambridge, 1995.

[43] K. Edwards, E. D. Mynatt, and K. Stockton. Access to graphical interfaces for blind users. *Interactions*, 2(1):56–67, 1995.

[44] K. W. Edwards and E. D. Mynatt. An architecture for transforming graphical interfaces. In *Proceedings of UIST 94*, volume 1, pages 39–47, Marina del Ray, California, 1994. ACM Press.

[45] J. Edworthy and A. Adams. *Warning Design: A Research Perspective*. Taylor and Francis, London, 1996.

[46] M. Fernström and L. Bannon. Explorations in sonic browsing. In *Procedings of HCI 97*, volume 1, pages 117–131, Bristol, UK, 1997. Springer Verlag.

[47] H. Fletcher. Auditory patterns. *Reviews of Modern Physics*, 12:47–65, 1940.

[48] D. J. Folds and J. M. Gerth. Auditory monitoring of up to eight simultaneous sources. In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, pages 505–509, 1994.

[49] Freedom Scientific. http://www.freedomscientific.com/, 2003.

[50] G. W. Furnas. The fisheye view: a new look at structured files. Technical Report 82-11221-22, Bell Laboratories, 1982.

[51] G. W. Furnas. Generalized fisheye views. In *Proceedings of CHI 86*, volume 1, pages 16–23, Boston, Massachusetts, 1986. ACM Press.

[52] W. W. Gaver. The sonicfinder: An interface that uses auditory icons. *Human Computer Interaction*, 4(1):67–94, 1989.

[53] W. W. Gaver. Synthesizing auditory icons. In *Proceedings of INTERCHI'93*, volume 1, pages 228–235, Amsterdam, The Netherlands, 1993. ACM Press.

[54] W. W. Gaver. Auditory interfaces. In M. G. Helander, T. K. Landauer, and P. V. Prabhu, editors, *Handbook of Human-Computer Interaction*, pages 1003–1041. Elsevier, Amsterdam, 2nd edition, 1997.

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

[56] S. A. Gelfand. *Hearing: An Introduction to Psychological and Physiological Acoustics*. Marcel Dekker Inc, New York, 3rd edition, 1998.

[57] J. M. Gerth. *Performance Based Refinement of a Synthetic Auditory Ambience: Identifying and Discriminating Auditory Sources*. PhD thesis, Department of Psychology, Georgia Institute of Technology, 1992.

[58] L. Halpern. The effect of harmonic ratio relationships on auditory stream segregation. Technical report, Psychology Department, McGill University, 1977.

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

[60] S. Hart and L. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. Hancock and N. Meshkati, editors, *Human Mental Workload*, pages 139–183. North Holland, Amsterdam, 1988.

[61] G. A. Heise and G. A. Miller. An experimental study of auditory patterns. *American Journal of Psychology*, 64:68–77, 1951.

[62] S. Holland, D. R. Morse, and H. Gedenryd. Audiogps: Spatial audio navigation with a minimal attention interface. *Personal and Ubiquitous Computing*, 6(4):253–259, 2002.

[63] D. M. Howard and A. James. *Acoustics and Psychoacoustics*. Focal Press, Oxford, 2nd edition, 2001.

[64] D. Howell. *Statistical Methods for Psychology*. Duxbury, Belmont, 4th edition, 1997.

[65] ICAD. http://www.icad.org, 2003.

[66] C. M. Kincaid, P. B. Dupont, and A. R. Kaye. Electronic calendars in the office: An assessment of user neeeds and current technology. *ACM Transactions on Office Information Systems*, 3(1):89–102, 1985.

[67] M. Kobayashi and C. Schmandt. Dynamic soundscape: mapping time to space for audio browsing. In *Proceedings of CHI 97*, volume 1, pages 194–201, Atlanta, Georgia, 1997. ACM Press.

[68] G. Kramer. An introduction to auditory display. In G. Kramer, editor, *Auditory Display: Sonification, Audification and Auditory Interfaces*, pages 1–77. Addison-Wesley, Reading, Massachusetts, 1994.

[69] G. Kramer. Some organizing principles for representing data with sound. In G. Kramer, editor, *Auditory Display: Sonification, Audification and Auditory Interfaces*, pages 185–221. Addison-Wesley, Reading, Massachusetts, 1994.

[70] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner, J. Neuhoff, R. Bargar, S. Barrass, J. Berger, G. Evreinov, W. T. Fitch, M. Göhn, S. Handel, H. Kaper, H. Levkowitz, S. Lodha, B. Shinn-Cunningham, M. Simoni, and S. Tipei. Sonification report: Status of the field and research agenda. Technical report, ICAD, 1999.

[71] Lake Technology Limited. http://www.lake.com, 2003.

[72] G. Leplâtre and S. A. Brewster. An investigation of using music to provide naviagtion cues. In *Proceedings of ICAD 98*, Glasgow, UK, 1998. BCS.

[73] G. Leplâtre and S. A. Brewster. Designing non-speech sounds to support navigation in mobile phone menus. In *Proceedings of ICAD 2000*, pages 190–199, Atlanta, GA, 2000. ICAD.

[74] J. M. Loomis, C. Herbert, and J. G. Cincinelli. Active localization of virtual sounds. *Journal of the Acoustical Society of America*, 88(4):1757–1764, 1990.

[75] J. M. Mandler, D. Seegmiller, and J. Day. On the coding of spatial information. *Memory and Cognition*, 5(1):10–16, 1977.

[76] D. L. Mansur. *Graphs in Sound: A Numerical Data Analysis Method for the Blind*. MSc Thesis, University of California, 1985.

[77] A. Marcus. Icons, symbols, and signs: Visible languages to facilitate communication. *Interactions*, 10(3):37–43, 2003.

[78] Microsoft DirectX. http://www.microsoft.com/windows/directx/default.aspx, 2004.

[79] D. K. McGookin and S. A. Brewster. Fishears - the design of a multimodal focus and context system. In *Proceedings of IHM-HCI 2001*, volume 2, pages 1–4, Lille, France, 2001. Cépaduès Éditions.

[80] D. K. McGookin and S. A. Brewster. Dolphin: The design and initial evaluation of multimodal focus and context. In *Proceedings of ICAD 2002*, volume 1, pages 181–186, Kyoto, Japan, 2002. ICAD.

[81] D. K. McGookin and S. A. Brewster. An investigation into the identification of concurrently presented earcons. In *Proceedings of ICAD 2003*, pages 42–46, Boston, Massachusetts, 2003. ICAD.

[82] D. K. McGookin and S. A. Brewster. Empirically derived guidelines for the presentation of concurrent earcons. In *Proceedings of BCS-HCI 2004*, volume 2, pages 65–68, Leeds, UK, 2004. BCS.

[83] 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.

[84] D. K. McGookin and S. A. Brewster. Understanding concurrent earcons: Applying auditory scene analysis principles to concurrent earcon recognition. *ACM Transactions on Applied Perception*, 1(2):130–155, 2004.

[85] MIDI Manufacturers Association. http://www.midi.org, 2003.

[86] G. A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63(1):81–97, 1956.

[87] E. N. Mitsopoulos and D. N. Edwards. Auditory scene analysis as the basis for designing auditory widgets. In *Proceedings of ICAD 97*, pages 13–17, Palo Alto, California, 1997.

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

[89] B. C. J. Moore. Controversies and mysteries in spatial hearing. In *Proceedings of AES 16th International Conference on Spatial Sound Reproduction*, volume 1, pages 249–256, Rovaniemi, Finland, 1999. AES.

[90] T. J. Moore. Voice communication jamming research. In *Proceedings of AGAD 331: Aural Communication in Aviation*, volume 2, pages 1–6, Neuilly-Sur-Seine, France, 1981.

[91] E. D. Mynatt. Designing with auditory icons. In *Proceedings of ICAD 94*, pages 109–119, Santa Fe, New Mexico, 1994.

[92] E. D. Mynatt. Designing with auditory icons: How well do we hear auditory cues? In *Proceedings of CHI 94*, volume 2, pages 269–270, Boston, Massachusetts, 1994. ACM Press.

[93] E. D. Mynatt, M. Back, R. Want, M. Baer, and J. B. Ellis. Designing audio aura. In *Proceedings of CHI 98*, volume 1, pages 566–573, Los Angeles, California, 1998. ACM.

[94] E. D. Mynatt, M. Back, R. Want, and R. Frederick. Audio aura: Light-weight audio augmented reality. In *Proceedings of UIST 97*, volume 2, pages 211–212, Banff, Canada, 1997. ACM.

[95] J. G. Neuhoff, G. Kramer, and J. Wayand. Sonification and the interaction of perceptual dimensions: Can data get lost in the map? In *Proceedings of ICAD 2000*, pages 93–98, Atlanta, Georgia, 2000. ICAD.

[96] A. F. Newell. Extra-ordinary human-computer interaction. In A. D. N. Edwards, editor, *Extra-Ordinary Human-Computer Interaction: Interfaces for Users with Disabilites*, pages 3–18. Cambridge University Press, Cambridge, 1995.

[97] BBC News. http://news.bbc.co.uk/1/hi/sci/tech/1924144.stm, 2002.

[98] C. Nichol, S. Brewster, and P. Gray. A system for manipulating audio interfaces using timbre space. In *Proceedings of CADUI 2004*, pages 366–379, Madeira, Portugal, 2004. ACM Press.

[99] D. A. Norman. Cognitive artifacts. In J. M. Carroll, editor, *Designing Interaction: Psychology at the Human-Computer Interface*, Cambridge Series on Human-Computer Interaction, pages 17–38. Cambridge University Press, Cambridge, 1991.

[100] D. A. Norman. *The Design of Everyday Things*. MIT Press, Boston, Massachusets, 1998.

[101] OpenAL. http://www.openal.org/about/, 2003.

[102] L. Palen. Social, individual technological issues for groupware calendar systems. In *Proceedings of CHI 99*, volume 1, pages 17–24, Pittsburgh, Pennsylvania, 1999. ACM Press.

[103] Palm. http://www.palm.com, 2001.

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

[105] A. L. Papp and M. M. Blattner. Dynamic presentation of asynchronous auditory output. In *Proceedings of Multimedia 96*, volume 1, pages 109–116, Boston, Massachusetts, 1996. ACM Press.

[106] R. D. Patterson. Auditory warning systems for high-workload environments. *Ergonomics International*, 85:163–165, 1985.

[107] R. D. Patterson. Guidelines for the design of auditory warning sounds. *Institute of Acoustics*, 11(5):17–25, 1989.

[108] R. D. Patterson, J. Edworthy, and M. J. Shailer. Alarm sounds for medical equipment in intensive care areas and operating theatres. Technical report, Institute of Sound and Vibration Research, June 1996.

[109] S. J. Payne. Understanding calendar use. *Human Computer Interaction*, 8(2):83–100, 1993.

[110] D. R. Perrott. Concurrent minimum audible angle: A re-examination of concept of auditory special acuity. *Journal of the Acoustical Society of America*, 75(4):1201–1206, 1984.

[111] H. Petrie, V. Johnson, S. Furner, and T. Strothotte. Design lifecycles and wearable computers for users with disabilities. In *Proceedings of the first international workshop of Human Computer Interaction with mobile devices*, Glasgow, UK, 1998. GIST.

[112] R. Pieraccini, K. Dayanidhi, J. Bloom, J.-G. Dahan, M. Philips, B. R. Goodman, and V. Prasad. Multimodal interfaces that flex, adapt, and persist: Multimodal conversational systems for automobiles. *Communications of the ACM*, 47(1):47–49, 2004.

[113] A. Pirhonen, S. Brewster, and C. Holguin. Gestural and audio metaphors as a means of control for mobile devices. In *Proceedings of CHI 2002*, volume 1, pages 291–298, Minneapolis, Minnesota, 2002. ACM Press.

[114] Polhemus Inc. http://www.polhemus.com, 2001.

[115] PURE Digital. http://www.puredigital.com, 2004.

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

[117] R. A. Rasch. The perception of simultaneous notes such as in polyphonic music. *Acoustica*, 40(1):22–33, 1978.

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

[119] J.-C. Risset and D. L. Wessel. Exploration of timbre by analysis and synthesis. In D. Deutsch, editor, *The Psychology of Music*, pages 113–169. Academic Press, San Diego, 2nd edition, 1999.

[120] RNIB. http://www.rnib.org, 2004.

[121] D. A. Ross and B. B. Blasch. Development of a wearable computer orientation system. *Personal and Ubiquitous Computing*, 6(1):49–63, 2002.

[122] T. D. Rossing, F. R. Moore, and P. A. Wheeler. *The Science of Sound*. Addison Wesley, San Fransisco, 3rd edition, 2002.

[123] N. Sawhney and C. Schmandt. Nomadic radio: Scaleable and contextual notification for wearable audio messaging. In *Proceedings of CHI 99*, volume 1, pages 96–103, Pittsburgh, Pennsylvania, 1999. ACM Press.

[124] N. Sawhney and C. Schmandt. Nomadic radio: Speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on CHI*, 7(3):353–383, 2000.

[125] C. Scaletti. Sound synthesis algorithms for auditory data representations. In G. Kramer, editor, *Auditory Display: Sonification, Audification, and Auditory Interfaces*, pages 223–251. Addison-Wesley, Reading, Massachusetts, 1994.

[126] R. D. Shilling and B. Shinn-Cunningham. Virtual auditory displays. In K. M. Stanney, editor, *Handbook of Virtual Environment Technology: Design, Implementation, and Applications*, pages 65–92. Lawrence Erlbaum, Associates Inc, 2000.

[127] B. Shinn-Cunningham, H. Lehnert, G. Kramer, E. Wenzel, and N. Durlach. Auditory displays. In R. H. Gilkey and T. R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 611–633. Lawrence Erlbaum, Mahwah, 1997.

[128] B. Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley, Reading, 2nd edition, 1998.

[129] P. G. Singh. Perceptual organisation of complex tone sequences: A tradeoff between pitch and timbre. *Journal of the Acoustical Society of America*, 82(3):886–899, 1987.

[130] P. G. Singh and A. S. Bregman. The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *Journal of the Acoustical Society of America*, 102(4):1943–1952, 1997.

[131] P. G. Singh and I. J. Hirsh. Influence of spectral locus and f0 changes on the pitch and timbre of complex tones. *Journal of the Acoustical Society of America*, 92(5):2651–2661, 1992.

[132] R. Spence and M. D. Apperley. Database navigation: An office environment for the professional. *Behaviour and Information Technology*, 1(1):43–54, 1982.

[133] M. F. Spiegel and L. Streeter. Applying speech synthesis to user interfaces. In M. G. Helander, T. K. Landauer, and P. V. Prabhu, editors, *Handbook of Human-Computer Interaction*, pages 1061–1084. Elsevier, 2nd edition, 1997.

[134] L. J. Stifelman. The cocktail party effect in auditory interfaces: A study of simultaneous presentation. Technical report, MIT Media Laboratory, 1994.

[135] E. Taylor. *The AB Guide to Music Theory*. Associated Board of the Royal Schools of Music, 1989.

[136] Y. Tougas and A. S. Bregman. The crossing of auditory streams. *Journal of experimental psychology*, 11(6):788–798, 1985.

[137] Turtle Beach. http://www.turtlebeach.com/site/products/santacruz/, 2003.

[138] J. Tzelogov, R. Srebro, A. Henik, and A. Kushelesky. Radiation search and detection by ear and by eye. *Human Factors*, 29(1):87–95, 1987.

[139] L. P. A. S. van Noorden. *Temporal Coherence in the Perception of Tone Sequences*. PhD thesis, Institue for Perception Research, 1975.

[140] M. L. M. Vargas and S. Anderson. Combining speech and earcons to assist menu navigation. In *Proceedings of ICAD 2003*, pages 38–41, Boston, Massachusetts, 2003. ICAD.

[141] A. Walker and S. Brewster. Sitting too close to the screen can be bad for your ears: A study of audio-visual location discrepancy detection under different visual projections. In *Proceedings of ICAD 2001*, pages 86–89, Helsinki, Finland, 2001. ICAD.

[142] A. Walker and S. A. Brewster. Spatial audio in small screen device displays. *Personal Technologies*, 4(2):144–154, 2000.

[143] 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*, volume 1, pages 531–539, Lille, France, 2001. Springer.

[144] B. N. Walker and J. T. Cothran. Sonification sandbox: A graphical toolkit for auditory graphs. In *Proceedings of ICAD 2003*, pages 161–163, Boston, Massachusetts, 2003. ICAD.

[145] K. Wang, V. Sundareswaran, C. Tam, P. Bangayan, and P. Zahorik. Efficient and effective use of low-cost 3d audio systems. In *Proceedings of ICAD 2002*, pages 239–243, Kyoto, Japan, 2002. ICAD.

[146] E. Wenzel. Spatial sound and sonification. In G. Kramer, editor, *Auditory Display: Sonification, Audification and Auditory Interfaces*, pages 127–150. Adison-Wesley, Reading, Massachusetts, 1994.

[147] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman. Localization using nonindividualized head-related transfer functions. *Journal of the acoustical society of America*, 94(1):111–123, 1993.

[148] E. M. Wenzel, F. L. Wightman, and D. J. Kistler. Localization with non-individualized virtual acoustic display cues. In *Proceedings of CHI 91*, volume 1, pages 351–359, New Orleans, Louisiana, 1991. ACM Press.

[149] F. L. Wightman and D. J. Kistler. Headphone simulation of free-field listening ii: psychological validation. *Journal of the Accoustical Society of America*, 85(2):868–878, 1989.

[150] S. M. Williams. Perceptual principles in sound grouping. In G. Kramer, editor, *Auditory Display: Sonification, Audification, and Auditory Interfaces*, pages 95–125. Addison-Wesley, Reading, Massachusetts, 1994.

[151] Xybernaught Corporation. http://www.xybernaught.com, 2004.

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

[153] W. Yu and S. A. Brewster. Evaluation of multimodal graphs for blind people. *Universal Access in the Information Society*, 2(2):105–124, 2003.

[154] P. Zahorik. Auditory display of sound source distance. In *Proceedings of ICAD 2002*, pages 326–332, Kyoto, Japan, 2002. ICAD.

[155] D. N. Zotkin, R. Duraiswami, and L. S. Davis. Customizable auditory displays. In *Proceedings of ICAD 2002*, pages 167–176, Kyoto, Japan, 2002. ICAD.