Interactive non-speech auditory display of multivariate data

Author: Sandra Pauletto
Thesis Submission for the Degree of PhD

University of York, Department of Electronics
Intelligent Systems Group, Audio Systems

March 2007

THESIS CONTAINS

CD

Abstract

This research investigates the combined use of non-speech sound and user interaction as an effective and efficient way of displaying and extracting meaningful information from large data sets. Three experiments form the main focus of this thesis. The first two employ data from the domains of helicopter engineering and physiotherapy, research fields in which there is a strong need for new techniques to display and analyse data. The first experiment compares an auditory display of helicopter flight data with an equivalent visual display. The results show that, on average, people rank the presence in the data set of specific structures (periodicities, noise, etc.) similarly using either the auditory or the visual display. The second experiment verifies if a particular auditory display of electromyography (EMG) data portrays effectively information which we know to be present in the data. The results show that the auditory display is effective in displaying such information, that the roughness of the sound is a good indicator of the level of presence of this information in the data and that this auditory display is an appropriate candidate for the future development of a real-time auditory feedback system for EMG data.

The third experiment focuses on user interaction with auditory displays. Three ways of interacting with a known auditory display are compared in terms of usability, namely: effectiveness, efficiency and user satisfaction. These three interacting methods differ from one another regarding the amount of real-time control they offer the user. Results from this experiment show that the method that allows the least real-time control is the least efficient and effective. The other two interaction methods score similarly for efficiency and effectiveness. The method allowing a medium level of real-time control, however, scores more highly in user satisfaction.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisor Andy Hunt for his guidance and support throughout this experience.

I would like to thank my parents Antonietta, Giancarlo and my sister Marta for their love and support which always reaches me even if I live far away from home.

Thanks to Tom who has been there in the worst moments.

Thanks to Frank for his help in correcting my English mistakes.

Thanks to all my friends, in particular the ones in this island and in that peninsula, for being supportive and always encouraging.

Declaration

The author of this thesis has worked as a research assistant in the EPSRC funded project *Data mining through an interactive sonic approach* (GR/S08886/01, Principal Investigator: Dr Andy Hunt) between April 2003 and August 2005. The main results of this thesis have been published in the final EPSRC project report.

The literature survey of Chapter 3 has been published in [Hunt, Hermann and Pauletto, 2004].

The literature review of Section 6.2.1 of Chapter 6 has been published in [Edmonds, Candy, Fell, Knott, Pauletto, Weakley, 2003]

The description of the Interactive Sonification Toolkit developed in this research project, here described in Chapter 6, has been published in [Pauletto & Hunt, 2004].

The results of experiment one have been published in [Pauletto & Hunt, 2005].

The results of experiment two have been published in [Pauletto & Hunt, 2006].

The results of experiment two have been published in [Pauletto & Hunt, 2007].

This thesis only exploits those parts of collaborative papers that are directly attributable to the author.

"Ogni manifestazione della nostra vita è accompagnata dal rumore. Il rumore è dunque familiare al nostro orecchio, ed ha il potere di richiamarci immediatamente alla vita stessa"
Luigi Russolo

"Every manifestation of our life is accompanied by noise. Noise is therefore familiar to our ear, and has the power to remind us immediately of life itself"
Luigi Russolo

L'arte dei rumori. Manifesto futurista. Milano, 11 Marzo 1913
[English translation by Sandra Pauletto]

Chapter 1 - Introduction

1.1 Aims and objectives

Today's computers allow the gathering and storing of large amounts of digital data. These data can be produced by any type of complex system such as an aircraft, the planet (e.g. seismology or climate data) or a human being (e.g. medical data). These data contain important information about the systems that produced them and this information can improve our knowledge of the systems. Essential to this process is the ability to extract meaningful information from large data sets. The sheer amount of data produced, however, can be imagined as a lock on a treasure chest that does not open easily. In fact, there is no simple way to analyse such large data sets and the precious information, therefore, seems to hide away at the bottom of the chest behind the lock. Various methods are used to analyse data and all of them allow us to "see" particular characteristics of the data. Statistical methods calculate the probability of the presence of specific data in the context of the overall data set. Data mining techniques highlight the presence of patterns. Visualisation techniques exploit human visual perception characteristics, such as our ability to distinguish different colours, shapes and to group similar visual objects together, to display data in such a way that the information contained in it is clearly presented to the analyst.

This research is concerned with the study of a novel method to display large multivariate, time-based data sets. This method aims to exploit the characteristics of human perception of sound to present important information to the analyst. A display that uses sound to present data is referred to as an auditory display (AD) and the particular auditory displays that present data sets through non-speech sound are called sonifications. The display studied here seeks also to exploit the intrinsically interactive nature of sound. Being sound an event that evolves in time, sounds' interesting instances continuously appear and disappear in the listener's perception. A display however needs to be persistent, i.e. it must exist for as much time as the analyst needs to explore and understand it. Sound can persist if someone keeps interacting with it, e.g. keeps playing back parts already elapsed in time. There are many ways of interacting with sound. The action of pressing the 'Play' and 'Stop' button on a

CD player can be considered a simple form of interaction. Playing an instrument, and therefore producing and controlling the output sound, is another example of more complex interaction. The aim of this research is to study whether displays that use sonification and interaction to present data sets can perform in an effective and efficient way. The main objective of this thesis is to extend the knowledge of how sound can be used to display data sets and how such a display can be improved by the addition of user interaction.

1.2 Thesis hypothesis

1.2.1 Statement of hypothesis

The following hypothesis is to be investigated within this thesis:

> *The combination of user interaction and sonification allows the display of multivariate time-based large data sets effectively and efficiently.*

1.2.2 Discussion of hypothesis

The above hypothesis states that data sets containing many variables varying simultaneously (multivariate), each containing thousands of data points (large), which have been sampled in time (time-based), can be displayed effectively. That is, their main characteristics are shown clearly and efficiently, and the information is communicated rapidly, using a display that combines sound and user interaction.

The International Standard Organisation (ISO) defines effectiveness in [ISO 9241-11, 1998] as:

> accuracy and completeness with which users achieve specified goals [Ibid., p.2]

and efficiency as:

> resources expended in relation to the accuracy and completeness with which users achieve goals [Ibid., p.2]

Guidelines are given on how to measure effectiveness and efficiency.

> Measures of effectiveness relate the goals or subgoals of the user to the accuracy and completeness with which these goals can be achieved [Ibid., p.4]

For example, if a user can achieve the same goals (e.g. recognising data structures) either using a visual interface or an auditory interface, it can be said that the two interfaces are similarly effective. If the aim is to establish if one interface is more effective than the other, then a typical measure of effectiveness would be to compare the number of mistakes a user commits in performing a task with the first interface and then with the second [see example in Ibid., p.7].

> Measures of efficiency relate the level of effectiveness achieved to the expenditure of resources. Relevant resources can include mental or physical effort, time, material or financial cost [Ibid., p.7]

This research focuses on evaluating if the addition of a certain level of user interaction to an auditory display of a large data set diminishes the expenditure of time in performing a task. Resources such as mental or physical effort have not been of primary consideration when using the interaction methods studied here. For this reason the main measure used to assess efficiency in this research is the length of time users spend when performing a given task using different interaction methods.

Effectiveness and efficiency are two of the three elements that need to be evaluated when measuring the usability of a product. Usability is defined as:

> [the] extent to which a product can be used by specified users to achieve specified goals [Ibid., p. 2]

The third element is user satisfaction which is defined as:

> freedom from discomfort, and positive attitudes towards the use of the product [Ibid., p.2]

User satisfaction can be measured using subjective ratings on scales such as how much they like the product, by asking users to express free comments on how they found the experience of using the product, by measuring the users' cognitive overloading, etc. User satisfaction is not fully explored in this research. However, the subjective results presented here, particularly in the third experiment, presented here do give some information about user satisfaction.

Sonification is the use of non-speech sound to convey information [Kramer, 1994]. In particular data sonification is

> [...] the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation. [Kramer et al., 1999, p. 3; in Neuhoff, 2004]

Research in data sonification is relatively recent, the first international conference dedicated to auditory displays was established in 1992, only fifteen years ago [ICAD], and, even if there are examples of used and well known sonifications such as the Geiger counter, standardised sonification techniques have yet to become well established. Some design principles were derived by reviewing the research of the last decade [Flowers, 2005], however these principles apply mainly to relatively small data sets. More evaluations are needed to advance the knowledge in this field. In particular, evaluating large data sets' sonifications present complex and varied challenges which are specific to the fact that these data sets are large. In order to explain what is meant in this thesis by a 'large' data set, it is useful to refer to the following concepts proposed by De Campo in [2007]. De Campo proposes a *Sonification Design Space Map* to help sonification designers to decide which sonification method can be meaningfully applied to a data set on the basis of the data set dimensions and perceptual concepts. First of all, De Campo defines two useful categories of sonification techniques: Continuous Data Representation and Discrete Point Data Representation.

> "Continuous Data Representation treats data as quasi-analog continuous signals, and relies on two preconditions: equal distances along at least one dimension, typically time and/or space; and sufficient (spatial or temporal) sampling rate, so that interpolation between data points is meaningful. Both simple audification [play back of data points, appropriately scaled, as audio samples] and parameter mapping [using the data points, appropriately scaled, to modulate a sound parameter] onto continuous sounds belong in this category. [...]
> Discrete Point Data Representation creates individual [sound] events for every data point." [Ibid. p. 342]

Another useful consideration in sonification design is to determine the time scale to be used. The listener needs to perceive the data characteristics sought to be displayed as identifiable sound events. De Campo proposes to consider the duration of echoic memory (1- 3 seconds) [Snyder, 2000] as the time frame for an identifiable audio event to be perceived. Echoic memory refers to our ability to 'hear', for a few moments after hearing a sound, a trace of that sound in our mind's ear [Levitin, 1999]. In this context the number of data points estimated to be involved in representing one data characteristic as an identifiable audio event becomes a dimension of the Sonification Design Space Map. De Campo suggests that the Discrete Point Data Representation zone ranges between 1 and 1000 data points per time frame (the minimum time frame is 1 second) and the Continuous Data representation ranges between 1000 and 100,000 data points per time frame. These are only suggested ranges as there is no clear cut-off between these two zones. In this thesis a data set is considered large if it has so many data points that Discrete Point Data Representation is not a feasible choice for sonification. Therefore a large data set will normally have a number of points in the order of thousands. On the other hand a small data set is a data set for which Discrete Point Data Representation is a feasible choice for sonification. There are clearly established principles to solve the problem of sonifying simultaneously three data streams with ten samples each, i.e. a small data set. A particular sound timbre and pitch range can be mapped to each data stream. A relationship can be established between the samples values and the chosen pitch scale so that the highest sample value corresponds to the highest pitch and vice-versa. The three streams can then be clearly separated in the stereo field (left, centre and right). Finally an appropriate, slow playback speed can easily be determined that allows a listener to distinguish the three streams and understand their relative relationships. This type of set up exploits what we know about how our perception groups or separates sounds and streams of sounds. This set up, however, would not work well if applied to a data set with thirty streams with one hundred thousand samples each, i.e. a large data set, because our perception would not be able to make sense of so many audio streams played simultaneously and our concentration would struggle to listen to the duration of this sonification played at slow speed. To create a sonification of reasonable duration, the data will have to be updated in the sonification algorithm at a rate that will not allow the perception of a discrete sound event for each data point. Instead, due to the high update rate of the data, complex sounds displaying timbral changes will be created and our understanding of these sounds is very different from our understanding of streams of discrete sound events. This shows that what can be valid for a small data set does not necessarily apply to a large data set. Sonification of large data sets presents specific problems and their evaluation presents specific challenges. This research sets out to find valuable information about a complex problem without reducing it to a simpler case because the knowledge derived from a simpler case may not be applicable to the more complex one. Many questions arise when preparing to study such a complex problem. For example: which data sets should be used in such research? Which sonification techniques should be used? What should the sonification methods be evaluated against? Which methods of interactions, to be used in conjunction with such sonification algorithms, should be used? Decisions have been made in this research to answer these questions and set up meaningful experiments. This thesis presents three experiments and the combination of their results produces new knowledge that contributes to prove the research hypothesis. The data sets used in the first two experiments come from real systems and from fields where data analysis is a known problem. The first data set comes from the field of helicopter engineering. A helicopter is a very complex dynamic system which needs continuous parameter adjustment to remain in stable equilibrium during flight. Engineers record large amounts of data while the helicopter is flying and then analyse it in order to understand better how different parameters interact with each other and how the system can be made easier and more secure to control. The second data set comes from the field of physiotherapy. Researchers in this field attempt to learn more about muscle activation during an action by detecting the muscle firings using electrodes attached to the skin. These data are believed to contain a lot of information which have not yet been extracted using traditional signal processing techniques. These two data sets can be considered as two case studies from fields in which the need for new analysis methods is highly desired because traditional analysis methods have so far shown limitations in extracting information from the data. Two different sonification techniques were applied to these data sets. They were chosen because they are general (applicable to any data set) and because they produce sounds that, even if complex, can be interpreted relatively easily because they exploit the innate ability of our auditory system to perceive structures such as noises and frequencies as separate sonic elements. Each of these sonification techniques was evaluated against a traditional analysis method used to analyse the original data. In the first experiment the sonification of helicopter flight data was evaluated against a corresponding visual display. In the second experiment the sonification of physiotherapy data was evaluated by verifying that information, known to be present in the data and extracted using standard techniques, was effectively displayed. Finally, in the third experiment, three ways of interacting with a sonification algorithm were compared. These three methods differ from each other by allowing different degrees of control during the interaction. In this experiment, in order to make the interaction methods the only variable, the data set was synthesised, and therefore was known, by the researcher and the sonification algorithm was simple and well understood by the test subjects so that also the interpretation of the sonification was not a variable in itself.

These experiments tell us:

1) whether the sonification of large multivariate, time-based, data sets is effective in displaying information about the data sets when compared to traditional data analysis methods;
2) whether being allowed to explore a known sonification with a high degree of interaction control improves the effectiveness and efficiency of the sonification display.

The combination of the results from these experiments contributes to answer the issues posed by the hypothesis stated in this thesis.

1.3 Thesis structure

The following section briefly explains the structure of this thesis introducing the content of each chapter sequentially.

Chapter 2 reviews the current state of auditory display research. First the advantages and disadvantages of using sound to display information are explained. Then the different types of auditory displays are reviewed. Finally the review focuses on data sonification and the recent research literature.

Chapter 3 begins by introducing the need for multimodal interfaces and the issues surrounding their use and evaluation. Then the concept of interactive sonification (the combination of sonification and interaction) is described and the literature on the subject reviewed.

Chapter 4 introduces the main perceptual parameters of sound such as pitch, loudness and timbre. General perceptual mechanisms are described and their effects in the auditory domain summarised. Finally the implications of the existence of these perceptual mechanisms for the domain of auditory displays are discussed.

Chapter 5 describes the research methodology used in this research project and summarises the main experimental methods and statistical tests used in the research.

Chapter 6 describes the software application (Interactive Sonification Toolkit) that was programmed as part of this research in order to experiment with different sonification algorithms and to create an application prototype that could be used in the experiment about interaction.

Chapter 7 describes the first experiment of this research. First the issues surrounding the analysis of large data sets in the field of helicopter engineering is presented. Then the experiment conducted is detailed. Finally the results of this experiment are analysed and discussed.

Chapter 8 describes the second experiment of this research. The problems surrounding the analysis of physiotherapy data, electromyography (EMG) data in this project, are explained. A section describing the standard analysis methods used to understand this type of data follows. Then the experiment is detailed and its results presented, analysed and discussed.

Chapter 9 describes the third and final experiment of this research. Firstly a review of how interaction is designed and evaluated in the field of human computer interaction (HCI) is presented. Then the experiment is described in detail and the results analysed and discussed.

Chapter 10 concludes this thesis bringing together the main concepts reviewed in the previous chapters and the results obtained from the experiments conducted during this research. This chapter also shows how the combination of these results address the issues posed by the hypothesis stated in this thesis and contribute to advance the general knowledge on interactive sonification displays.

Chapter 2 - Auditory displays

2.1 Introduction

This chapter introduces the field of auditory displays. It presents the advantages and disadvantages of the use of sound in displays and the contexts in which there is a strong need for such systems. The different types of auditory displays are described paying particular attention to the field of sonification research which is the focus of the hypothesis explored in this thesis.

2.2 Auditory displays: applications and research

2.2.1 Eyes and ears

When sound is introduced into a human-machine interface, the result is called auditory display. Research on auditory displays is recent and a lot of information can be found in the Proceedings of the International Conference on Auditory Display (ICAD) which was first established in 1992 [ICAD].

The majority of today's human-machine interfaces are based on visual representations of information and the use of the computer keyboard and mouse as interface devices. The addition to this standard interface of an auditory display can make the whole interface more efficient and complete. This is possible since the eye and the ear have very different perceptual characteristics and they can complement each other. Eyes and ears are sensitive to different frequency and wavelength ranges. The speed of light and the speed of sound are different. One cannot perceive a ratio between two colours (a colour cannot be said to be 'double' another), while ratios of pitches are perceivable (we can say that one sound is an octave higher than another sound). Notice, though, that shapes can be perceived as one being half of another: e.g. a line half the length of another. Looking at something requires one to turn his/her head towards what he/she wants to look at. Sound, on the other hand, is omni-directional. Eyes can be shut or open, while ears do not have an in-built mechanism to shut down. Since looking is directional and can be easily shut down, it is linked to a conscious decision of wanting to look. Looking requires the person to direct attention to whatever needs to be explored. Hearing is omni-directional, cannot be easily avoided, and is often used to perceive danger effectively. Colour unfolds in space, while sound unfolds in time. However, to look at something requires time (the time the eyes require to explore the entire picture or object), while sound travels in space and is highly influenced by the space where it is produced [Sonnenschein, 2001].

As a consequence of these differences between eyes and ears, we can deduce benefits and problems that can be encountered in designing auditory displays.

A first important distinction to be made in auditory displays is that between displays that use speech sounds and those that use non-speech sounds [Kramer, 1994]. In the first case the display can exploit all the meaning that words have historically and culturally acquired in a language, while in the second case the sounds have no linguistic connotations. This research is only concerned with auditory displays based on non-speech sound: this should be assumed also when it is not explicitly stated.

Kramer in [1994] lists various advantages and disadvantages of using auditory displays. The most relevant points are discussed here. Using auditory displays is of obvious benefit for impaired users. For sighted users, the presence of an auditory display gives the eyes freedom to do other tasks at the same time as hearing the auditory information. Audio is used to convey alarms, as it is omni-directional and, if properly designed (i.e. it is loud enough, has an appropriate frequency, etc.), it can be heard in noisy environments. In a monitoring or exploratory task that uses sound, changes in sound can inform the user of changes in the data and orient him/her on when and where to look for more information on this change. We have an acute temporal perception of sound changing in time. For example, we can hear and distinguish between two different rapid rhythms. There are perceptual mechanisms that allow us to group streams of sounds as belonging together (see Chapter 4 for more detail on auditory perception). This means that we can also segregate between groups of sounds playing together, allowing for the possibility of portraying more than one piece of information at once.

From these considerations it follows that the development of auditory displays is of fundamental importance when designing interfaces for visually impaired users, or when designing interfaces of tools that are used in environments where the user's eyes are busy doing something else (e.g. medical environments). They also are applicable when one needs to portray more than one piece of information, or when designing real-time feedback systems (due to the fast detection of changes in sound) and mobile systems (where the eyes must be free and sound can provide the necessary information, e.g. a mobile phone ring tone). There are also difficulties which have to be considered when designing auditory displays. The auditory system has a low spatial resolution comparing to the visual system. Some sound parameters are not independent of each other, making it difficult to create an equivalent of the vertical and horizontal axes we use widely to represent information visually through graphs. Sound can be pleasing, but can also be annoying. Attention needs to be given to this aspect to create usable displays. Sound can interfere with speech communication, e.g. if we listen to sound on loudspeakers, communication with other colleagues is impaired. Sound is not a persistent display, i.e. because it has its own duration, it does not necessarily present the information for the duration the user needs it to complete the analysis. One way to obviate to this problem is to allow user interaction with the display so that the relevant parts of it can be heard in many ways and many times. Sound cannot be printed out. However in recent years the possibility of sharing sound files and distributing them on the internet has partially reduced this as a problem. Finally, users can have auditory limitations such as not being able to hear the whole of the audible frequency range, without necessarily being aware of it.

Among the problems mentioned above, probably the most relevant and the one that makes the auditory domain very different from the visual one, is the absence of persistence. One way to obviate to the lack of persistence is to introduce some form of user interaction, i.e. the user can playback the sound many times and this action creates persistence. This research looks at adding interaction to an auditory display as a way to introduce persistence.

2.2.2 Different types of auditory displays

A generic auditory display can be divided in three parts [Kramer, 1994] as follows:

The three parts of a generic auditory display		
Information generators	**Communicative medium**	**Information receiver**
databases or real-time data generators	which includes: a) data receiving means, b) intermediary structures such as the means of mapping the data to sound, c) sound generating means	the listener

Table 2.1: Table summarised from Kramer, 1994 by Sandra Pauletto

Auditory displays can also be divided into different types. A fundamental distinction among auditory displays is between a symbolic representation and an analogic representation [Kramer, 1994]. A symbolic representation categorically denotes the thing being represented (e.g. an alarm sound), while the analogic representation directly displays relationships (e.g. a Geiger counter, where each sound event directly represents a radiation event) [Ibid.]. It is important to notice that the symbolic and analogic representations are the extremes of a continuum. Auditory displays can in fact have both symbolic and analogic characteristics. For example, a sonification based on the mappings of data to sound parameters can be considered to be both analogic and symbolic. It is analogic because it directly represents the data relationships, and it has symbolic characteristics because, after having learnt how the sonification works, the user can perceive a particular sound in the sonification as the signal that a certain data configuration is present in the data.

There are many types of auditory display, some very common and some slowly becoming important. Examples of well known auditory displays are: the stethoscope, the Geiger counter and the sonar, together with the auditory icons of our desktops, the ring tones in our mobile phones, and the sounds of computer games. These can all be considered auditory displays which are either analogic or symbolic, or a mixture of the two.

Auditory displays can be roughly divided into five types [Walker and Kramer, 2004]:

1) *Alerts, notifications, warnings*: examples are police sirens, fire alarms, traditional telephone ring, etc. These displays are simple and effective in portraying a status. They contain only a small amount of information.

2) *Auditory icons*: these are generally the auditory equivalent of visual icons or actions to be performed in a computer. Gaver [1986] defines an auditory icon as "a sound that provides information about an event that represents desired data" [Ibid., p. 168]. Therefore the sound of the auditory icon evokes the visual object and/or action it represents: e.g. the auditory icon for the recycle bin of the computer is the sound of paper being crumpled up.
3) *Earcons*: these are audio messages used in the user-computer interface to provide information and feedback to the user about computer entities such as messages, functions, states and labels [Blattner *et al.*, 1989]. Earcons usually portray more than one piece of information, employ a simple musical language to portray the information and the relationship between the earcon. The entity to be displayed is metaphorical at most [Walker and Kramer, 2004]. Various actions and statuses in the computer interface do not have a real world correspondent and require more than one piece of information to be portrayed. For example, to represent the "minimise" and "maximise" button of a program window one type of sound can be used. This sound (earcon) will have two versions: one representing the "minimise" button (e.g. a two note scale descending in pitch) and one representing the "maximise" button (e.g. a two note scale ascending in pitch).
4) *Audification*: this refers to the direct translation of a data waveform into sound [Kramer, 1994]. For example, large amounts of data that present a high degree of variation can be directly transformed into sound: the data will usually need to be scaled, time compressed and possibly pitch shifted to be "transported" into the audible range.
5) *Sonification*: this refers to "the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation." [Kramer *et al.*, 1999, p. 3; in Neuhoff, 2004]. In this case the data to be represented is mapped onto sound parameters (e.g. pitch, loudness, etc.). The resulting sound represents the structures present in the data.

Audification can be considered a special case of sonification. However, while any type of sonification requires a "translation" stage to be able to interpret the data displayed by the sound (i.e. we need to know how the data was mapped onto the sound parameters), in audification particular events in the sound directly signal the data structure represented. For example, if in audification we hear a frequency we immediately know that there must be a data structure that repeats at high rate (audible frequency) in the sonified data set. For this reason, the audification technique is often listed separately from sonification in literature. Finally, an introduction to auditory displays and an account of their historical development can be found in the book edited by Gregory Kramer (founder of the International Conference of Auditory Displays, ICAD) *Auditory Display Sonification, Audification and Auditory Interfaces* [Kramer, 1994].

2.3 A particular auditory display: sonification

This research focuses only on the particular types of auditory display which map directly data to sound parameters in order to display information. These auditory displays are often referred to, in literature, as sonifications and are highly analogic. Among all the different types of auditory display, sonifications are among the least used (apart from a few famous cases such as the Geiger counter) despite being rich in information. One possible reason for this situation is that research into sonification techniques is relatively recent and general guidelines have not yet been identified.

In sonification literature, three main types of research are present:

- research describing the implementation of systems that allow experimentation with various data-to-sound mappings (such systems are usually software-based and are called toolkits);
- research describing particular sonification algorithms, usually developed with a specific type of data set in mind;
- research describing evaluations of sonifications.

2.3.1 The toolkits

There are many applications that allow visualisation of data whatever the source and nature of the data, examples being Excel and Matlab. In these applications data can be very quickly displayed in a standardised way and a quick overview of the data is easily obtained. These applications provide also the possibility of displaying the data in various ways so that the user can choose the best one depending on the data. To date, the domain of sonification does not have an application equivalent to Excel or Matlab. Auditory display research is not as advanced as visualisation research and therefore standard techniques, popular and well understood by everybody, are not yet established. Since current analysis packages do not have sonification facilities, many researchers feel the need firstly to create their own sonification environment in order to experiment with different mappings and techniques. The need for such environments is well expressed in [Ben-Tal *et al.*, 2002] when the authors describe their toolkit:

> [The Toolkit] will provide researchers with the means of exploring parameter mapping with the same high-level control afforded by many data visualization packages. [Ibid., p.1]

Various toolkits have been presented at the ICAD Conference since the conference was established and they normally have a similar architecture and structure. In particular, all the toolkits provide:

- a way to read data sets and scale them;
- a method for assigning data to sound parameters (these can either be low level sound synthesis parameters or higher level musical parameters);
- a way to produce sound and therefore the sonification;
- some toolkits allow ways of interacting with the sonification, e.g. [Bruce & Palmer, 2005].

All toolkits also attempt to be flexible, extendible and portable. The first toolkits to be presented at the ICAD Conference were *Personify* [Barrass, 1995], *Listen* [Wilson and Lodha, 1996], and *MUSE* [Lodha *et al.*, 1997]. Already in these early toolkits, there is an attempt to make them interactive, quick to use, flexible in their options of sonifications, extendible with new sonifications, and portable. Later toolkits also try to have these characteristics while improving the availability of sonification techniques. In particular, toolkits tend to divide into those that attempt to create more musical sonifications, allowing the user to map high level parameters of sound (e.g. tempo, timbre, etc.), which also tend to map data to MIDI data because they are considered higher level musical parameters, and toolkits that map data directly onto low level sound synthesis parameters, which are part of some synthesis algorithm. It could be said that in the first type of toolkit the sonification happens at a musical score level, whilst in the second it happens at a sound/timbre creation level.

The MUSE toolkit [Lodha *et al.*, 1997], the Musart [Joseph & Lodha, 2002], the Sonification Sandbox [Walker & Cothran, 2003] and TrioSon [Cullen & Coyle, 2005] are examples of the first type of toolkit, while SonART [Ben-Tal *et al.*, 2002], and SIFT [Bruce & Palmer, 2005] are examples of the second type.

Various programming environments have been used to program sonification toolkits, e.g. C++, Csound [CSound], Java, Pure Data (PD) [Pure Data programming environment], etc. and it is probably true that each of these programming environments are best at implementing some parts of the toolkits (e.g. CSound for synthesis, PD for interaction, Java for the Graphical User Interface, etc.). In the author's opinion, after the experience of building a toolkit for sonification (see Chapter 6 of this thesis for its description), environments such as PD and CSound are in themselves already good environments in which to experiment with various sonifications and user interactions. It is not necessary, unless a specific prototype sonification application is requested for user experimentation, to build new sonification environments for experimenting with algorithms. Sonifications algorithms could be created and evaluated in environments such as PD or CSound, etc. and, if found to be appropriate, the particular algorithms could then be added to the already existing and used data analysis environments such as Matlab, Excel, or PD, CSound, depending what is the main application expected for the sonification. Examples of sonifications being added to data processing environments are described in [Miele, 2003; SKDTools], in which the implementation of a Sonification Toolkit for Matlab is described, the addition of sonification functions in the data mining environment Neo/NST [Neo/NST] by Thomas Hermann and the addition of sonification functions for spreadsheets in [Stockman, 2004; Stockman *et al.*, 2005].

The architecture of the Interactive Sonification Toolkit described in Chapter 6 derives from the study of this literature about previous toolkits. The Interactive Sonification Toolkit provides a data scaling section, a mapping section, a sound generating section and allows interaction with the sonification created. In this toolkit the mapping happens at a sound/timbre level as this was considered more appropriate for the sonification of large data sets. However, the toolkit could be extended to include a MIDI section. The most important use of the

Toolkit in this research has been as a sonification application prototype for experiment three (see Chapter 9) on interaction.

2.3.2 Sonification algorithms

The field of sonification is relatively new and, as mentioned above, not many basic sonification principles have been clearly established yet. However, there are a few data to sound mappings that have been shown, in the short history of the discipline, to work better than others in portraying data streams. At the Limerick Symposium on Auditory Graphs [AGS2005, 2005], Flowers [2005] summarised the mapping approaches that seem to work:

- mapping numerical values to pitch;
- mapping data to sound durations;
- mapping to loudness levels in order to create rhythmic patterns;
- conjunction of loudness and pitch mapping;
- mapping different streams to different timbres to facilitate streams segregation;
- maintaining time relationships when sonifying time-based data sets;
- sequential comparison of short data streams (as supposed to simultaneous comparison).

These approaches are successful, in particular when working with relatively small data sets. They have been used and evaluated with various types of data sets by researchers working on auditory graphs, earcons and on general sonification projects.

Here follow examples of studies that support Flowers' summary of approaches. Mansur *et al.* [1985] compared how blind users and sighted users understand sound graphs and tactile graphs. The sound graphs were created by mapping the *y* coordinate of the graph to the pitch of a sine wave (the pitch changed continuously), while the tactile graphs were made by tracing the lines with a pen over special plastic sheets which left behind rough lines that could be felt by the fingers. The task involved recognising mathematical characteristics such as differences in slopes, straight lines and exponentials, etc. The results showed that: there was no difference in accuracy between sighted and blind users; that the accuracy was slightly better using tactile graphs, but overall accuracy was in both cases higher than 83%; and finally that exploration using sound graphs was significantly faster. This final result was considered very encouraging with regard to the idea of using sound to display line graphs, especially for blind users. More examples of the use of the mapping approaches listed above can be found in Brewster's research on earcons and auditory graphs starting from his PhD work [Brewster, 1994] up to his current work with colleagues [e.g. Kildal & Brewster, 2006]. In Brewster's guidelines for earcons [Brewster, 1994] timbre, pitch, rhythm, duration, loudness and spatial location are the main parameters used and the programmer is advised to use them in such a way as to improve either the grouping of the same family of earcons or the segregation of the different families of earcons (e.g. the same family of earcons should have same timbre and register, while a different family of earcons can be separated by having different spatial location). The work of Brewster and colleagues is supported by many evaluation studies. Keller *et al.* [2003] presented a sonification of gamma ray data from Mars in which the principal sonification parameters were timbre, tempo, pitch and volume. This study presented also an evaluation of sonification as an educational aid. Peres and Lane [2003] presented a sonification of statistical graphs that used pitch, panning and timing as sound parameters. In this project different mapping strategies are compared and the condition that uses timing results to be the best. Malandrino *et al.* [2003] presented a sonification used to monitor a computer network system in which various MIDI tracks, using different timbres, are associated with events happening in the network. Janata and Childs [2004] presented a sonification of real-time financial data that uses pitch, timbre, stereo location, tempo and volume as mapping parameters. Evaluation of the auditory display showed that it augmented data monitoring accuracy. Zhao *et al.* [2005] presented the sonification of geographic statistical data to improve the access of blind users to these data. Mapping parameters were again pitch, timbre, localisation and tempo. A pilot study shows that the display allows recognition of the information. Hinterberger and Baier [2005] presented an auditory feedback of electroencephalography (EEG) data that mapped reduced data (by pre-processing) to timbre, pitch, volume and stereo localisation. The auditory feedback allowed for acquisition of self-regulatory skills of brain activity. Stockman presented in 2004 and 2005 [Stockman, 2004; Stockman, Hind & Frauenberger, 2005] a sonification of spreadsheets that mainly used pitch and timbre as mapping parameters. Baier *et al.* [2006] presented a two-channel sonification of EEG data of epileptic patients. Maxima present in the channels are mapped to an event produced by a synthesised sound. The volume of the event is proportional to the distance between the maximum and the previous minimum in the channel. The duration of the event is proportional to the time between the maximum and the previous maximum. The number of harmonics is proportional to the time between the maximum of the current time series and the previous maximum in the second time series. The two channels are panned hard left and right to facilitate separation. More examples of recent research on these sonification methods can be found in the papers of the International Symposium on Auditory Graphs 2005 [AGS2005, 2005].

The approaches summarised by Flowers, and described in the examples above, are based on human perception of sound characteristics such as our relatively high resolution for pitch (if compared to other sound parameters), high resolution for temporal patterns, stream segregation, high memory for sequences, etc., and work well in particular when data updates relatively slow. This survey highlighted various mapping ideas that have been used in this research. These ideas are: maintaining the sequential relationship between samples; comparing channels by presenting them sequentially; helping channel segregation by assigning a different frequency range to each channel or by spatial separation; mapping data to amplitude; and differentiating data structures by exploiting timbre segregation. In the sonification of the helicopter data (Chapter 7) the time scale is highly sped up, but the sequential relationship between data samples is maintained. This allowed a very short sound sample for each channel (2.5 seconds) to be created, which can then allow sequential comparisons of the different channels. In the sonification of EMG data (Chapter 8), the time relationship between samples is maintained, as one aim of this sonification is to create a real-time display of the data. Each channel of data is mapped to a sine oscillator having a different frequency. Specifically, the data samples are mapped to the amplitude of the sine and various data structures can be heard as changes in the overall timbre of the sonification. In Experiment 3 on interaction (Chapter 9), the first channel of the data set is mapped to the left audio channel and the second data channel is mapped to the right. This was done to help segregation of the two channels.

As already mentioned, when the data's update rate becomes close to the audio frequency range, then mapping the data directly onto a sound parameter normally creates overall timbral changes, instead of separable parameter changes. For instance, mapping data that change at audio rate to the frequency of an oscillator will create a 'frequency modulation' sound with a complex timbre. When data updates so quickly, in order to design a meaningful display, we need to have a good knowledge of the timbral and perceptual effects created by the particular synthesis algorithm used. This knowledge will also be employed to interpret the finished display. A very effective sonfication algorithm used when data update at audio rate is audification [Hayward, 1994; Dombois, 2001; Dombois, 2002; Hermann *et al.*, 2002; Baier *et al.*, 2005]. In audification, the data samples (appropriately scaled, time compressed and, if necessary, frequency shifted) are directly heard as audio. The resulting sound of an audification of complex data has a complex timbre in which certain structures of the data such as noise, periodicities, etc. are perceived very clearly as noise, frequencies, rhythmic patterns, etc. The separation of these data structures does not require a great conscious effort, because they exploit our auditory system and brain's innate ability to separate noise, frequencies, rhythmic patterns, etc. This approach was adopted here for the sonification of a helicopter's flight data. This allowed the creation of a very fast and very general display of each data channel, which exploits only the innate abilities of the auditory system to segregate the different data structures.

Other mapping strategies are being developed in which a strong interest is shown towards the manipulation of the evolution of timbral characteristics of sound, instead of parameters such as pitch, loudness, etc., that can be perceived as separate. An example of how particular changes in timbral characteristics of the same source sound can result as producing very separate sounds are vowel sounds. Vowel sound, such as 'a', 'o', 'u', can be produced from the same sound source (a band-limited pulse) just by varying the energy peaks (formants) in its spectrum. In literature there are examples of sonifications using formant synthesis [Cassidy *et al.*, 2004; Fox & Carlile, 2005; Hermann *et al.*, 2006], AM synthesis [Baier *et al.*, 2005], waveguides [Lee *et al.*, 2005], general spectral characteristics [Sturm, 2002; Sturm, 2003] and psychoacoustical models of timbral characteristics such as roughness, sharpness, etc. [Ferguson *et al.*, 2006]. It can be noticed that the sonifications which exploit timbral evolutions in time are normally sonifications of large data sets. This approach was chosen in this research for the sonification of EMG data. The EMG data are sonified using a synthesis technique called amplitude modulation. This technique allows sounds to be created which exhibit timbral changes when the signal modulating the amplitude of an oscillator varies at audio rate. This synthesis technique can also create sounds that are perceived as more or less rough, depending on the variation of the modulating signal. In this choice of sonification the timbral changes represent the information to be displayed.

When mapping data directly onto sound parameters, two different approaches can be found: one in which the amount of data is relatively small and the sonification tries to exploit the segregation abilities of human perception of sound to perceive various channels together, and a second in which the amount of data to be studied is large and the sonification approach is to create an overall evolving timbre which can provide information about the data holistically. The approaches to sonification described so far, are referred to as *parameter mapping*. In this thesis the author will differentiate the parameter mapping approach that leads to a sound with a time evolving timbre by calling it *timbral parameter mapping*.

2.3.3 Model-based sonification

Mapping data directly to sound parameters is not the only possible sonification approach. Generally, it can be said that there are two main approaches to data sonification:

- parameter mapping (mentioned so far);
- model-based sonification.

In parameter mapping the data are appropriately scaled and then directly mapped onto a sound parameter. The understanding of the sonification relies on how well the evolution of the mapped sound parameters can be perceived and/or on the understanding of how the timbre of a particular synthesised sound evolves when the parameters change.

Model-based sonification was first developed by German researcher Thomas Hermann and the approach is explained in *Listen To Your Data: Model-Based Sonification for Data Analysis* [Hermann and Ritter, 1999] or in his PhD thesis [Hermann, 2002]. Sound examples of many of Hermann's sonifications can be found on his web site [Thomas Hermann]. The aim of model-based sonification is to find a solution to the limitations of parameter mapping sonification methods when applied to large data sets. The main idea of model-based sonification is that our ear/brain system is naturally optimised to extract valuable information from sounds that are produced as a consequence of physical processes.

The idea is to consider the data under observation as points in a multidimensional space and to impose physical laws acting between the points in the data space. Then, if the user excites the data points in some ways, they will react to the excitation. The reaction of the points is then mapped onto sound's parameters producing a sonification of the data.

The main difference between this method and parameter mapping is that in parameter mapping the data sets could be thought of as the 'score' of the sonification, while in model-based sonification the data sets become the 'instrument' of the sonification. In model-based sonification there is still data mapped onto sound variables, but the data which are mapped onto sound are not the original raw data anymore, instead they are a manipulation of that data on the basis of the physical laws imposed on the data space. Hermann has produced a large amount of model-based sonifications using a variety of physical relationships between data. In *Principal Curve Sonification* [Hermann, Meinicke and Ritter, 2000], the authors describe a method for sonifying data that are not naturally time-variant. In the areas of data mining and exploratory data analysis, techniques such as Multidimensional Scaling [Cox and Cox, 1994] and Principal Component Analysis (PCA) are used to reduce the dimensions of a data set to the most important dimensions. Principal Curves are continuous one-dimensional trajectories that approximate the data set and pass through the 'middle' of it. The Principal Curve can therefore be chosen as the curve of time in which the sonification of the data set evolves. This means that the Principal Curve is mapped directly onto time. We can imagine a virtual listener walking along the timeline (principal curve) hearing the scenario (data). The action of the virtual listener moving along the timeline becomes the excitation of the data that produces the sound. In this context the distance between the listener and the point source determines the perceived volume of the source sound (high distance – low volume, small distance – high volume). This is a method used in model-based sonification: what is sonified is not the features of the data, but the relationship between the data and the listener who, just by going through the Principal Curve, is exciting the data space. Features of the particular data point can be mapped onto audible parameters such as pitch using the parameter mapping method.

The paper *Markov Chain Monte Carlo Simulations* [Hermann, Hansen and Ritter, 2001] describes an example of model-based sonification applied to the simulation of a Markov chain (stochastic process evolving in discrete time steps). To create the model-based sonification, the N data points of the chain are interpreted as fixed masses in a multidimensional space. The existence of the masses creates a virtual global potential. This space is explored by introducing virtual test particles with a given kinetic energy. What is sonified is the interaction between the test particle and the data space, i.e. the test particle's kinetic energy and trajectory are sonified.

Crystallization Sonification of high-dimensional data sets [Hermann and Ritter, 2002] describes another model-based sonification based on a law that determines crystal growth. The user can select a place in the high-dimensional data-space where the crystal growth starts. As the virtual crystal grows, it includes the data-points present in the space. The sound summarises the temporal evolution of this crystal growth process by mapping evolving parameters of the process onto the amplitude of the partials of a sound created by additive synthesis and the size of the crystal to the overall pitch of the sound. Other, more recent examples of model-based sonifications can be found in [Baier & Hermann, 2004] and [Bovermann, Hermann & Ritter, 2005].

Model-based sonification has also been applied by Eslambolchilar and Murray-Smith to other, very different applications, such as target searching and browsing for mobile devices [Eslambolchilar *et al.*, 2004; Eslambolchilar & Murray-Smith, 2006]. In the first of these examples, the targets are music sound files and the physical model used to sonify the presence and closeness of the target is the Doppler effect. This effect is perceived when a sound source is moving with respect to the listener or vice-versa and influences the pitch of the perceived sound. In this study the Doppler effect is used to give the idea to the user of how close he/she is to the target. In the second example, French sentences have to be identified while browsing a text with a majority of sentences written in English. Here the exploratory task is modelled using the image of a fisheye lens. The behaviour of the lens is modelled like the movement of an elastic ball in water. Various parameters of the sonification (which gives information about the language of the sentences explored) are mapped to the dynamics of this physics model.

This separation between model-based sonification and parameter mapping is quite useful when categorising different approaches, however, in reality, the separation is not so clear cut. As mentioned above, in parameter mapping the choice of associations between data and sound parameters largely depends on the dimensions of the data sets and on the data update rate. Model-based sonification is an interesting approach to the problem of the sonification of large data sets which attempts to eliminate the problem of the quantity of channels to map and attempts to be general, i.e. independent of the data set (however one could argue that a parameter mapping technique such as audification is as independent from the largeness and type of data set as model-based sonification).

It is the opinion of the author that the major advantage of model-based sonification is the use of a high level metaphor (arbitrary physical laws that link the data points) to connect the data together which helps the understanding of the sonification. The major disadvantage of the technique is the manipulation of the data prior to mapping. Often, prior to mapping, the data are manipulated and relationships are first found using standard analysis methods and then displayed with a sonification which, in itself, can result to be quite simple. This is a very different approach from that of mapping the raw, unprocessed data and trying, by exploiting sound perception properties, to display correlations and/or properties present in the raw data just using the sound. The research presented in this thesis is based on the second of the above approaches: i.e. the correlations and data structures should emerge from our perception of the sonification, not be calculated before mapping and then be displayed with sound. In Chapter 7 of this thesis, an example of model-based sonification of the helicopter flight data is described. Although model-based sonification is not the focus of this research, this example is described so that the differences between the parameter mapping and model-based approaches are clearer.

2.3.4 The evaluation of sonifications

Sonification is a relatively new discipline and therefore evaluation is important to create a large database of experimental results from which fundamental and general trends, useful for the constructions of design guidelines, can emerge. In the last few years various evaluation studies have been done especially in the fields of auditory graphs and earcons [see a good summary in Bonebright, 2005]. One difficulty for any evaluation of this kind is the extent to which any study can be considered general. In a sonification study there are at least three elements that tend to be very specific:

- the nature of the data to be sonified;
- the sonification technique;
- the task of the experiment.

While the second and third element can be made general in some experiments, the first element can never be made completely general because there is no way of producing a truly general data set.

Some types of data sets and sonifications have been evaluated more often than others. Various experiments of displays created specifically for blind people have been evaluated and there is a large number of studies for the sonification of relatively small data sets (auditory graphs and earcons). However, there is less evaluation of sonifications of large data sets. As mentioned above, the sonifications of large and complex data sets tend to use timbral parameter mapping. Creating experiments to evaluate how well these displays show the data characteristics, i.e. how well we can hear, in this case, changes in timbral characteristics, is a difficult task. Timbre is a sound parameter not yet well understood and is also multidimensional. Furthermore the dimensions of timbre have not yet been identified therefore there are no clear scales to vary and labels for the timbre's dimensions to use in experiments. For example, we can vary a spectral characteristic of a sound, but we do not know how to name (label) the relative timbral change and how to produce equally spaced perceptual steps in that timbral dimension. Moreover, there are fewer examples of large data set sonifications. Another field where the evaluation of sonification is sparse is when the sonification involves interaction.

Depending on what the sonification is used for, evaluations use different techniques:

a) sonifications are compared to an equivalent, normally visual, displays [e.g. Fitch & Kramer, 1994];
b) different sonifications of the same data are compared [e.g. Peres & Lane, 2003].

Here follows a summary of recent examples of evaluations, first in the domain of auditory graphs (small data sets), then in the domain of large data sets and auditory display with interaction.

In *Smith-Kettlewell display tools: a sonification toolkit for Matlab* [Miele, 2003; SDKTools] the development of tools that allow blind users to operate the software Matlab is described. SKDtools implements two main methods of data sonification: a Continuous Mode, which modulates a single sinusoidal tone in proportion to the input value, and a Discrete Mode, which uses a set of fixed-frequency sinusoids presenting only the tone or tones nearest to the corresponding output frequency. From the conducted experiments the first method seems to give more easily qualitative information, while the second allows the portrayal of information in a more quantitative fashion because comparisons between tones are used to indicate the values of data with respect to one another. From this study we learn that continuous sonification seems to be better at portraying qualitative, overall information. Similarly, in this thesis, continuous sonifications were created to portray overall information about data.

In *Drawing by ear: interpreting sonified line graphs* [Brown and Brewster, 2003], the authors conduct experiments on how line graphs, containing two data series, can be best sonified. The experiments aimed to discover the level of accuracy with which sighted people were able to draw sketches of the graphs after listening to them. It aimed to identify any differences in performance when the graphs were presented using different combinations of instruments, i.e. either with piano representing two data sets (same-instrument condition) or with a piano and a trumpet representing the two data sets (different instrument condition). High accuracy was found and not much difference between the two conditions. It was also found that using the same instrument makes it easier to find correlations between the data streams (because the two sounds become one when data coincide), while using two instruments makes it easier to identify the features of each line. From this experiment it is clear that to sonify different channels using the same set-up allows more similarities between the sounds to be more easily identified, which can be an indication of similarities in the original data. This idea is used here in the sonification of the helicopter flight data where each channel is sonified separately, but using the same mapping. Two channels can be compared by listening to them in sequence (they are only 2.5 seconds long) and by listening for similarities and differences in the sound.

In *Sonification of statistical graphs* [Peres and Lane, 2003], two experiments are described that compared the effectiveness of different parameters of sound for the auditory representation of plots. The data sets were either mapped to pitch, panning or the timing between sound's onsets. Temporal mapping (the last described above) was found to be better than pitch or panning mapping. The redundant condition, using both pitch and panning, was preferred by the subjects, but did not produce better experimental results. Subjects' performance did not improve in time as much as expected. This study shows that the subjects' subjective perception of a display can be different from the objective measurement of the effectiveness of the same display. This highlights the importance of recording user satisfaction when evaluating a display. In experiment three of this thesis a similar situation emerges where two interaction methods perform similarly in terms of objective measurements of effectiveness and efficiency, but the majority of the users report a preference of one method over the other.

In [Walker, Kramer & Lane, 2000] the authors designed experiments to attempt to answer three general questions that arise when designing auditory displays. The questions were:

- which sound parameter should be used to represent a certain data dimension (e.g. temperature/pitch or temperature/loudness?);
- what is the right polarity of the sound parameter in relation to the data dimension. In this context the term polarity refers to how the sound parameter's scale and the scale of the variable to be represented should be associated with each other. For example, should the weight of an item be associated with a high pitched sound because the numerical value of the weight is high, or should it be associated with a low pitched sound because heavy items normally produce lower pitch sounds than lighter objects?
- what is the correct scaling factor between data and sound parameter.

The experiment answering the first question was found not to be informative enough. In the second experiment subjects were presented which sounds with different pitches randomised and they were asked to relate one sound with a certain data parameter value and then assign the other data values to the rest of the sounds. It was verified that different data parameters have different polarities (e.g. high temperature/high pitch while high weight/low pitch) and that sound parameters perceptual scalings should be used to create clear auditory displays. This study highlights the importance of the relationship between the sound created to portray specific information and the meaning of that information. Particular attention is given to this aspect in the second experiment of this thesis, where the subjects are asked to rate the appropriateness of the sonification produced as an "audio image" or "audio metaphor" of the data represented.

In [Janata and Childs, 2004] a sonification of financial data is explained and evaluated. Two experiments were run in which a monitoring task was involved and a comparison between the use of a visual display on its own or with audio feedback. In the first experiment the subjects had to say when the data points were changing direction in the two conditions of the visual display with and without audio feedback. In the second a visual task was added while users had to continue monitoring direction change. The auditory feedback was found to improve the ability of performing the monitoring task in particular when a second visual task was added to the original task. This experiment shows that the presence of real-time auditory feedback can improve the monitoring of data when the eyes are busy doing other tasks. This supports the idea that the creation of a real-time auditory feedback for EMG data can improve the current visual monitoring of the signal.

Bonebright [2001] studied whether the presentation of sonified graphs to students increased the comprehension of graphs and if stereo presentation was better than mono presentation. Subjects were presented with the sonified graph and were asked to match them to the respective visual representations. Pitch and timbre were the main sound parameters used for the auditory graphs. The subjects matched the auditory graphs with high accuracy and the mono presented sonified graphs worked better and took less time to understand. This is an example of a study which focuses on the comparison between the effectiveness of an auditory display and an equivalent visual display. A similar comparison is done in this thesis in experiment one where the information portrayed by the audifications of the data is compared to the information portrayed by the spectra of the data.

Brown *et al.* [2006] describe the evaluation of the use of audio glances (kind of earcons) to give an impression of size, complexity and topology of abstract graphs. The graphs are a series of lines that meet at nodes. The sonification scans the graph from left to right and plays a tone when a node is encountered. The vertical axis is mapped to panning so that the higher the node the more left the panning and vice versa. Sighted subjects where asked to match the auditory graph with the visual correspondent. The auditory glances were found to be working well. This is again an example of a study which focuses on the comparison between the effectiveness of an auditory display and an equivalent visual display.

Examples of evaluation of auditory displays with interaction and sonification of large data sets follow here. The following three studies focus on the use of sonification to provide real-time auditory feedback to help self-regulatory abilities. Similarly, the EMG sonification presented in Chapter 8 of this thesis also aims to be used as a real-time auditory feedback of muscle activity. Hinterberger & Baier [2005] presented a sonification of EEG data in which the data, after some processing, were mapped to different timbres. Durations were also used as a parameter in the sonification. An experiment evaluated the sonification as a self-regulatory audio feedback. Subjects were asked to focus on particular sounds and then task dependent variations where analysed. The results showed that participants revealed self-regulatory abilities as significant changes in amplitude of the EEG signals were recorded. Effenberg [2005] presented a sonification of movement and in particular of countermovement jumps (CMJs). Visual only, auditory only and audio visual feedback methods were compared by asking the subjects to judge the difference between two CMJs in each of the three conditions. The audiovisual display came out best. In a second experiment, subjects were asked to reproduce a CMJ after having experienced it using either the visual only or the audiovisual display. The audiovisual condition was shown to be the best feedback method. In *An auditory display system for aiding interjoint coordination* [Ghez *et al.*, 2000], a real-time auditory feedback of joints and muscle motion is described. The goal was to create an effective auditory feedback system for patients who are unable to maintain "internal models" of their limbs and therefore monitor their limbs' movements. The auditory display used MIDI sounds and the environment used included a data acquisition system and a program implemented in Max/MSP. A first approach aimed at creating an auditory feedback for the hand position. In this case, the subject used the mouse to control pitch and amplitude of a sound (where the pitch was mapped to the vertical axis and the amplitude to the horizontal axis). The task was to match regularly appearing target tones, and therefore to match target hand positions. The results from this exercise showed that target matching tasks in space through auditory feedback could be learned, but target matching through vision was much more accurate and involved much less training. Two other experiments were run to verify if the auditory display could portray the timing structure of the movement. Rhythm and timing of the sonification would provide information about the timing of the movement, while the melody of the sonification would provide spatial information for the movement. The authors conclude that the most important cues the sonification is able to provide are about the timing of the action and that the melodic contour seems to be less important. In all these examples, the presence of audio helps self-regulatory abilities. These are encouraging results which support the idea of creating EMG auditory feedback.

The following two studies evaluate the usability of interactive navigation methods in sonification. Zhao *et al.* [2005] presented the sonification of geographical data of the United States. A virtual map was created by mapping data/sounds corresponding to east states to the left of the spatial domain and data/sounds corresponding to the west states to the right. Pitch and timbre were the principal parameters used in the sonification. Various ways of navigating the map, using the keyboard as the interface device, were also compared. The main results were that the sonification allowed understanding of the data and that vertical navigation (as supposed to diagonal) was found to be the easiest. Subjects also preferred to hear entire columns at once, instead of proceeding data point by data point. This method required them to input less keystrokes while obtaining a reasonable amount of information with each one.

Kildal and Brewster [2006] evaluated a sonification that allows the exploration, through a tablet interface device (i.e. a 2D interface) of large matrices of data using pitch. The display was found to be effective and efficient and independent of the size of the data set at least for the dimensions tested in the study. Also in this study, entire columns and rows could be heard by performing only one action. The experiment described in Chapter 9 of this thesis compares the usability of three interactive navigation methods to be used with sonification. The two methods that allow a higher level of real-time user control score very similarly. However, the method that allows the user to have moments of non-interaction while the sonification still plays is the preferred method by the users. This result confirms the observations of the above studies where users prefer to use an interaction method that does not always require continuous action from the user to produce a sonic result.

Here follow a few studies that have evaluated the sonification of large data sets either against an alternative sonification or against a visual display. Krishnan *et al.* [2001] describe the use of audification to represent data related to the rubbing of knee-joint surfaces. In this case though the audification is compared to another sonification technique and it is not found to be the best at showing the difference between normal and abnormal signals. This study is an example of use of audification being used to sonify large data sets. However, the use and evaluation of audification in the study is very different from the way this technique is employed in this thesis (Chapter 7) and therefore the results cannot be compared. In this thesis, audification is used for the sonification of a helicopter flight data set, its effectiveness is compared to that of a visual display and it is found to be effective in portraying general information about the data.
Nesbitt and Barass [2004] compared a sonification of stock-market data with a visual display of the same data and with the combined display (audio-visual). The combined display proved to be more effective in providing information. Valenzuela [1997] compared the sonification of impact-echo signals (a method for non-destructive testing of concrete and masonry structures) with a visual display of the signal. The use of sound proved to be effective in this field compared to the visual methods. Fitch and Kramer [1994] compared the efficacy of an auditory display of physiological data with a visual display by asking the subjects (who play the role of anaesthesiologists) to try to keep alive a 'digital patient' by monitoring his status with each display. In this case, the auditory display was found to be better than the visual display. In these studies the audio displays were more effective than the equivalent visual displays, a result which is similar to that obtained by the first experiment described in this thesis (Chapter 7).
More and more often sonification techniques are presented accompanied by a series of experiments that evaluate their effectiveness. This is very useful for the advance of the research field in general because the combination of results can reveal general trends for choosing the best mappings and sonification algorithms. However, evaluation studies and their results are often presented as specific for a particular data set or task. The author believes that the derivation of general knowledge could be facilitated if evaluation studies and their results where designed and presented always keeping in mind the larger context of data sonification.

2.4 Conclusions

The human perception of sound is very different from human visual perception. For this reason auditory displays have different advantages and disadvantages to visual displays. In many cases, auditory displays and visual displays can complement each other, allowing the implementation of more complete and powerful (information rich) displays. The auditory display field is much newer than its visual counterpart and not many techniques and development environments are well established. There are various types of auditory displays, of which sonification (the focus of this research) is one. In sonification the data have a direct relationship with the sound produced. The literature reports various attempts to create new development environments specifically for experimenting with sonification, however no single one has yet managed to become established. The author advances the hypothesis (after having created one of these toolkits/environments) that there is no need for the implementation of a completely new environment. Musical programming software such as PD, Max/MSP or Supercollider already allows experimentation with various mappings and interactions. After a sonification technique has been implemented and evaluated using one of such environments, it can be programmed into the specific application it was created for. Established musical programming software can be also used to create prototypes of interactions with sonifications to allow evaluation of the interaction. The principal technique for the sonification of data sets is parameter mapping where each data parameter is mapped onto a sound parameter.

This survey highlighted the dependency of the specific sonification technique on the size of the data set. Techniques used for small size data sets (e.g. in auditory graphs and earcons) exploit the principles of grouping and segregation of auditory scene analysis. When data sets become large, both in terms of number of channels and number of samples, this approach shows its limits as there are no more sound parameters to use and the sonification becomes very long in time because in this context parameters change relatively slowly. However some principles used in this domain can be used also in the sonification of large data sets, e.g. the sequential comparison of short complex sounds and the mapping of different data channels to different frequency ranges to help segregation.

The focus in the hypothesis explored in this research is the sonification of large data sets for which two main approaches were identified when reviewing the literature: timbral parameter mapping and model-based sonification. In timbral parameter mapping, the data parameters are still mapped to sound parameters, but the result of the mapping is a complex timbre evolving in time. In model-based sonification, the data set is seen as an array of points in an n-dimensional space. An arbitrary physical law is then imposed onto this space which creates a reaction when energy is introduced in the system via some sort of excitation (usually driven by the user). The physical reaction of the system is the element that is sonified and that gives information about the system. In timbral parameter mapping usually the raw data are directly mapped onto sound parameters, while in model-based sonification the raw data are usually manipulated before being mapped onto sound parameters. One of the main advantages of model-based sonification is the fact that it makes explicit the importance of interaction in sonification.

Different approaches to mapping data to sound parameters are summarised in Figure 2.1:

- the 'raw data' are mapped directly onto sound parameters that have an almost directly correspondent 'perceptual parameter' (such as frequency/pitch, amplitude/volume, time/duration, etc.);
- the 'raw data' are mapped onto 'sound variables' that do not have direct correspondences in the perceptual domain (e.g. the bandwidth of a filter in subtractive synthesis, etc.). In this case, normally, the timbre of the resulting sound is a function of the raw data, but not in an explicit way;
- the 'raw data' are manipulated and 'new data' are created which are functions of the original 'raw data' (e.g. in model-based sonification). This 'new data' are then mapped onto 'sound perceptual parameters';
- the 'raw data' are manipulated and 'new data' are created (as in point 3). The 'new data' are then mapped onto 'sound variables' (e.g. can happen in model-based sonification).

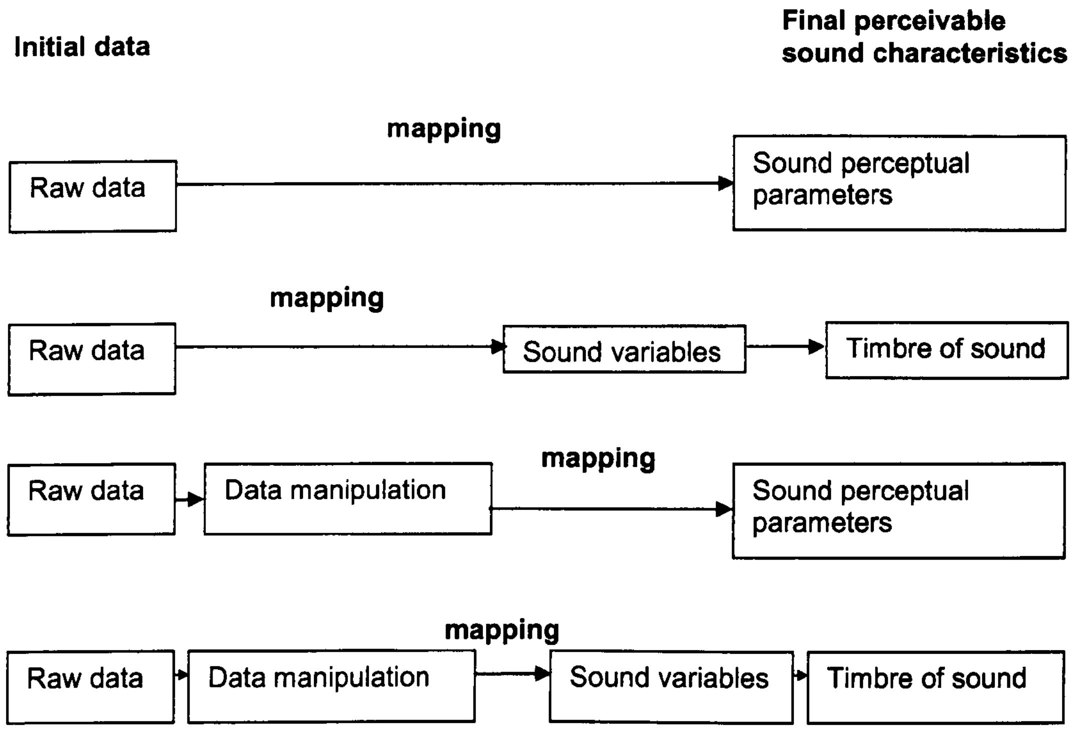


Figure 2.1: Four different ways of mapping data to sound

Sonification is a new field and therefore is in need of evaluation so that from experimental results design guidelines can be drawn. Evaluations often compare auditory displays to visual displays, or to audiovisual displays, or compare different sonifications of the same data. From this survey, we can see that the majority of evaluation studies are concerned with auditory graphs and earcons, i.e. relatively small data sets. In contrast, there are very few examples of evaluations of complex timbral sonifications, created from large data sets, or of auditory feedbacks (which imply real-time sonifications). Reasons for this could be that it is more difficult to have access to large data sets than small ones, and that designing experiments that investigate changes in timbral characteristics is a multidisciplinary, challenging and complex task. This thesis aims to improve our knowledge about the sonification of large data sets by presenting new and original evaluation studies in this particular field.

Chapter 3 - Interaction

3.1 Introduction

A focus of this research is to evaluate how the addition of a high degree of user interaction can improve the usability of a sonification display. This chapter describes the role of interaction in multimodal displays in general and in sonification in particular. Firstly it argues the reasons why there is a necessity for interaction in computer displays and sonification, and then it reviews the research done so far in interactive sonification.

3.2 Multimodal and interactive displays

Digital technology has developed very rapidly in the last few decades. This means that today's cheap home computers can easily process and store large amounts of data and sound in real-time. This improvement in processing and storage power means that it has become realistic for scientists to research into increasingly demanding interfaces, in particular multimodal interfaces that include the use of sound. A multimodal interface is a communication channel between the user and the machine, which exploits many senses and/or interface devices. A multimodal interface can use auditory, visual and tactile stimuli to reach the user and can also use various interface devices to communicate users' commands to the computer. Therefore a display that uses sonification, a graphical interface and allows a high degree of user interaction can be considered a multimodal interface.

This research focuses on the design and evaluation of sonifications which are potentially to be used in the context of a multimodal interface. For example, the final application for the analysis of helicopter flight data could allow the user to simultaneously see the spectra of the data, hear the sound and have the possibility of navigating through the data with a high level of real-time playback control. The final application for the analysis of EMG data could be real-time auditory feedback with which both the patient and the physiotherapist could hear sound reflecting the muscle activity of the patient. In this situation, the physiotherapist could also look at the patient posture or at the graphical display of the data while listening to the sonification.

The reasons for developing research on multimodal interfaces are many. Two important ones follow here.

1. To use the different characteristics of the senses so that information can be communicated more efficiently.

   A person could be seen as a "multimodal" being who experiences the world through all the senses at all times. It seems logical therefore to think that this intrinsic multimodality of the human being should be used and that, perhaps, if the number of modalities of a display increases, then the dimensionality of the information transmitted can be increased as well.

2. To create a more "natural" relationship between the human and the machine.

   Since the computer is a particular machine that allows one to quickly process information coming from a variety of systems, it is an indispensable tool for understanding, monitoring and exploring these systems. However, the computer interface is far removed from the physical perception of the actual system studied and often relatively poor. Multimodal interfaces could bridge the gap between the system explored and its perception.

The inclusion, in displays, of many modalities opens up many questions about how to integrate them. Hermann and Hunt [2005] summarise the main questions as follows:

- What are general requirements for a multimodal system?
- How do these requirements affect the system design?
- How should the information be distributed to different modalities to have best usability?
- What synchronisations between modalities are most important?
- How does the user's interaction influence what is perceived?
- How do we process multimodal stimuli?

These are relatively new questions with respect to computer systems and they have not yet been thoroughly researched. Technology has only recently acquired the possibility of being truly multimodal and in real-time so experimentation and evaluation studies have only started relatively recently. Furthermore designing such experiments offers difficulties in itself. Controlled experiments are based on the assumption that different variables can be isolated and controlled. In a multimodal system, the object of the observation is the interplay of the different modalities, which could lead to better system performance than what might be expected from summing the individual contributions.

In any case, before evaluating it, a multimodal interface has to be designed. No single approach to designing a multimodal interface has been established. Actually completely different, almost opposite, approaches seem reasonable a priori. For example, an obvious approach to design any multimodal system is the following. If we are able to reproduce in the system the mixture of redundant information (e.g. think of the process of cooking food where the combination of look, smell and taste give us the necessary information about the quality of the dish) and disjointed information (e.g. think about our ability to follow more than one stream of information at the same time, like when driving while listening to the radio) that the world proposes to us through the many senses, and to which we are accustomed, then the system should feel more natural.

On the other hand, we should not forget that our brain is quick to adapt and find new connections, meanings and metaphors between different stimuli, i.e. what is not perceived as "natural" now can become "natural" with use. This means that systems do not need necessarily to reproduce reality to be perceived as "natural". A clear example of this phenomenon can be found in the cinema. A film, though perceived as realistic, not only does not reproduce reality, but cannot. For example, the perception of time in the cinema is different from reality and half an hour can be transformed into one hour, five minutes or a week in the viewer's perception. Another example could be that non-realistic sounds are in many cases better at portraying actions in the scene (think about the Foley or effect sounds which are seldom produced using the object shown in the film). Finally the effect of synchresis [Chion *et al.*, 1994] supersedes our experience of the real world. Synchresis refers to the phenomenon by which a sound occurring at the same time as a visual action, even if not coming from the right direction, is directly associated to that action. This happens in the context of the cinema, where our brain prefers rather to believe that the association is true, and so the flow of the story is maintained, than to doubt that association and not be able to follow the story. Phenomena like this could occur in a multimodal system as well, where new associations between sound and vision, for example, can become established even though they are not realistic.

Finally, it is important to understand how we interact with multimodal stimuli in order to make sense of them and how important it is for the user to be in control of the interaction. The understanding of this factor also contributes to improve multimodal interface design. In the context of HCI (Human Computer Interaction) to be in control of the interaction is referred to as being at the centre of the control loop. For example, driving a car requires, for the overall task to be performed satisfactorily, watching the road, listening to the car engine and the noises of surrounding cars and using feet and hands to move the pedals and steering wheel. All of these actions need to be monitored constantly and adjustments have to be made quickly in particular when something is not working properly. Being in the control loop of an action means being able to have real-time feedback (that can be visual, auditory and tactile) of any change in the situation. This allows corrections to be made continuously and quickly to reach the optimised action. Being in the control loop also means being highly active and involved, and it requires attention, concentration and engagement.

The third experiment presented in this thesis evaluates the usability of a multimodal interface that uses sonification, to a limited extent graphics (a red rectangle on the screen represents the borders of the data set to be explored) and user interaction to display information. This experiment is used to evaluate the effect of adding user interaction with different levels of real-time playback control to a multimodal, in this case mainly sonification-based, display. The experiment is task-based, meaning that users are asked to perform a realistic task with this interface and their performance is the main indicator of how well the whole interface portrays the information to be displayed. The task, the data and the sonification remain constant, while the interaction methods differ in the degree of user control they allow. Changes in usability scores are then attributed to the interaction method used each time and we can infer whether or not allowing a higher degree of user interaction control improves the usability of the overall display.

3.3 Interactive sonification

As mentioned above, interactive sonification can be considered as a particular case of a multimodal system. Hermann and Hunt gave their definition in the introduction of the IEEE Multimedia issue [Hermann & Hunt, 2005] on interactive sonification that they guest edited:

> We define interactive sonification as the use of sound within a tightly closed human-computer interface where the auditory signal provides information about data under analysis, or about the interaction itself, which is useful for refining the activity. [Ibid., p. 20]

Interaction in sonification can be implemented in two main ways:

- *Interactive sonification of recorded data*: a sonification of data can be interactively explored, e.g. the sound can be navigated along the time-line and/or the mapping and scaling between data and sound can be interactively modified;
- *Interactive sonification of data gathered in real-time*: the sonification can be produced as the data are gathered, i.e. in real-time. In this case, changes in the data are instantly transformed into changes in sound and the user could be involved in producing the data. This case of interactive sonification can be used to create an *auditory feedback* of evolving systems or actions.

In both cases the interaction should allow the user to learn more about the information being displayed and therefore improve the display. These two ways of interacting with the sound and the information in it, conceptually, are not new: they are only “transported” into a new context. We can find these two approaches in music studying and playing.

A complex musical piece, to be learnt and understood, needs to be taken apart, to be heard at different speeds, different ranges, and in different orders. In other words, to extract and understand the form and structure of the piece, exploration is necessary. Similarly, exploration is necessary to understand the data structures and the information displayed by a sonification.

The second approach is akin to playing an instrument. When a guitarist tries many guitars in order to buy one, he/she assesses the guitar properties by playing with different techniques to assess the physical properties of the instrument. He/she hears the frequency and dynamic responses of the physical object by plucking, hitting, etc. Therefore knowledge about the object that produces the information (guitar in this case) can be obtained by injecting energy in different ways and listening to the physical responses of the object. Considering all these similarities between music playing and interactive sonification, it seems logical that the designer of an interactive sonification system can learn a lot from the research produced in musical interface design for electronic instruments.

Producing an electronic instrument requires designing both the interface and its relationship to the sound source. Hunt *et al.* [Hunt, Paradis and Wanderley, 2003; Hunt, 2000] showed that the mapping between the input and the output of a musical interface has an important role in determining the success of the interaction. The mapping can also influence if the user perceives the system as a musical instrument or as a system that produces sound, but that cannot be used to make music.

A good musical interface and sonification interaction should also facilitate what Csikszentmihalyi [2000] calls *interaction flow*. When a good musician plays, he/she does not have to concentrate on the detailed movements of the fingers: they somehow go on their own. The musician is then free to concentrate on higher-level musical and expressive thoughts. In fact to be too self-conscious of the fingers' movement can stop the "flow" and ruin the performance (like driving a car and being more conscious of the small movements of the feet, than the traffic in the street). Hunt *et al.* [2004] call the quality of control that allows interaction flow *control intimacy*. An interactive interface should allow good *control intimacy*.

Learning also has an important role in the playing of musical instruments. A musical performance requires a good deal of time spent learning how to play the instrument (i.e. exploring how to best interact with the system). An easy to learn interface and interaction can give quick satisfaction, however an interface and interaction that need time to be learnt can be much more effective in the long run [Hunt, 2000]. Learning can also be an important factor in interactive sonification use.

Another important element in sonification that requires interaction is personalisation (or "tuning-in" a system to suit a particular person). Sound perception is felt as a highly subjective experience. This is probably due to the fact that sound activates our emotions strongly and very directly. Our reactions to sounds are very personal and sounds that are considered too loud for one person are too soft for another, or sounds that are considered beautiful to some are considered horrible by others. Details that a person is able to hear in a piece of music, that other people cannot hear, are difficult to indicate to another person unless a playback system allows playback from different points in the piece, filtering of concurrent frequencies, etc. Being able to "tune" a sonification to personal settings is very important [Hunt *et al.*, 2004] - rather like being able to use spectacles if one does not see well, i.e. in order to see the same information as well as perfectly sighted people, a person with myopia needs to customise the interface (wear glasses). Finally, to investigate a sonification means to learn about the data which are being sonified. Learning cannot be done passively: it involves interacting with whatever is being investigated and a certain level of engagement.

3.3.1 Interaction in parameter mapping and model-based sonification

As mentioned in Chapter 2 (Sections 2.3.2 and 2.3.3), there are two main ways of sonifying data sets: parameter mapping and model-based sonification. When a sonification is built using parameter mapping, one can create an interface that allows scrolling the sound backwards and forwards at different speeds and that allows the rapid change of mapping and scaling.

When data are fed through the sonification in real-time there could be two levels of interaction.

1. The user is not involved in producing the data to be sonified and therefore the only level of interaction involves being able to adjust sonification parameters as he/she listens to the sonification.
2. The user is involved in producing the data to be sonified (e.g. data produced by the users' muscles' activation) and therefore can modify interactively the data being fed through the sonification, which changes accordingly.

In model-based sonification, interaction has a particular role. There is a conceptual difference between parameter mapping and model-based sonification. In the context of parameter mapping, the interaction is generally something that can be added after the sonification (even though as it was explained above it seems a quite necessary add on for thorough exploration), whereas in model-based sonification the interaction is part of the sonification system. In MBS (model-based sonification) what is sonified is not the original data in itself, but the interaction between the excitation (it can be a mouse click or another event) and the original data (which is usually thought of as a "data space"). Therefore in MBS no interaction means no sonification. As before, also in this case the interaction can be used to explore pre-recorded data or be part of the process that produces the data to be explored. It can be said that MBS integrates interaction as a central constituent of the sonification model definition. This method aims to exploit our ability to interpret sound as the response of an object being excited (e.g. a bottle of water being shaken). Since we understand many properties of the object by listening to its sound, we could understand the characteristics of a data set being 'played' in MBS [Hermann & Ritter, 2004]. Many questions arise about which physical laws should be imposed onto the data sets to create meaningful reactions from an object that is, in any case, a virtual object (a data set), however the concept is very interesting and will certainly produce important results.

3.3.2 Review of research on interactive sonification

Interest in interaction has been shown by the sonification community since the early 90s. The concept of a *Sound Probe* for sonification was presented in 1992 [Gröhn, 1992] and then used again in [Barrass, 2000] for the exploration of well-logs. Here the user holds the probe in one hand and points it towards a region of interest. The probe is tracked in 3D space by a 6 degrees of freedom sensor and then is displayed as a virtual probe visually on screen. Fernström *et al.* [1998] describe the importance of *direct manipulation*, concept analysed by Shneiderman and others [Shneiderman, 1983; Hutchins *et al.*, 1986], and apply it to an application for browsing musical tunes. *Direct manipulation* is characterised by the continuous representation of the objects of interest and by immediate feedback [Shneiderman and Plaisant, 2005]. Jovanov *et al.* [1998] presented the use of audio as feedback for precise manual positioning of a surgical instrument and referred to it as *tactile audio*. Saue [2000] presented a general model for the sonification of large spatial data sets in which

> [...] the interpreter is walking along paths in areas of the data set, listening to locally and globally defined sound objects [Ibid. p.1].

The "virtual walking" could be done using the mouse or another similar input device. The mouse has been the first choice of interface device for many researchers working in interactive sonification as it is the most common computer interface device in use to this day. Winberg *et al.* [2001] use the mouse as a *virtual microphone* and in [Brazil *et al.*, 2002] an extension of the work on sonic browsing is presented and multiple musical tunes are navigated using the mouse. Hermann used the mouse to interact with data spaces [Hermann, 2002] in his early examples of MBS. In this research also the mouse has been chosen, because of its popularity, as one of the two interface devices used to allow interaction in experiment 3 (Chapter 9).

In the last few years, with the higher processing power of computers, more research on new interfaces can be found. DiFilippo in [2000] presented a new computer interface that renders simultaneously audio and haptic stimuli in a tightly coupled way. The interface, a three degrees of freedom haptic device that renders the force of the manipulation as output, attempts to reproduce real-life contact events which produce both sounds and reaction forces. Beamish *et al.* [2003] present a system that uses a haptic turntable for controlling the playback of digital audio. In the paper it is argued that the system, initially intended for DJs, could be used for the exploration of sonifications. Fox [2005] presented the *SoniMime* system for the sonification of hand motion, where the data coming from sensors attached to the hand are mapped to timbre parameters. This auditory feedback aimed to assist the user in learning a gesture with minimal error. Hermann and colleagues since 2001 have been exploring the use of novel interfaces to interact with MBS systems. The aim is to create a truly continuous control of the interaction. In [Hermann *et al.*, 2001], a custom-built 'hand box' interface is described in which the hand posture is analysed and reconstructed as a hand model with 20 joint points. This new complex interface (the hand model) is then used for sound/data space exploration. In [Hermann *et al.*, 2003] the 'gesture desk' interface is described. This interface tracks the free movements of the hands and uses them to interact with data spaces. The inspiration for these interfaces is the way we use our hands to explore new physical objects. However, if it is true that in the real world we use our hands to explore what is around us, it is usually true that the hands are manipulating a real physical object, they are not just moving in air. This problem can be solved if data spaces can be "transformed" into physical objects that we can manipulate. In 2002, Hermann presented the "audio-haptic ball". The ball is made of plasticine, and equipped with various sensors (acceleration and force sensitive sensors) which send data to the computer when shaking, scratching, squeezing, rotating, and hitting the interface ball. These interactions are then used to excite the MBS system, simulating the scenario in which one shakes a box to guess what is inside. Milczynski *et al.* [2006] presented a malleable interaction surface for continuous and localised exploration (using fingers' movements) of MBSs, and Bovermann [2006] presented a project whereby moving an object (for example a stick) one is allowed to scan a data set which is virtually positioned around the stick. Although the study of novel interface devices for interactive sonification is a very interesting area of research, in the author's opinion there has so far not been enough evaluation of the performance of common interface devices for sonification-based data analysis applications (in particular when large data sets are sonified). Common interfaces, such as the mouse, would probably be the first choice of interface devices used in a sonification-based data analysis application and therefore their effects in this context need to be clearly studied. In this research, three ways of interacting with a sonification of large data sets are compared (see Chapter 9) which involve the use of the mouse and the jog wheel/shuttle interface [Contour Design], both popular devices used in audio and video editing applications.

Few recent studies have evaluated the use of the more common interface devices such as the mouse [Holmes, 2005], the keyboard [Stockman *et al.*, 2005] and the tablet [Kildal & Brewster, 2006] for the navigation of two dimensional data sets. In particular, in [Holmes, 2005] the mouse was used to navigate a two-dimensional map and locate specific points in the map. This task was done using the auditory display only and the visual display only. The interactive visual display turned out to be better for this task. It was also found that generally people would use recognisable search patterns with the mouse in order to complete the task. Although this study, like in experiment 3 of this thesis, uses the mouse as the interaction device, this experiment task is very different from experiment 3 of this thesis and therefore the results cannot be compared.

Another area of research connected to interactive sonification is the study of physical modelling sound synthesis algorithms. A logical consequence of considering a data set as a virtual object to interact with is to make it react acoustically in the way that a physical object would. In this context, it makes sense to use sound synthesis algorithms based on physical modelling in which the synthesis parameters are already directly related to the parameters used in the interaction with the object. Sound synthesis methods based on physical modelling are of high interest for interactive sonification and although the research presented in this thesis does not employ physical modelling sound synthesis, a few references are mentioned here for completeness. Avanzini *et al.* [2004] presented a study on physics-based sound models of continuous contact and on their ability to convey the physical properties of an object and Van Den Doel [2004] presented a physical model-based synthesis technique for the production of liquid sounds.

The real world, and the way we interact with it, is certainly a great inspiration for the design of interactive sonification, however it is not the only context from which the designer can learn. As mentioned in Section 3.3 of this Chapter, the design of interactive sonification can learn from the development of new musical instruments and from the way interfaces are used in more general artistic contexts. Examples of studies related to this area are the already mentioned use of turntables for audio exploration [Beamish *et al.*, 2003] and a study presented by Beilharz [2005] in which it is argued that the investigation of digital art works that use gestural controllers yields important knowledge on how to create successful interactions between the user, the controllers and the sonified information.

Finally, a good source of information on the various aspects of interactive sonification can be found in the special issue of IEEE Multimedia which was dedicated, in 2005, to interactive sonification. In this issue we can find various examples of interactive sonifications from the navigation of maps [Zhao *et al.*, 2005] using a keyboard and a tablet, to a discussion of the use of interactive sonification in mobile devices [Fernström *et al.*, 2005; Williamson & Murray-Smith, 2005], from the evaluation of an auditory feedback of movement [Effenberg, 2005], to the use of physical modelling sounds to create a realistic sonic feedback [Rath & Rocchesso, 2005], from an auditory biofeedback of EEG data that allows self-regulation of the brain activity [Hinterberger & Baier, 2005], to the use of auditory cues in a virtual environment [Lokki & Grohn, 2005].

3.4 Conclusions

Multimodal interfaces have the potential to make the relationship between the user and the information contained in a computer more "natural" and more efficient. However the design, evaluation and use of multimodal interfaces propose to researchers many, so far, unanswered questions such as how should the information be distributed to different modalities, how do we process multimodal stimuli, should the multimodal interaction be modelled on the way we interact with reality or can new associations between modalities be established.

Interactive sonification is a particular type of multimodal system which aims to improve the understanding of the information being sonified. A review of the literature in the field shows that sonification researchers are more and more involved in studying the effects produced by the addition of interaction to sonification systems. Different interface devices and interaction approaches are in the process of being developed. The review also found that such developments need to be accompanied by thorough evaluation studies which, at the moment, are still very few. This thesis presents in Chapter 9, an original study which evaluates the role of interaction in a sonification system of large data sets. This study aims to contribute to our thus far limited knowledge of how interaction can improve a sonification system.

Chapter 4 - Sound perception

4.1 Introduction

The focus of this thesis are three experiments (described in Chapters 7, 8, and 9) that test chosen sonifications based on data to sound parameters' mappings. Whether in the scientific domain or in the creative domain, the key to being able to meaningfully design and manipulate mappings between data and sound is the knowledge of how humans perceive sound. In this chapter a brief summary of some of the main principles of sound perception is presented and the main sources of information for this summary are [Howard and Angus, 1996], [Cook, 1999] and [Bregman, 1994].

Furthermore, the knowledge of sound perception is based on psychoacoustic experiments carried out in very controlled and limited situations. This chapter also reports the challenges, discussed in [Neuhoff, 2004], posed by the field of auditory displays to the methods of psychoacoustic experimentation.

4.2 Understanding sound

To create an auditory display means to use information (e.g. data) to create a sound (e.g. a physical event) that is meaningful to our perception and represents the initial information. Key to this process is to understand what sound is, how to produce it, and how the sound event is perceived and understood by our brain. Sound is heard when a certain physical process (e.g. hitting a drum skin, plucking a string of a guitar, etc.) makes the particles in the air vibrate at a certain rate. Sound is therefore air pressure waves. When the waves hit the ear-drum of a human being, the vibrations are transmitted to the hearing system which is then able to analyse and send the appropriate information to the brain so that the sound is perceived. An example of a simple sound is a sine wave (see Figs. 4.1 and 4.2).

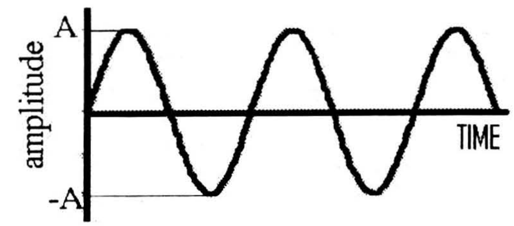


Figure 4.1: a sine wave in the time domain

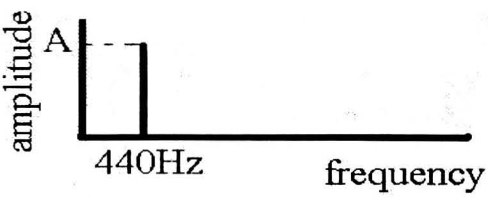


Figure 4.2: a sine wave in the frequency domain

A sine wave has only one frequency that tells us how many times a cycle of the sine is repeated in one second. A more complex periodic sound contains more than one frequency in its spectrum. A periodic sound is said to be harmonic if the frequencies present in its spectrum have frequencies which are multiples of the first frequency (called the fundamental frequency). If the spectral frequencies are not multiples of the fundamental frequency, then the sound is said to be inharmonic. The French mathematician and physicist J.B.J. Fourier (b.1768-d.1830) demonstrated that any periodic sound can be built by adding a certain amount of sine waves with different amplitudes: therefore the sine wave is considered a building block for creating any periodic sound.
Sounds can also be non-periodic (or have non-periodic components). For instance white noise is a non-periodic sound. In this case, the waveform of the sound does not have regular repetitions in the time domain, and in the frequency domain all the frequencies, not only discrete frequencies, are present in a continuous way.

A generic sound wave, as a physical phenomenon, has the following characteristics:

- frequency;
- amplitude;
- spectral components;
- duration.

The sound that is perceived when this generic sound wave is played has the following characteristics:

- pitch;

- loudness;
- timbre;
- duration.

A fifth characteristic, which is not discussed in this thesis because it is not relevant to the main experimental work presented here, is the direction of the sound wave and the perception of localisation. The physical characteristics of the sound wave and the perceptual characteristics of the sound are related to each other. However they are not related in a simple, linear way because the hearing system (which "transforms" the physical event into what we perceive) is not a linear system. The relationship between physical characteristics of the wave and perception of the sound is only partially understood so far. Here follows a summary of essential information about the sound perceptual components relevant to this research.

4.2.1 Pitch

Pitch is defined by the American National Standards Institute (1960) [ANSI, 1960, in Howard and Angus, 1996, p. 119] as:

> "[Pitch is] that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high."

Pitch is a subjective measurement in the sense that a subject is required to judge if a pitch is high or low. A subjective scale for the measurement of pitch of simple tones was established by Stevens *et al.* in [1937]. Subjects were asked to adjust the frequency of a tone until it sounded just half as high in pitch as a reference tone (keeping loudness constant). On the basis of the results of this experiment, a subjective pitch scale and a measurement unit was established called 'mel'. In this scale the number 1000 is arbitrarily assigned to the pitch of a tone of frequency 1kHz, then 2000 mels are assigned to the pitch that is considered twice as high than the 1kHz tone, 500 mels are assigned to the tone perceived to have a pitch that is half the 1kHz tone, etc. The result is that the relationship between perceived pitch and frequency of a tone in Hertz is not linear, it is actually logarithmic. More in general, sounds produced by a periodic air pressure variation are perceived to have a pitch, while sounds that have a non-periodic wave are perceived to have no pitch or a not well defined pitch.

4.2.2 Complexities in pitch perception

While the pitch of a harmonic sound corresponds to its fundamental frequency, the pitch of a sound with non-harmonic or non-periodic components is a more difficult matter. Various perceptual phenomena, due to the way our hearing system works, can create situations in which a pitch correspondent to a certain frequency is perceived even though no component in the spectrum of the sound has that frequency. For example, if a sound has three components at 300Hz, 400Hz and 500Hz, it can be perceived has having a 100Hz pitch because the brain interprets the three components as being the 3rd, 4th and 5th harmonics of a sound with a fundamental frequency of 100Hz. This phenomenon is called *virtual pitch*. Another example of virtual pitch was created by Schouten [1940], who produced a sound with components at 1040Hz, 1240Hz and 1440Hz and with perceived pitch at 207Hz. Notice that the three components are the 26th, 31st and 36th components of a sound with fundamental frequency 40Hz, so if the brain only simply "calculated" whether or not sound components were harmonics of a certain fundamental, then the perceived pitch should be 40Hz. In this case the brain seems to "interpret" 1040Hz as the 5th harmonic of a fundamental frequency of 208Hz, 1240Hz as the 6th harmonic of a fundamental frequency of 207Hz and 1440Hz as the 7th harmonic of a fundamental frequency of 206Hz. Then the brain "interprets" the average of these three fundamental frequencies as the overall pitch. Also some non-periodic sounds can be perceived as having a lower or higher pitch with respect to another sound. The frequency of this pitch is difficult to define, but it is related to where, in its spectrum, the sound presents high energy. Other difficulties come from the fact that the perception of pitch depends on the loudness of the sound. For example, Stevens [1935] found that with tones of long duration the pitch depends on the loudness of the sound. For sounds less than 1kHz the pitch becomes flatter as the loudness is increased, while for pitches higher than 2kHz the pitch becomes sharper as the loudness increases [Howard & Angus, 1996; p. 135]. Also the timbre has an effect on the pitch. Brighter sounds can be perceived as sharper than dull ones [Howard & Angus, 1996]. The minimum perceivable difference between two pitches, called JND or Just Noticeable Difference, also depends on frequency: it is smaller at lower frequencies and higher at high frequencies. In musical terms, the JND is about one twelfth of a semitone [Howard and Angus, 1996; p.125]. The duration of a sound can also influence the perception of pitch. A certain number of sound wave cycles needs to be played in order to perceive the correct pitch (about 10 cycles for a 1kHz sound, 100 for a 10kHz sound, etc.).

4.2.3 Loudness

Loudness is the measure of the volume of a sound. It relates to an objective measurement called Sound Pressure Level (SPL) that can be performed on a sound wave and which depends on the amplitude of the sound wave. Sound pressure levels are expressed in a logarithmic scale and are based on the ratio between the sound pressure of the threshold of hearing and the sound pressure of the sound that is being measured.

$$SPL = 20\log_{10}\left(\frac{p_{actual}}{p_{ref}}\right) \tag{4.1}$$

where p_{ref} = 20μPa (microPascals) which is the threshold of hearing for a 1kHz frequency.

The units of measurement for SPL are decibels (dBs). The threshold of hearing in this system is represented by 0dBs and the threshold of pain reaches 120dBs.There is a relationship between the perceived loudness and the measured SPL of a sound: when SPL is increased the loudness increases too. Doubling the sound pressure of a single sound results in a 6dB increase in SPL while a tenfold increase of the sound pressure results in a 20dB increase in SPL. It is important to notice also that listening to two sounds at the same level does not double the overall loudness level because the SPL will result in the logarithm of the sum of the two sound pressures, not the sum of the logarithms of the two sound pressures.

4.2.4 Complexities in loudness perception

Loudness perception depends on the frequency of the sound. A subjective scale subdivided in units called 'phons' represents the loudness of sine waves as a function of frequency and sound pressure level. This is a subjective scale based on judgements of listeners when asked to match the loudness of given tones to reference tones at a frequency of 1kHz. By definition the curve for N phons intersects 1kHz at N dB(SPL) [Howard and Angus, 1996]. The Fletcher and Munson diagram [Fletcher & Munson, 1933] of equal loudness contours shows the relationship between perceived loudness, frequency and sound pressure levels for sine waves. Low frequency sounds with high sound pressure levels can be perceived to be as loud as higher frequency sounds with relatively low sound pressure levels (see Fig. 4.3).

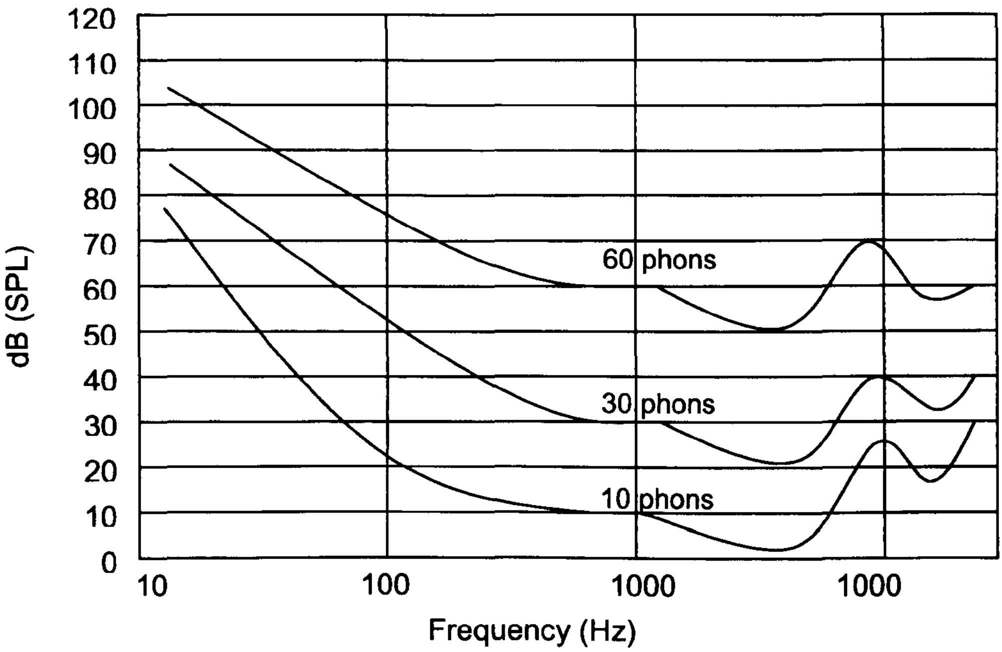


Figure 4.3: Equal loudness contours of the human ear. The figure intends to be only a schematic representation of the real contours

The reason why these contours have humps and bumps depends on how our ears and hearing system are physically made. Another fact that affects the perception of the loudness of a sound when played together with another sound, and that depends highly on the loudness and frequencies of the sounds involved, is *masking*. When two or more sounds are played together, or in close sequence, the presence of one sound can "hide" or mask the presence of the other. Care should be taken in setting the levels of different sounds in an auditory display to ensure that all the sounds are perceivable, i.e. not masked.

4.2.5 Timbre

Timbre refers to the quality of a sound that makes us describe it as bright, dull, dark, harsh, light, etc. The definition of the American National Standards Institute (1960) [ANSI, 1960, in Howard and Angus, 1996, p. 198] says:

> "Timbre is that attribute of auditory sensation in terms of which a listener can judge two sounds similarly presented and having the same loudness and pitch as being dissimilar."

In other words timbre is that quality of a sound that makes a piano note, at a certain pitch and loudness, different from a violin note played at the same pitch and loudness. Timbre is a multidimensional parameter and there are no clear perceptual, and correspondent acoustical, scales defined for it. There are certain characteristics of the sound wave that are known to affect the perception of timbre.

- The spectral components present in the sound and their harmonic or non-harmonic relationship. For example, harmonic sounds have a completely different timbre to noise.
- The concentration of sound energy. For example, the main difference between vowel sounds such as 'a' and 'o', which are clearly different in timbre, is the different position of high spectral energy in the spectrum of the sounds.
- The onset and offset of the sound amplitude envelope. For example, instrument sounds with a similar steady part cannot be told apart if the onset and offset part of the sound are eliminated [Howard & Angus, 1996, p.203].

The main problem with timbre is that it is difficult to define descriptors for it, and related acoustic parameters. Some experiments have been carried out that have defined some multidimensional timbre space for certain type of sounds. For example, Grey [1977] defined a three dimensional space which described common classical instruments, however the totality of descriptors that, if varied, can allow to create any type of sound has not yet been found.

There are important and widely accepted descriptors that have been defined, such as *brightness* and *roughness* [Zwicker & Fastl, 1999], and that have been related to acoustic parameters so that less or more rough or bright sounds can be synthesised. However also these perceptual models are limited and describe only certain types of sounds. From the point of view of auditory displays, timbre is a very interesting attribute that can be used because a lot of complex information can be portrayed manipulating this sound component. The efficacy of the use of timbre in an auditory display can be evaluated in each single case. One has to be careful to generalise the experimental results to other data contexts, environmental context, etc. because timbre parameters, as it was just stated above, are difficult to manipulate in a controlled way.

4.2.6 Duration

The duration of a sound refers to how much time passes from when the sound starts to be played to when it stops. The perception of the duration of a sound (time between when the sound starts to be heard and when it disappears) is related to the objective time elapsed from when the sound starts and when the sound finishes. Human beings have also a great resolution in distinguishing two separate sounds played in sequence one after the other. For instance, two sounds (clicks) starting 2ms apart in time are perceived as separate [Pierce, 1999, p. 102].

4.3 General perception

When attempting to directly represent data with sound, one assumes that it is possible to create a representation of the data in sound so that the information perceived by listening to the sound is somehow 'equivalent' to the information encoded in the data. This equivalence assumes that there are certain categories in our brain with which one compares data and retrieves information, that these categories are independent of the type of sensory data received (audio, visual, etc.) and that they allow this sort of 'translation' of information from one sense to another. Cognitive psychology is the discipline that studies how humans interpret the information received from their senses. Firstly, there is generation of energy by some external object, then transmission of the energy through the space in-between the event and the observer, then the reception and processing of the energy by the observer's sensory receptors, and finally transmission to the brain where some more processing takes place [Pierce, 1999]. Then somehow the brain forms a representation of the external world. Humans perceive the world through many different senses and it is possible that the brain operates a sort of confluence of these signals to create the representation of what is happening. Here follows a summary of some important perception mechanisms.

4.3.1 Unconscious inference

If we see a chair drawn on a piece of paper which is two dimensional, we directly interpret the drawing as the representation of a three dimensional chair, not just as lines in a piece of paper. In our brain there are certain mechanisms that allow us to make sense of those lines on paper and to recognise the relationship between those lines and the physical object 'chair'. In particular the ability of seeing three-dimensional objects on a piece of paper is due to what is called *unconscious inference*, which was first described by physicist and physiologist Hermann von Helmholtz (b.1821-d.1894).

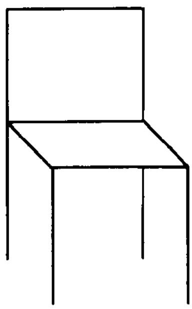


Figure 4.4: Unconscious inference: this drawing is interpreted as the image of a chair, not just as a group of lines

4.3.2 Size and loudness constancy

Size and loudness constancy encapsulates the idea that small things or low volume events appear to be far away (other cues are needed to actually be sure that the object or event is far or not).

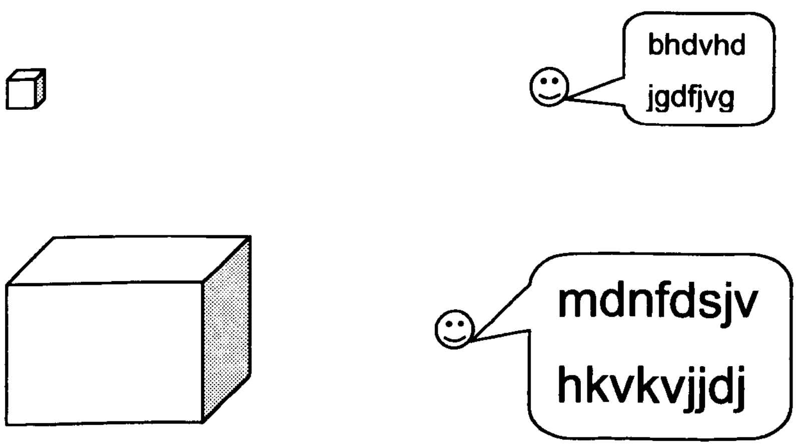


Figure 4.5: Size and loudness constancy

4.3.3 Perceptual completion

Perceptual completion describes the fact that humans tend to complete the information that is not there both in the visual and in the aural field. For example, in this picture it is easy to see a triangle that is actually not there.

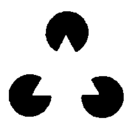


Figure 4.6: Perceptual completion

An equivalent example in the aural field is the following. Short bursts of noise and short sinusoidal signals are presented alternating. Depending on the relative intensities of the two signals, we can perceive the sine wave as pulsating or as continuous (even though the signal is not continuous) [Houtsma *et al.*, 1987; p. 33, CD track 26].

4.3.4 Gestalt Grouping Principles

At the end of the 19th century, a School of psychology born in Germany and Austria, called Gestalt psychology, created the basis for the modern study of perception. This School emphasised the difference between the perception of the whole thing and the sum of the perceptions of its single parts. It described how humans tend to perceive patterns and configurations rather than single bits and pieces. *Gestalt Grouping Principles* describe the way humans perceive objects (audio or visual) and group them.

These principles are:

Proximity: things closer together are likely to be grouped as being part of the same object

Figure 4.7: Proximity: in this image, we see three pairs of circles, not only six circles

In audio, this principle contributes to stream segregation due to pitch (see section 4.4.2 of this Chapter for more on stream segregation).

Similarity: if objects are equally spaced, the ones that appear similar tend to be grouped as related

Figure 4.8: Similarity: in this image we see three pairs of different objects, not just six objects

In audio, this principle contributes to stream segregation due to timbre (see section 4.4.2 of this Chapter for more on stream segregation).

Symmetry: random selected objects are unlikely to be symmetrical, so symmetry relates objects to each other

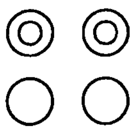


Figure 4.9: Symmetry: in this image, we see four circles and the top two are different from the bottom two. The left part of this image is equal to the right part of the image and if we swap the left circles with the right circles the image remains the same, i.e. the left and right sides are symmetrical.

In audio, a listener can be presented with four sounds. One coming from the front left corner, one from the front right corner, one from the back left corner and one from the back right corner. The two sounds presented at the front are the same and the sounds presented at the back are the same, however the sounds presented at the front are different from the sounds at the back. Also in this case, if we swap the sound sources from the left to the right side, the overall effect does not change because there is symmetry.

Good Continuation: if objects appear to continue on one another then they are grouped together

Figure 4.10: Good Continuation: in this image, we see one segmented line, not many lines grouped together

An example of good continuation in audio could be the Shepard-Risset glissando [Howard and Angus, 1996, p. 228; Houtsma *et al.*, 1987, p.59, sound example of the Shepard-Risset glissando: CD Track 52]. This glissando appears to increase continuously in pitch, while in fact it only increases within an octave and then it is repeated continuously. This effect is obtained by successively introducing low harmonics and eliminating high harmonics which are spaced by octaves. The overall effect sounds like a single evolving sound object, while in reality sound elements are continuously introduced and eliminated.

Common Fate: objects that move together are likely to be related.

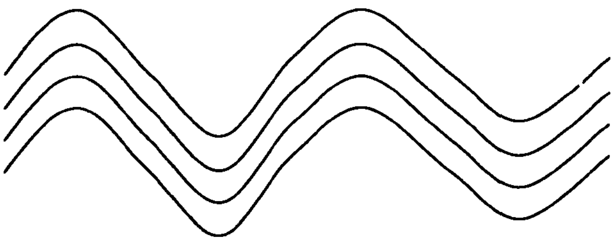


Figure 4.11: Common Fate: in this image, we see four lines following one shape and forming one object, we do not see the lines as unrelated

The last principle, in particular, is considered 'strong' compared to the others. It is with this principle that humans can distinguish two instruments or voices playing together as the brain groups the movements of the partials of one sound (which move together) separately from the other. More on how these general Gestalt principles apply to the perception of sound is described in the following section of this chapter.

4.4 Auditory scene analysis

Albert Bregman, in the book Auditory Scene Analysis [Bregman, 1994], brings together the knowledge of how the brain is able to analyse sounds. The brain appears to have two main ways of analysing, identifying and grouping features in sound:

- the first uses *primitive processes* of auditory grouping, which are based on the way the auditory system is naturally built and are independent from learning and experience;
- the second is called *schemas-based grouping* and it uses the humans' knowledge of familiar sounds and experience of listening.

The human auditory system, when listening to sound, receives a pressure wave, the spectrum of which could be called an 'energy array'. This array is broken down by the auditory system into many separate analyses in particular moments in time and frequency regions. Each region is characterised by parameters such as intensity, fluctuation patterns, etc. The brain needs to group, or separate, all this information in order to recognise the various events in the sound.

4.4.1 Primitive processes of auditory grouping

In primitive grouping there are two main dimensions along which the brain groups sounds: time and frequency. The process of grouping while time is passing is called by Bregman *sequential integration* and the process of 'instantaneous' grouping is called *spectral integration.*

4.4.2 Sequential integration

In sequential integration, pitch and timbre are the two parameters that strongly determine separation between sounds.

Sequential integration and pitch

It can be seen that if the same sound varies only in pitch, then pitches close together in frequency are perceived as grouped together (stream forming) and pitches farther apart in frequency are separated (stream segregation). The speed of the sequence also highly influences this type of grouping: when the sequence is played fast, the grouping is more easily perceived. Pitch grouping seems to follow the Gestalt principle of 'proximity' in frequency and time.

Sequential integration and timbre

Sounds are perceived as separate (stream segregation) also if they are very different in timbre. In particular, humans distinguish very well between noisy sounds and pure tones. Tones with equal or similar timbres are grouped together (stream forming). Timbre grouping seems to follow the Gestalt principle of 'similarity' in timbre and time. In particular, tones that are similar in brightness (parameter related to the mean frequency of the sound's components when all frequency components are weighted according to their loudness) are grouped together. Also peaks and valleys (formants) in the spectra of sounds may affect the grouping, since they affect the timbre of sounds.

Other parameters that affect sequential grouping are spatialisation of sound, the loudness and the rhythm.

Sequential integration and spatialisation

Spatialiasation is a weak parameter in sequential grouping. Normally sounds coming from the same point in space are grouped together, but spatial differences seem to have their strongest effects when they are combined with other differences.

Sequential integration and loudness

Loudness differences mainly help to reinforce groupings created by other parameters.

Sequential integration and rhythm

Rhythm does not create grouping by itself, instead stream formation, by some other criteria (timbral segregation for instance), identifies a rhythm. Therefore stream segregation can affect the perception of rhythm: when two different streams are perceived separately (for example two different instruments playing simultaneously), then two independent rhythms are heard.

Sequential integration, collaboration and exclusive allocation

Finally, two parameters can be said to collaborate to the same grouping when they both contribute to the perception of a certain grouping. In the same sequence two elements could have parameters participating to two different groupings. In this case there is a competition between the parameters. Either one of the two parameters is much more important for sound grouping and therefore it imposes a grouping to perception, or ambiguity is perceived. Our brain is biased against perceiving a sound in two streams at the same time (this phenomenon is called *exclusive allocation*) and therefore, if we introduce in the sequence a third sound that agrees and reinforces one of the two groupings, then the ambiguous sounds will be 'captured' by that grouping and will be perceived as being part of that stream and not as being ambiguous. The fact that our brain uses exclusive allocation is very useful for creating experiments that show characteristics of sounds that are not clearly perceivable and it is potentially very useful in the exploration of sonified data.

4.4.3 Spectral integration

As already mentioned, frequency is the second dimension along which primitive processes, in the brain, group sound. This phenomenon is called spectral integration. If humans were not able to group sound elements at each instant in time, then they would hear only one sound at the time and the qualities of that sound would represent the sum of all events going on around them. In order to separate the components of one sound from another at each instant in time, the brain 'looks' for correlations and correspondences among parts of the spectral content of the sound that would be unlikely to have occurred by chance.

Spectral integration and the harmonicity principle

For example, the brain uses the harmonicity principle, i.e. it listens to the relationship between sound components, and it groups partials that are harmonics of the same fundamental. A different pitch is then derived for each group of partials, so that different and separate pitches are heard at the same time. When spectral components are grouped together also different timbres can be derived and perceived as separate.

Spectral integration and the common fate principle

The brain also uses the Gestalt principle of common fate by looking at correlated changes in the frequency or amplitudes of different components. For example, an unconscious micro-vibrato in voices helps to distinguish them from each other. Also the correlated changes in components' amplitude, at the beginning and at the end of sounds, help in separating those sounds.

Spectral integration and spatialisation

The spatial characteristics of the spectral components strongly influence spectral integration since the direction and position of sound sources influence the spectrum of sounds. Sounds coming from the same direction and position are grouped together.

4.4.4 Schema-based grouping

The second method the brain uses to group sounds together is called schema-based grouping. The description of this grouping method takes into account the influence of culture and knowledge in the way we group sounds. The theory of schema-based grouping argues that particular classes of sounds such as speech, music, machine noises and other familiar sounds, are captured in units of mental control called *schemas*. It also argues that humans use certain constant properties of the acoustic world such as the fact the most sounds tend to be continuous, to change location slowly, to have components that start and end together and the fact that a schema (a particular class of sound) is characterised by a particular pattern. For example, the word 'apple' and the word 'fruit' are two schemas. In schema-based grouping, humans store in their brain the pattern of the world 'apple' and then expectation and association are used to recognise the sound as the schema we have stored in the brain.

4.4.5 Differences between primitive grouping and schema-based grouping

The main differences between primitive and schema-based grouping are:

Primitive grouping	**Schema-based grouping**
Primitive grouping employs neither past learning nor voluntary attention, it is probably **innate**, and therefore its **principles are general** (used by every person).	Schema-based organizational methods have been developed for particular sounds, so they are **not general** and depend on the **used language**, they rely on knowledge and use of **voluntary attention**, i.e. the person, when using schema-based grouping, is 'listening for' something.
Primitive processes **partition** the **sensory evidence**.	Schema-based processes **select from the evidence** without partitioning it.
The fact that primitive processes partition the sensory evidence means that the effects of primitive segregation are **symmetrical**. For example, when the brain segregates high notes from low ones, we can listen more easily to either the high ones alone or the low ones alone.	The fact that schema-based process select from the sensory evidence without partition it means that schema-based organizational principles are **asymmetrical**. For example, when our own name is mixed with other sounds the fact that it is our name makes it easier for us to hear it in the mixture, but it does not make it easier for us to tell what the reminder of the mixture is.

Table 4.1: Main differences between primitive grouping and schema-based grouping

Finally, schema-based groupings are linked with our ability to learn to recognise a sound and to recognise patterns.

4.4.6 Basic information on memory in general and memory of musical attributes

Since sound develops in time, an important aspect of sound's perception to review is human memory in general, and memory for musical attributes in particular. The following section is a summary of the information found in *Memory for musical attributes* [Levitin, 1999].

Humans have different types of memory:

- *Echoic memory* refers to our ability to 'hear', for a few moments after hearing a sound, a trace of that sound in our mind's ear;
- *Short term memory* refers to the ability of holding a thought in our mind, for example what we are going to say next in a conversation;
- *Long term memory* refers to remembering things experienced or learned a long time ago. This can be divided into:
    - *Episodic*: remembering a fact: e.g. my last birthday;
    - *Semantic*: remembering a word: e.g. the capital of Italy;
    - *Procedural memory*: remembering a process: e.g. how to ride a bike.

The brain uses various methods to remember large amounts of data. A method used to memorise large amounts of data is 'chunking' things together. For example, it is difficult or even impossible to remember a sequence of many numbers, but if they are three telephone numbers of three very good friends, then one can remember them as three distinct objects.

The brain is also able to make educated guesses for missing information on the basis of the existing information. For example, if one sees the following word 'coff_e', he/she will guess that the missing letter is an 'e' (obviously the subject is assumed to know the English language). This is a visual example of this ability of the brain, but equivalent sound examples can be created as well. One example is to play a sine wave interrupted by very short bursts of noise. The brain will not hear the interruption, it will hear a continuous sine wave and some noise bursts.

There is evidence that memory preserves both details and the gist of experiences and we can just access the information appropriately. Our ability to remember music and sound is a particular part of our general ability to memorise concepts, images, odours, touches, sentences, etc.

Looking just at our memory for musical and sound attributes, it is found that we remember melodies better than any other musical attribute such as timbre, rhythm, tempo, loudness, contour (general pitch direction: up or down) and spatial location.

Melodies have particular characteristics:

- a melody can be recognised even when each of the other perceptual attributes are changed, with the exception of contour and sometimes rhythm;
- a melody is also independent from the actual absolute values of the pitches of the tones played;
- a melody is defined by patterns of tones and the relationships between pitches;
- when a melody is not well known (such as a melody we only have heard a few times), then the contour is remembered better than the actual intervals.

4.5 Auditory displays and sound perception

The field of auditory displays poses many, partially new, questions to the science of sound perception. The task of using, and therefore perceiving, an auditory display can be divided into three subtasks [Walker and Kramer, 2004]:

1. *perceiving* the display;
2. *segregate and group* multiple streams of sounds appropriately;
3. use our *cognition* to make the appropriate *association* between the sound and the meaning of the information portrayed.

From the point of view of the auditory display designer this means that:

1. *perception*: an auditory display needs to be perceived to be of use. For example, an alarm needs to be above a certain frequency and loudness thresholds to be heard. Resolution is also important: if what conveys information is changes in sound's timbre, then the minimum perceivable change in timbre needs to be known;

2. *segregation and grouping*: one needs to know how different streams of sounds are grouped or separated in our perception to be able to exploit these characteristics and convey more complex information;
3. *association and cognition*: one needs to know which sound is more appropriate for conveying a certain type of information, and which sound parameter and polarity facilitate user understanding of the meaning intended for the sound.

Traditional psychoacoustics research has answered many questions in the first subtask (determination of thresholds, determination of dependencies between sound parameters, study of masking, etc.) which is also the most fundamental, since if the auditory display is not perceived in the first place no meaning, either correct or incorrect, can be attributed to it.

Traditional psychoacoustics, however, does not answer the questions raised by subtasks two and three, which are fundamental for the correct interpretation of an auditory display. The designer of auditory displays needs to take into consideration the environment in which the display will be used, what concurrent environmental sounds are competing for the attention of the user, which task the user needs to complete using the display, which sounds are the most appropriate for the task and for the usability of the display (e.g. complex sounds or simple sine waves), which are the sound parameters' polarities that best portray the information, etc. The environment, the task, the meaning and the ergonomics of the display are all issues that need to be understood and known, together with thresholds of hearing for different parameters, etc. Investigating these issues requires designing evaluation studies to identify the optimal sound parameters to be used, polarities, etc. However, to create a meaningful experimental set-up poses many challenges. Psychoacoustics experiments require all the possible variables to be kept constant but one: the one we want to learn about. This isolates the effects of that specific variable which can then be studied deeply as the variable is manipulated. In psychoacoustics this translates to studying the properties of pitch using single sine waves, carrying out listening experiments in unnaturally silent rooms (e.g. anechoic chambers), simulating environmental masking noise with synthetically produced noise, etc. All of these experimental situations are far from the reality of environments where one perceives sound and where one interacts with an auditory display. It seems therefore quite difficult to generalise the results from these situations to real situations. On the other hand, it is not clear how to deduce, from a complex experimental situation in which many variables evolve simultaneously, what are the real causes of the perceived effects.

In the field of psychoacoustics, a new approach called the *ecological* approach [Neuhoff, 2004] is addressing these questions by investigating the perception of complex sounds using a variety of experimental techniques borrowed from HCI, psychology, etc. This ecological approach to psychoacoustics, which puts at the centre of the evaluation a realistic situation, i.e. a realistically complex sound and environment, is beneficial to the field of auditory display because it is the perfect candidate field for exploration. Walker and Kramer [2004] observe that the field of auditory display is also beneficial for ecological psychoacoustics because it provides many experimental scenarios and realistic tasks.

4.5.1 Examples of perception studies in relation to auditory displays

Sounds created in auditory display design can be complex, i.e. have various frequencies, noises, etc. in their spectrum. This is the case, for example, for the sounds produced for this thesis. These types of sounds are not the typical stimuli used in psychoacoustics experiments and the study of their effectiveness as auditory displays requires a new approach. There is another category of complex sounds which is studied in order to understand how these sounds portray information and how they can be used in auditory display design. These sounds are referred to as everyday sounds and they are the sounds we all hear in our lives which are produced by our interaction with common objects in use. Examples of these sounds can be the sound of keys jingling, a clock ticking and a record being scratched [Bonebright, 2001]. We use everyday sounds all the time in our life to monitor situations and to understand what is happening around us. We also would like to have a method to establish if a synthesised complex sonification can be effective in portraying information. Learning how complex everyday sounds portray information can give us knowledge about what characteristics a synthesised complex sonification should have to be an effective auditory display. A very interesting study on everyday sounds identification was presented by Ballas in *Common factors in the identification of an assortment of brief everyday sounds* [1993]. This paper firstly points out how little we know about the perception of everyday sounds and then describes a

study on the many factors that contribute to our ability to identify everyday sounds. 41 everyday brief sounds were analysed from a variety of perspectives: acoustic, perceptual and cognitive. The main acoustic parameters were analysed, perceptual and cognitive ratings were obtained, identification responses were analysed for uncertainty and accuracy, and ecological frequency (frequency of occurrence of sound in normal life) were obtained and related to the other results.
Some important knowledge was obtained in this study, in particular:

1. the variance in identification time depends on acoustical variables and ecological frequency;
2. identification relates to the ease with which a mental picture of the sound is made, the familiarity of the sound and the ease of using words to describe the sound.

The main difference between complex synthesised sonifications and everyday complex sounds is that the latter have by definition a high ecological frequency and they normally have a known cause, whereas the former do not have these attributes. However, synthesised sonifications may aspire to become "everyday" sounds through frequent exposure to them and they may become easily identifiable and highly representative of the information being sonified. From this study we can deduce that ecological frequency (which can be created with repetition and training) and acoustical variables influence the identification of a complex sound and that identification is improved if we can easily create in our mind a mental picture of the sound and we can describe it. Similar results are found in [Brazil & Ferström, 2006]. In this study, the presentation of concurrent everyday sounds that can be used as auditory icons is examined. The hypothesis is that if the concurrent sounds have clearly distinct descriptors, then the identification of the concurrent sounds improves. The experiment confirmed the hypothesis.
Following these findings, in experiment 2 (Chapter 8) of this thesis, attention is paid to finding out if the sonification chosen for the EMG data creates an appropriate mental picture in the listeners' mind and if an easy language descriptor can be used to communicate the main changes in the sonification. The results from Ballas' study also support the idea that a theory of sounds' identification will incorporate a variety of factors and therefore will be a hybrid

theory that combines acoustical information with, for example, contextual information.

A multidimensional approach was used by Bonebright [2001] in another study of the perception of everyday sounds. Subjects were presented with 74 everyday sounds which they were asked to sort into different groups of similar sounds (minimum six groups). Subjects could listen to the sounds as many times as they wanted and were asked to fill in a questionnaire about which sounds were the most difficult to sort and which attributes were used to sort the sounds. Another group of participants were then asked to rate the sounds (from 1 to 7) in terms of attributes such as pleasant/unpleasant, interesting/uninteresting, rough/smooth, etc. All the sounds were played once before the rating task began. Sounds were also analysed using acoustic measurements and then these acoustic measurements were correlated to the subjects' ratings. As a result, a 3D timbre space was created that could be used in auditory display design. These results are important for the sonification of EMG data created in experiment 2 (Chapter 8) of this thesis, which produces complex sounds with different levels of roughness. Bonebright's study provides us with knowledge about how we perceive everyday rough sounds and how the roughness relates to other acoustic attributes. In this study significant correlations were found between the following perceptual scales: rough/smooth, tense/relaxed, unpleasant/pleasant. In experiment 2 of this thesis, the roughness of the sonification is related to the activity of a leg muscle. The correlation between rough/smooth, tense/relaxed confirms that representing a muscle movement, and therefore tension, with a rough sound makes sense. The correlation rough/smooth with unpleasant/pleasant is confirmed in experiment 2 as the sonification was not judged to be particularly pleasant.

Since the 90s various studies have been presented that attempt to answer the specific questions posed by auditory displays to the field of sound perception. An example of one of the first studies, that aimed at studying the use of sound in the context of an auditory displays, is Walker and Kramer [1996b] *Mappings and metaphors in auditory displays: an experimental assessment* [for later comments on this paper see also Walker and Kramer, 2005]. In this study, a general control-process task was created for evaluating various mappings and metaphors in terms of accuracy (i.e. effectiveness of the display) and reaction times (i.e. efficiency of the display). The user was presented with a scenario in which he/she had to control the process of crystal-making in a virtual factory. Four parameters of the process were represented by different parameters of sound. A training session allowed the subject to learn how the normal situation sounded and how the problematic situation sounded. During the experimental task, the user (by listening to the display) had to recognise if the situation was critical (e.g. temperature too low) and fix it by pressing the right button (e.g. heater on). The study confirmed that the choice of polarity and data to sound mappings influenced the performance greatly and perhaps surprisingly, the study showed that the mappings and polarities that were thought to be the best by the designer did not result in best performance. This demonstrated the necessity of empirical assessment. This study also represents one of the first investigations that uses a task-based experiment to extrapolate information about the perception of sound in the particular task context. All the experiments presented in this research are task-based as they require the subjects to analyse the sonifications like a real analyst from that specific data domain would be requested to do if the sonification was in use.

Walker [Walker, 2000; Walker, 2002] and colleagues at the Sonification Lab [Sonification Lab] of the Georgia Institute of Technology are building up a large quantity of experimental research in sound perception explicitly related to the context of auditory displays.

Their studies touch various areas of auditory display perception:

- *scalings*: studies on how much a sound parameter should vary to represent a certain variation of the variable it represents. In [Walker, Kramer & Lane, 2000] users were asked to write what change in, say, pitch would represent a certain change in, say, temperature. With these results, perceptual scales could be made which show that different mappings need different scalings and polarities;
- *polarities*: in [Walker & Lane, 2001] it was found that sighted and blind users need different polarities to display some data variable because they can have different mental images of the variable being displayed. For example, sighted people can associate a high amount of money with high pitch, while blind people can associate a high amount of money with low pitch because a lot of money is thought of as being heavy and therefore something that, if left to fall to the ground, would make a low pitched sound when hitting the floor;

- *cognitive abilities*: individual differences and cognitive abilities are studied in the interpretation of auditory graphs [Walker & Mauney, 2004]
- *contextual audio*: contextual audio refers to audio that does not represent data, but helps the interpretation of the display. For example, in an auditory graph, contextual audio would be the audio that represents the tick marks of the graph. Contextual audio was found to be beneficial when it does not clutter the overall auditory information [Smith & Walker, 2002; Nees & Walker, 2006];
- *training*: studies are produced on the usefulness of training to improve auditory display interpretation [Walker & Nees, 2005]

Another example of the investigation of high level sound parameters and meaning can be found in [Palomäki, 2006] in which the meanings conveyed by simple rhythms are studied by asking the subjects to rate them for categories such as pleasant/unpleasant, fast/slow, easy/difficult, familiar/unfamiliar, etc. Relationships between some adjectives and types of rhythms were found and their possible use in auditory earcons design discussed. Similar rating scales (pleasant/unpleasant, uninteresting/interesting, etc.) are used in the experiments presented in this research to verify the appropriateness of the sonifications with respect to what they represent.

From this survey we have learnt that the understanding of how complex sounds portray information requires a multiple approach based on objective and subjective measurements. The study of complex everyday sounds can produce knowledge useful for the design of complex sonifications. Furthermore, task-based experiments are a useful method to evaluate the effectiveness and more in general the usability of an auditory display.

4.6 Conclusions

The design of auditory displays requires knowledge of the acoustical properties of sounds, of their perceptual characteristics and of the typical associations humans make between sounds and meanings. While the understanding of the perception of single simple sounds is quite thorough and well established, knowledge of our perception of complex sounds and concurrent sounds is still insufficient. A multidisciplinary approach to the study of sound perception can help the advancement of knowledge of these complex issues and the field of auditory displays can offer the perfect experimental environment to test hypotheses because it presents the experimenter with urgent problems to solve and goals to achieve in specific and realistic contexts. This type of multidisciplinary and specific research is growing since the 90s and it will contribute greatly to the identification of useful design guidelines for auditory displays. The experiments presented in this thesis aim to improve our knowledge of how much useful information we can derive from complex, and at the same time, realistic sounds and tasks. The research methodology used to design the experiments is detailed in the next chapter.

Chapter 5 – Research methodology

5.1 Introduction

This chapter describes the research methodology used in the work presented in this thesis. Firstly the principal hypothesis of this research is re-stated, then the research approach is explained, and finally the evidence gathering methods and the data analysis techniques are described. The final part of this chapter describes the basic statistical concepts used in this research and the statistical tests used to analyse the experimental data. The aim of this final section is to facilitate the understanding, in Chapters 7, 8 and 9, of the data analysis sections for the reader who is not an expert in statistics.

5.2 The hypothesis

The principal aim of this research is to contribute to the knowledge of how well sound can display multivariate data sets and how the possibility of interacting with such a display can improve the understanding of the data. As a reminder, here is the hypothesis of this research:

> *The combination of user interaction and sonification allows the display of multivariate time-based large data sets effectively and efficiently.*

On the basis of this hypothesis, a research approach was chosen and used to find out if the hypothesis could be accepted or not. Here follows a description of how the specific research method was chosen and why.

5.3 Research approach

The first problem that arises from the hypothesis is the nature of a multivariate time-based data set. In fact, there is an infinite amount of different multivariate data sets and there is no single way to create a representative data set that can be known and used in an experiment. One way to select meaningful data sets is to choose data sets that are produced in domains where there is a request for new displays and analysis systems. As was shown in the previous chapters, there are various such domains: e.g. seismology, vulcanology, electromyography (EMG) analysis, electroencephalography (EEG) analysis, helicopter engineering, etc. Choosing real data (as opposed to artificially constructed data) has the advantage that the results can be at the same time meaningful for the specific domain and also useful for a general understanding of the problem. In this thesis, data from two domains (helicopter engineering and physiotherapy) are studied as examples of fields in which there is a need for improving data analysis and where a sonic display is regarded as a possible alternative to standard techniques. These two domains represent the case studies of this thesis. The generalisation of this research resides in the techniques implemented on the data and their investigation, which are kept totally general and independent from the specific data set studied.

The primary research approach used in this work is an experimental one in which controlled experiments are designed, quantities are measured and the results are statistically analysed. The techniques used to design the experiments come from the fields of psychology, psychoacoustics, social sciences, and usability studies in HCI [Greene & D'Oliveira, 1982; Levitin, 1999; Burns, 2000; Preece *et al.*, 2002]. A secondary qualitative approach complements the experimental one and allows forming a more complete understanding of the research results. Qualitative and subjective information is collected in the form of subjective data, free comments from the test subjects and comments from the domain experts.

5.3.1 Reasons for choosing this approach

The field of sonification is relatively new. Some techniques are well established in the research community (e.g. audification), but they are far from being well established in the domain of information display. This is partially due to the lack of evaluation studies. More quantitative evaluation, together with qualitative evaluation, is necessary to push knowledge and practice forward. This research aims to add a valuable contribution to the studies that evaluate the extent to which sonification, combined with interaction, is effective when multivariate data are displayed. This approach also satisfies two requirements at once: it provides results that can be generalised and used in other domains, and provides results useful in the two specific domains studied in this research. There are three main experiments in this research. The collection of the results from these experiments, plus the qualitative information, contributes to the evaluation of the principal hypothesis of this research.

5.3.2 The experiments

The first two experiments are concerned with evaluating the effectiveness of the sonification displays of multivariate data sets when compared to standard displays and analyses (e.g. visual displays and signal processing analysis), while the third experiment is concerned with evaluating the effect of adding human interaction to a sonification display.

In particular:

1. the first experiment compares the audification of helicopter sensors' data with a standard visual display (spectogram) of the same data;
2. the second experiment compares trends known to be present in electromyographic (EMG) data (discovered through standard signal processing analysis) with trends perceivable by sonification;
3. the third experiment compares three different ways to interactively navigate a complex sonification in terms of their efficiency and effectiveness.

Each experiment is described in detail in Chapters 7, 8 and 9.

5.3.3 Evidence gathering

The multivariate data sets were provided for this project by experts in the two domains. Two general sonification techniques, which could be applied to any large multivariate data set, were used for representing the first and second data sets used in the first and second experiment. Experimental hypotheses for each test were formulated and the experiments designed. The results of the tests, the experimental evidence, were gathered mostly in an automatic way by the software developed by the author for the tests. Additional evidence was gathered using questionnaires. For the third experiment, a prototype application was developed, and objective and subjective efficiency and effectiveness measured. More evidence was gathered with a questionnaire. Qualitative information was gathered also through informal interviews with the domain experts.

5.3.4 Data analysis

The gathered data underwent statistical analysis. Specific software was used to calculate the statistical analyses: SPSS [SPSS software] and G*Power [GPower software]. Comments from test subjects and domain experts were analysed mainly by reflecting on their general agreement or disagreement with the experimental results.

5.4 Summary of main statistical tests and concepts used in this thesis

In the following paragraphs the fundamental statistical concepts used in the analysis of the experimental results are summarised in order to facilitate the reader's understanding of the chapters that follow which describe the experiments in detail. The information is summarised from [Field, 2005].

5.4.1 Experimental designs

There are two main types of experiments:

1. those in which the differences in results between two or more conditions are compared ("*comparing means*" experiment);
2. those in which the existence of a relationship between two variables is verified ("*correlational*" experiment).

Both types of experiments can be designed in two ways:

- *between-subjects* or *unrelated* design;
- *within-subjects* or *related* design.

In the between-subjects design, different groups of subjects perform the experiment in each experimental condition. The disadvantage with this design is that often a large number of people are required to perform the experiment for the results to be statistically significant.

In the within-subjects design, the same group of subjects performs the experiment under all the conditions. This design has the advantage that a smaller number of subjects is required for the experiment. The disadvantage is that the results can show *order* effects if the experimental conditions are presented to each subject in the same order. One way to minimise this type of error is to randomise the order of presentation of the conditions each time they are presented to a new subject. All the experiments conducted in this thesis use the within-subjects design.

5.4.2 Basic statistical concepts

A researcher running an experiment attempts to find out if a phenomenon that is hypothesised to be present in a certain population is really present or not. Usually the whole population cannot be tested (e.g. the whole world population cannot be asked to do a listening test!), therefore only a small sample of the population is tested. The results produced by the small sample are used to infer the results for the whole population. Obviously a large sample gives a better approximation of the whole population, while a small sample can be insufficient to give us information on the whole population. As we will see later in this chapter, statistical tests tell us if the results found using a small number of subjects can be used to infer results for the whole population.

Once the researcher has gathered the results from all the people in the sample, he/she can calculate two fundamental statistical parameters: the *mean* and the *standard deviation*. The *mean* is calculated with the formula:

$$mean = \bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{5.1}$$

where:

N is the number of subjects in the sample;

x_i is the experimental result for subject 'i'.

The mean is a simple statistical model which represents the summary of the sample results. To determine if the mean is an accurate model of the sample results, it is necessary to calculate how much each individual result deviates from the mean in average. The measure of this is the *standard deviation* 's':

$$s = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}} \tag{5.2}$$

where:

$x_i - \bar{x}$ is the *deviance*;

$SS = \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})$ is the *sum of squared errors*;

$s^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}$ is the *variance*.

The standard deviation measures the average deviation from the mean. Thus, a small standard deviation relative to the mean tells us that the data points are close to the statistical model (i.e. the mean), therefore the mean can be considered a good summary of the sample results.

Another useful way to look at the results of a test is to plot how many times a certain result has occurred. This creates a *histogram* or *frequency distribution* graph. Frequency distributions can be shaped in various ways, but the most important one is the *normal distribution* which is bell-shaped and symmetrical around the central result. The *normal distribution* is very important because, in many cases, results of measurements of a variable will have this distribution, if the only factors that make the measurements differ from one another are small independent effects.

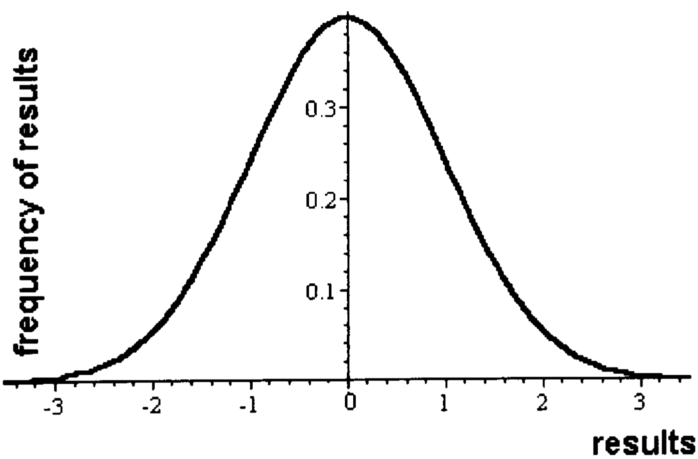


Figure 5.1: Normal distribution

A distribution can deviate from the *normal* in two main ways:

- by not being symmetrical (*skewness*), i.e. the most frequent scores are at one end of the scale;
- by having too narrow a peak (*kurtosis*), i.e. distribution tails off too rapidly or not rapidly enough.

The standard deviation of a sample's results tells us also about the shape of the distribution of these results and how much it deviates from the normal distribution. For instance, if we have two distributions with the same mean and one has a large standard deviation while the other has a small standard deviation, then the first distribution will be more spread and flatter, while the second will be pointier.

The frequency distribution (given that we have tested enough people and the distribution is a good representation of the population results) also tells us about the probability of a new result, or score, occurring.

Statisticians have calculated the probability of new scores occurring when the already obtained results have a normal distribution with mean 0 and standard deviation 1. These probabilities are published in tables within statistics books.

Fortunately the results of any test can be linearly transformed into results having mean 0 and standard deviation 1. This allows us, by looking at the probability table, to know how probable it is for a new result to occur for any test.

The transformed results are called *z-scores* and the linear transformation is the following:

$$z_i = \frac{x_i - \bar{x}}{s} \tag{5.3}$$

where:

z_i is the i[th] z-score;

x_i is the original i[th] result;

$\bar{x}$ is the mean of the original results;

s is the standard deviation of the original results.

We can see from the demonstration in 5.4 that, after this transformation, the *mean* of the z-scores is 0:

$$\bar{z} = \frac{\sum_{i=1}^{N} z_i}{N} = \frac{\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{s}\right)}{N} = \frac{1}{s}\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)}{N} = \frac{1}{s}\left(\frac{\sum_{i=1}^{N} x_i}{N} - \bar{x}\right) = \frac{1}{s}\left(\bar{x} - \bar{x}\right) = 0 \tag{5.4}$$

From the demonstration in 5.5, we can see that the *standard deviation* of the z-scores is 1:

$$s^2 = \frac{\sum_{i=1}^{N}\left(z_i - \bar{z}\right)^2}{N-1} = \frac{\sum_{i=1}^{N} z_i^2}{N-1} = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{s^2(N-1)} = \frac{1}{s^2}\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N-1} = \frac{1}{s^2}s^2 = 1 \tag{5.5}$$

From the pre-calculated tables, it is found that 95% of the z-scores lie between

-1.96 and 1.96, and 99% of the scores lie between -2.58 and 2.58. This means that, after linearly transforming the z-scores back to the original values, a new score will have 95% probability to lie between ($\bar{x}$ - 1.96*s) and ($\bar{x}$ + 1.96*s) and 99% probability to lie between ($\bar{x}$ - 2.58*s) and ($\bar{x}$ + 2.58*s).
This result has an important role in determining if our experimental results are significant. Let us see how.
To know if the result obtained by testing a sample of the population is representative or not of the population as a whole the *standard error* or SE (not to be confused with the standard deviation) needs to be calculated. To understand the concept of SE let us imagine the following situation: if the whole population is divided into different groups and the experiment is run on all the groups, then different groups of the population will produce different means of their respective results. The *standard error* is the standard deviation of the groups' means from the whole population real mean. If the standard error is small then the particular population group is representative of the population because the mean of that sample is close to the real mean of the population. The standard error of a population 'x' can be calculated from the standard deviation of the group and the number of subjects in the group.
The *standard error* of a group is:

$$SE_x = \frac{s_x}{\sqrt{N}} \qquad (5.6)$$

where:
s is the standard deviation of the group;
N group size.
This means that, if we test another group, its *mean* has 95% probability to lie between ($\bar{x}_{real}$ - 1.96*SE) and ($\bar{x}_{real}$ + 1.96*SE), i.e. the grey area in Figure 5.2.

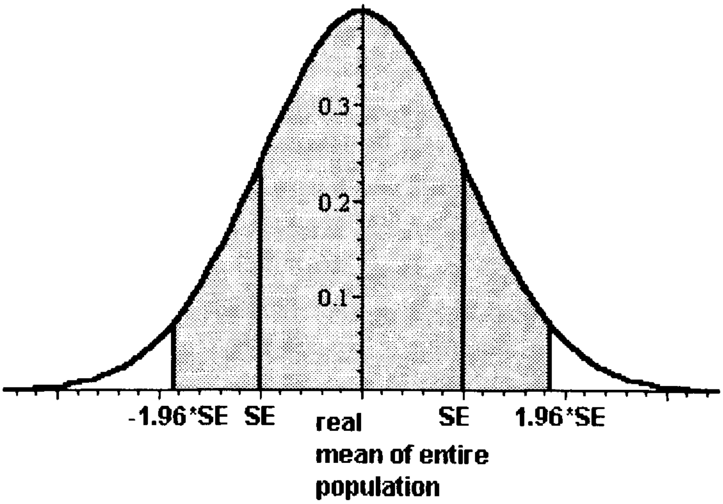


Figure 5.2: Distribution of *means* of groups from a population

It is now useful to look at this probability interval from the point of view of the single group tested. We call *confidence interval* of a distribution of a group's results an interval of values which has the following boundaries:

$$Lowerboundary = \bar{X} - (1.96 * SE) \tag{5.7}$$

$$Upperboundary = \bar{X} + (1.96 * SE) \tag{5.8}$$

Where:

$\bar{x}$ is the mean of the group;

SE is the standard error of the group.

The 95% *confidence interval* is the interval of values within which the real mean of the population has a 95% probability of being found.

In Figure 5.3, we can see that the groups that have their *mean* lying in the grey area, e.g. point B, (remember that groups have 95% probability of lying in the grey area) include the real mean of the population within their confidence interval. Groups that have their *mean* lying outside the grey area, e.g. point A, (remember that groups have 5% probability of not lying in the grey area) do not include the real mean of the population within their confidence interval.

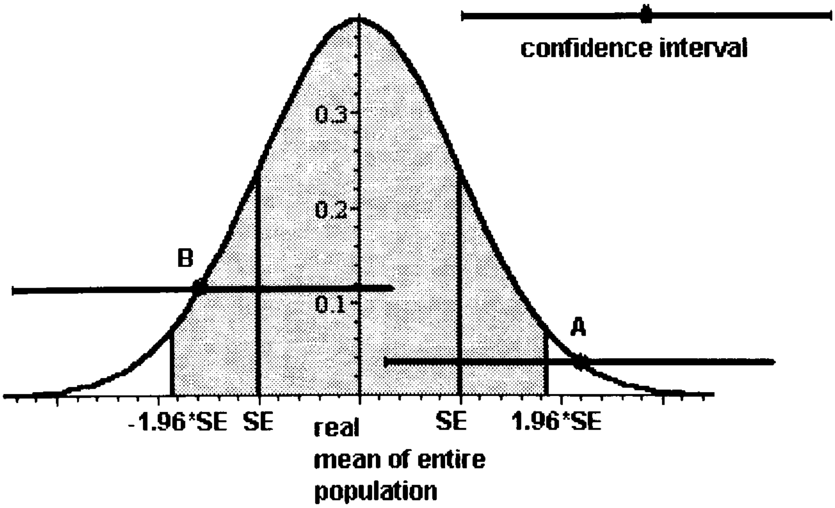


Figure 5.3: This figure shows the cases when the real mean of the population lies within the confidence interval (case B) and when it does not lie within the confidence interval (case A)

95% is the *confidence interval* usually used in experiments that do not require a high confidence interval, while a 99% *confidence interval* is normally used when a stricter confidence interval is necessary. In this thesis, all the experiments use the 95% confidence interval.

When displaying a mean result in a graph, it is very important to show the confidence interval for that mean result. The way to do this is to show the mean result of the test with two error bars (one negative and one positive) of value '1.96*SE' (absolute value). We can be then 95% confident that the real mean of the population is within the error bar's interval. As a consequence, when two mean results are compared, we can only be sure that they are significantly different when the error bars of the two means do not overlap [Streiner, 1996]. In fact, if the error bars do overlap then there is a possibility that both the two real means lie within the overlap and therefore they could have the same value, i.e. we cannot be confident that they differ.

More in general, at this point, we need to have a method to establish if differences in results, obtained when we measure the same variable under different conditions, are definitely due to the changes in conditions and not to chance.

If there is less than 5% probability that an effect occurs by chance, then the effect is considered *statistically significant*. The test that measures the probability of an effect occurring by chance is called the *significance test*.

To calculate the significance test a *test statistic* is calculated.

The *test statistic* can be explained as follows: in an experiment there are two types of variance:

- *systematic variance*, i.e. the variation that can be explained by the model that was fitted to the data;
- *unsystematic variance*, i.e. variation that is not explained by the model.

A *test statistic* (there are various test statistics e.g. T or Chi-square, etc.) generally can be described as:

Test statistic = (systematic variance) / (unsystematic variance) (5.9)

In a significance test, the *test statistic* is calculated and is expected to be higher than 1, i.e. the systematic variance should be higher than the unsystematic variance.

For example [example taken from Greene & D'Oliveira, 1982], a group of people is asked to read two texts: one simple and one complex. Ten minutes later they are asked to write down the words that they remember from the two texts. The hypothesis is that they will remember more words from the simple text than from the complex one. To calculate the *test statistic*, the variance only due to the presence of the two conditions is calculated (systematic variance) and it is compared to the overall variance in the test (unsystematic variance). If the systematic variance is much higher than the unsystematic variance then the difference in results between the two conditions can be considered significant.

Statisticians have already calculated the probability of getting a certain *test statistic* by chance (i.e. only due to unsystematic variance). Therefore once the *test statistic* of an experiment has been calculated, it can be looked up on the table that reports the probability of that *test statistic* occurring. If the probability of getting that *test statistic*'s value is less than 5%, then the model (in the previous example, the difference in text's complexity) is said to explain a sufficient amount of variation and the experimental hypothesis can be accepted, while the null hypothesis (negation of the experimental hypothesis) can be rejected.

Significance tests are very important for judging our results; however, a significant test statistic does not tell us anything about the importance of the effect. Very small and unimportant effects can be significant. Furthermore, a non-significant *test statistic* does not mean that the null hypothesis is true. A

non-significant *test statistic* only tells us that the effect is not big enough to be anything other than a chance finding. Therefore a significant test only tells us that the null hypothesis is highly unlikely not that it is completely false.
Test statistics can be said to be one-tailed or two-tailed.
When the experimental hypothesis has one precise direction (e.g. eating makes you fat), the test statistic is said to be *one-tailed*. When the experimental hypothesis has two possible directions (going to the cinema will affect the intention of going to the cinema in the future: the intention will either increase or decrease), then the test statistic is said to be two-tailed.
In a one-tailed test, the 5% probability that the *test statistic* value was produced by chance is all considered in one tail of the *test statistic* probability curve.
In a two-tailed test the 5% probability that the *test statistic* value was produced by chance is split in a half between the two tails of the *test statistic* probability distribution.
Now, once the results of an experiment are analysed, two errors can occur:

- *Type I error*: we believe there is a genuine effect in our population when in fact there is not. The probability of this error occurring, after the result was found to be significant, is 5%. This value is called the *α-level*.
- *Type II error*: we believe there is no effect in a population when in fact there is. This would occur when the test statistic is small, i.e. there is a lot of natural variation. The maximum acceptability of this type of error should be 20% [Cohen, 1988] and this is called *β-level*.

An effect can be significant, but its size can be very small. There are a standardised ways to measure the size of an effect.

Common measures of effect sizes are:

- Pearson's correlation coefficient *r*
- Cohen's *d*

Cohen (1988, 1992) made widely accepted suggestions of what constitutes a large or small effect [Field, 2005; p. 32].
The effect sizes reported in this thesis were calculated using the free software G*Power [GPower] recommended in [Field, 2005]. The effects are considered

small, medium or large following Cohen's guidelines and the limits are the following.

Correlations effect size r:

$0.10 < r < 0.30$ small

$0.30 < r < 0.50$ medium

$r > 0.50$ large

Comparison of two means effect size d:

$0.20 < d < 0.50$ small

$0.50 < d < 0.80$ medium

$d > 0.80$ large

Comparison of more than two means effect size f:

$0.10 < f < 0.25$ small

$0.25 < f < 0.40$ medium

$f > 0.40$ large

The effect size in a population is dependent on:

- the sample size (number of participants) on which the effect size is based;
- the probability level at which a result is accepted to be significant: α-level (normally 5%);
- the ability of a test to detect an effect of that size, i.e. the *statistical power*;
- whether the test is one or two-tailed.

The statistical power of a test is the probability that a given test will find an effect assuming that one exists in the population. Its probability is the opposite of the probability of missing an effect when there is one, i.e. 1-β (remember that β is the probability of missing an effect when there is one). Given α, the effect size based on the sample and the number of participants (sample size), the statistical power can be calculated. Cohen in (1988) gives a guideline on the level of statistical power a test should have. A test that aimed at obtaining a result with a 0.05 significance level, should have a statistical power equal or higher than 80%. This means that the chance of making error Type II (not detecting an effect that actually exists) is 4 times bigger than the chance of making error Type I (considering an effect significant when it actually does not exist). This 4 to 1 ratio between the two types of error means that falsely detecting an effect is considered a mistake four times more important than not detecting an effect that exists. This makes sense because claiming that an effect exists when it does not is more dangerous than not detecting an effect that can anyway be detected in a following experiment that uses a larger sample (and therefore has a bigger power).

Obviously the size of the sample tested is an important parameter to achieve accurate and relevant results from the test we are performing.

Given α and β, past research can be used to estimate the effect size one hopes to detect in an experiment. These three values allow one to calculate a-priori the sample size necessary to detect that effect. Appendix B of this thesis shows the a-priori power analyses for the experiments presented here. Often compromises have to be made when choosing the real sample size used in an experiment due to subjects availability for the test. In this research, tests with a statistical power higher than 60% that produce effects with a significance level of 5% (Type I error/Type II error ratio of 8) are considered marginally significant if they are consistent with results, that have a statistical power of 80% and significance level of 5%, from other similar tests done with the same subjects

Now that all the important statistical parameters have been described, we need to describe the different types of statistical tests available.

There are two main types of statistical tests: *parametric* tests and *non-parametric* tests.

Parametric tests can be run when the experimental results satisfy three main assumptions:

- they are *normally distributed data*;
- they have *homogeneity of variance*, i.e. the variance can be considered the same for all data;
- they are *interval data*, i.e. distance between points of the scale used to measure the data should be the same at all points along the scale.

The third assumption is tested only by common sense, while statistical tests can be used to test assumptions 1 and 2.

One objective way of testing if a distribution is normal or not is to compare the scores in the sample with normally distributed scores that have the same mean and the same standard deviation. The *Kolmogorov-Smirnov test* (*K-S test*) and the *Shapiro-Wilk test* do this. The tests are similar, but the Shapiro-Wilk test is considered more accurate than the K-S test [Field, 2005; p. 560], so if the K-S test says that the results are normally distributed, it is better to run the Shapiro-Wilk test also to check this in more detail.

When the normality tests are not significant ($p > 0.05$), it means that the results' distribution is not significantly different from a normal distribution, if the tests are significant ($p < 0.05$) then the distribution is significantly different from a normal distribution.

Standard way of reporting the results of the normality tests:

> $D(df) = value, p < value$
>
> e.g. $D(100) = 0.10, p < 0.05$
>
> D = test statistic
>
> df = degrees of freedom = sample size
>
> p = significance

For groups of results (e.g. when comparing various means), homogeneity of variance means that the variance in each group should be the same.

For correlation experiments, homogeneity of variance means that the variance of one variable should remain the same at each level of the other variable.

The *Levene's test* allows measuring the homogeneity of variance for groups of data. The Levene's test verifies the hypothesis that the variances in the groups are equal. If the Levene's test is not significant ($p > 0.05$) it means that the variances are not significantly different and homogeneity of variance can be assumed. If the Levene's test is significant ($p < 0.05$), then homogeneity of variance cannot be assumed.

Standard way of reporting the results of the Levene's test:

> $F(df1,df2) = value, p < value$
>
> e.g. $F(1, 98) = 7.37, p < 0.01$
>
> F = test statistic
>
> df1 = degrees of freedom 1 = number of conditions -1
>
> df2 = degrees of freedom 2 = total number of observations – number of conditions
>
> p = significance

Finally, if the experimental results do not satisfy the above parametric assumptions, then non-parametric significance tests can be used. For example, when the data are measured in an ordinal scale (as supposed to an interval scale) or are not normally distributed, then non-parametric tests need to be used [Field, 2005; p.129]. Non-parametric tests make fewer assumptions on the data than their parametric counterparts. In non-parametric tests the data are ranked, i.e. given an order value, and the test is run on the ranks not on the absolute value of the data.

Using ranks instead of the actual data means that the information about the difference between scores is lost. The results of non-parametric tests can be less powerful than the results from parametric tests on data that are normally distributed. Therefore, if the data are normally distributed and satisfy all the other requested assumptions parametric tests should be preferred to non-parametric tests. However, if the data do not satisfy the requested assumptions and in particular, they are not normally distributed, then the parametric tests should not be used and the non-parametric tests are the only tests that can be used [Field, 2005; p. 533].

All the significance tests calculated in this thesis are non-parametric tests as the majority of the results' scales are not interval scales or the data are not normally distributed. Here follows a summary of the tests used in this thesis and a description of the way they are reported.

Comparing two related conditions: the Wilcoxon's signed-rank test

This test is used when there are two sets of scores to compare and they come from the same group of participants (within-subjects design). This test is based on calculating the differences in score between the two conditions for each subject and their ranking. The test statistic of this test is called T before it is linearly transformed into its z-score Z (i.e. is normalised). After Z is calculated its significance is determined.

The standard way of reporting the results of the Wilcoxon's signed rank test is:

Z = value, $p <$ value

e.g. $Z = -2.53$, $p < 0.05$.

Z = test statistic

p = significance

Comparing several related conditions: Friedman's ANOVA

This is a non-parametric test used to test differences between several experimental conditions when the same participants have been used under all conditions. This test is also based on ranked data. The test statistic of this test is called Chi-Square.

The standard way of reporting the results of the Friedman's ANOVA test is:

Chi-Square(df) = value, p < value

e.g. Chi-Square(2) = 0.2, p > 0.05

Chi-Square = test statistic

df = degrees of freedom = number of conditions -1

p = significance

Post-hoc tests for Friedman's ANOVA

If the Friedman test is significant, then the next step is to go into more detail and find out which pair of conditions' comparisons are significant. One way to do this is to run a Wilcoxon signed-ranked tests for each pair.

In this case, the level of significance for the post-hoc Wilcoxon tests needs to be adjusted so that the overall significance level for the Friedman's ANOVA remains 5%.

For example, if the number of pairs is 3, the level of significance used in each Wilcoxon test is 5% and the tests are considered independent, then the overall probability of not having Type I errors is 0.95 to the power of 3 which is 0.857 and the overall significance for the Friedman's ANOVA becomes 14.3% which is too high (much higher than 5%).

One way to correct this is by accepting a level significance for each pair of comparisons of 0.05 divided by the number of comparisons (i.e. divided by 3, if we have 3 conditions, divided by 6, if we have 4 conditions, etc.). This is referred to as *Bonferroni correction*.

Pearson's and Spearman's correlations

Correlation is a measure of the linear relationship between variables. Important concepts in the study of correlations are the *covariance* and the *correlation coefficient*. If two variables are related, a change in one variable should correspond to a change in the second variable. Therefore if a variable deviates from the mean the second variable should also deviate from the mean in a similar way. One way of checking that this occurs is to calculate the covariance.

$$\mathrm{cov}(x,y) = \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{N-1} \qquad (5.10)$$

where x_i is the i^{th} score for one variable, y_i is the i^{th} score for the second variable, $\bar{x}$ is the mean of the scores of the first variable, $\bar{y}$ is the mean of the scores of the second variable, N is the number of scores.

A positive covariance indicates that when one variable deviates from the mean the second variable does the same in the same direction. A negative covariance indicates that the two variables deviate similarly, but in opposite directions.

The covariance depends on the measurement scales used to measure the data. The covariance can be made to be independent of the scales, i.e. can be standardised, and the standardised version of the covariance is the *Pearson correlation coefficient*:

$$r = \frac{\mathrm{cov}_{xy}}{s_x s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(N-1) s_x s_y} \qquad (5.11)$$

where the 's_x' and 's_y' are the standard deviations of the two variables.

r ranges between -1 and 1, with 1 maximum positive correlation, -1 maximum negative correlation, 0 no correlation.

The Pearson's correlation coefficient assumes that the data used to calculate it is normally distributed.

The Spearman's correlation coefficient r_s is the non-parametric version of the Pearson's correlation coefficient. It is used when the data do not satisfy the parametric assumptions. In this case, the data are ranked and then the Pearson's correlation coefficient is calculated on the ranks.

The standard way of reporting the results of the Pearson's correlation coefficient is:

r = value, p(one-tailed) < value

e.g. r = 0.56, p(one-tailed) < 0.05

r = Pearson's correlation coefficient

p = significance for two-tailed test

p(one-tailed) = significance for one-tailed test

The standard way of reporting the results of the Spearman's correlation coefficient is:

r_s = value, p(one-tailed) < value

e.g. $r = 0.80$, $p < 0.05$

r = Spearman's correlation coefficient

p = significance for two-tailed test

p(one-tailed) = significance for one-tailed test

5.5 Conclusions

This chapter described the research methodology adopted in this work. This methodology is based both on the objective results of experimental tests and the analysis of gathered qualitative information. The experiments that were run as part of this work are described in relation to the main hypothesis of this thesis. The final section of this chapter described the relevant statistical concepts and tests used in this thesis with the aim of facilitating the understanding of the test results analysis which follows in Chapters 7, 8 and 9.

Chapter 6: The Interactive Sonification Toolkit

6.1 Introduction

This chapter describes the Interactive Sonification Toolkit, a prototype data-analysis software application, based on sonfication, which was built during this research project. The purpose of this prototype was to create an environment in which various sonification mappings could be quickly demonstrated and tested, and in which various interactive navigation techniques could be evaluated. More specifically, this toolkit has been used to create the sonifications for experiment two (Chapter 8) and to create the sonifications and interaction methods for experiment three (Chapter 9).

6.2 A review of software environments suitable for the implementation of interactive sonification systems

A general interactive sonification system is a complex, open system that consists of several components [Hunt *et al.*, 2004] which require tight interaction, are computationally expensive, and often demand special platforms. The main components of such a system are:

1. A component that allows *data-related computations.*

   The data to be analysed could enter the system either in real-time or as a recorded file. The simplest form of data computation that the system should perform is linear rescaling of the data, but much more complex computations could be necessary such as, for example, the computation of the principal dimensions of a multidimensional data set. Data computations need to be done as fast as possible so that the interaction with the data and the sonification can be perceived by the user as real-time. A simple example could be that as the user selects different sections of the data, these sections are automatically re-scaled to the maximum perceivable range for their immediate sonification.
2. A *sound synthesis component.*

   This component can also be subdivided into two main parts. One part should manage the events that happen at control rate, i.e. a data update rate less than 1kHz. Examples of such events are note triggering, changes in volume, changes in cut-off frequency of a filter, etc. The other part should manage events that happen at audio rate, i.e. an update rate higher than 1kHz. To be more precise this rate is normally 44100 events per second in audio. Examples of such events are: audio signal playback, amplitude and frequency modulation data in AM and FM synthesis respectively, etc.

3. An *interactive component*.

   This part allows the user to interact with the sonification via more or less complex external interfaces. An example of this is to use the mouse to navigate the sonification.
4. A *Graphical User Interface* or *GUI component*.

   This component guides the user on how to use the system.
5. An *optional visual component*.

   If the system allows data visualisation as well as data sonification, then a visual component would be necessary.

It is clearly difficult to find a programming environment which allows the efficient and rapid implementation of all these components to build a prototype. Often programming environments have been developed with one main application in mind and whilst they are appropriate to be employed for the development of one component, they are not as appropriate to be employed for other components. The Toolkit to be used in this research needs to:

- allow some basic data computations, such as scaling data;
- allow sound synthesis;
- allow easy connection with external interface devices for interaction;
- allow the programming of a simple GUI;
- allow some simple data visualisation.

The Toolkit should also preferably be:

- platform independent;
- extendible;
- free and open source.

Below follows a description of various programming environments that could be suitable to create such a prototype.

6.2.1 Pure Data, Max/MSP and Jmax

In the 1980s, a particular type of software for music was developed. Central to the development of this application is the concept of mapping. The original software was called Max. From this software other very similar applications developed: Pure Data (or PD) [Pure Data programming environment], Max/MSP [Cycling74] and Jmax [JMax].

The development of the Max graphical environment began at IRCAM [IRCAM], a music research institute in Paris (France) by Miller Puckette in the mid-1980s. Puckette programmed an application called 'Patcher' for the realisation of composer Philippe Manoury's piece Pluton [Puckette, 1988]. This piece needed a system to allow the computer, and therefore the electronic part of the piece, to follow in real-time the score during the performance [Puckette & Lippe, 1992]. While the Max system was in general a set of classes that defined objects and how they interacted, 'Patcher' represented the graphical environment in which Max objects could communicate by passing messages. It was created for making real-time computer music, initially, only using MIDI controllable synthesisers [Puckette, 1988]. In Max, a 'box' represents an object or function and the 'inlets' and 'outlets' of the box represent its inputs and outputs. Inlets and outlets of the various objects can be connected, and strings represent these connections graphically (see Figure 6.1). The connections represent the passage of messages between objects. 'Patcher', running on a Macintosh, did the MIDI and the control processing, while the computer did the audio processing [History of Max].

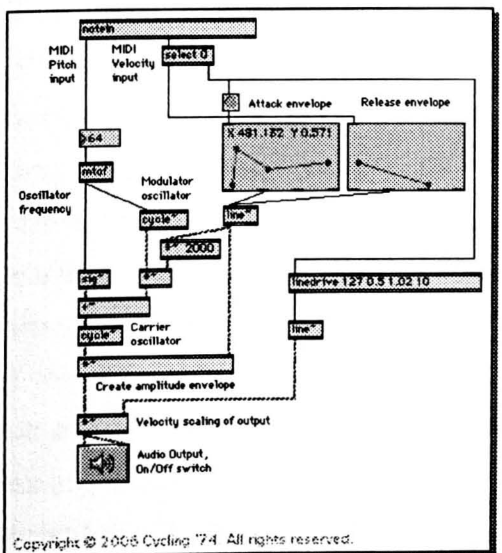


Figure 6.1: A patch in MaxMSP. Image taken from [Cycling '74]

Various composers such as Michael Jarrel and Philippe Manoury contributed to the development of Max, using it since the beginning of its development [Puckette, 1988]. Puckette's program was licensed to the company Opcode Systems, Inc. and became known as Max/Opcode. At this point, developer

David Zicarelli introduced many enhancements in the GUI and the 'external objects' development kit. In 1989, IRCAM started work on a real-time synthesiser card for the NeXT computer called the Ircam Signal Processing Workstation (ISPW). This was under the IRCAM Musical workstation project. Puckette ported Max to the NeXT and ISPW and added a set of objects to process audio on the card. Max combined with the audio processing was known as Max/FST [History of Max]. In 1996 Miller Puckette began developing another program called PD (Pure Data) [Pure Data programming environment] at the University of California (San Diego, California). A goal of PD was to do real-time signal processing by connecting Max objects together. Shortly after the PD development started, David Zicarelli decided to add audio processing for PowerPC computers to the existing Opcode Max environment. Max/MSP is the result of these additions. MSP represents a set of objects in Max that deal with audio processing in real-time. Since 1999, Max/MSP has been published and supported by Cycling '74 [Cycling74]. At IRCAM, the development of the 'Max concept' continued through the development of the software JMax [JMax]. At present Max/MSP is a commercially available program for Macintosh and for Windows, while PD and JMax are free programs working on both PCs and Macintosh under Windows, OS and Linux operating systems.

These programs have characteristics that make them particularly effective when prototyping an application that requires real-time sound processing and control. First of all, they have built-in functions for real-time sound synthesis and connections with peripherals such as MIDI devices or devices that connect to the computer via the serial or USB port. They also have functions for sending messages through a network. Moreover, these programs allow visual and real-time programming [Edmonds *et al.*, 2003].

They are visual programming environments because functions, messages and data flows are represented graphically by boxes and strings on screen. This means that a PD program (called a 'Patch') looks like a data flow diagram. An example of a visual programming environment which has similar characteristics is LabVIEW (National Instrument Corporation, 1990) [LabVIEW] which is designed for engineers and scientists for the development of data acquisition, analysis, display and control applications. PD, Max/MSP and JMax are also real-time programming environments because the result of any change in the program can be heard immediately: there is no need for a compiling phase.

These two characteristics support fast prototyping so that many algorithms or prototype configurations can be tested rapidly and final design decisions can be made on the basis of many test results. These characteristics also support creativity [Edmonds *et al.*, 2003], which is a fundamental element not only in the artistic field, but also in software design. PD, in particular, is freely available, it is open source and available for both PC under Windows and Linux, and Macintosh under OS. PD is extendible as new objects can be programmed in C, and the code of existing objects can be used and changed. PD is a widely used environment and has a large community of users that help supporting any development of PD.

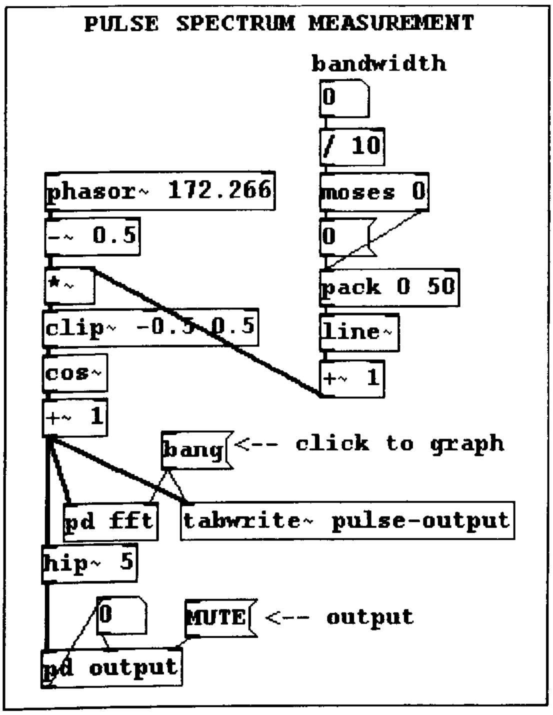


Figure 6.2: A patch in PD. Screen shot of a PD example patch

6.2.2 Neo/NST

Neo/NST [Neo/NST] is a graphical simulation environment especially well suited for applications in the domain of artificial neural networks and data mining. Complex data computations are easily done in this environment which is, however, less powerful in other domains. It was developed by Prof. Dr. Helge Ritter at Bielefeld University in Germany. It provides a graphical programming environment, similar to the Max concept, which allows for rapid prototyping. It is extendible because new functions/objects can be programmed in C or C++. Collection of objects can be organised in libraries. It provides very good visualisation and 3D graphics. It contains functions for data uploading, processing and data mining. Some sound functions for sonification have been developed by researcher Thomas Hermann [Hermann], however this environment allows only basic real-time sound synthesis. This environment is available for Linux and is free.

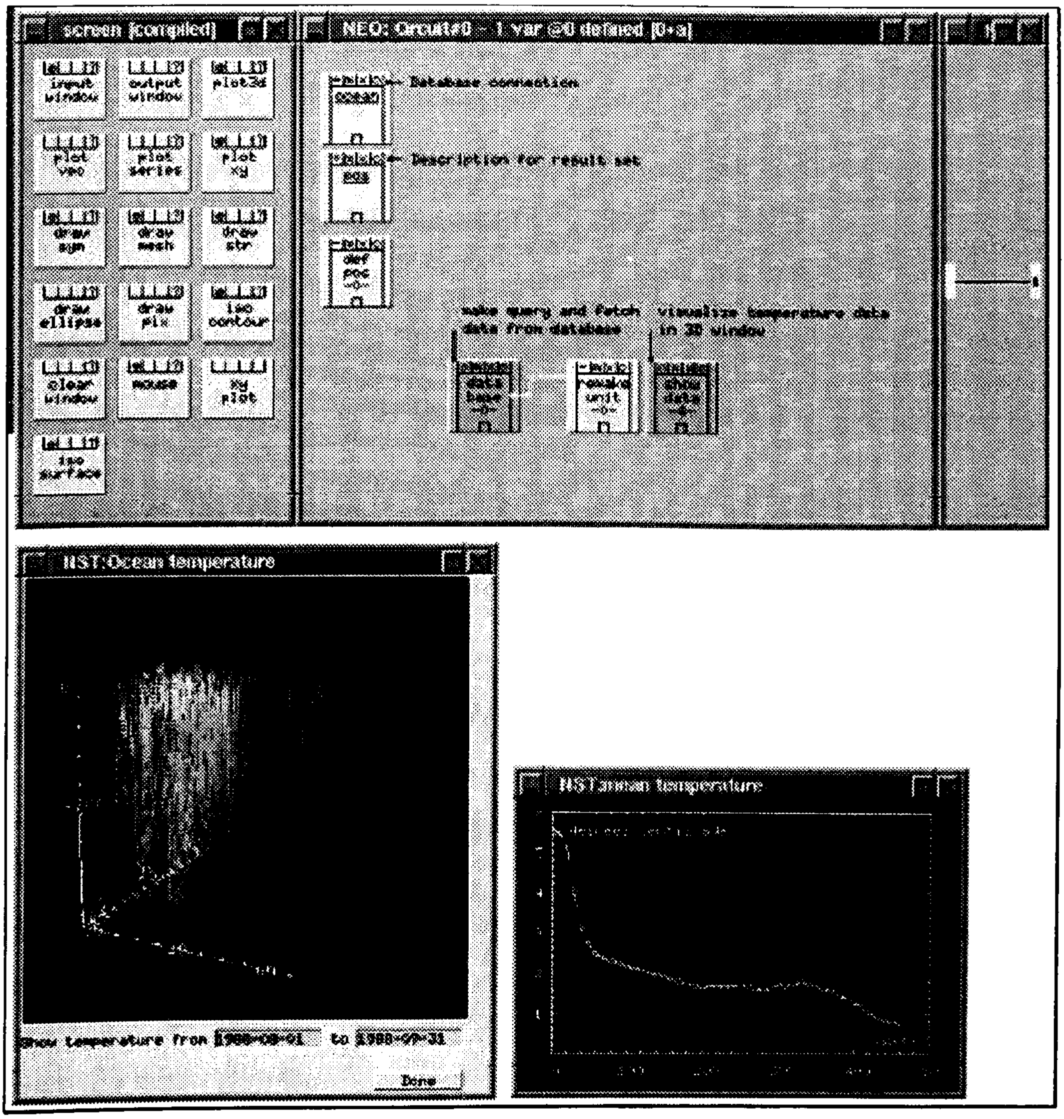


Figure 6.3: Neo/NST. Image from [Neo/NST]

6.2.3 Csound

Csound [Csound] is a text-based programming language. Real-time programming is not available and a compiling phase is needed. It is freely available for Windows, Mac and Linux. A program compiled in Csound can be controlled in real-time via MIDI. Csound programming can be quite time consuming as it is text based. Csound was developed to create complex digitally synthesised sounds. There are no in-built complex mathematical functions for generic data.

```
<CsoundSynthesizer>

<CsOptions>
-A -o csound_1.aif
</CsOptions>

<CsInstruments>
        instr   1
a1      oscil   p4, p5, 1
        out     a1
        endin
</CsInstruments>

<CsScore>
; Table
f       1       0       8192    10      1
; Note  Instr   Start   Dur     Amp     Freq
i       1       0       1       10000   440
</CsScore>

</CsoundSynthesizer>
```

Figure 6.4: Csound file. Image from [Andy Kopra]

6.2.4 Supercollider

Supercollider [Supercollider] is a high level programming language similar to Smalltalk [Smalltalk]. It has a program text editor and a GUI builder so that panel controls can be created. It is free and available for Mac and recently for Windows. This programming platform is used in various sonification projects (see recent projects by Thomas Hermann [Hermann] and colleagues), however complex data computations are generally performed in other programs.

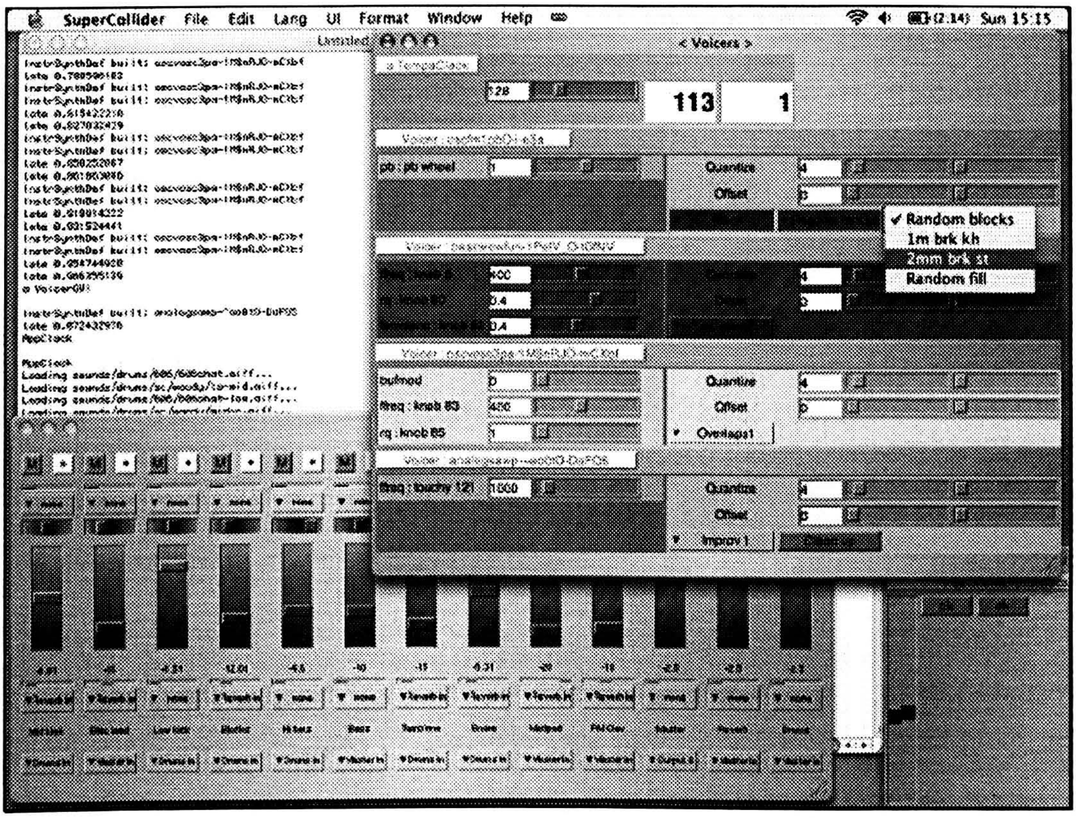


Figure 6.5: Supercollider. Image from [Super Collider]

6.2.5 Matlab

Matlab [Matlab] is a commercially available programming environment from Mathworks. It is relatively expensive and does not allow real-time programming. However, it is very effective for data processing, sound processing and data visualisation. It is platform independent and available for PCs, MACs and UNIX systems.

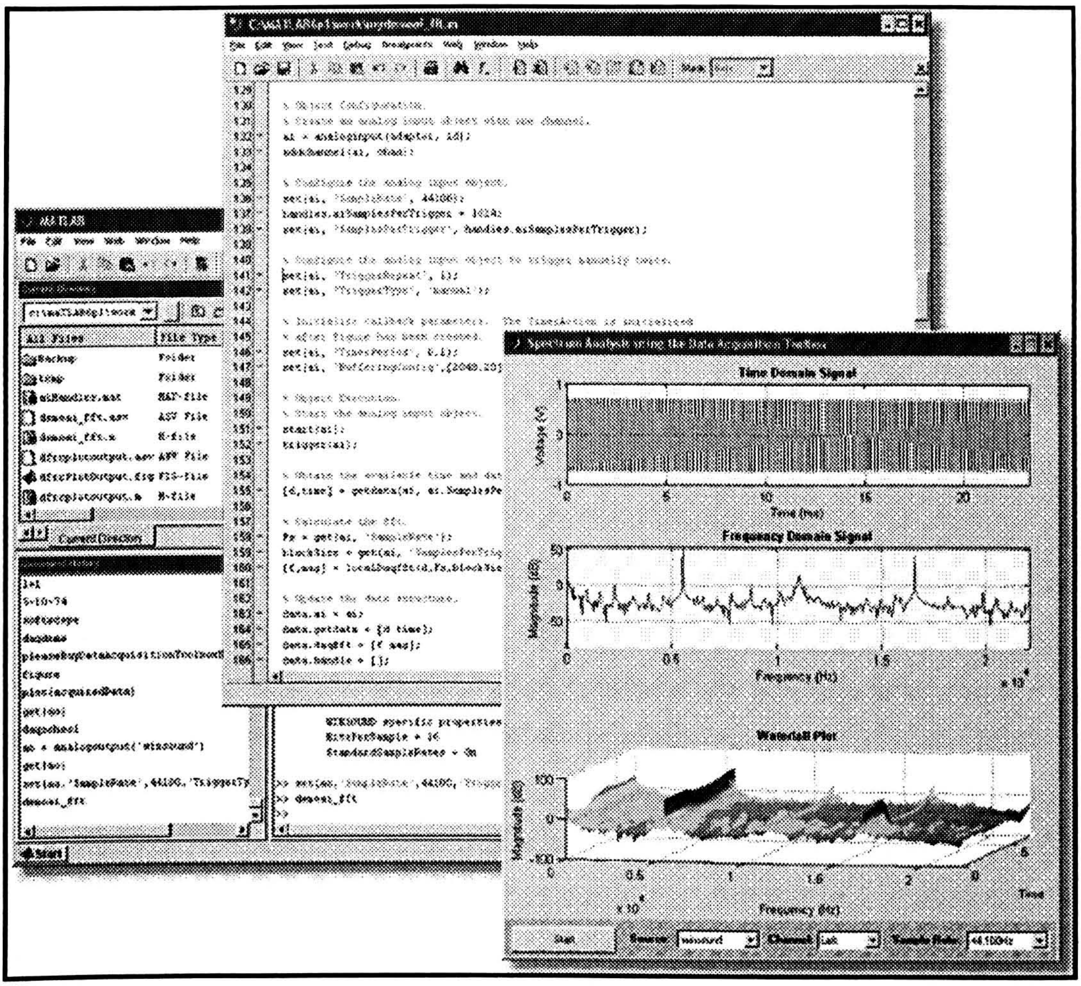


Figure 6.6: Matlab. Image from [Matlab]

6.2.6 The chosen programming environment

In conclusion, the only programming environments that have all the desired characteristics (see Section 6.2 for the list) are PD and Supercollider. Since PD has a very stable version running on Windows (the operating system used in this research), it was chosen as the programming environment in which to build the Interactive Sonification Toolkit.

6.3 The developed toolkit

As part of the experimental research conducted in this project, an Interactive Sonification Toolkit was programmed which allowed the researcher to investigate and evaluate various sonification mappings and various interaction methods applied to large multivariate data sets. Experiments were conducted using various types of data. Some data sets were purposely devised by the researcher so that they contained known structures. Other data sets were gathered from real dynamic systems. Sonification mappings were applied to these data in order to verify whether particular structures could be perceived via the sonifications. The Interactive Sonification Toolkit, programmed using PD, can be found in folder *Interactive Sonification Toolkit* of the accompanying DVD. The Interactive Sonification Toolkit uses existing PD libraries. In particular, it uses the basic PD objects that come with the 0.39.0 Version of PD, the Zexy Library 1.3 by Iohannes Zmoelnig [Zexy library], the GrIPD Library Version 0.1.0 by Joseph A. Sarlo [GrIPD library] which is used to build the Graphical User Interface (GUI) and the GEM library originally by Mark Danks [GEM library].

6.3.1 Data to sound mappings and toolkit architecture

Prior to programming a sonification toolkit, we need to consider what kind of data to sound's variables mappings should be available. There are different types of data to sound mappings. For instance, in the literature regarding Toolkits for sonification (see Chapter 2, Section 2.3.1 of this thesis), researchers distinguish between mapping data to sound perception parameters (e.g. pitch or brightness) or mapping data to sound synthesis variables that do not necessarily have a correspondent perceptual parameter (e.g. mapping data to the modulating frequency of a FM algorithm).

For the toolkit to be a highly flexible tool, it was decided that the programmer should not impose which sound's variables are available for mapping. The toolkit should allow scaling the different channels of data independently and preparing them as different inputs ready to be fed into a synthesis algorithm. Furthermore, the synthesis algorithm should be programmed and inserted in the toolkit by the person who is using it. The toolkit should then give the possibility to easily route the output of the synthesis algorithm to the sound outputs as necessary.

Notice that the number of inputs to the synthesis algorithm is not usually equal to the number of outputs of the synthesis algorithm and the number of perceivably separate audio streams at the output do not necessarily correspond to the number of data channels to be sonified. The relationship between the number of data channels and the number of separate audio channels depends on the sonification algorithm and therefore can only be specified by the user.

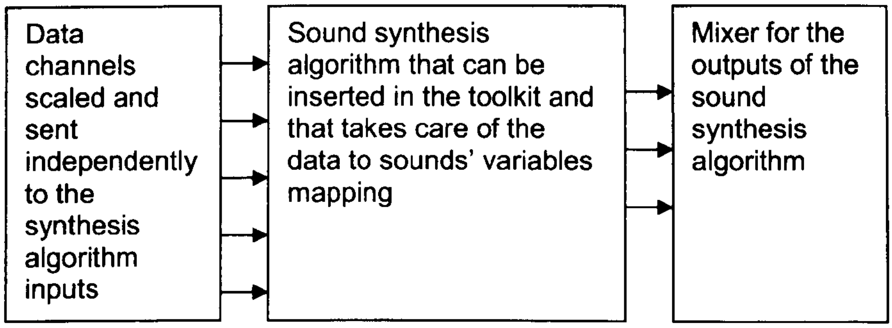


Figure 6.7: Schematic diagram showing the mapping between data channels, sound synthesis algorithm and final sound mixer

In summary, the relationship between the number of data channels, the number of mapped sound variables and number of separate audio streams at the output can be categorised as follows:

1) *One-to-one*: one channel of data is mapped onto one sound variable. The result is one *single sound* (notice that a sequence of notes perceived as a unified melody is considered here a *single sound*)

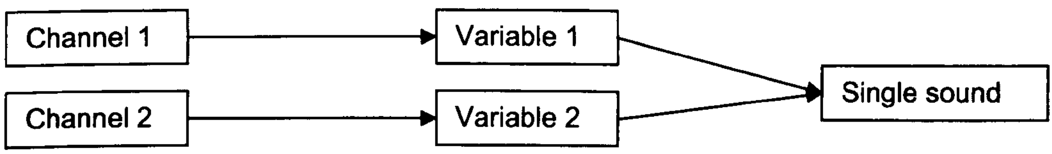


Figure 6.8: One-to-one mapping producing a single sound

2) *One-to-many*: one channel of data is mapped onto two or more variables of the same sound. The result is one *single sound.*

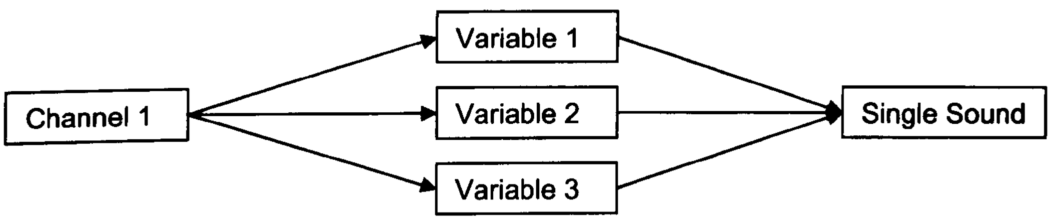


Figure 6.9: One-to-many mapping producing a single sound

3) A *combination of one-to-one and one-to-many that creates a single sound*: a *single sound* can also be created by various data channels mapped using the *one-to-one* criterion and some other channels mapped using the *one-to-many* criterion. For example, one channel

could be mapped onto frequency and amplitude (one-to-many) and another to notes durations (one-to-one): the result is a single sound.

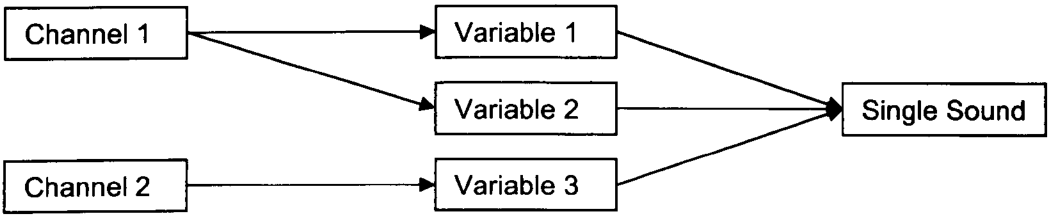


Figure 6.10: Combination of one-to-one and one-to-many mappings to produce a single sound

4) A *combination of one-to-one and one-to-many that creates a mixed sound*: single sounds, created using the above criteria, can be mixed together creating a mixed sound.

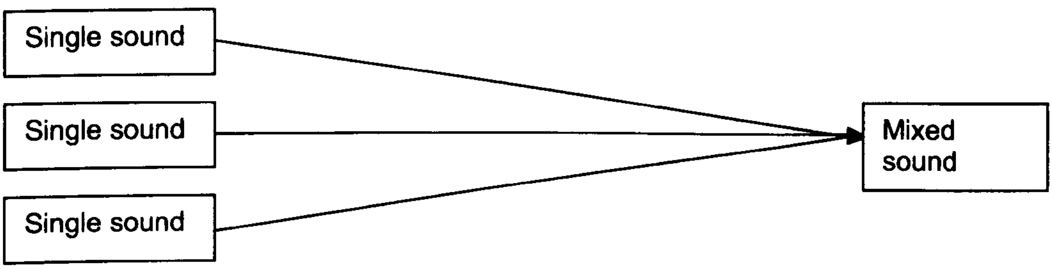


Figure 6.11: Combination of single sounds (results of any mapping) to produce a mixed sound

6.3.2 The perception of the data structure in the sonifications

We have just seen that the relationship between number of data channels and number of perceivable separate audio streams depends on the choice of sonification algorithm. In turn, the choice of sonification algorithm is determined by what we know about the perception of different data structures through a sound parameter and our knowledge of sound perception in general.

Here there follow some considerations about the perception of data structures in data to sound mappings.

a) *Mapping of one single data channel.*

If only one data channel is mapped to one or more variables of a sound (while the rest of the sound variables are kept constant) the listener can perceive the structure of the data focusing on the variables driven by the data channel. For example, a data channel can be mapped to both frequency and amplitude of an oscillator. The perception of that data channel is strongly dependent on the perceivable ranges of the variables chosen to represent the data.

b) *Mapping of two or more data arrays.*

If two or more data channels are mapped onto sound variables to produce a single sound, often the listener will not be able to perceive the action of the different data channels separately. Instead, the overall evolving timbre of the sound, which represents the combination of the data channels, will be perceived. Notice that this is not always true. For instance, the pitch of notes and the duration of notes are very separable variables and they can be perceived separately.

c) *Perception of two or more sound streams.*

If mixed sounds are considered, then the perception of single data channels will normally depend upon grouping criteria. For example, one channel can be mapped to the pitch of a guitar sound and another to the pitch of a piano sound. The two streams will be perceived as separate because the human brain distinguishes the two different timbres and can focus separately on the evolution of the guitar sound or of the piano sound. If both channels are represented by the same timbre, then it would be impossible to distinguish the evolution of one channel separately from the other unless they are mapped onto two very different frequency ranges.

d) *Restrictions of the number of data arrays to be mapped onto a sound synthesis algorithm.*

Not all mappings' methods allow for the mapping of more than a certain number of channels to the sound variables. For instance, when mapping a data channel to the samples of the waveform of a sound (this is done in audification), no other variables of that sound are available for mapping because this would alter the wave form and therefore alter the initial data channel.

6.4 A description of the Interactive Sonification Toolkit and its functions

There are two versions of the Interactive Sonification Toolkit. In both versions a screen area represents the data that can be navigated with the mouse or the jog wheel/shuttle interface [Contour Design], however only in one version the data array is also represented as a graph in the screen area. The description that follows is valid for both versions of the Toolkit, the only difference is that in one version the data are also represented visually as a graph. The Interactive Sonification Toolkit consists of two main programs (called 'patches' in PD): 'uploadAll.pd' and 'mouseNavigation.pd'. Each of these is now described. A user guide (see the file 'readme.pdf') can be found in the accompanying DVD in the Interactive Sonification Toolkit folder. A video showing a basic use of the Toolkit can be found in the accompanying DVD (see folder "Interactive Sonification Toolkit" ToolkitVideo). This video was created for the presentation at the International Conference on Auditory Display 2004.

6.4.1 The patch uploadAll.pd

The function of this part of the program is to upload the desired data set. A data set is represented by a matrix with various columns and rows. In the case of Helicopters and Physiotherapy data, each column of numbers in the matrix represents the measured, sampled in time, values of a certain sensor. For instance, in the helicopters data there will be a column representing the speed of a rotor, another representing the temperature of the air, etc. Sensors normally are very different from each other and measure various parameters. Their ranges and measure units are also very different. In this context, a column of data is referred to as a 'channel'. This part of the program scales each channel separately and displays the data of each channel as an array in PD which shows up as a low resolution graph [see Figure 6.12].

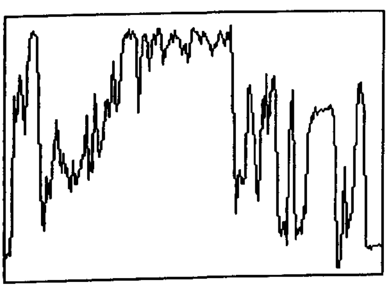


Figure 6.12: A data array in PD

The number of rows in the matrix corresponds to the number of samples in each channel. If a sensors records data for 10 minutes at the rate of 1 measurement per second, then there will be 600 rows or samples per channel. The data set or matrix is written in a text file that can be edited using a text editor such as Notepad.

A matrix file looks like in Figure 6.13:

```
matrix 5 2
1 3
2 4
1 45
1 4
344 46
```

Figure 6.13: A matrix file

The first line of the file must contain the word 'matrix' followed by the numbers of rows and columns. A *return* separates the lines. Separators between values can be either *spaces* or *tabs*. The file has to be saved with the file extension '*.mtx*'.

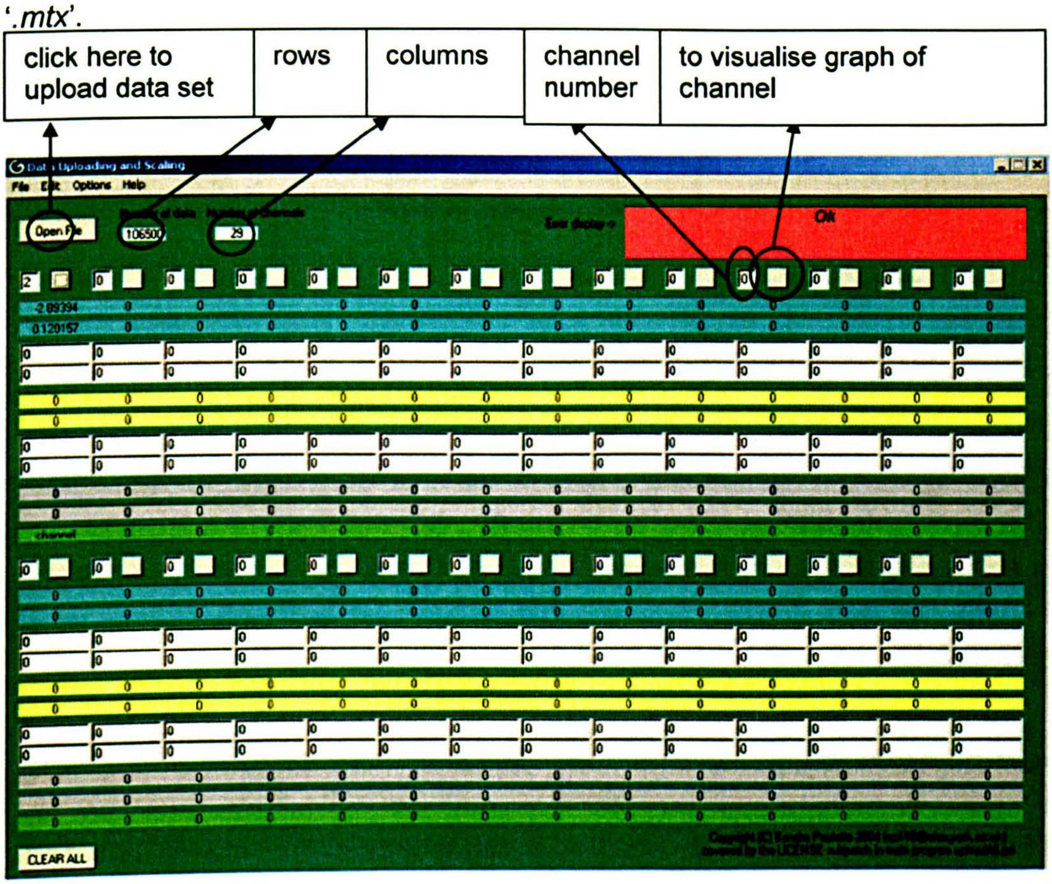


Figure 6.14: UploadAll.pd Graphical User Interface (original in colour)

When a data set is uploaded using the *Open File* button (see Figure 6.14), the number of data samples per channel and the number of channels appear in the two text boxes next to the *Open File* button. Each column of the Graphical Interface represents the main properties of a channel (see Figure 6.15).

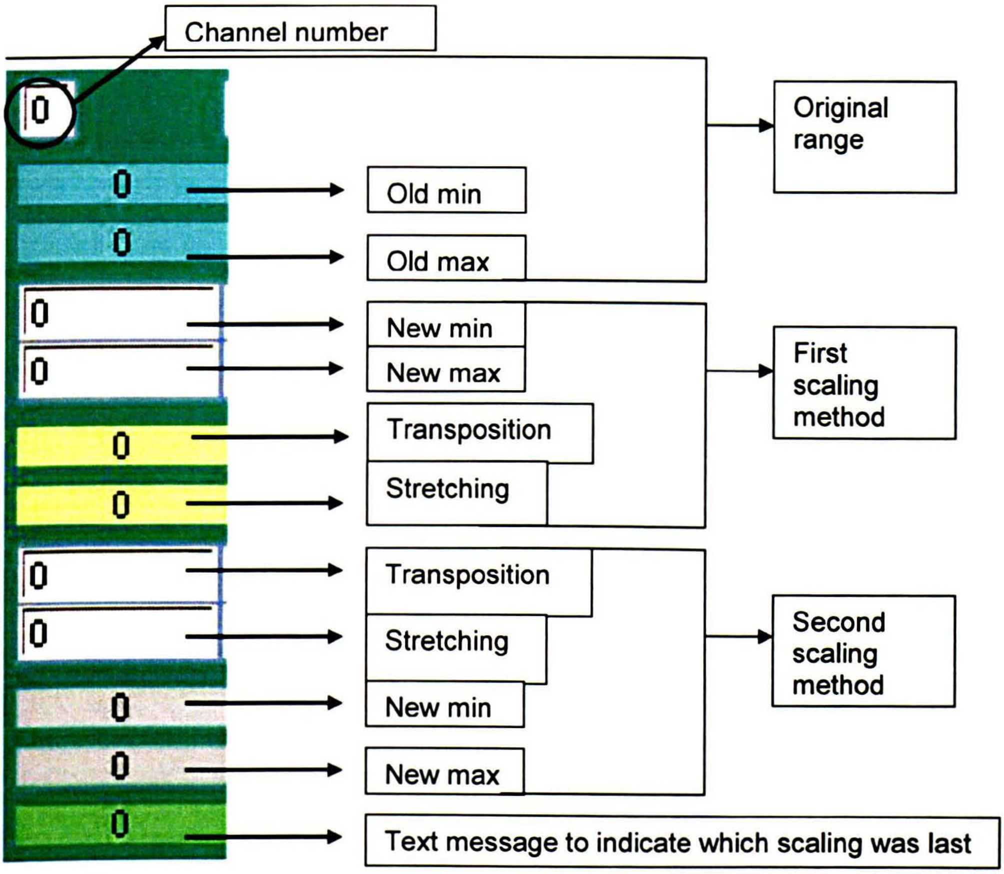


Figure 6.15: Channel column in the GUI of uploadAll.pd (original in colour)

In the Interactive Sonification Toolkit there are 28 channels available. In a future version of the toolkit, the number of channels should be made a variable that the user can specify at the beginning of the process.

When a channel number is typed into the Channel Selection Box and the *Enter* key is pressed, its maximum and minimum values are calculated and displayed below it. The user has two ways to linearly scale the data:

A. by defining *new minimum* and *maximum values*: this allows two different channels to be scaled to the same range;
B. by defining *new transposition* and *stretching factors*: this allows two different channels to be scaled so that their new values are measured in the same reference system.

Mathematically the first way of scaling uses the formula (6.1) below, while the second way of scaling uses formula (6.2) below.

Scaling by defining the new extremes of the range:

$$scaled\_data = new\min + (original\_data - old\min) * \frac{(new\max - new\min)}{(old\max - old\min)} \qquad (6.1)$$

Scaling by defining new transposition and stretching factors:

$$scaled\_data = trasposition\_factor + (stretching\_factor * original\_data) \qquad (6.2)$$

The relation between *Transposition_factor*, *Stretching_factor*, *Newmin*, *Newmax* can be found by equating formulae (6.1) and (6.2).
It follows that:

$$new\min = trasposition\_factor + (old\min * stretching\_factor) \qquad (6.3)$$
$$new\max = transposition\_factor + (old\max * stretching\_factor)$$

and

$$transposition\_factor = \frac{(new\min * old\max - old\min * new\max)}{(old\max - old\min)} \qquad (6.4)$$

$$Stretching\_factor = \frac{(new\max - new\min)}{(old\max - old\min)}$$

In the Interactive Sonification Toolkit, when the user employs the first scaling, the program automatically calculates and displays the corresponding transposition and stretching factors (using formula 6.4). When the user employs the second scaling, the program calculates and displays the corresponding new minimum and maximum values of the channel (using formula 6.3).
Displaying *Newmin*, *Newmax*, *Trasposition_factor* and *Stretching_factor* has proven to be very useful in particular in two cases:

1. When the user employs the second scaling, looking at the corresponding minimum and maximum values shows if the new range takes the sound outside the perceivable range. For example, imagine that there are 2 channels, one with a range between 1 and 5 and one with a range between 0.1 and 8. We want to map both channels to frequency. We need to scale them using the second scaling because we want to be able to compare the pitches and therefore maintain the proportional relationships between the data of the two channels. A stretching factor of 4000 would be ok for the first channel because it would bring the frequency up to a maximum of 20kHz. On the other hand a stretching factor of 4000 would take the second channel frequency up to 32kHz which is outside the audible range. Being able to see the new minimum and maximum values corresponding to a particular stretching and transposition factor helps to select meaningful values for the scaling.

2. When the user needs to display two channels using the same reference system, then the second scaling has to be used. However, it is not immediately obvious how to calculate what transposition and stretching factors need to be used in order to scale the channel to a range that is perceivable. An easy solution to this problem is to scale channel 1 with the first scaling (i.e. by entering perceivable minimum and maximum values), write down the corresponding Transposition and Stretching factors, and finally scale the second channel using the second scaling and the Transposition and Stretching factors produced by the first scaling of channel 1.

Each scaled channel is displayed in a PD graph (see Figure 6.16). The resolution of the PD graphs is low, however this visual display has proven useful while testing and programming the Toolkit prototype.

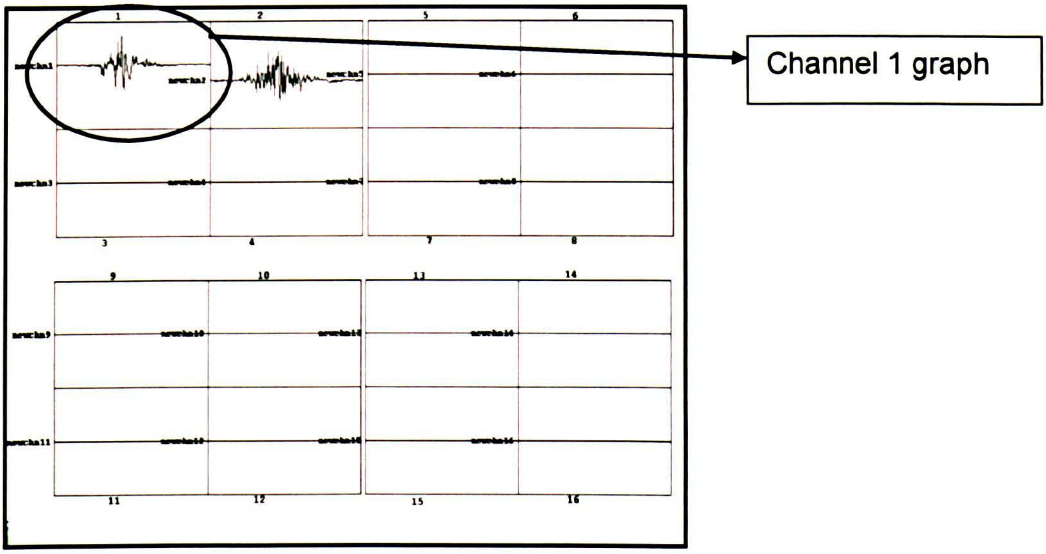


Figure 6.16: Arrays in PD represented as graphs

6.4.2 The patch mouseNavigation.pd

The mouseNavigation.pd patch constitutes the second part of the Toolkit and it has its own separate GUI window (see Figures 6.17, 6.18 and 6.19).

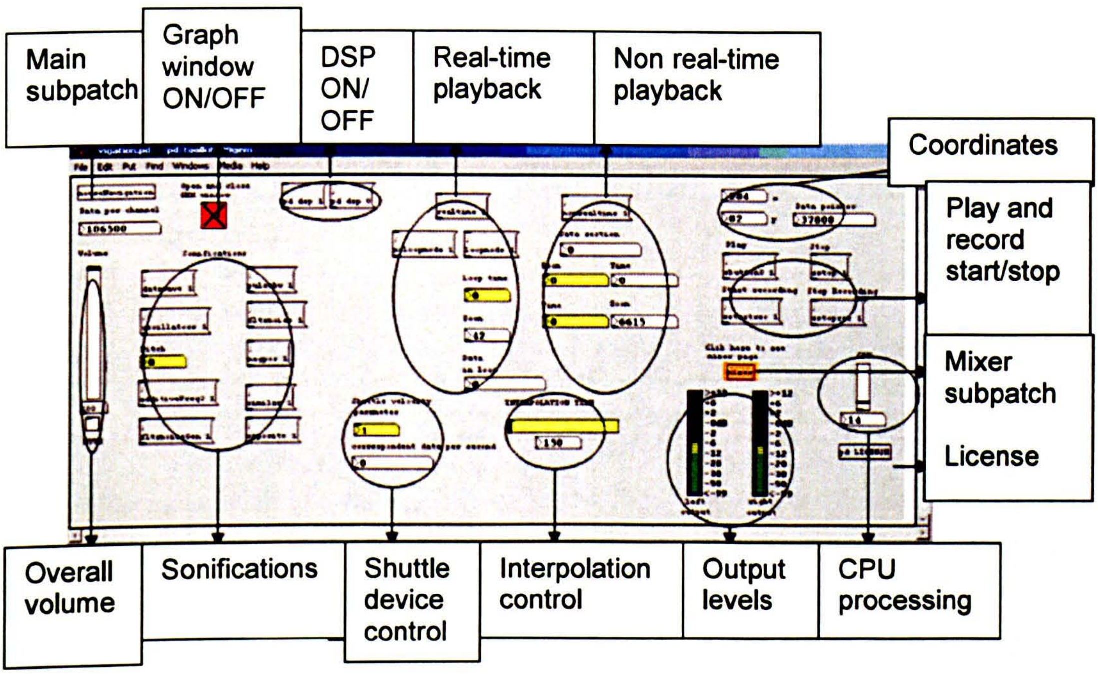


Figure 6.17: Navigation GUI (original in colour)

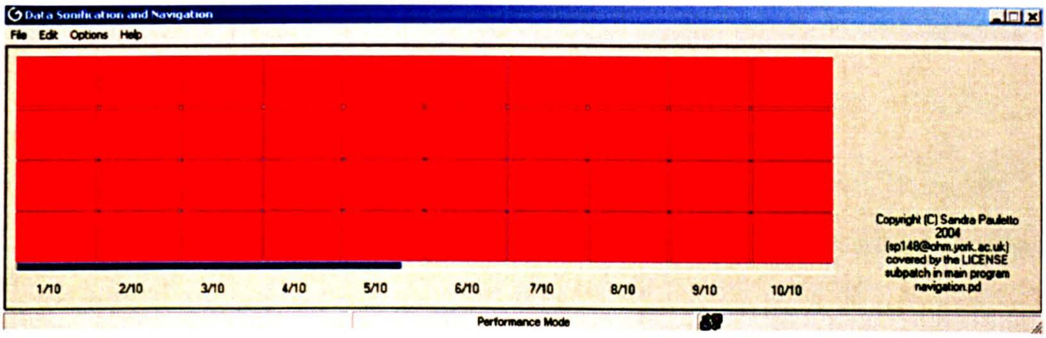


Figure 6.18: Interaction area in version 1 which does not visually display the data array (original in colour)

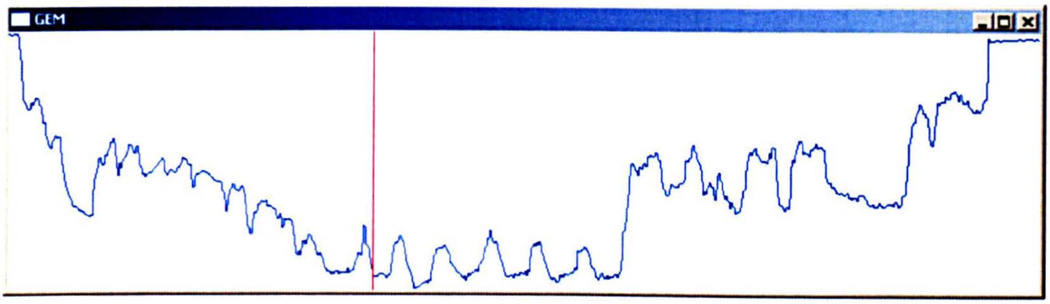


Figure 6.19: Interaction area in version 2 which also visually displays the data array

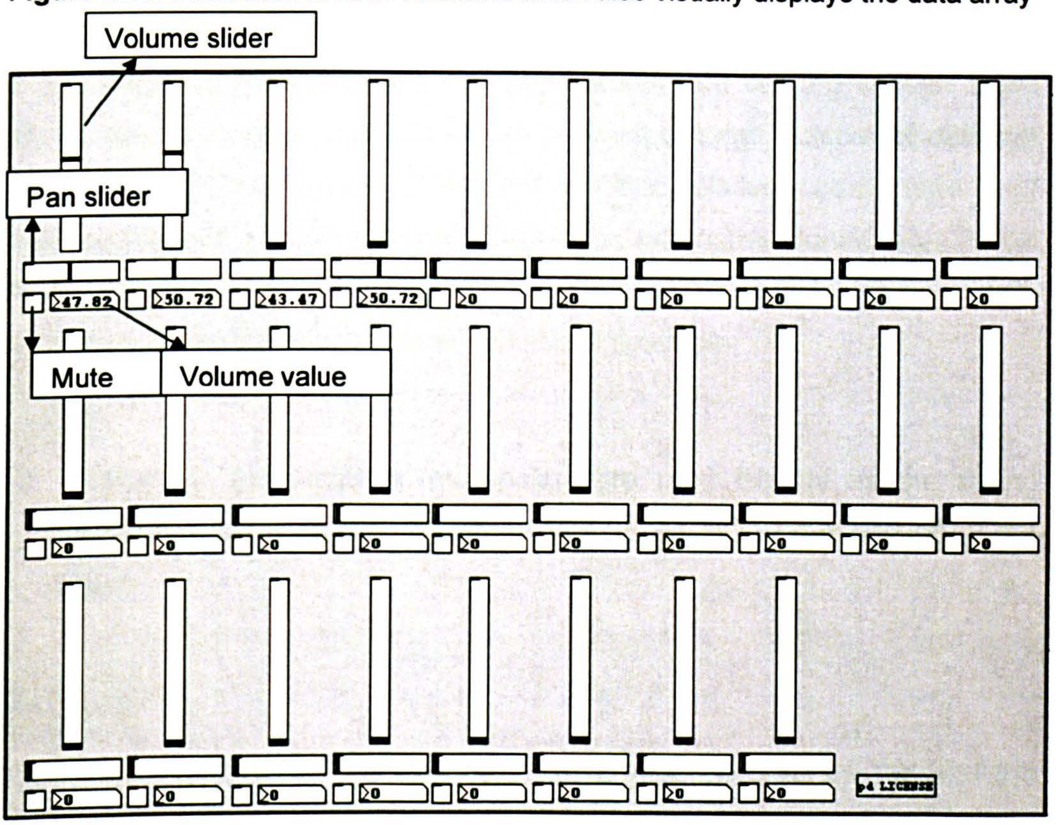


Figure 6.20: Mixer subpatch (original in colour)

The two main patches are connected. When a data set is uploaded in uploadAll.pd, the dimensions of the matrix are passed on to the mouseNavigation.pd patch and the data from the scaled channels are used by the mouseNavigation.pd patch.

The two main functions of the mouseNavigation.pd patch are:

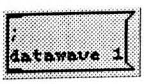
- to allow selecting among types of data-to-sound mappings;
- to allow navigating the sonified data with a mouse or jog wheel/shuttle interface.

The navigation can be done in three different ways: two real-time ways and one non real-time way. The navigation can be recorded onto the hard drive as a wave file.

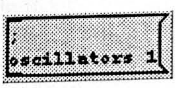
6.4.3 The sonifications

Various types of sonifications have been explored during this project. This experimentation helped the author to develop her understanding of the issues surrounding the choice of a sonification algorithm to display data structures. In this version of the Toolkit several sonifications are available, but only two have been used in the experiments reported in this research. The sonifications available cover all the possible mappings mentioned in sections 6.3.1 and 6.3.2 and are simple examples of different approaches that can be chosen when sonifying one or more data channels with a small or large number of data per channel. Sonifications can be added and modified relatively easily (see User Guide 'readme.pdf' in accompanying DVD in the Interactive Sonification Toolkit folder).

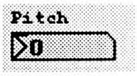
Here follows the description of the sonifications available.



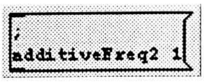
1. *DataWave*: the rescaled data points are read directly as the sound samples of a waveform. This mapping results in the *Audification* of the data.



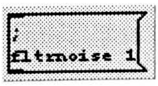
2. *FrequencyOsc*: the rescaled data points of a channel are fed into the frequency input of an oscillator producing a sine wave.



3. *AdditiveSynthesis with constant pitch*: the data points of each channel drive the amplitude envelope of a harmonic partial of a sound synthesised by additive synthesis with constant fundamental frequency chosen by the user (to input in the yellow text box).

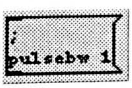


4. *Additive synthesis 2*: the data points of one channel control the fundamental frequency of an additive synthesis sound; the other channels drive the amplitude envelopes of the sound harmonic partials and thus control the harmonic content.

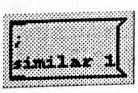


5. *FilterNoise*: one channel's data points control the centre frequency of a band pass filter applied onto noise; the data of a second channel drive the Q factor of the filter (Q = CentreFrequency/BandWidth). This sonification works only using channel 1 and 2.

6. *FilteNoiseGen*: the data channels control the centre frequency of band pass filters applied to noise while the bandwidth of the filters is constant.

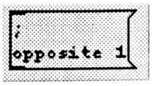


7. *PulseBW*: one channel's data control the frequency of a pulse sound; the data from the second channel drive the duty cycle of the pulse. This sonification works only using channel 1 and 2.



8. *Similarity Test*: the data of two channels are rescaled between -1 and 1; the two channels are then subtracted data point by data point and the result (scaled to audible frequency range) fed into the frequency input of

an oscillator. If the two channels' data are identical, then the frequency is zero and therefore no sound is heard.



9. *Opposition Test*: the data of two channels are rescaled between -1 and 1; the two channels are added point by point and the result (scaled to audible frequency range) fed into the frequency input of an oscillator. If the two channels' data are completely opposite to each other, then the frequency is zero and therefore no sound is heard.

10. *NegPos*: the positive data values from a data channel are mapped to the amplitude of an oscillator of constant frequency, while the negative data values from a data channel are mapped to the amplitude of a second oscillator of constant frequency (different from the first oscillator).

In earlier versions of the Toolkit sonification mappings using MIDI were also implemented.

11. *Data as MIDI notes' durations*: the data points control the durations of successive midi notes.
12. *Data as pitches of MIDI notes*: different instruments are assigned to different channels and the data points of each channel drive the pitches of the MIDI notes. In this mapping method, a threshold can be applied to the data so that only data with pitch above a certain threshold are heard. This allows hearing if the channels reach the same threshold simultaneously or not.
13. One channel is mapped onto *durations of MIDI notes*, the other onto the *pitch* of the notes.

6.4.5 The navigation and interaction methods

In the Interactive Sonification Toolkit, the user can navigate the sonification using three main methods.

The *interaction area* (see Figure 6.18 and Figure 6.19) represents the sound in its totality. It can be explored by moving a mouse over it or by using the jog wheel/shuttle interface. The coordinates of the cursor moving in the area are tracked by the program and used as pointers to the sound/data array. In Version 1 of the Toolkit, ten gradations are superimposed to the *interaction area* (see Figure 6.18) which indicate to the user approximately where he/she is positioned in the sound/data array at each instant. In Version 2 of the Toolkit, the visual display (see Figure 6.19) can be navigated in the same way.

The navigation's methods:

1. *Non Real-Time navigation*: in this case the user needs to select a zone of the sound he/she wants to hear. The user right clicks the mouse on the left side of the interaction area, then drags the mouse to the right, and let the mouse button go where the selection ends. The user can either input the *zoom* factor (a zoom factor equal 1 means that 44100 data samples are played back in 1 second, a zoom of 2 mean that 44100 data samples are played back in 2 seconds, etc.), and then the *playback time* of that selection is calculated by the program, or he/she can select the *playback time* in milliseconds and the program will calculate the corresponding *zoom* factor. To hear the sound the user presses the *Play* button.
2. *Real-Time Navigation with No Loop*: the user moves the cursor, indicated by the blue horizontal line in Fig. 6.18, or the red vertical bar in Fig. 6.19, around the interaction area using mouse or shuttle. At each instant the coordinates of the cursor drive a pointer to the sound/data array. The 'y' axis of the interaction area does not have a function. The sound is played in immediate response to the cursor movement.
3. *Real Time Navigation with Loop*: the user interacts with the interaction area in the same way as above, however this time the cursor x position drives the beginning of a sound loop. The user inputs the length of this loop in milliseconds before navigation. The 'y' axis has a function: if the mouse is at the top the *zoom* is 1, which means that in a loop of 1000ms there are 44100 samples of sound; at the bottom of the interaction area the *zoom* is 100, i. e. in a 1000ms long loop there are 441 samples of sound. Therefore, moving the cursor down causes zooming into the sound. This method can only be used using the mouse.

All navigations and sounds can be recorded as Wave files and saved in the hard drive.

6.4.6 Interaction methods

The Toolkit uses two interfaces to navigate the sonifications: the mouse and the jog wheel/Shuttle XPress interface [Contour Design].

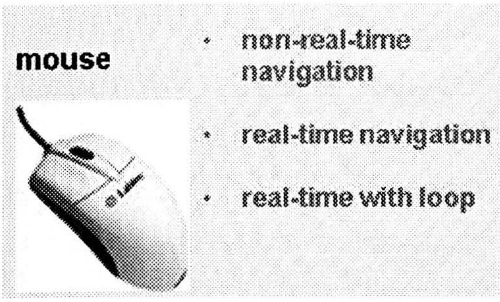


Figure 6.21: Mouse interface

The mouse allows producing all navigations' types.

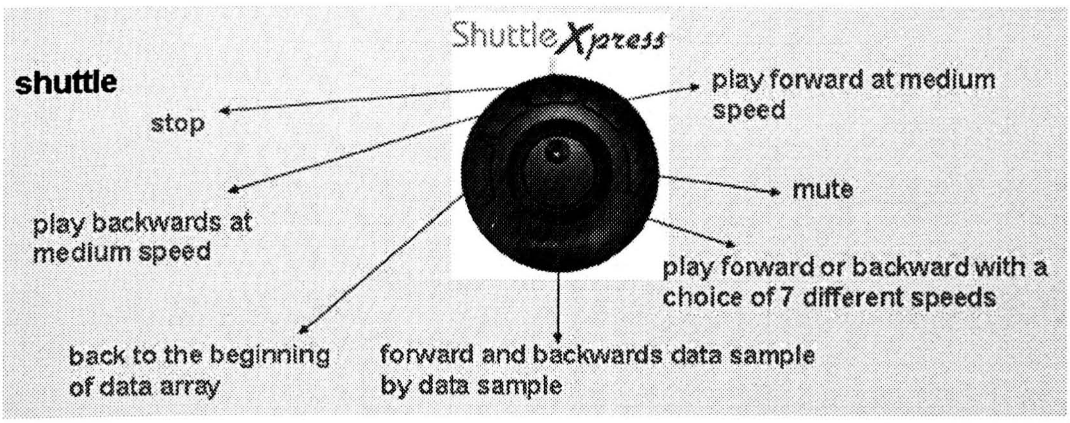


Figure 6.22: The Shuttle XPress interface

The jog wheel/shuttle interface allows performing only the real-time navigations and does not use the 'y' axis of the interaction area.

As can be seen in Figure 6.22, the Shuttle interface allows the cursor to play the sonification backwards and forwards at set speeds and data sample by data samples, while the mouse plays the sonification back and forward at a speed directly dependent on the hand movement. In this research three interaction methods are distinguished and compared in the experiment described in Chapter 9 of this thesis:

1. *Low interaction method.* This method corresponds to the non real-time navigation method.

2. *Medium interaction method.* This method corresponds to the real-time navigation method without loop using the jog wheel/shuttle interface.
3. *High interaction method.* This method corresponds to the real-time navigation method without loop using the mouse.

6.5 Conclusions

This chapter described the development and implementation of the Interactive Sonification Toolkit. This Toolkit was used to create the sonifications for experiment two (Chapter 8) and also used to allow subjects in experiment three (Chapter 9) to interact in different methods with a sonification. The first section of this chapter discussed the general requirements for an interactive sonification application and listed the possible programming environments that could be used to develop such an application. The third and fourth sections justified the choice of programming environment used in this research project and described the data-to-sound mapping considerations that influenced the architecture of the Toolkit presented here. Finally, in Section 5 all the functions of the Toolkit were described.

Chapter 7 – Experiment one: a comparison of audio and visual analysis of complex time-based data sets

7.1 Introduction

This chapter describes the first experiment produced during this research. Experiment one aims to show that the sonification of a multivariate time-based large data set is an effective display. In particular, this chapter focuses on the comparison between auditory and visual displays of large multivariate data sets. Visual displays have been used for a lot longer than auditory displays and therefore represent the natural comparison benchmark for auditory displays. A specific experiment is described here which compares audifications and spectrograms of the same data. The data represent the values gathered from various sensors located in a helicopter during a flight. Furthermore, section 7.5 of this chapter, describes the implementation of a model-based sonification (MBS) of the same helicopter's data. MBS is not the focus of this thesis, however it is the main alternative approach for the sonification of large data sets to parameter mapping. The description of this MBS example aims to clarify the MBS approach by showing a concrete working example. This example of MBS was produced in collaboration with researcher Thomas Hermann of Bielefeld University. Finally, in the conclusion, the author summarises the experiment's results, and briefly comments on the two specific techniques (audification and MBS) described in this chapter.

7.2 Background

7.2.1 Comparing auditory and visual displays

Visual representations of data have been used for a lot longer than auditory representations. In fact, visual displays can be said to be the norm, and particular visual displays (graphs, diagrams, spectrograms) are widely understood. It is therefore natural when evaluating new auditory displays that their effectiveness in portraying information is compared to that of an equivalent visual display. In literature there are various studies which compare audio and visual displays. Nesbitt and Barass [2004] compared a sonification of stock-market data with a visual display of the same data and with the combined display (audio-visual). Brown and Brewster [2003] designed an experiment to study the understanding of sonified line graphs. Peres and Lane [2003] evaluated different ways of representing statistical graphs (box plots) with sound. Valenzuela *et al.* [1997] compared the sonification of impact-echo signals (a method for non-destructive testing of concrete and masonry structures) with a visual display of the signal. Fitch and Kramer [1994] compared the efficacy of an auditory display of physiological data with a visual display by asking the subjects (who play the role of anaesthesiologists) to try to keep alive a 'digital patient' by monitoring his status with each display. In all these examples, the effectiveness of a new auditory display is measured relative to the effectiveness of a standard visual display of the same data. The same approach is taken in this experiment, in which the effectiveness of the auditory display of a helicopter flight data is measured relative to that of an equivalent visual display. In the examples mentioned above, the auditory display was effective in portraying information. However these results and the specific experimental designs cannot be generalised to any other auditory display as they depend on the type of data set used, on the sonifications used, on the specific tasks to be performed, etc.

In this study the sonification method used is audification, i.e. where data are appropriately scaled and used as sound samples. There are some studies in the literature about the effectiveness of audification of complex data. Audification is often used for the sonification of data that are produced by physical systems. Hayward [1994] describes audification techniques of seismic data. He finds that audification is a very useful sonification method for such data, but he stresses that proper evaluation and comparisons with visual methods are needed. Dombois [2002; 2001] presents more evidence of the efficacy of audification of seismic data which appears to complement the visual representations. The results presented in these studies are the product of observations made by Dombois. There is no attempt here to verify how generally perceivable these effects are to the ear of an average analyst. Furthermore, the observations reported are highly specific to the data domain explored. That is, Dombois' research tells us if characterstics such as the presence of tectonics, explosions, etc. are audible or not, but does not tell us about the effectiveness of an audification display in representing general data structures that could be present in any data domain. Krishnan *et al.* [2001] describe the use of audification to represent data related to the rubbing of knee-joint surfaces. In this case the audification is compared to other sonification techniques (not to a visual display) and it is not found to be the best at showing the difference between normal and abnormal signals. Also, in this case, the audification display is evaluated for a specific task (identifying normal and abnormal signals) which is domain dependent. This study does not give us knowledge about what kind of general information the audification was able to display. These studies show that audification is effective at portraying useful information at least in the seismology domain. In none of these studies, however, is there an attempt to verify the effectiveness of audification in general.

The experiment presented in this chapter, on the other hand, aims to evaluate the effectiveness of audification compared to the use of spectra for general data characteristics. The results of this experiment are therefore not only specific to the data set used, but they are valuable for understanding the effectiveness of audification in general.

7.2.2 The data domain: helicopter flight analysis

The domain of helicopter engineering is not completely new to auditory display research. In particular, there is research that focuses on representing some information through sound for flight simulators or in the cockpit. Rizzi *et al.* [2003] presented a technique to render aircraft flyover in a virtual reality system that uses recordings of flyover noise. Kazem *et al.* [2003] presented a method for enhancing cockpit transparency through a background auditory environment. Williamson and Smith-Murray [2005] presented a sonification of probabilistic feedback applied to flight control in a flight simulation package.

These examples are, however, very different from the auditory display presented here, which mainly aims to create an effective and efficient display of helicopter flight data for post flight analysis. The data used in this experiment was given to the author by flight analysis engineers working at Westland Helicopters (UK). Our collaborators reported that these engineers are routinely required to handle flight data and analyse it to solve problems in the prototyping process. The flight data are gathered from pilot controls and many sensors around the aircraft. The many large data sets that are collected are currently examined off-line using visual inspection of graphs. Printouts of the graphs are laid across an open floor and engineers walk around this paper display looking for anomalous values and discontinuities in the signal. The paper is considered more useful than the limited display on a computer monitor. The analysis of these data is mainly used to improve flight control laws. The issues posed by flight control are well described in Williamson and Murray-Smith [2005]:

> "Controlling a flight is difficult for several reasons: pilots must coordinate controls with 4 degrees of freedom, there is a significant lag between input and aircraft response, and the aircraft is dynamically unstable (that is, it will not tend to return to a steady state and must be continuously controlled to remain stable)." [Ibid. p. 51]

The aim of the auditory display described here is to improve the standard analysis techniques by providing a sonic rendition of the data which can be heard rapidly, and therefore will save valuable technician time and speed up the analysis process. Sound representation also provides the added benefit of allowing the presentation of several time-based data series together or in sequence, for dynamic comparison of two (or many more) signals. The flight engineers are often given the task of analysing these data because a pilot has reported something wrong in a test flight. The analysts now have a huge amount of data to sift through in order to look for unusual events in the data. These unusual events could be, for instance:

- unwanted oscillations;
- vibrations and noise superimposed on usually clean signals;
- unusual cyclic structures (data repeated, where it would normally be expected to progress);
- drifts in parameters that would normally be constant;
- non-standard variations in power or level;
- a change in the correlation between two parameters (e.g. signals which are normally synchronized becoming decoupled);
- discontinuities or 'jumps' in data which is in general smooth or constant.

Identification of such events helps to pinpoint problems in the aircraft, and can provide enough information to launch a further, more focused, investigative procedure. If an auditory display of these data has many practical advantages, it is important to determine whether any information from the data series is going to be lost when rendered sonically rather than graphically. For the purposes of this experiment, five basic attributes of data were identified to be studied both visually and aurally.

These are:

a) noise;
b) repetitive elements;
c) oscillations at fixed frequencies;
d) discontinuities;
e) signal power level.

If a human analyst perceived the presence of one or more of the first four attributes, or a change in overall signal strength, in an area of the signal where it would not be expected, this would prompt further investigation. This experiment determines whether subjects rate the presence of the first four attributes, and the average level of the signal power, to the same degree using a) visual and b) aural presentation.

7.3 The experiment

The aim of the experiment is to compare how users rank the above five attributes when a series of data sets is presented visually or aurally. The goal is to see whether aural presentation allows the identification of each attribute to the same degree as visual presentation. This research focuses on the average response across a large group of subjects, rather than identifying whether an individual subject can use visual or audio presentation equally well. The aim of this experiment is to compare a visual and an audio display of the same data set.

7.3.1 The aim of this experiment

The experimental hypothesis of this experiment is that for each data series, there will be a strong correlation between the recognition of each of the five data attributes in the visual domain and the audio domain. If this hypothesis is proved, then there is a strong basis for trusting the analysis of the data using sound alone, i.e. this sonification of a large multivariate data set can be considered effective. This experiment only attempts to verify whether the sound portrays the data attributes at least as well as the visual display. If there is poor correlation, from this experiment, the reasons cannot be inferred. Other experiments would be necessary to discover the reasons for a poor correlation.

7.3.2 The data

In consultation with the flight handling qualities group at Westland helicopters 28 sets of time synchronized data taken from a half hour test flight were gathered. Each data set is taken from a sensor on the aircraft under test. Each data set contains 106500 samples, which were originally sampled at 50Hz. The helicopter's parameters measured are of highly differing natures: from the speed of the rotors, to the engine power, etc., and they change over time. For this experiment, the knowledge of what each channel represents in the helicopter system is not important, only whether the user perceives the presence of noise, frequencies, repetitions, discontinuities and the strength of signal power in both the visual and audio displays. However, the list of sensors used in this test can be found in Appendix A of this thesis.

7.3.3 Overview of the experimental task

The visual display used in this experiment is the spectrogram of each data series. The audio display is the audification of the data. The subjects were presented with a screen containing thumbnail pictures of the spectrograms of all the data series. After having had an overview of all the spectrograms, they were asked to examine and score each spectrogram (now visible on a larger scale) for the following characteristics on an integer scale from 1 to 5:

a) presence of noise;
b) presence of a repetitive element in time;
c) presence of oscillations at fixed frequencies;
d) presence of discontinuities or jumps in amplitude;
e) signal power.

For the sonic display, the subjects were presented with buttons, one for each data series, which played the audification when clicked. Subjects were asked to listen to all the sounds at least once. Then they were asked to listen to each sound as many times as required, then score it using the same categories as for spectrograms.

7.3.4 The audifications

Kramer [1994] describes the audification of data as

> "[...] a direct translation of a data waveform to the audible domain" [Ibid. p. 186]

The audifications in this experiment were created by linearly scaling the 28 data arrays between -1 and 1 and by converting each array into a wave file of sampling rate 44.1kHz in Matlab. Each audification was therefore around 2.5 seconds long. The audifications of these data result in complex sounds that contain a combination of rhythmical and frequency patterns and noise. The audifications used in this experiment can be found in the accompanying DVD in Experiment 1 (helicopters)/heliExperiment/sounds.

7.3.5 The spectra

The spectrograms, of the same data channels, were created by using the Matlab function 'specgram'. The arguments of the specgram function are as follows: sampling frequency 50Hz (original sampling frequency), Hanning window of 256 samples of length with a 50% overlap, as by default. The minimum and maximum values of the colour scale of the spectrograms were set the same for each spectrogram so that the spectrograms were comparable to each other. All the spectrograms were saved as .jpg files. The spectra used in this experiment can be found in the accompanying DVD in Experiment 1 (helicopters)/heliExperiment/spectra. Four images of channels' spectra follow below.

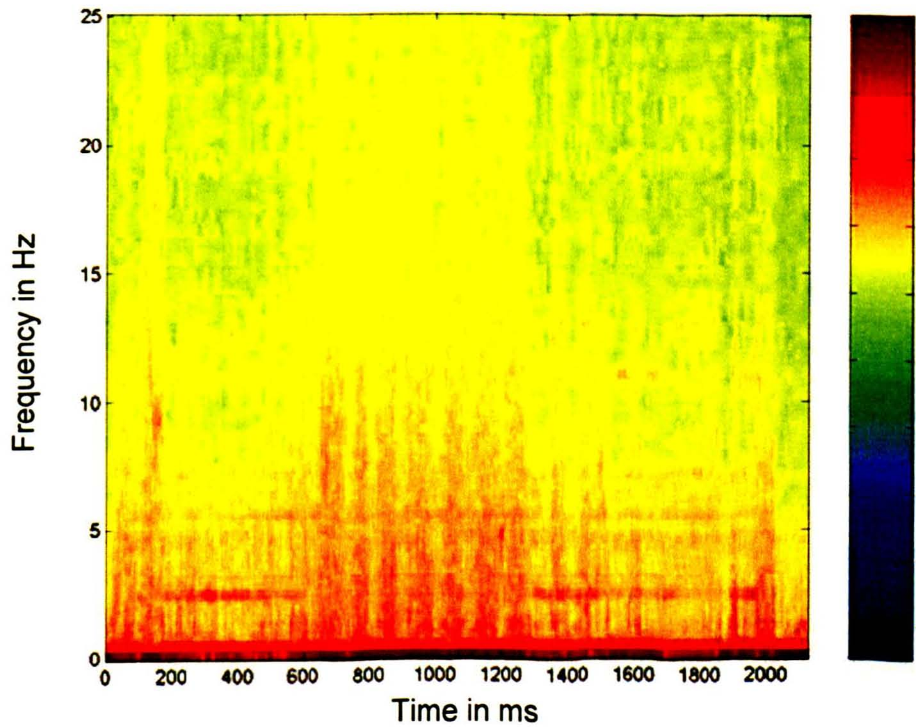


Figure 7.1: Spectrogram of channel 1 (original in colour)

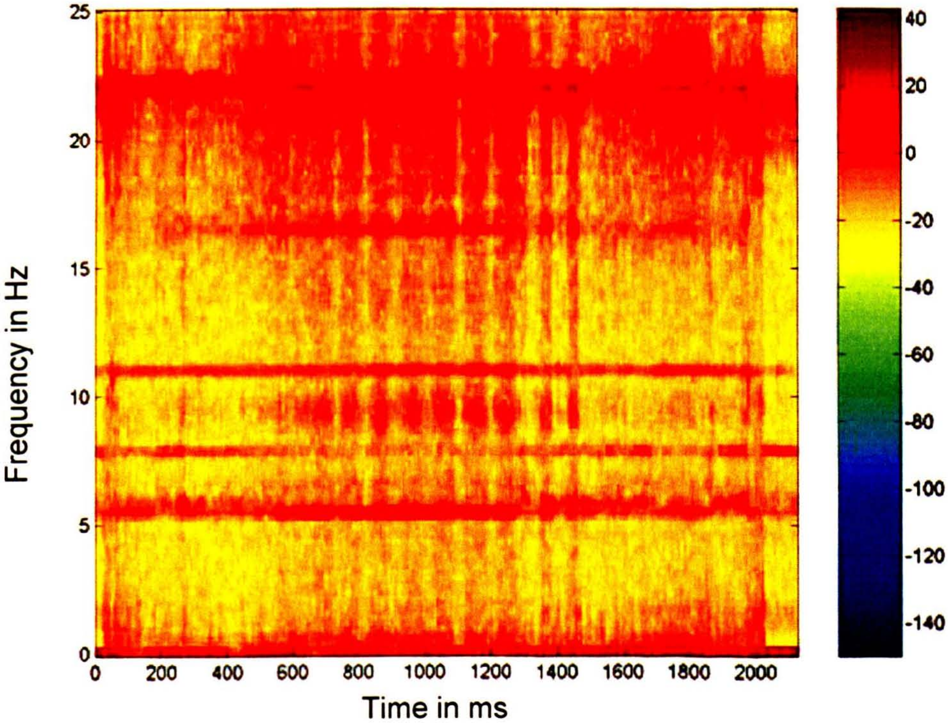


Figure 7.2: Spectrogram of channel 13 (original in colour)

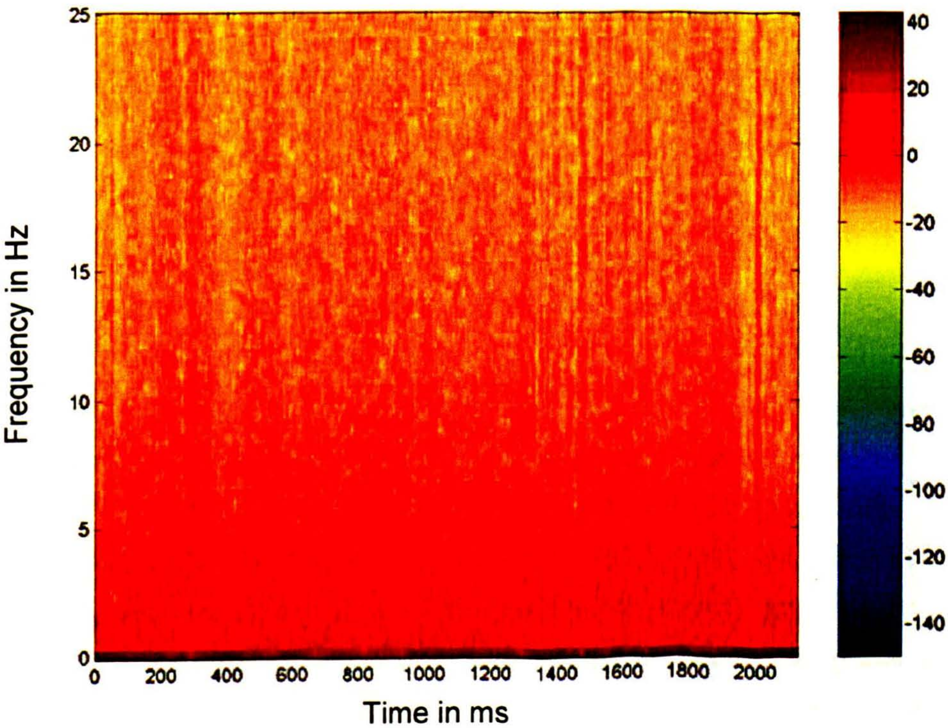


Figure 7.3: Spectrogram of channel 11 (original in colour)

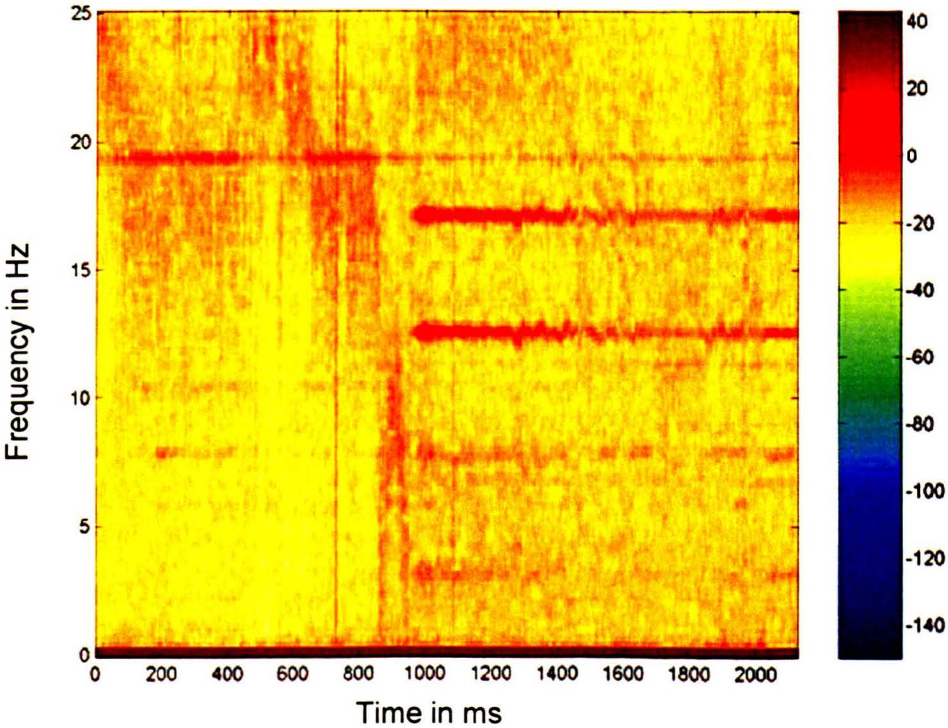


Figure 7.4: Spectrogram of channel 15 (original in colour)

7.3.6 The subjects

The subjects for this test were selected according to the following criteria. It was considered that the end user of such an auditory display would be an experienced analyst, able to interpret spectrograms and able to distinguish various characteristics in a sound's signal such as noise, repetitions, frequencies, discontinuities and signal level. Apart from this specific knowledge, the user could be any gender or age or from any cultural background. A between-subjects design, in which there are two groups of different subjects (one of which scores the spectra and the other the sounds), would have been ideal for this experiment. This would have required the recruitment of too many subjects, which was not realistic. Instead, a group of 23 subjects was chosen and a mixed within-subjects / between-subjects design was used, in which mostly the same group of people scored both the spectra and the sounds, but some only did one or the other. This design was due to the fact that some subjects were available only for a short time. In order to minimize the errors in the results due to the order of presentation of the task, the order in which the spectra and the sounds were presented to each person was randomised between subjects and tasks. Out of the 23 subjects tested, 21 were men and 2 were women. The average age of the subjects was 33. All the subjects were lecturers, researchers or postgraduate students in media and electronic engineering (with a specialisation in audio and music technology) and one person was a computer music composer. They all had experience in working with sounds and spectrograms. The subjects' understanding of sounds and spectrograms was considered to be similar to the expected understanding of the ideal end user. Subjects were from different nationalities. All the subjects declared that had no known problems with their hearing and that they had good sight or, if the sight had some defect, it was fully corrected by spectacles.

7.3.7 Procedure

Firstly, each subject was given a single-page written document which explained the task. Then the subject was asked to fill in a questionnaire to gather the information about occupation, gender, age, nationality, his/her familiarity with spectrograms and sound interpretation, and any known hearing or sight problems (see Appendix E for the questionnaire). The audio test was carried out in a silent room (mostly in the recording studio performance area at York). Good quality headphones (DT990 Beyerdynamic) were used with a wide frequency response (5 – 35kHz). This minimised the errors that could be due to external sounds. The volume of the sounds was maintained the same for all subjects. The spectrogram test was also conducted in a generally quiet room which allowed concentration. Subjects who were able to do both tests in one sitting were asked to take at least a two minutes rest between the visual and the audio parts of the test. The total test, for each subject, lasted about 45 minutes. Subjects were also asked to record on a piece of paper any comments about the test they thought could be valuable. For the experiment, a program was created in PD [Pure Data programming environment] which recorded all the results of the test automatically in a text file. Before presenting the spectrograms and the sounds to each subject, the order of presentation of each data series on the screen was randomised, so as to minimise errors due to the order of presentation. The test began with an overview of all the spectrograms (see Figure 7.5). Then by clicking on each thumbnail image a larger version of the spectrogram appeared (see Figure 7.7). A click on the 'Test' button brought up a further window (see again Figure 7.6), consisting of a series of radio buttons (labelled from 1 to 5) for each parameter being scored (noise, repetition, frequency, discontinuity and signal power).

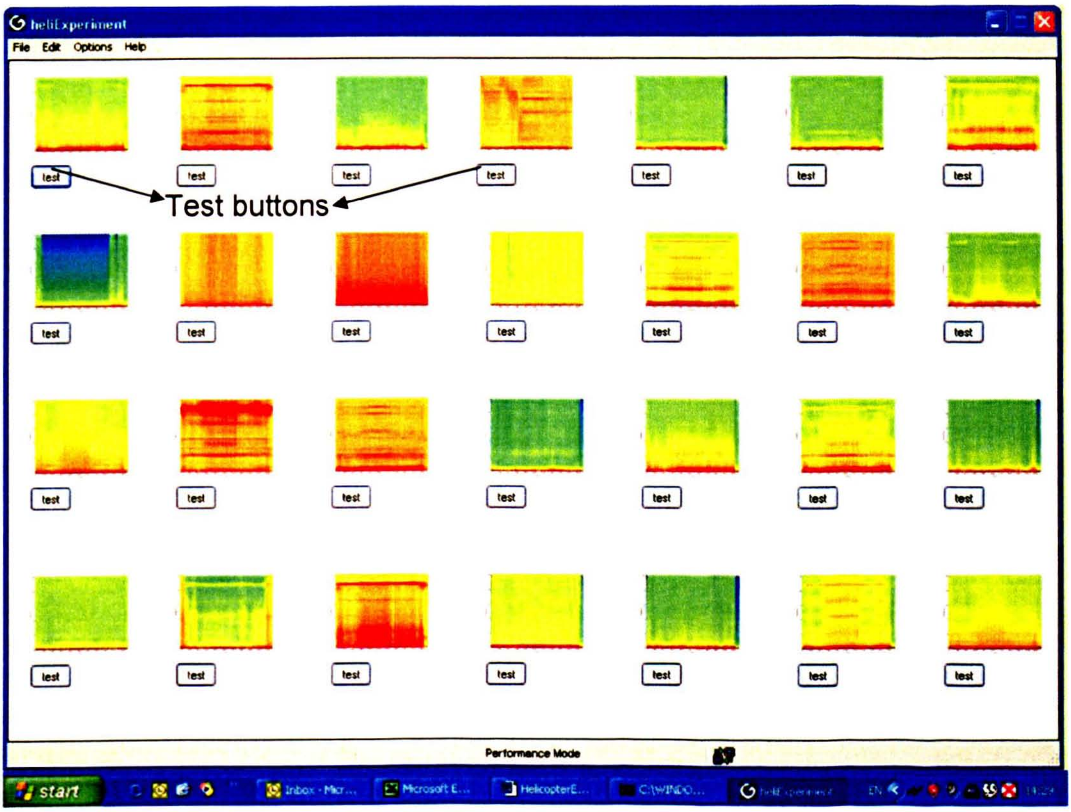


Figure 7.5: Experiment 1 software GUI showing the thumbnails of the spectra (original in colour)

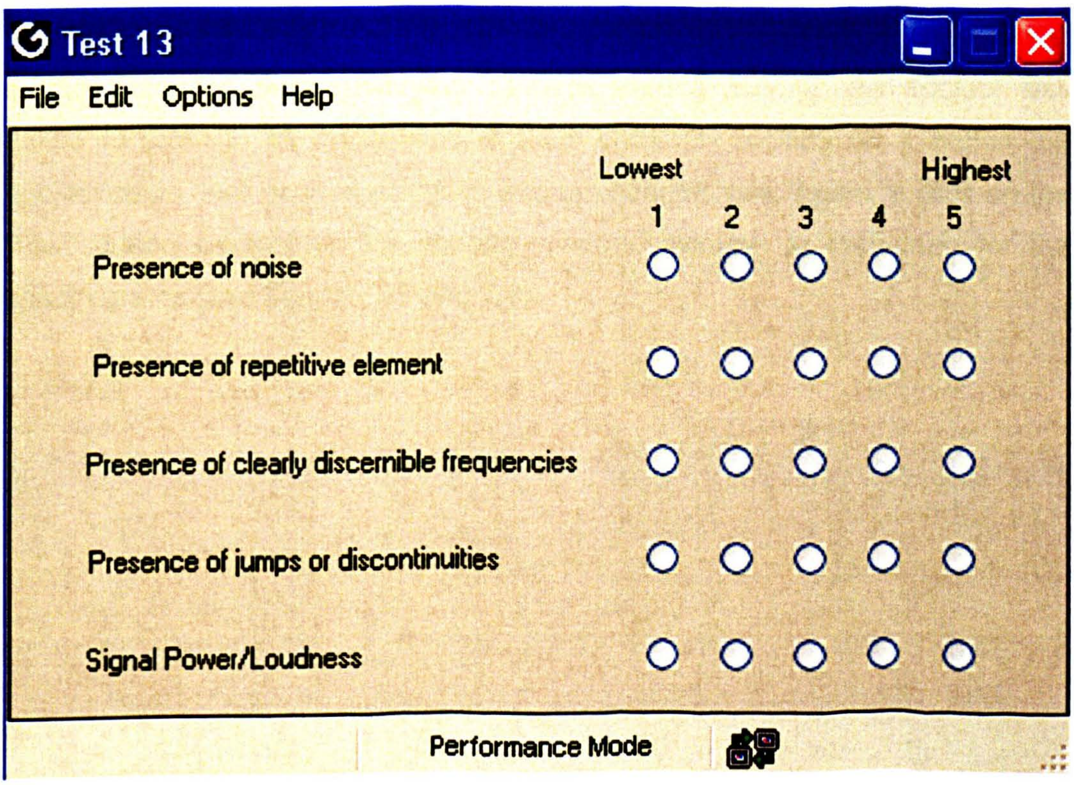


Figure 7.6: Experiment 1: the Test window (original in colour)

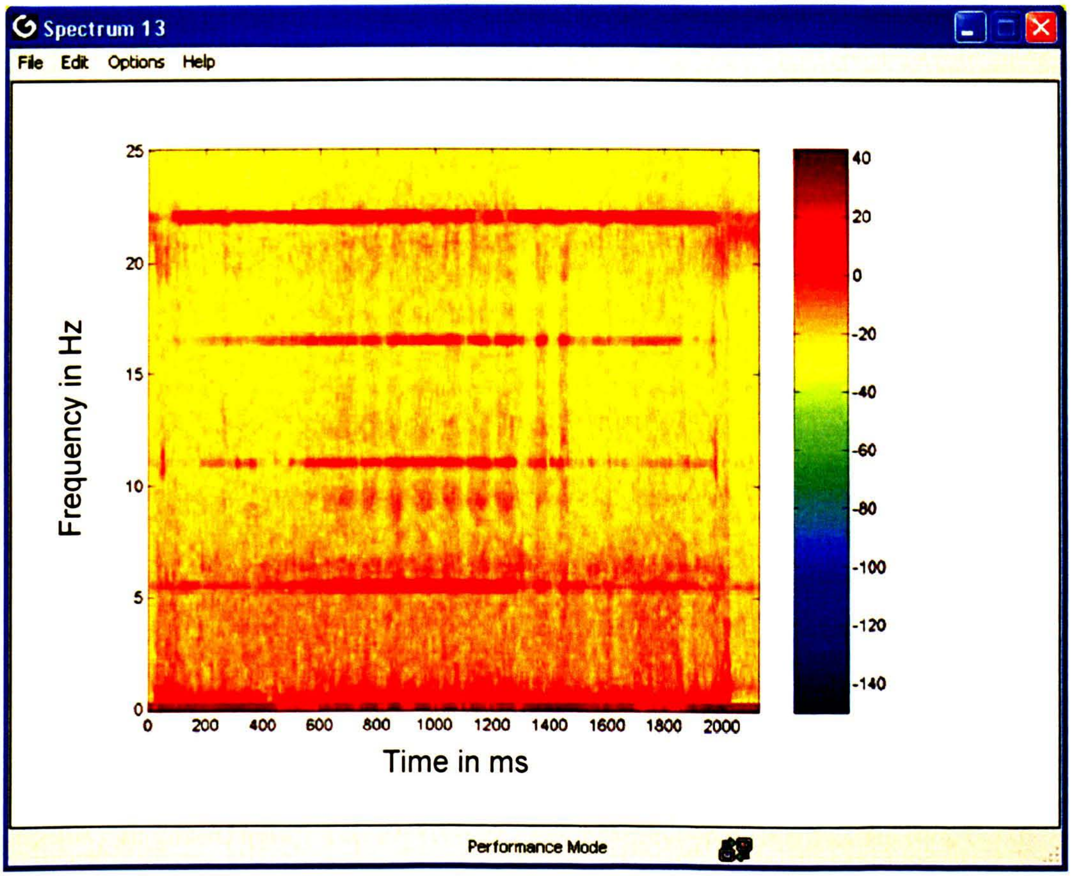


Figure 7.7: Experiment 1: the detailed spectrogram window (original in colour)

For the second part of the experiment the subject was presented with a set of buttons, one for each audification. Before starting scoring, the subject was asked to listen to all the sounds at least once. By clicking on a button the subject could hear each audification through headphones. Again, a click on the 'Test' button brought up the scoring window, identical to that used for the spectrograms (see Figures 7.6 and 7.8).

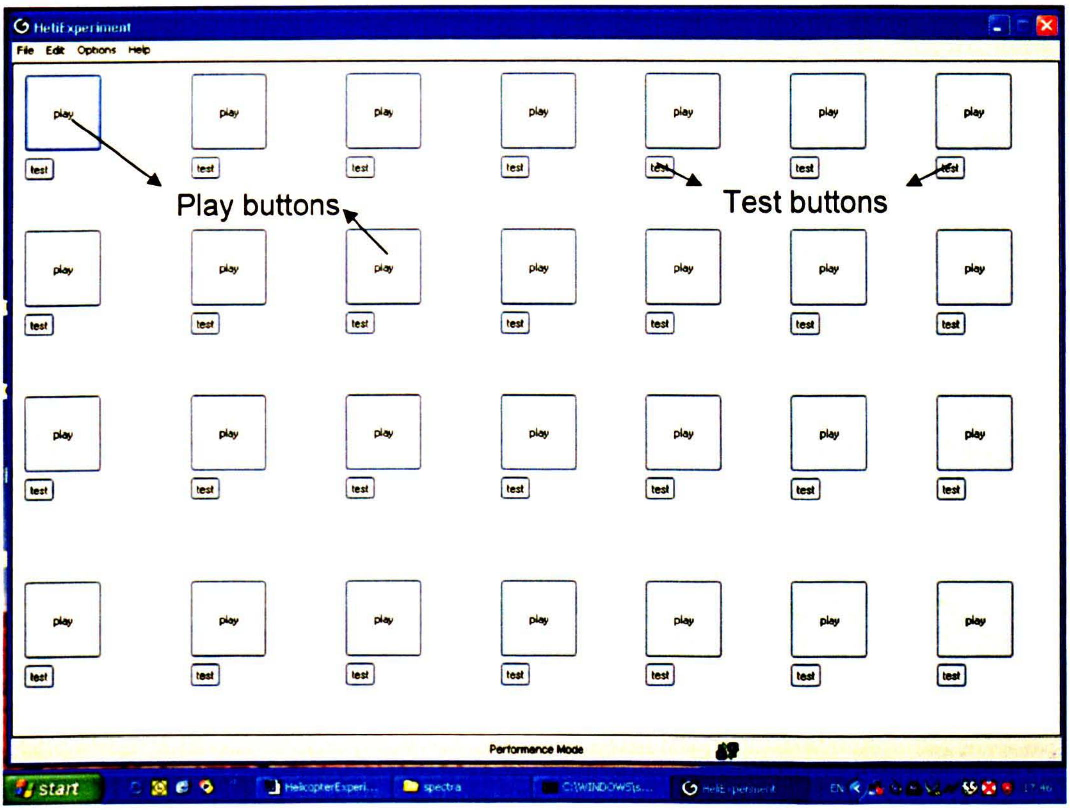


Figure 7.8: Experiment 1 software GUI showing the sonifications buttons (original in colour)

The Test window is the same as in Fig. 7.6. All the results were recorded in a text file saved in the hard drive which had the following format:

```
Data channel ID/noise score/repetitions score/
frequency score/discontinuity score/power score
```

For example:

`28 2 4 1 2 5`

means:

data channel 28
noise score 2
repetitions score 4
frequencies score 1
discontinuities score 2
signal power score 5

Before presenting the spectrograms and the sounds to each subjects, two randomised orders of the numbers 1 to 28 (number of channels) were generated with a program created in PD. In pseudo-C code this program could be written as follows:

```
m = 28;
i = 1;
a[m] = (a1, a2, a3,…, am) = (1,2,3,…,28); //channels' numbers
y[m] = (y1, y2, y3,…, ym) = (0, 0,…, 0);  //new random order

loop for (m >= 1, i <= 28, m=m--; i=i++)
{
n = rand(1, m);                  //random number between 1 and m
yi = an;                         //put in yi the channel number an
a = [1, 2, 3,…, n-1, n+1,…, m]; //eliminate an from the
channels'
                                 //numbers array so it cannot be
                                 //selected again
}
```

When the loop has reached the end the array `y[m]` contains the channels' numbers in a randomised order. This randomisation of the order of presentation of the sounds and the spectra, together with the fact that half the subjects were presented with the spectra first and half with the sounds first, ensured that errors due to order of presentation were minimised.

7.4 The results

The scores were divided and analysed by the five attributes being tested (noise, repetition, frequency, discontinuity and signal power). For each of the 28 data series (i.e. channels of sensor information from the helicopter) two mean scores were calculated across all subjects: one for the sound display and one for the visual display. Therefore for each attribute being tested (noise, repetition, etc.) there are two arrays of average scores (one for the sound and one for the spectra), with an average score across all subjects for each data series. If the two displays portray information in exactly the same way, then the two arrays of scores would be expected to be exactly the same. A scatter plot (x axis = spectra scores, y axis = sounds scores) is plotted for each of the five attributes under test (see Figures 7.9, 7.11, 7.13, 7.15, 7.17). This helps us to see if a relationship exists between the spectra scores and the sound scores.

Before verifying if there is a relationship between the audio and spectra average scores for each attribute being tested (noise, repetition, etc.), two non-parametric Friedman's ANOVA tests were calculated, one on the audio average scores and one on the spectra average scores. This was done to make sure that the differences between the audio average scores were overall statistically significant and that the differences between the spectra average scores were overall statistically significant. If these tests were not significant, then it would mean that all the audio average scores could have the same value and/or that all the spectra average scores could have the same value. In that case it would not make sense to calculate the correlation between the audio average scores and the spectra average scores. Finally, where the results of the Friedman's ANOVA tests were significant, the Spearman's rank correlation between audio average scores and spectra average scores was calculated.

7.4.1 Noise results

Noise results: significance of means	*Friedman's ANOVA Chi-Square(27)*	*significance*	*effect size*
Audio scores	184.050	$p < 0.01$ significant	0.73 (large)
Spectra scores	220.899	$p < 0.01$ significant	0.83 (large)

Table 7.1: Noise results: significance of means

The differences between the average audio scores of the different channels of data are statistically significant. Similarly, the differences between the average spectra scores of the different channels of data are statistically significant.

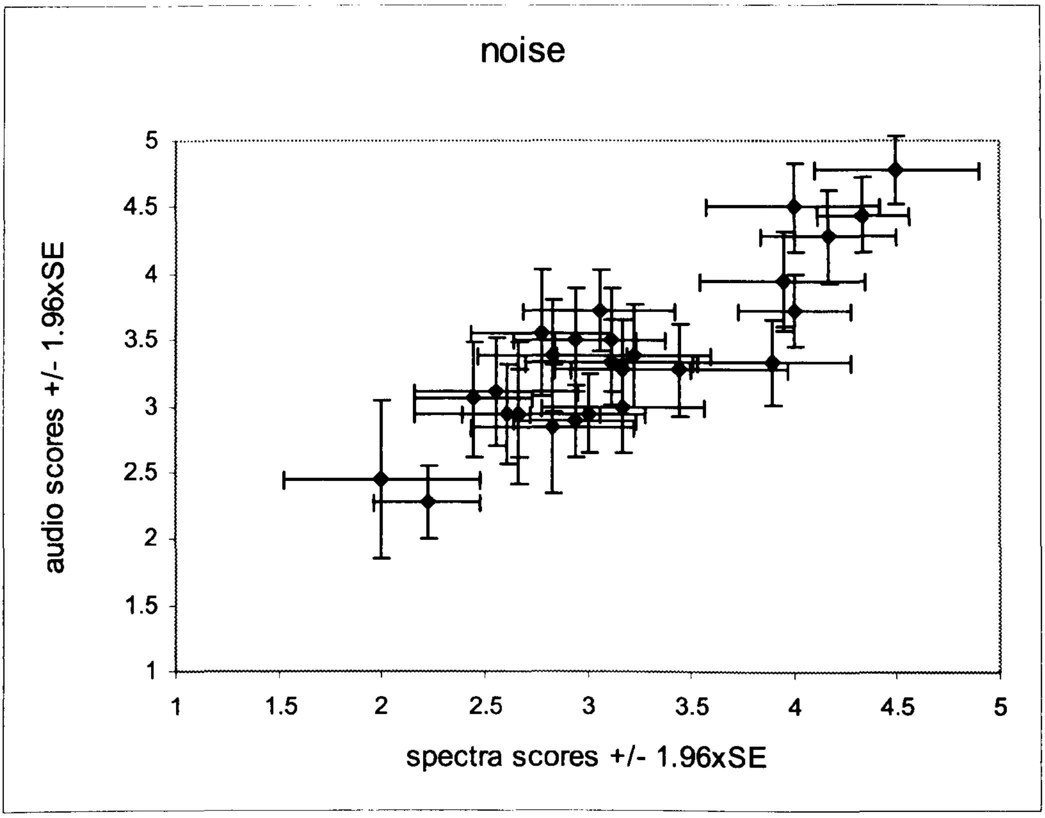


Figure 7.9: Scatter plot for *noise* parameter

Spearman's rank correlation

$r_s = 0.79$, p(one-tailed) < 0.01, significant, effect size = large

The Spearman's Rank correlation is significant which means that the test subjects have scored the presence of noise in the sonifications of the 28 channels of helicopter flight data similarly to how they have scored the spectra of the same channels. This means that the presence of noise in these data channels is similarly perceived by either listening to a short audification of the channels or by looking at the spectra of the channels.

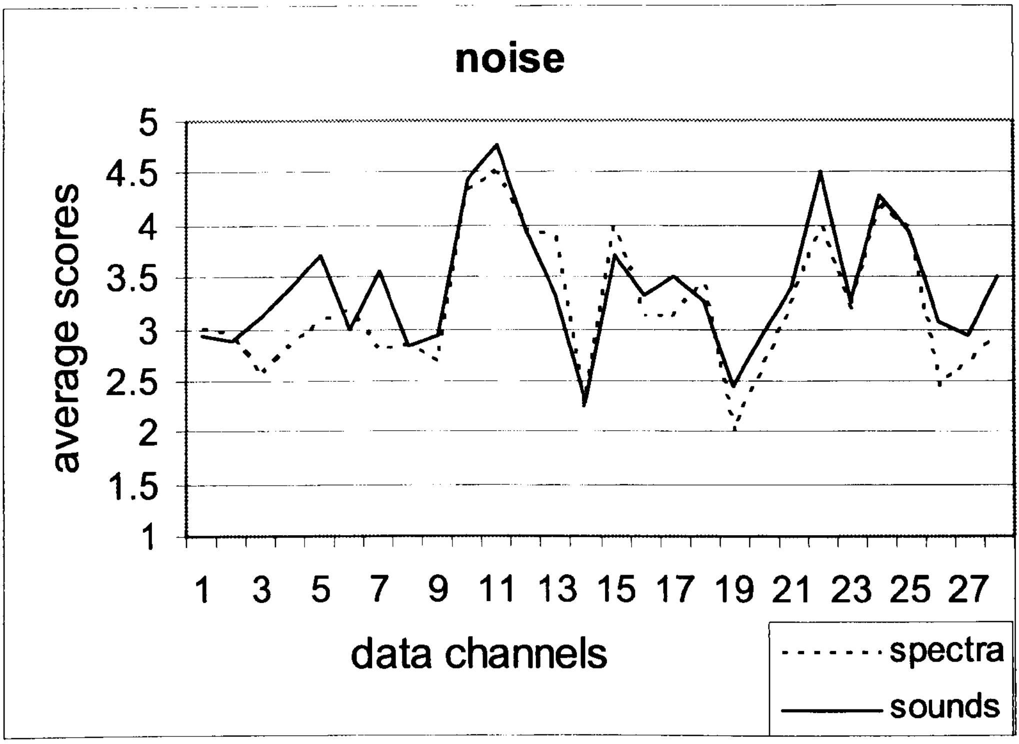


Figure 7.10: Sounds and spectra average *noise* scores vs. data channel number

Figure 7.10 shows how similar the average spectra scores are to the average sound scores. Another way of looking at this is as follows. The average scores are rounded up to the nearest integer (remembering that people were asked to score with a 5-step integer scale) and the absolute value of the difference between the rounded spectra scores and rounded audio scores for each channel is calculated (see Table 7.2).

Difference in rounded scores		
Rounded spectra scores	Rounded sounds scores	Abs(difference between scores)
3	3	0
3	3	0
3	3	0
3	3	0
3	4	1
3	3	0
3	4	1
3	3	0
3	3	0
4	4	0
5	5	0
4	4	0
4	3	1
2	2	0
4	4	0
3	3	0
3	4	1
3	3	0
2	2	0
3	3	0
3	3	0
4	5	1
3	3	0
4	4	0
4	4	0
2	3	1
3	3	0
3	4	1

Table 7.2: Difference in rounded scores

Only 7 data series out of 28 are scored differently in the visual display than in the audio display (for the degree of noise present) and the difference is only 1 point. The rest of the results will be presented in the same formats for each of the remaining attributes.

7.4.2 Repetitive element results

Repetitive element results: significance of means	*Friedman's ANOVA Chi-Square(27)*	*significance*	*effect size*
Audio scores	140.155	$p < 0.01$ significant	0.50 (large)
Spectra scores	259.831	$p < 0.01$ significant	0.90 (large)

Table 7.3: Repetitive element results: significance of means

The differences between the average audio scores of the different channels of data are statistically significant. Similarly, the differences between the average spectra scores of the different channels of data are statistically significant.

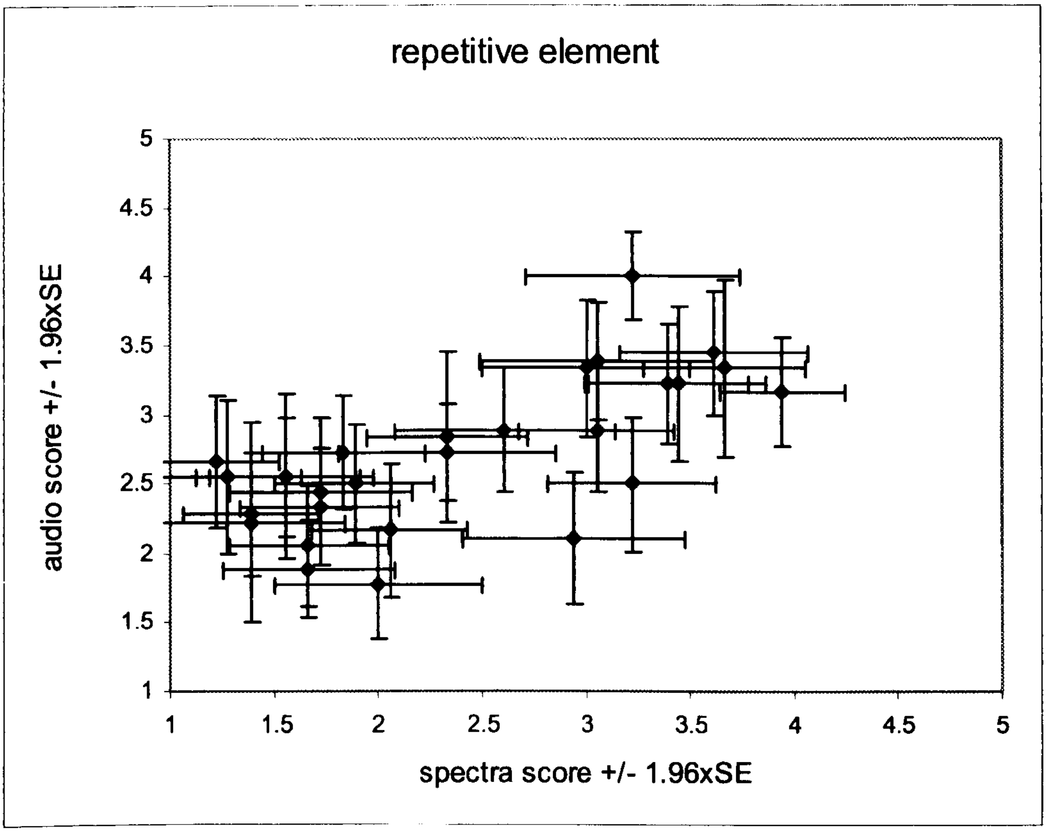


Figure 7.11: Scatter plot for *repetitive element* parameter

Spearman's rank correlation

$r_s = 0.65$, p(one-tailed) < 0.01, significant, effect size = large

The Spearman's Rank correlation is significant which means that the test subjects have scored the presence of repetitive elements in the sonifications of the 28 channels of helicopter flight data similarly to how they have scored the spectra of the same channels.

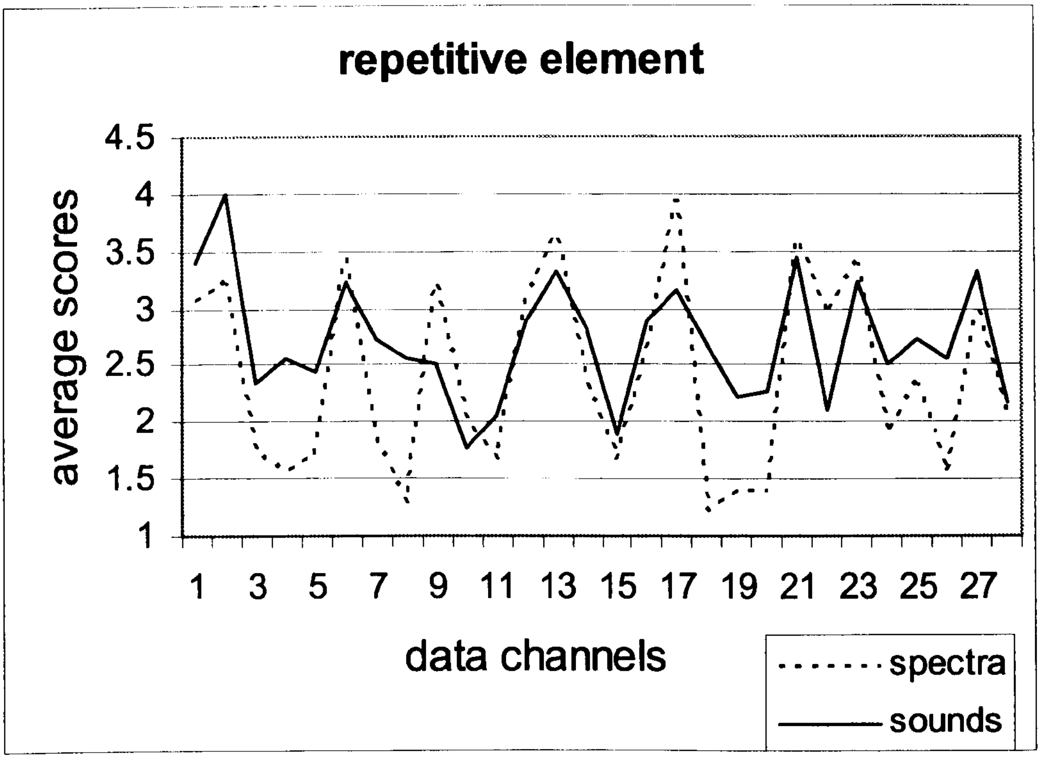


Figure 7.12: Sounds and spectra average *repetitive element* scores vs. data channel number

The Spearman correlation value is lower than that for the noise attribute, showing that there is less agreement between the audio scores and the spectra scores than in the case of the presence of noise. 15 out of 28 rounded average scores are different between the visual and the audio display. In 13 the difference is by 1 point and in 2 by 2 points.

7.4.3 Distinguishable frequencies results

Distinguishable frequencies results: significance of means	*Friedman's ANOVA Chi-Square(27)*	*significance*	*effect size*
Audio scores	229.166	$p < 0.01$ significant	0.76 (large)
Spectra scores	397.775	$p < 0.01$ significant	1.67 (large)

Table 7.4: Distinguishable frequencies results: significance of means

The differences between the average audio scores of the different channels of data are statistically significant. Similarly, the differences between the average spectra scores of the different channels of data are statistically significant.

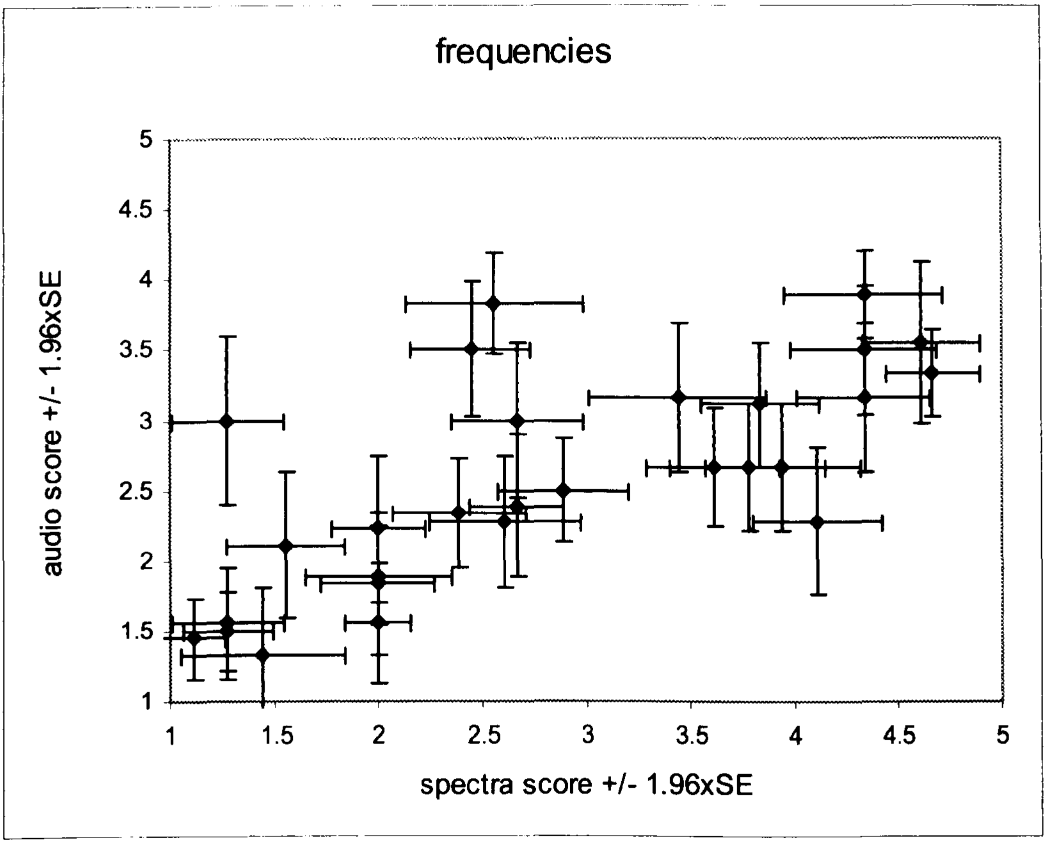


Figure 7.13: Scatter plot for *distinguishable frequencies* parameter

Spearman' s rank correlation

$r_s = 0.74$, p(one-tailed) < 0.01, significant, effect size = large

The Spearman's Rank correlation is significant, which means that the test subjects have scored the presence of distinguishable frequencies in the sonifications of the 28 channels of helicopter flight data similarly to how they have scored the spectra of the same channels. This means that the presence of distinguishable frequencies in these data channels is similarly perceived by either listening to a short audification of the channels or by looking at the spectra of the channels.

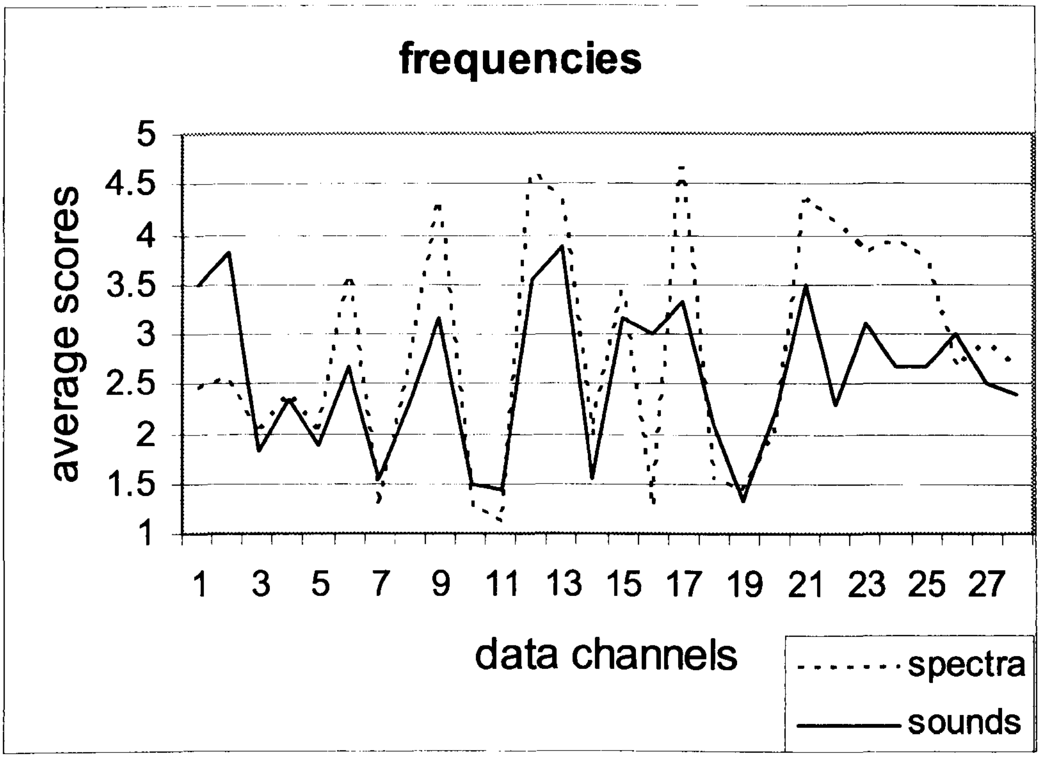


Figure 7.14: Sounds and spectra average *distinguishable frequencies* scores vs. data channel number

This Spearman's rank correlation result is similar to that obtained for the presence of noise. This could indicate that noise and frequencies are characteristics which are more clearly portrayed by both audifications and spectra, than the presence of repetitive elements. 15 out of 28 rounded average scores are different between the two displays: 11 have a difference of 1 point, and 4 by 2 points.

7.4.4 Discontinuities in amplitude results

Discontinuities in amplitude: significance of means	*Friedman's ANOVA Chi-Square(27)*	*significance*	*effect size*
Audio scores	144.039	p < 0.01 significant	0.69 (large)
Spectra scores	207.660	p < 0.01 significant	0.79 (large)

Table 7.5: Discontinuities in amplitude: significance of means

The differences between the average audio scores of the different channels of data are statistically significant. Similarly, the differences between the average spectra scores of the different channels of data are statistically significant.

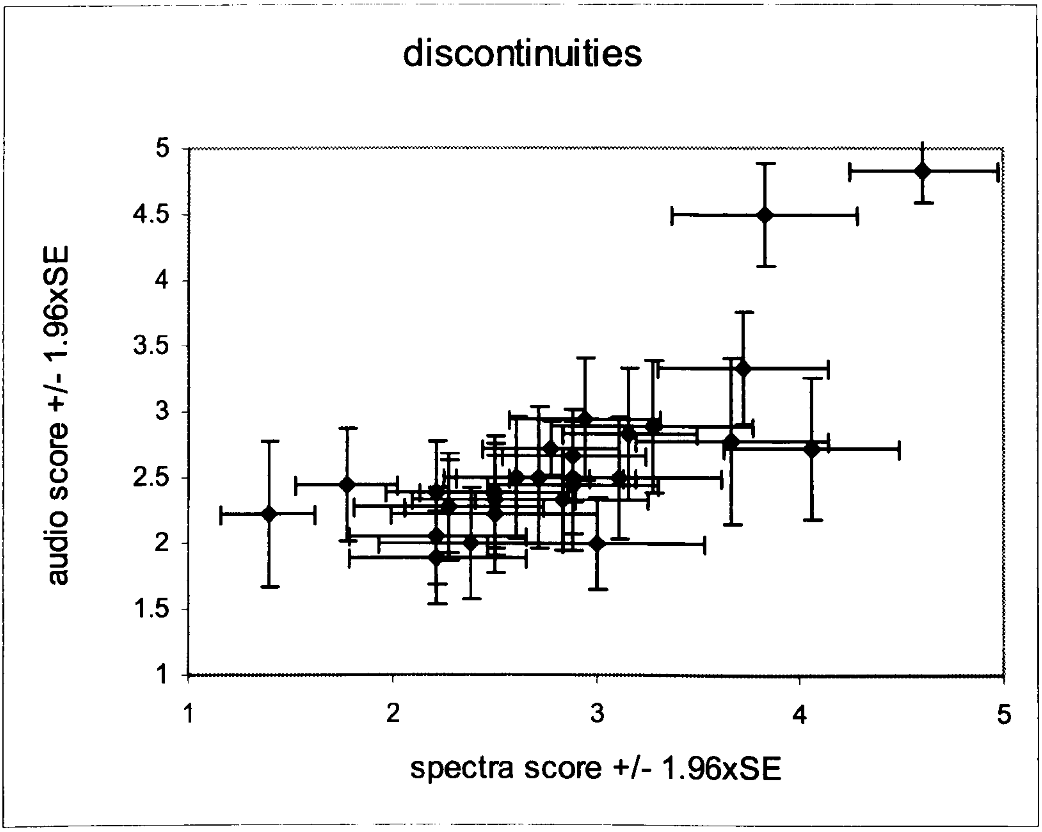


Figure 7.15: Scatter plot for *discontinuities in amplitude* parameter

Spearman's rank correlation

$r_s = 0.78$, p(one-tailed) < 0.01, significant, effect size = large

The Spearman's Rank correlation is significant which means that the test subjects have scored the presence of discontinuities in amplitude in the sonifications of the 28 channels of helicopter flight data similarly to how they have scored the spectra of the same channels. This means that the presence of discontinuities in amplitude in these data channels is similarly perceived by either listening to a short audification of the channels or by looking at spectra of the channels.

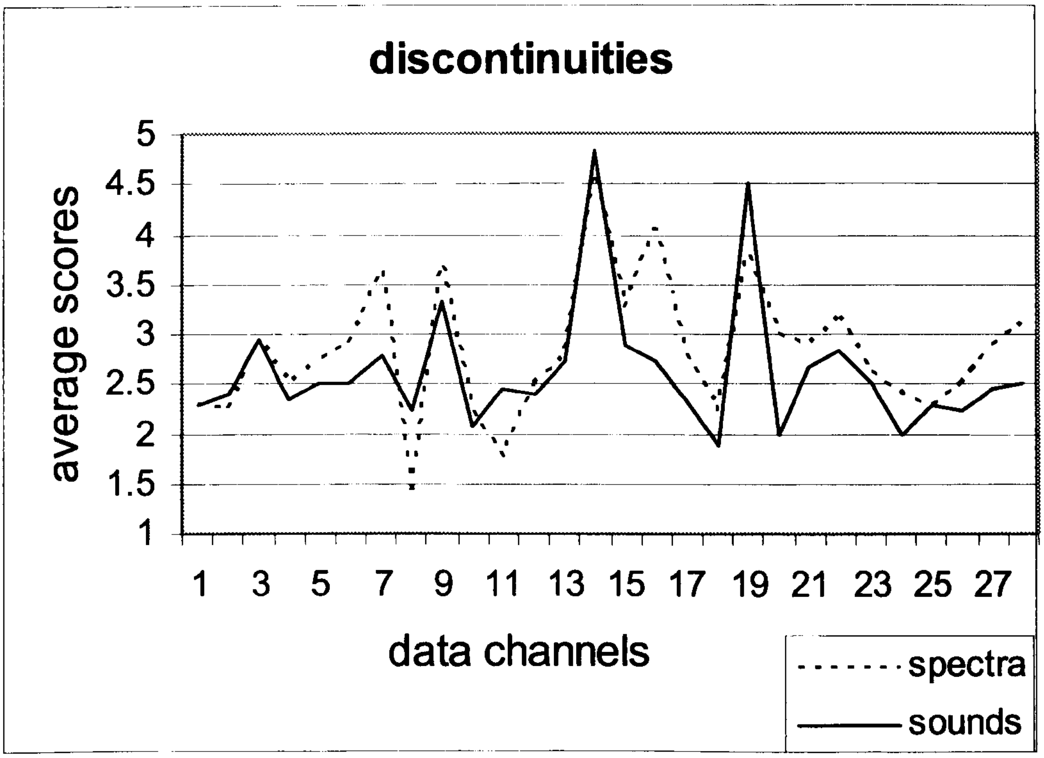


Figure 7.16: Sounds and spectra average *Discontinuities in amplitude* scores vs. data channel number

This Spearman's rank correlation result is similar to that obtained for the presence of noise and the presence of distinguishable frequencies. This again could indicate that noise, frequencies and discontinuities in amplitude are characteristics which are more clearly portrayed by both audifications and spectra than the presence of repetitive elements. 11 out of 28 rounded average scores are different between the 2 displays: in all cases the difference is by 1 point.

7.4.5 Signal power results

Signal power results: significance of means	*Friedman's ANOVA Chi-Square(27)*	*significance*	*effect size*
Audio scores	315.862	p < 0.01 significant	1.19 (large)
Spectra scores	373.666	p < 0.01 significant	1.39 (large)

Table 7.6: Signal power results: significance of means

The differences between the average audio scores of the different channels of data are statistically significant. Similarly, the differences between the average spectra scores of the different channels of data are statistically significant.

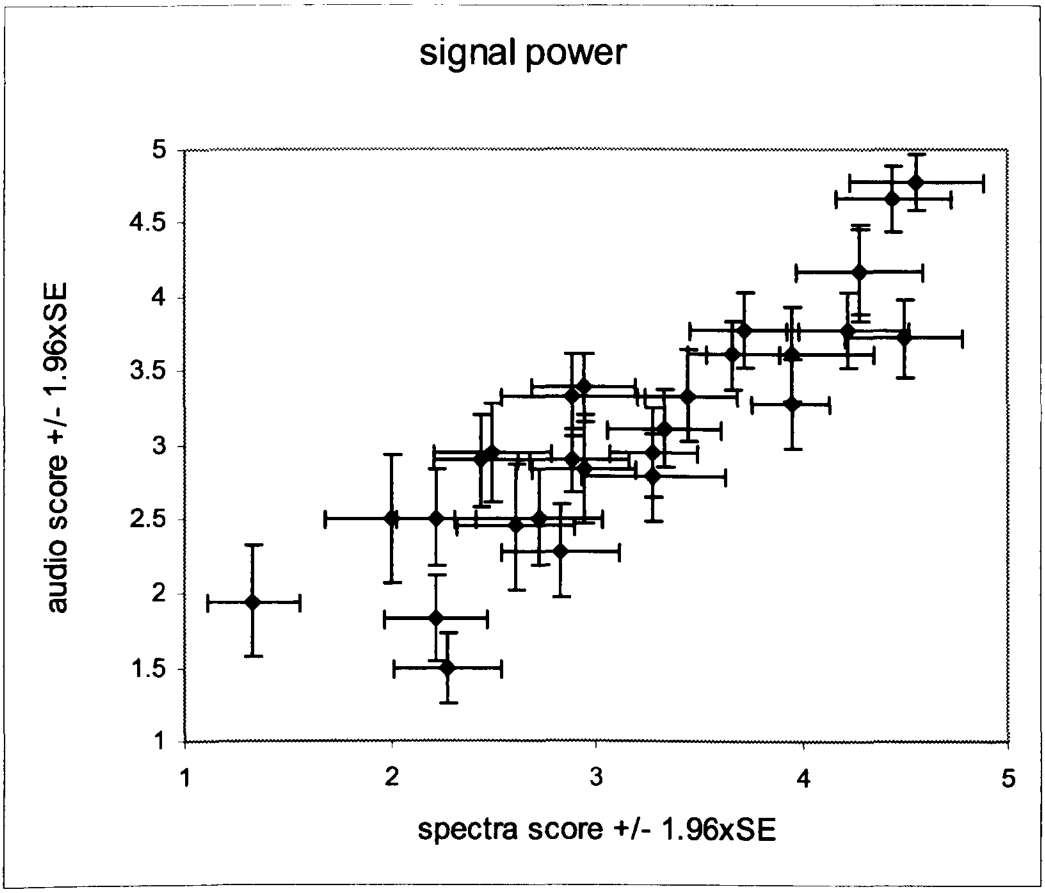


Figure 7.17: Scatter plot for *signal power* parameter

Spearman's correlation

$r_s = 0.89$, p(one-tailed) < 0.01, significant, effect size = large

The Spearman's Rank correlation is significant which means that the test subjects have scored the signal power in the sonifications of the 28 channels of helicopter flight data similarly to how they have scored the spectra of the same channels. This means that the signal power in these data channels is similarly perceived by either listening to a short audification of the channels or by looking at the spectra of the channels.

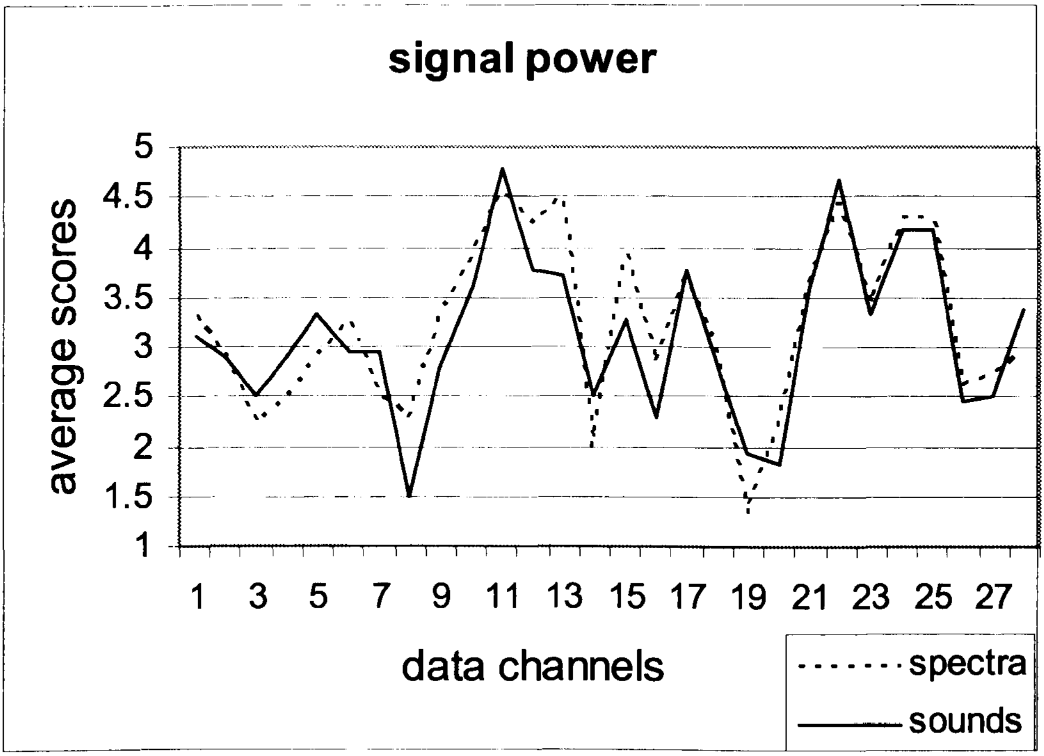


Figure 7.18: Sounds and spectra average *signal power* scores vs. data channel number

This Spearman's rank correlation result is the highest among the results obtained in this experiment. This indicates that the signal power is the most similarly portrayed characteristic by both audifications and spectra. Only 9 out of 28 average rounded scores are different between the displays and these are by only 1 point.

7.4.6 Discussion of results

For each of the five attributes the average scores for the spectra show high rank correlation with the average scores for the sounds. This means that the two displays do indeed allow users to gather some basic information about the structure of the data to a similar degree. In particular, from this experiment it appears that signal power is very similarly portrayed by audifications and spectral observations. Noise, distinguishable frequencies and discontinuities in amplitude are also similarly portrayed using audifications or spectra, though to a slightly lower degree than for the signal power characteristic. The presence of repetitive elements is still similarly portrayed by using either audifications or spectra. However, from the results this similarity appears to be less strong than for the other characteristics. These are overall positive results which mean that data analysis through audification in this field is a feasible proposition. Further work in this field could concentrate on finding out if some information is better displayed as audio than through a spectrogram and vice-versa. This could then lead towards the development of a data analysis application that employs both audio and visual displays to represent data and that exploits the best properties of both these two ways of representing data to produce a more effective and efficient display.

From this experiment, it is reasonable to think that the degree of similarity of the two displays could be improved by considering the following:

- the audio display could be improved by choosing a different data scaling better informed by sound perception principles;
- the subjects were presented with a complex task. They had to score 28 data series for each of the 5 attributes, both in the visual mode and in the audio mode. It is possible that an easier task (e.g. score 10 channels for one category only at the time) could show an even higher similarity between the data;
- the subjects had to score very complex sounds containing (to varying degrees) noise, clicks, the presence of many frequency components, and often a complex evolution of the sound over time. Again with easier sounds, i.e. simpler data structures, the similarity in the scores could be higher;
- the test questions were often ambiguous. For example, subjects often wondered if the noise of the clicks, produced when there is a discontinuity in amplitude, should count for 'presence of noise' since it was already accounted for under 'presence of discontinuity'. These ambiguities have surely contributed to the increase in variance in the results. Less ambiguous questions could yield better results.

The correlation between average scores for the visual display and the auditory display in the Noise and Signal Power attributes is higher than that for the other three. The reason for this difference can probably be found in the nature of the data displayed and the way the displays were built. For instance, the perception of frequency influences the perception of loudness, e.g. to perceive a 100Hz and a 1kHz sound with the same loudness, the level of the 100Hz sound needs be higher than that of the 1kHz sound [Howard and Angus, 1996]. It is possible, therefore, that frequencies that can be seen in the spectrogram are not easily perceivable in the audification. The difference could also be due to the different characteristics of the visual and the auditory sense: one could be better at picking up certain elements than the other. For instance, the ear could be better at perceiving repetitive elements in time, since we are used to recognizing rhythmic structures in sound, while repetitions could be harder to spot in a spectrogram.

Finally, during the test, the subjects were free to write down any comments about the spectra or the sounds or the test procedure. 13 out of 23 subjects chose to comment and here is a summary of the most common observations:

- "It is difficult to score in particular noise, discontinuities and repetitions" (7 comments);
- "I can hear more detail in the sounds than in the spectra" (2 comments);
- "I feel that I get better at scoring as I go along" (4 comments);
- "Some data sets actually sound like a helicopter" (2 comments).

These comments highlight questions that could be investigated in further experiments such as:

- are some characteristics more difficult to define and therefore to score?
- which details are best portrayed with sound and which ones in a spectrum?
- does training improve our ability to analyse complex sounds and therefore complex data?

7.5 A different point of view: an example of model-based sonification

During the time available for this research project, the author had the opportunity to create a second, very different interactive sonification of this helicopter's data set. This second sonification is a model-based sonification (MBS) and it is presented here for two main reasons:

1) it clarifies with a practical example the MBS concept already presented in Chapter 2;
2) it allows for the discussion of the pros and cons of parameter mapping and MBS sonifications.

This MBS example was realised during a two week research visit funded by the European Project Congas-COST287 [Congas] in Bielefeld University. The author worked with German researcher Dr Hermann to apply MBS to the helicopter's data set.

The aims of this visit were:

- to create a MBS for a multivariate time series;
- to create a multimodal and interactive interface for it;
- to apply the system to the helicopters' data set.

The programming environments used for this project were:

- Neo/NST [Neo/NST]: an application for data analysis and data mining developed in Bielefeld University, for the data-related computations;
- Pure Data [Pure Data]: for the programming of the sonification and the multimodal interactive interface;
- OSC (Open Sound Control) [Open Sound Control]: for the communication between NeoNST and PD.

In MBS, any data set is considered as a series of points, or vectors, in the multidimensional state space of the system being studied. These points are then interpreted as particles of a material that, if excited in some way, produces sound as a reaction to the excitation. In the case of the helicopter's data, the points of the data set represent the trajectory in the state space travelled by the helicopter during a half an hour flight. In this example, the main idea was to select some important and representative points in the state space of the system and consider them as the material to be "played". Once this material was defined, the helicopter, represented by a moving point in the space state, would travel in the state space following the trajectory described by the data set. When the helicopter travelled closer to the particles of the 'playing' material, it would excite them and make them sound. This way, the evolution in time of the sonification would follow the evolution in time of the original flight.

7.5.1 Data computations

The intrinsic dimensionality of a data set, i.e. the number of independent dimensions sufficient to describe each point of the data set, is often much lower than the number of channels present in the original data and can be found using a technique called Principal Component Analysis.

Using Neo/NST, a Principal Component Analysis was performed on the data to find out how many of the 28 dimensions were sufficient to describe the variance present in the data. It was found that only 5-6 dimensions contained more than 90% of the variance. The data set was then rotated so that the first six columns of the rotated data set represented the projections of the original data points onto the principal components. This rotation allowed in particular the creation of a meaningful two-dimensional visual display just by plotting the projection of the original data on the plane made by the first and second principal component.

7.5.2 Vector Quantisation

As mentioned in Section 7.5 of this chapter, a certain amount of meaningful points of the state space were to be selected to create the "sounding material". These points were selected using a technique called Vector Quantisation (VQ). With VQ, a fixed number (30 in this case) of points (called in VQ prototypes) can be placed in the state space in such a way that the average distance between the prototype and all the original data points lying in the Voronoi cell (a particular way of dividing a space based on the distances between points in the space) associated with the prototype is minimised. Practically such a configuration is obtained by starting with a set of random points as prototypes and iteratively:

(i) re-assign the original data points to the nearest prototype

(ii) move all the prototypes to the position in which the average distance to the nearest points is minimised.

After a certain amount of iterations the prototypes (centres of the cells) do not move anymore and their positions are highly representative of the "shape" of the original data set.

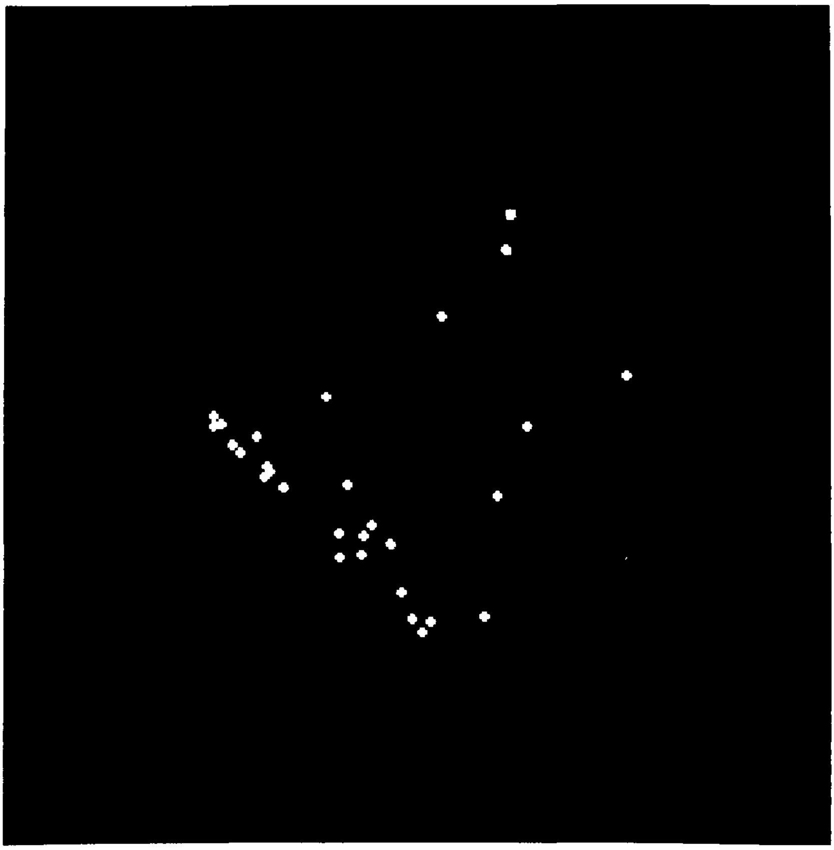


Figure 7.19: The image illustrates the projection of the 30 prototypes onto the plane made by the first two principal components of the data set. The display is programmed in PD using the GEM graphical library for the visual display

In Figure 7.19 the dots represent the projection of the 30 prototypes, calculated for the helicopters data, onto the plane made by the first two principal components. The visual representation is programmed in PD using the graphical library GEM. This reduced data set constitutes the "material" to be excited and to produce sound in the model.

7.5.3 The sonification model

In model-based sonification, five elements need to be defined [Hermann, 2002]:

1. *Setup*: what physical metaphor is attributed to the data set;
2. *Dynamics*: what physics laws apply to the data set given the metaphor of the setup;
3. *Excitation*: how is the system excited;
4. *Interaction*: by what means does the user interact with the system;
5. *Sound*: how is the sound constructed.

In this model:

1. *Setup*: around the prototypes we imagine a sphere of radius R.
2. *Dynamics*: if the prototypes are excited, there will be an oscillatory behaviour with a damping term. In sound this will be represented by an oscillator with decaying amplitude. To distinguish the spheres (prototypes) sonically, a different pitch is assigned to each prototype.
3. *Excitation*: when the system's (the helicopter in this case) trajectory (represented by the yellow dots in Figure 7.20) passes inside the prototypes' spheres, the prototypes are excited and produce sound. The sound's volume depends on energy of impact and how long the system remains inside the sphere.

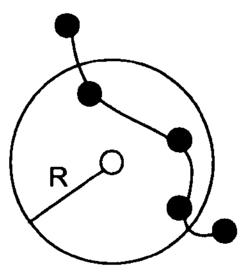


Figure 7.20: Representation of when the helicopter trajectory passes near a prototype

4. *Interaction*: the user can guide the point representing the helicopter through its trajectory at his/her own speed using the mouse or the jog wheel/Shuttle XPress [Contour Design] interface. The point/helicopter can also be set to travel through its trajectory at a set speed.
5. *Sound*: each time a prototype is excited an oscillator plays and if more than one prototype is excited at the same time, then the sounds of the different prototypes are mixed together.

7.5.4 Observations

The sonification, a video showing the graphical interface and the PD program can be found in the accompanying DVD in folder Model-based sonification.

This sonification provides an overview of the whole data set. In the audio-visual display, take off, middle section and landing sections of the flight are distinguishable. These sections are audible, however the visual display helps distinguishing these sections much more clearly. In the middle section of the flight, repetitions are heard which imply that the helicopter goes through similar states at various times. This model is general, can be applied to any time based data set and does reproduce the evolution of the system in time.

7.6 Conclusions

In this chapter, an experiment was presented that compared the effectiveness of a sonification display (audifications) of a large multivariate data set with that of an equivalent visual display (spectra) of the same data set. The results of this experiment showed that the sonification display can be considered to be as effective as the visual display in portraying five basic characteristics of the data. The audification and spectrogram of each channel of data were scored in terms of presence of noise, distinguishable frequencies, discontinuities in amplitude and signal power by a group of test subjects. High Spearman's correlations were found between the average scores of the audifications and spectra for all the five characteristics explored. In particular, the scores for signal power show the highest rank correlation. That is, this seems to be the characteristic portrayed most similarly by the audio and visual display. Noise, distinguishable frequencies and discontinuities in amplitude also showed high correlation, though slightly lower than that for signal power. The presence of repetitive elements still showed a significant correlation between audio and visual average scores, however the correlation value is lower than that of the other characteristics, implying that the two displays are slightly less similar in portraying this characteristic than in portraying the others. Overall these results are positive and support the idea that audification can be an effective tool for the analysis of these data. Further experiments could be conducted to improve this auditory display and to explore more deeply its potential as an informative tool on its own and in the context of a multimodal analysis application which includes both audio and visual displays.

In section 7.5 of this chapter an example of interactive model-based sonification of the same data was described. It was observed that the MBS example displays an overall summary of what is happening in the data set. However, the implementation of this summary does require a certain amount of data pre-processing and the creation of an appropriate data/sound relationship. On the other hand, audification is an effective method for displaying the basic information as well as more complex details and requires only a very limited amount of data pre-processing.

Chapter 8 – Experiment two: the sonification of electromyography (EMG) data

8.1 Introduction

This chapter describes experiment two which aims to show that the sonification of a multivariate large data set is effective in portraying information. The particular data set used in this experiment contains electromyography (EMG) data. To set the context, in the first section of this chapter a background is given that summarises the final aim of this particular display, how sound has been used in the past to evaluate muscle activity and what standard techniques are used to date to analyse EMG data.

The experiment and the sonification of EMG data are then described in detail and the quantitative and qualitative results discussed. One important characteristic of the sonification described here is that it can be produced in real-time, as the EMG data are gathered. The final section of this chapter describes how, in this project, the author implemented a real-time working version of this sonification which can be used as a prototype of an auditory feedback of EMG data.

The data used in this experiment were provided by collaborators from Teesside University. The results of this experiment contributed to submission for a new research proposal (a collaboration between the Universities of York, Teesside and Huddersfield) to the EPSRC in February 2007.

8.2 Background: the overall aim of this display

Physiotherapists use EMG sensors to monitor the electrical activity of the muscles of patients. Electrodes attached to the skin of the subject detect the electrical signals from the muscles below the skin, and send it to a computer where the signal is transformed into digital information. The computer typically runs an application that receives the data, performs some basic statistics on them and displays them in graphical form.

Such applications nowadays display the data in real-time as they are gathered and can also store it for later analysis. The physiotherapist will try to spot irregularities as the data are gathered, but the data can only be thoroughly analysed at a later time because of its complexity.

EMG signals are believed to be full of information about the muscle activity and it is hypothesised by our collaborators at Teesside University that this visual analysis does not exploit to the full the information contained in the data.

The fact that the technology allows us to display the data in real-time, i.e. as they are gathered, creates the possibility for the patient to be involved in a deeper understanding of his/her movement. If the patient was involved in both producing the data, understanding their properties and then adapt the next movement to obtain a better result, then the EMG display would work as a real-time biofeedback for the patient and could become a very useful tool for rehabilitation.

Two main obstacles do not allow the current EMG displays to work as biofeedback:

1) most of the graphical display is meaningless to an non expert;
2) there is little shared and non-technical language that allows the transfer of clear and simple information between the physiotherapist and the patient.

It is believed that a sound display of EMG data could help surpassing these two obstacles and allow creating a real-time biofeedback.

Some overall characteristics of sound are experienced and easily understandable by all: for example, we all understand the difference between a loud or a soft sound, between a high pitched and a low pitched sound, between a not noisy and a noisy sound, etc. Therefore if we could produce an auditory display of EMG data that maps correct movement to a clear and overall characteristic of a sound, and incorrect movement to a clear, overall and opposite characteristic of the sound, then the patient could have a clear understanding of what a good movement sounds like and could use that sound as the target to aim for in rehabilitation when exercising the muscle's movement. Sound has also the advantage of being omni-directional so the patient can easily interiorise the relationship between a certain physical movement and a sound. Examples of research on auditory bio-feedbacks are [Ghez *et al.*, 2000], [Fox & Carlile, 2005], [Effenberg, 2005] and [Hinterberger & Baier, 2005] and were described in Chapter 2, Section 2.3.3 of this thesis.

8.3 Background: muscle sounds

The idea of using sound to monitor and display muscle activity is not new. Actually, muscles emit sound naturally. The oscillation of the muscles' fibres creates vibrations, many of which are audible and detectable at the body surface. An interesting description of what the muscles sound like is given in the forward of Stokes and Blythe's book *Muscle Sounds in physiology, sport science and clinical investigation* [Stokes and Blythe, 2001] and the information about muscle sounds that follows here is summarised from the same book.

Astronaut Dan Barry describes the first time he heard the sound of his muscles in the following way:

> "I first recognised muscle sounds late one night when I was a resident in rehabilitation medicine, trying to get some sleep. I was lying in bed with my right arm under my ear and heard a low frequency rumbling. I found that I could modulate the loudness by increasing and decreasing the intensity of contraction of my biceps brachii muscle." [Ibid., p. vi]

Another way of listening to muscle sounds is through a stethoscope. These sounds are a direct display of the mechanical performance of the muscles.

They were first discovered and observed around 1660 by Italian scientist and Jesuit priest Francesco Maria Grimaldi who reported the existence of muscle sounds in *Physico-mathesis de lumine, coloribus, et iride, aliisque adnexis libri duo*, published in Bologna in 1665.

In history, interest in the study of muscle sound appeared and disappeared over time. Information on this research subject was never widespread over the whole of the scientific community. For this reason, the same discoveries were repeated by different people over time. Nineteenth century physiologists were fascinated by the subject, but the lack of clear practical applications consigned these studies to obscurity. Scientists such as Helmholtz made leading contributions to the field. Technology played an important role in keeping interest in the subject alive throughout history.

In 1880, a fine plot of the acoustic spectrum of muscle sounds was produced, and in 1923 the sounds were recorded electronically. Finally, the technological innovations of the twentieth century allowed for a resurgence of interest in the subject. With modern computers these sounds can be easily recorded, stored and analysed.

As the research in this field was so scattered, a unified terminology was not agreed upon for a long time. Only recently has the scientific community decided on unified terms. In the mid 1990s, at an International CIBA (now Novartis

Foundation [Novartis Foundation]) Foundation Symposium in London, researchers put their studies together on this matter.

In literature, many terms were used to refer to the sound of muscles: muscle sounds, muscle vibrations, acoustic myography, sound myography, phonomyography, vibromyography, etc. Finally, in 1995, the term mechanomyography (MMG) was accepted as the term to refer to studies on the sound of muscles. [Stokes and Blythe, 2001]

The data used in this chapter is not MMG data, but EMG data. However, this background on the sound of MMG data is relevant for various reasons.

The main difference between MMG and EMG data is that the first signal gives information about the mechanics of the muscle and the second about the electrical characteristics of the muscle.

The aim of both mechanomyography and electromyography is to record these signals and analyse them. The signals' data are similar and the traditional analysis techniques used to analyse them are the same. Sound produced by the MMG signals can be amplified and used as a display of the data. Similarly, EMG data can be transformed into sound by means of sound synthesis techniques and the produced sound can give information about the signal. MMG and EMG can be analysed and treated similarly.

The study and understanding of both MMG and EMG signals have potential applications in many fields such as sport science, physiotherapy, diagnostic ability and routine clinical assessment.

Examples of applications are assessment of force and fatigue in rehabilitation, biofeedback in sports science and rehabilitation, physiological investigations and diagnosis of muscle disorders.

8.3.1 Standard analysis of EMG signals

As mentioned above, with today's technology, EMG (and MMG) signals can be recorded and stored in computers for analysis. This section describes the standard ways of displaying and analysing the signals. The main information of this section is summarised from [LeVeau and Andersson, 1992].

The two most important parameters in the study of EMG signals are amplitude and frequency. The changes in these two parameters can be quantified and used to classify the electrical activity level that produces a certain muscular tension.

> "The change in the myoelectric signal is based on the recruitment and firing rate of motor units within the muscle. In general, as more force is needed, more motor units are recruited, and the motor units already firing increase their frequency of firing. [...] The interpretation of the changes in recruitment and the changes in firing rate can provide information concerning the muscle's level of force or its level of fatigue." [Ibid., p. 70]

There are various methods used to extract information about frequency and amplitude of the raw signal: a brief description of the ones most used follows.

1) Normalisation

Before analysing EMG data, a standard of reference must be found to allow for comparisons.

The amplitude of the EMG signal is an indirect measure of contraction force [Ibid, p. 70]. The fact that the relationship is indirect means that there is no one to one relationship between amplitude and contraction force. In order to be able to compare how this relationship varies between muscles, individuals and activities, a standard of reference must be found. The process of finding this standard is called *Normalisation*.

During a MVC or MVIC (Isometric Maximal Voluntary Contraction), i.e. a voluntary contraction of the muscle with no change in length of the muscle, the relationship between force and amplitude is near linear (i.e. near one to one). The measure of how amplitude varies with force in MVIC can be taken as reference for normalisation.

Once this standard of reference has been established, there are many methods to reduce the data and present them in numerical form.

The most important information that EMG analysis can produce is:

1. whether or not the muscle is active, when exactly it becomes active and what is the pattern of muscle activity during a movement, position or force production;
2. the relative amount of activity of the muscle;
3. when a peak of activity occurs;
4. whether fatigue has occurred.

2) Analysis method 1: Monitoring the raw signal

The raw signal is the basis of the analysis. It contains all the possible information we can have from EMG. While the EMG signal is recorded, it is displayed in a graphical form on the screen.

> "[The analyst] should monitor the raw signal, even though other signal processing may be used, so that artefacts can be detected and controlled as necessary. In the past, probably the most common way to interpret EMG was by visual inspection of the raw signal. The observer should be able to identify when the raw signal indicates that a muscle is active and when it is relaxed. The relative amount of activity may be classified either by words, such as nil, negligible, slight, moderate, marked or very marked, or by numerical values, such as 0-5, with 0 being no activity and 5 being maximal activity. Such visual observations are based on signal amplitude and frequency." [Ibid, p. 74]

The raw signal should be monitored for all investigations, because the investigator can pick out major artefacts and eliminate that area or part of the signal. To monitor the raw signal in real-time means to look at a graph that contains a lot of noise. The expert analyst is used to check for anomalies in the signal, but this monitoring requires a lot of experience and focus (the analyst cannot look away). Sound can be a good alternative for monitoring the raw signal. The analysis methods that follow can be considered higher order level analyses because they require transforming (e.g. scaling, produce averages, etc.) the raw signal before extracting information.

3) Analysis method 2: Linear envelope

The profile (envelope) of the EMG signal is found by rectifying the signal and passing it through a low pass filter that follows the peaks and valleys of the rectified signal. The combination of a rectifier and a low pass filter is known as a linear envelope detector. The characteristics of the filter must be stated because they determine what information can be extracted after the signal has been processed in this way. The linear envelope gives information about the onset and the duration of muscle activity and the pattern of muscle contraction. This type of analysis has allowed discovering, for example, that the EMG signal lags behind the production of tension by about 60 to 100ms [Ibid, p.79].

4) Analysis method 3: Root mean square or RMS

This provides a measure of the electrical power of the EMG signal. It produces a linear envelope of the voltage or a moving average over time. The RMS wave form is similar to the linear envelope wave form. The electrical power of the signal is an indicator of the functional state of muscle tissue and other characteristics of the muscle [Ibid, p.79].

5) Analysis method 4: Integration

The total amount of muscle activity occurring during any given time interval is represented by the area under the curve during that time interval. The process of determining this area is called *Integration*. It represents the number of active motor units and their rate of firings [Ibid, p.83].

6) Analysis method 5: Frequency analysis

This analysis method looks at the power spectrum of the signal created using Fourier analysis. This method is useful to evaluate local muscle fatigue. With a sustained muscle contraction, the high frequency components of the signal decrease, but the low frequency components gradually increase. This change results in a shift in the power spectrum toward the lower frequencies. Both mean frequency and median frequency (the frequency having 50% of the frequency distribution at each side) are calculated. Shifts in frequency give information about fatigue because frequency is related, among other things, to synchronisation of motor units and increase and decrease of number of active motor units which relate to fatigue [Ibid, p. 84].

7) Analysis method 6: Zero crossings

The number of times the raw signal crosses the baseline (zero line) is related to muscle contraction force. The zero crossings are just counted [Ibid, p.87].

8) Analysis method 7: Spike counting

The number of positive and negative peaks in the raw signal is related to the amount of muscle activity. Spikes are counted [Ibid, p.87].

9) Analysis method 8: Turns

The number of times a signal changes direction (not necessarily passing through zero) is counted. The number of turns increases rapidly as muscle force of low levels increases, but increases very slowly at high levels of muscle force. Turns analysis discriminates well between low level muscle forces, but not between high level muscle forces [Ibid, p.89].

8.4 The sonification of EMG data

One aim of this research is to study if it is possible to meaningfully map EMG data to sound parameters in order to create an informative sound display of the data themselves. If found to be effective, this new display has the potential to become a new useful tool for the analysis and the monitoring of EMG, in particular when coupled with the above mentioned standard analysis methods. The sonification described here has also the advantage that it can display the data in real-time, i.e. the EMG data can be fed to the sound engine as they are gathered, and the resulting sound can be used as real-time audio-feedback, or biofeedback, for the use of the patient, the physiotherapist and the analyst.

In this section, the EMG data used for this study are described, the sonification technique is explained and a description is given of the experiment set up to evaluate the effectiveness of the sonification technique.

8.4.1 The data

The data used in this experiment were gathered by physiologist Dr John Dixon, lecturer at Teesside University (Middlesbrough), for his PhD research work [Dixon, 2004].

Dr Dixon measured the EMG data of three muscles of the leg: the Vastus Medialis, the Vastus Lateralis and the Rectus Femoris.

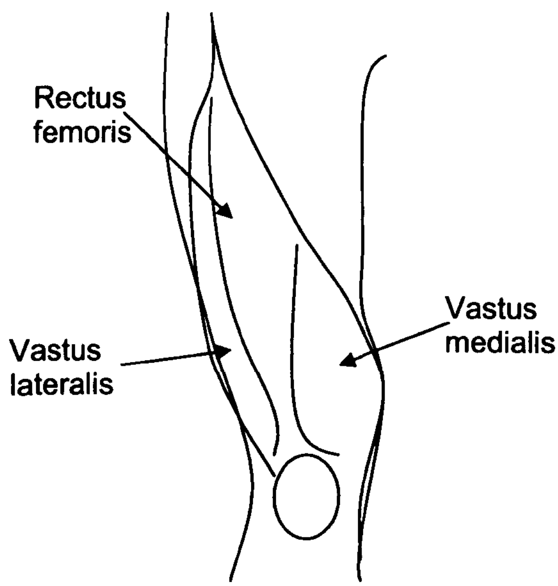


Figure 8.1: This diagram sketches the positioning of the vastus lateralis, vastus medialis and the rectus femoris on a right leg

All data were collected using a data acquisition system (BIOPAC Inc., USA) which includes a MP100 workstation with a high-level HLT100 transducer (converts one type of energy to a different type of energy) and dedicated analysis software (AcqKnowledge 3.5) [BIOPAC].

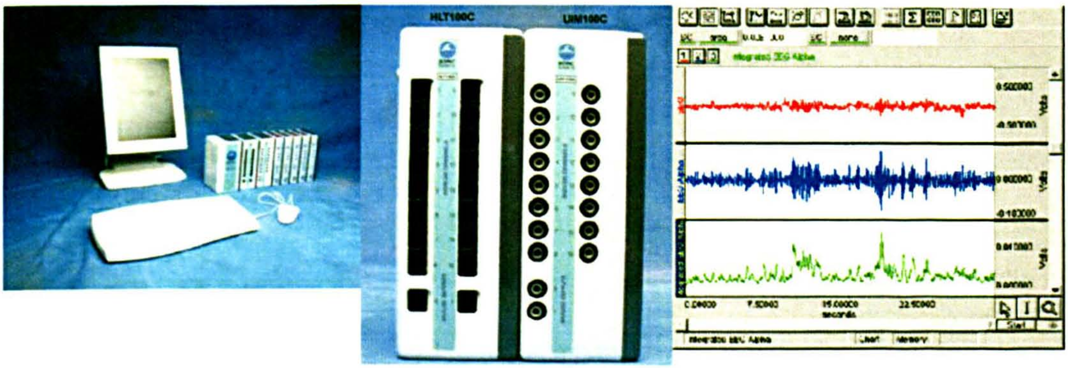


Figure 8.2: The BIOPAC MP100 workstation, the HLT100 transducer and the AcqKnowledge 3.5 software [BIOPAC] (original in colour)

To have an overall picture of the EMG activity on the superficial quadriceps, 6 pairs of recording electrodes were used: two pairs per muscle.

One electrode pair was placed in a low position with respect to the centre of the muscle and the other in a higher position.

Figure 8.3: Positioning of the electrodes' pairs. Still image from video made by the author at Teesside University in collaboration with Dr Dixon and colleagues. (original in colour)

The placement was standardised so that it could be reproduced in different subjects with different bodies. This standardisation allows for the data to be compared.

Electrode-muscle assignment	
Muscle	**Number of electrodes pairs**
Rectus femoris	2 pairs
Vastus lateralis	2 pairs
Vastus medialis	1 pair
Vastus medialis oblique	1 pair

Table 8.1: Electrode-muscle assignment

The subjects belonged to 3 particular groups: young (< 45 years old) asymptomatic (i.e. not known to be exhibiting symptoms in this case of osteoarthritis) participants (23 subjects), old (> 45 years old) asymptomatic participants (17 subjects) and old (> 45 years old) patients (17 subjects) with symptoms of osteoarthritis (OA) of the knee.

The subjects were asked to perform a Maximal Voluntary Isometric Contraction (MVIC). To perform this action, the patient is seated on a custom built chair padded and comfortable, without his/her feet touching the floor and with the back rest at 70 degrees (an angle of 90 degrees is uncomfortable). The leg due to perform the action is strapped to a fixed metal bar at the level of the ankle. The waist of the subject is also strapped to the chair to prevent extension at the hip joint. When the subject is asked to perform a MVIC, he/she will try to push his/her strapped foot forward as much as possible. The muscles of the thigh will therefore contract and produce firings detectable by the electrodes.

Each subject was asked, by Dr Dixon, to perform 5 contractions. Verbal encouragement was given using the phrase "push, push, push" in order to produce maximum contraction values. Between contractions the subjects were asked to rest for about 30 seconds [Dixon, 2004; p. 102]. The EMG data were gathered using the BIOPAC Acquisition System at a sampling rate of 2048 samples per second, i.e. one sample roughly every 0.5 ms.

The aim of Dr Dixon's thesis was to investigate whether the onset of EMG activity in vastus medialis oblique (VMO) was delayed relative to that of vastus lateralis (VL) in symptomatic OA knee patients compared to asymptomatic control subjects. In his investigation Dr Dixon could not find that such a delay existed. In the experiment described here we did not attempt to use sonification to verify the same hypothesis of Dr Dixon's thesis. Instead we decided to verify if trends that we knew existed in the data could be displayed using a sonification method. The reason for this choice is that if we can hear in the sonification a trend that we know exists then we can be sure that our display has a certain degree of resolution.

For the purpose of the experiment described here, 30 such contractions were selected, each one containing 6 channels of data (2 channels for each muscle). 12 data sets were selected from the young asymptomatic group, 9 from the old asymptomatic group and 9 from the old OA symptomatic group. Only the data of the third contraction was chosen among the five performed by each of Dixon's subjects. This data reduction was necessary to make the sonification experiment duration, for each test subject, reasonable (i.e. about 30 minutes). It was considered that the middle contraction should be the most balanced of the five in terms of force used and the occurrence of fatigue. The contractions data were chosen randomly from the existing groups of Dixon's subjects.

Subjects' age distribution						
Age ranges	**18-29**	**30-39**	**40-45**	**46-59**	**60-69**	**70-82**
Asymptomatic patients	19, 23, 24, 27	30, 33, 34,34	40, 40, 42, 43	52, 53, 54, 55	62, 64, 64, 66	74
OA patients				50, 54	60, 61	70, 71, 71, 75, 82

Table 8.2: Subjects' age distribution

8.4.2 Known characteristics of the data

There are some characteristics of these data sets which are known to change with the age of the subjects as reported by Dr John Dixon.

In particular:

1 - the overall amplitude of the signal tends to reduce with age;

2 - the slope or rise of the signal tends to reduce with age.

These two characteristics represent the muscles having less power with increased age and taking a longer time to reach the maximum power.

8.4.3 Aim of the experiment

The aim of the experiment described here is to verify that these two characteristics are clearly displayed by the chosen sonification algorithm.

The sonification algorithm will display the data maintaining the real-time rate of data gathering so that the algorithm could be used for a real-time auditory display.

8.4.4 The sonification

For each subject, there exists a data file with 6 channels: 2 channels (representing 2 electrode pairs) for the vastus medials and vastus medialis oblique muscles, 2 channels for the vastus lateralis and 2 channels for the rectus femoris [see Table 8.1]. The original sampling rate of the data is 2048 samples per second, i.e. one sample roughly every 0.5 ms.

The sonification was designed following the criteria listed here:

1. use the raw data, i.e. all the potentially useful information is mantained;
2. be a real-time sonification, i.e. the data must be fed into the sound synthesis algorithm at the same rate as they are gathered;
3. be easy to interpret.

Surface EMG signals do not contain frequencies above 500Hz [Dixon, 2004]. This means that if periodic patterns exist in the data they will vary at a rate that is either within a middle-low audio frequency range or at a rate that is lower than audio range, i.e. control rate.

Consequently, to create a real-time, easy to interpret sonification, we need to map the data to a sound parameter that can vary meaningfully at either audio rate or control rate. A sound parameter that has these characteristics is the amplitude of an oscillator. If the amplitude of an oscillator varies at control rate we will hear a slow increase and decrease of the level of the oscillator, if the amplitude varies at audio rate, energy will appear at specific frequencies in the output spectrum. These frequencies are in this context called *side bands*. The name of this sound synthesis technique is *Amplitude Modulation* (AM).

In this experiment each data channel was mapped to the amplitude of a sine wave oscillator (called carrier oscillator). Each sine oscillator had different frequencies. The frequencies of the different oscillators were set to be in harmonic relationship with each other with the intention of producing a more pleasing sound.

In AM, when the modulating signal varies at frequencies below the audio range, the amplitude of the carrier signal slowly increases or decreases following the modulating signal. When the modulating signal varies at audio range rate, then the spectrum of the output contains the spectrum of the modulating signal shifted over the carrier frequency and its mirror image below the carrier frequency. The reproduction of the input spectrum and its mirror image around the carrier frequency are the *side bands* [Dodge and Jerse, 1996].

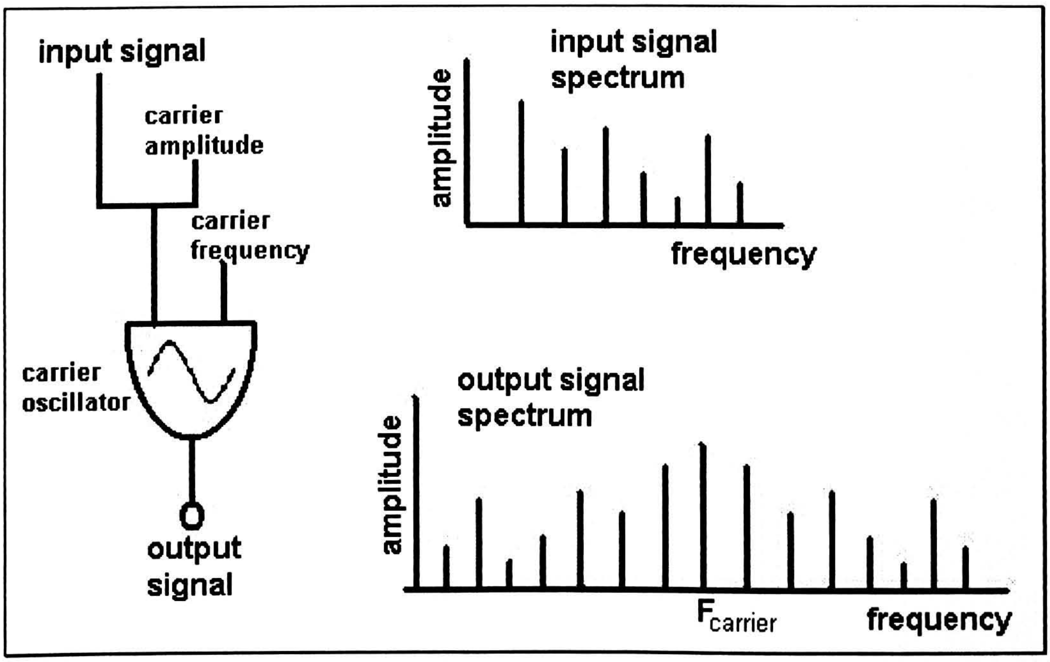


Figure 8.4: Amplitude modulation. Relationship between the modulating input signal and the AM output signal.

Table 8.3 shows the assignment of carrier frequencies for each electrode pair.

Electrode to carrier frequency assignment

Electrode pair	Oscillator frequency
RFUP	freq = 261.6Hz (mid C)
RFLOW	freq = 523.2Hz
VLUP	freq = 784.8Hz
VLLOW	freq = 1046.5Hz
VMUP	freq = 1308.1Hz
VMLOW	freq = 1569.7Hz

Table 8.3: Electrode to carrier frequency assignment

where:

RFUP = Rectus Femoris electrode pair positioned *high* with respect to the centre

RFLOW = Rectus Femoris electrode pair positioned *low* with respect to the centre

VLUP = Vastus Lateralis electrode pair positioned *high* with respect to the centre

VLLOW = Vastus Lateralis electrode pair positioned *low* with respect to the centre

VMUP = Vastus Medialis electrode pair positioned *high* with respect to the centre

VMLOW = Vastus Medialis electrode pair positioned *low* with respect to the centre

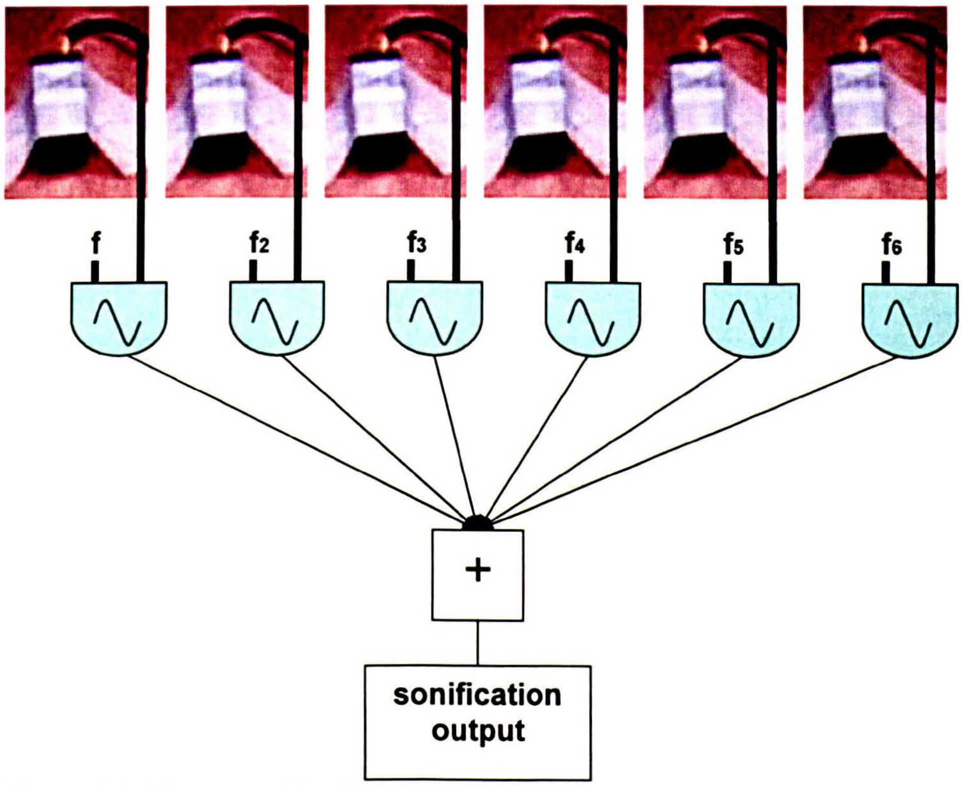


Figure 8.5: Diagram of the EMG sonification (original in colour)

To produce this sonification, all the channels were scaled using the same *transposition* and *scaling factors* (see Chapter 6, Section 6.4.1) and therefore the relative relationship between channels' values was maintained. The carriers' frequencies ranged between 200Hz and 1600Hz for two main reasons:

1. the loudness/frequency dependency (see Chapter 4, Section 4.2.4) is not very strong in this frequency range;
2. auditory display designers tend to centre their display in the 200-5000Hz frequency range because it corresponds to the typical range of various musical instruments and because it is the most effectively perceived frequency region (high sensitivity to frequency change) [Walker and Kramer, 2004].

8.4.5 The resulting sound

Typically a contraction data set sonified in this way contains an overall tone (produced by small offsets present in the 6 channels of data) and an enveloped noisy part. In some cases the distinct tones of some oscillators are perceived as separate from the overall tone and in some cases the overall envelope of the sound has a complex evolution in time (i.e. with various bumps and valleys), not simply an attack, a sustained and a release part. The resulting sound is normally perceived as one; i.e. one timbre (instead of a mix of different timbres playing simultaneously). Each data channel contributes to the output sound by introducing sound energy (the AM side bands) in a frequency band around the frequency of the carrier of that particular channel.

The presence or absence of each one of the six channels is perceived as a presence or absence of energy around the channel carrier frequency in the output sound. In Figures 8.6 and 8.7 we can see the visual display of one channel's contribution to the overall sound in isolation. In Figure 8.8 and 8.9 we can see the visual display of all the channels contributions to the overall sonification in the particular case of a subject of 23 years of age.

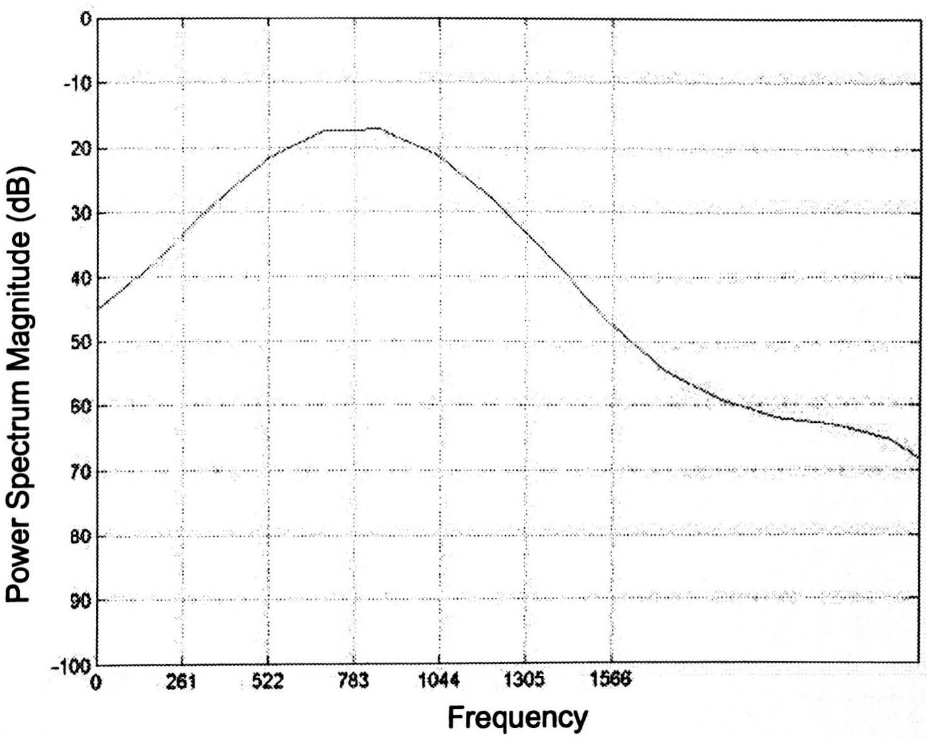


Figure 8.6: Power spectrum of data channel 3 (carrier frequency 783Hz) for subject 8 (age: 23)

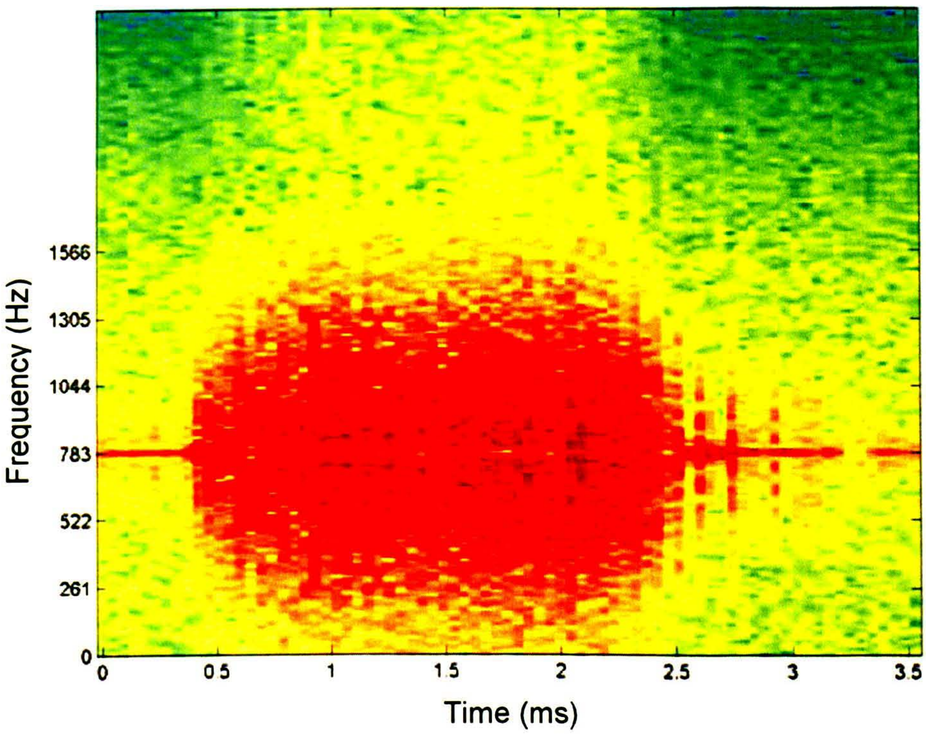


Figure 8.7: Spectrogram of data channel 3 (carrier frequency 783Hz) for subject 8 (age: 23) (original in colour)

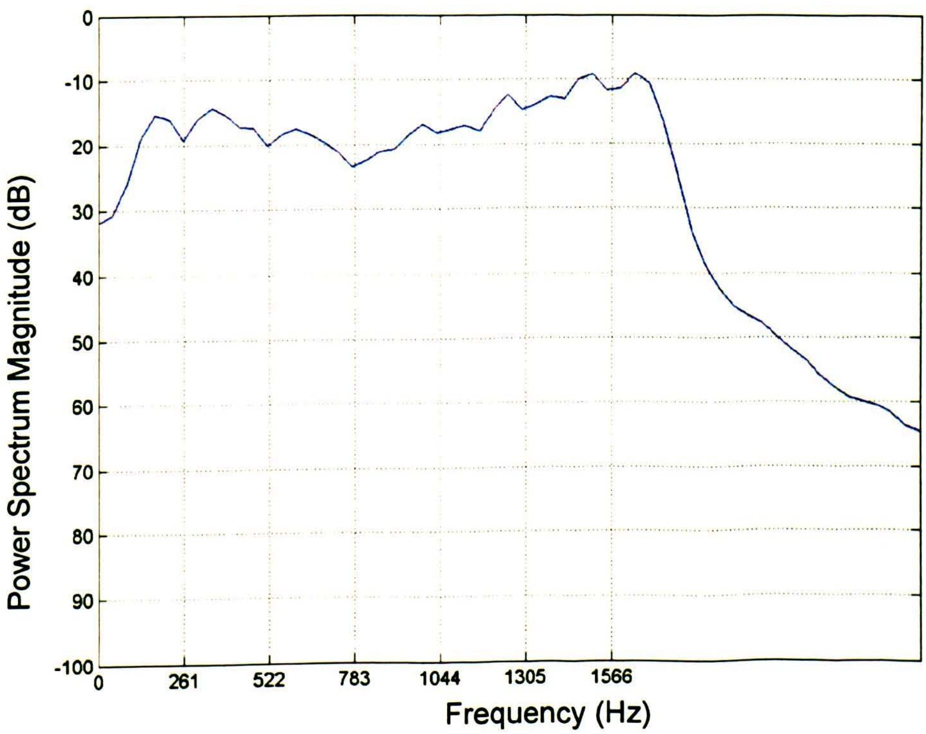


Figure 8.8: Power spectrum of all data channels for subject 8 (age: 23)

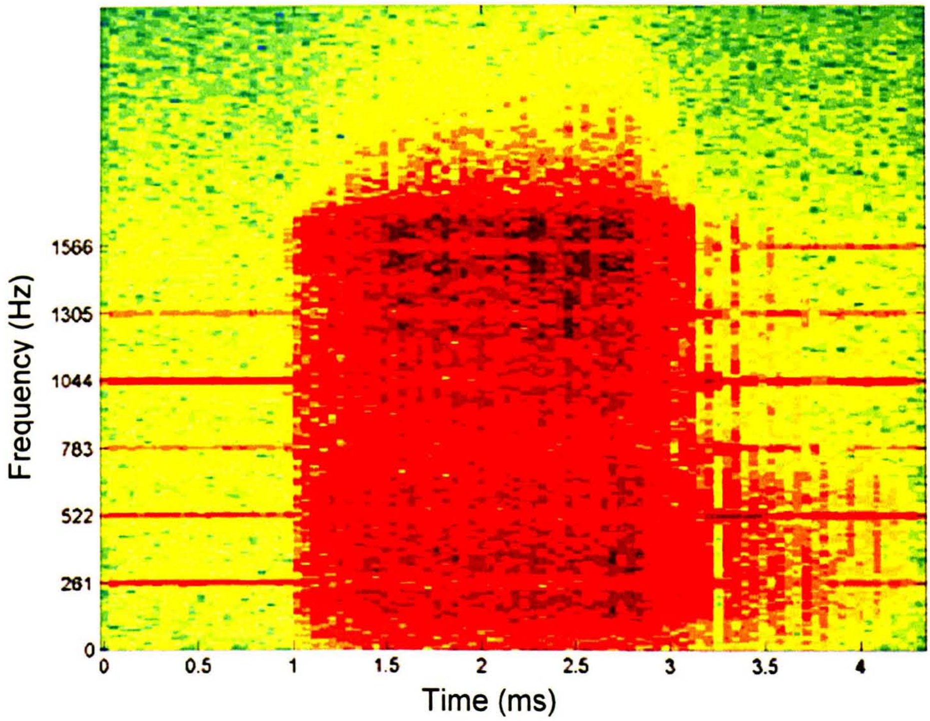


Figure 8.9: Spectrogram of all data channels for subject 8 (age: 23) (original in colour)

The overall sound can also be said to sound "rough". This fact is not surprising as the sensation of roughness can be caused by a signal with amplitude modulation [Zwicker & Fastl, 1999]. Roughness is created by a 100% modulation of the amplitude of a signal when the modulation frequency is in a range between 15 and 300Hz. The modulating signal does not need to be periodic, however its spectrum needs to be in the 15-300Hz region to produce roughness. For this reason narrow band noises can sound rough even though there is no periodicity. [Ibid, p. 257] This experiment will also verify if the perception of roughness in the sonifications can be related to any characteristic of the data.

8.4.6 The experimental procedure

After having transformed each contraction's data set into a sound (using the Interactive Sonification Toolkit), a program was developed in Pure Data in order to run the experiment and gather most of the experimental results automatically.

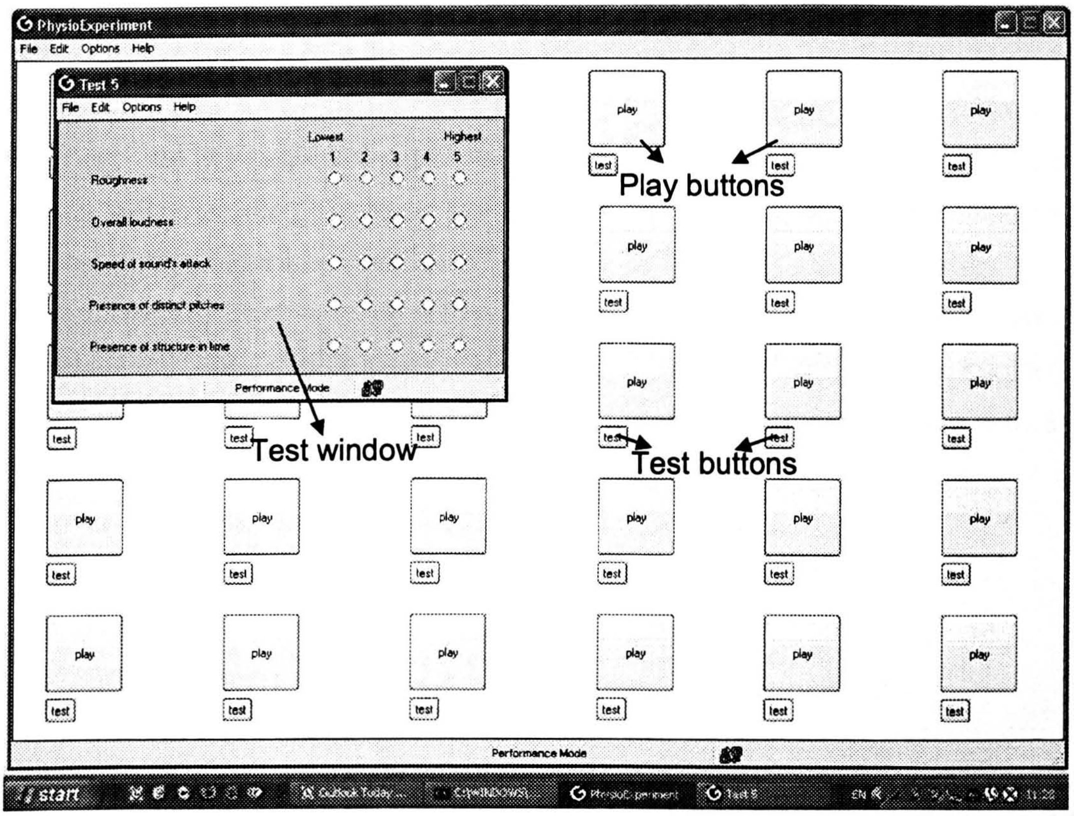


Figure 8.10: Experiment 2 software GUI

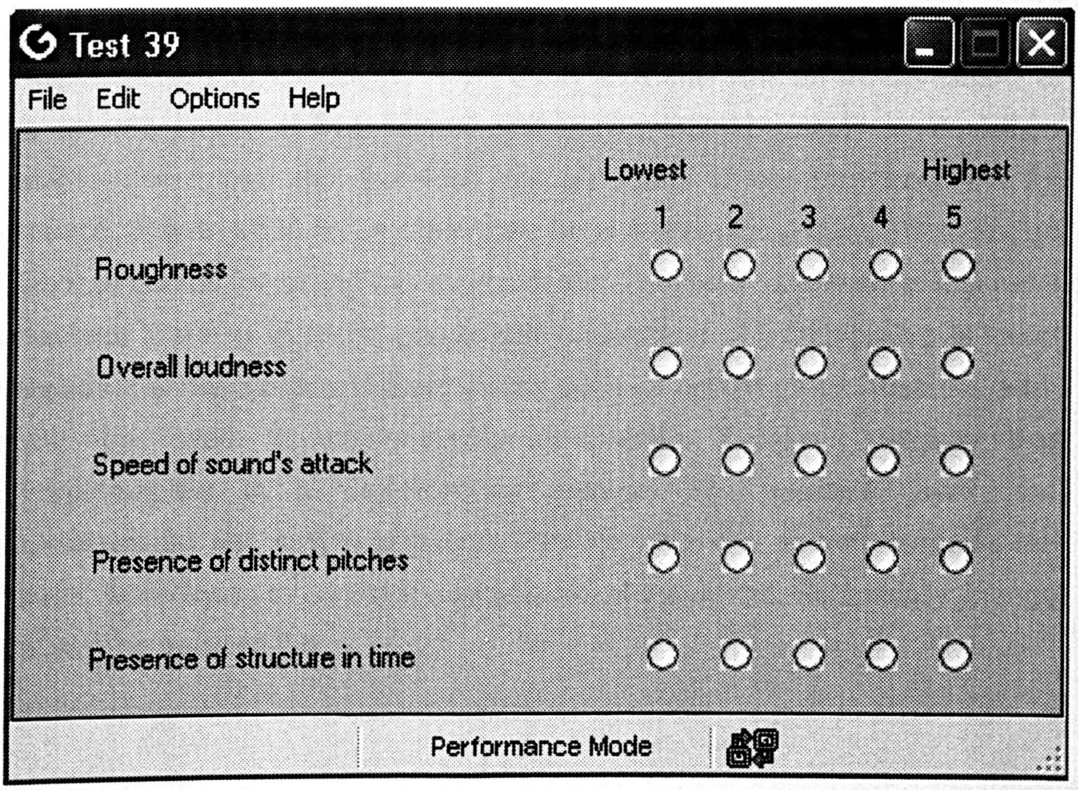


Figure 8.11: Experiment 2: Test window

When this PD program was opened, the subjects were presented with a screen containing 30 buttons. Each button, if clicked, played the sonification of one data set. In order to be able to score each sound in relation to the others, it was important that the subjects had an idea of the overall range of variation of the sounds before starting to score. The subjects were asked, at the beginning of the test, to listen to all the sonifications at least once before starting the test.

For each sonification, another button was present in the screen labelled 'test'. After having listened to a sonification as many times as desired, the subject clicked on the corresponding 'test' button and was asked to score from 1 to 5 (low to high) various characteristics of the sound.

The characteristics were:

1) Roughness;
2) Overall loudness;
3) Speed of sound's attack;
4) Presence of distinct pitches;
5) Presence of structure in time.

8.4.7 Reasons for choosing to test the above five characteristics

Roughness

Roughness can be considered a descriptor that indicates an overall quality of the sound's timbre. It was noticed that the sonifications could be described using this descriptor and, therefore, it was decided to test if it varied with, for instance, age or group (i.e. symptomatic or asymptomatic group). There is also a second reason for testing this characteristic. The word roughness is a general descriptor that non technical people can understand. If a relationship is found between this descriptor and a parameter such as age or group, then this term could be used to communicate information between patients and physiotherapists in a non-technical manner. For example, when the physiotherapist will say to a patient: "try to make the sound rougher"; the patient, while trying to control the roughness of the sound, will actually change the way the muscles fire.

Overall loudness and speed of sound's attack

The overall loudness and speed of the sound's attack are variables that, as mentioned in Section 8.4.2, should vary with age and they were the main variables to be tested in order to evaluate the effectiveness of the sonification.

Presence of distinct pitches and presence of structure in time

The two last characteristics, perception of distinct pitches and presence of structure in time, were noticed to be present in some of the sonifications. No relationship was hypothesised with regard to either age or group, and the range of variability of these characteristics was restricted. These characteristics were tested anyway to verify if any significant trend could be derived from the results or if we could consider them as part of the overall variability of the sonification of these EMG data.

The scores of each subject were written automatically into a text file saved in the computer. Each subject was presented with the sonifications in a new random order so that biases due to order of presentation were cancelled.

The test was carried out in a silent room (in the recording studio performance area at York University). Good quality headphones (DT 990 Beyerdynamic) were used with a wide frequency response (5 – 35kHz). This minimised the errors that could be due to external sounds and maximised the quality of sound reproduction. The level of the sounds was maintained the same for all subjects.

8.4.8 Questionnaire

After the test, additional information about the sonifications was gathered via a questionnaire (see Appendix E for the questionnaire). In the questionnaire the subjects were asked to score from 1 to 5 (low to high):

1) the *pleasantness* of the sonifications
2) how *interesting* they found the sounds

They were also asked to answer the following questions:

At the end of this test you found that:

You could have listened to more of these sounds

You could have listened to more sounds, but not to this type of sounds

You were tired of listening to any sound

The above questions aimed to find out if the sounds were tiring and inducing fatigue, something which might be detrimental if used in a clinical environment.

- *Did these sounds remind you of any natural sound? If yes which one?*
- *These sounds are synthesised from data produced by the activity of the leg's muscles. Do you find this sound appropriate to represent muscle's activity, i.e. movement? (please comment if you wish)*

The above questions intended to find out if the sound worked well as a sound metaphor (or 'audio image') of muscle movement.

8.4.9 The subjects

This experiment has a within-subjects or related design as the same group of listening subjects scored all of the sonifications. This is also a correlational study (see Chapter 5, Section 5.4.1 on experimental designs) because we primarily want to know if the perception of certain sonification parameters is correlated with the age of the subjects producing the EMG data.

The main criterion used to select the listening subjects for the test was that they should have some understanding of audio signals and their properties. The reason for having this condition was that the final user of this type of sonification would be someone (an analyst) used to interpreting sound. It was assumed that there would not be differences in judgment due to gender, and that, given a certain educational background, differences in cultural background could also be disregarded. 21 subjects performed the test. 8 of these subjects had done the experiment described in Chapter 7 a few months before doing this experiment. Their average age was 29. There were 3 females and 18 males. All the participants were British apart from a Malaysian and a French person. 19 participants were researchers, students, lecturers in engineering (specialisation in sound), 2 people were researchers in physiotherapy with sufficient understanding of sound and 4 people work with sound sporadically. The subjects declared having normal hearing capabilities.

8.5 The main results

For each one of the 30 sonifications and for each of the 5 characteristics, an average was calculated over the number of listening subjects. Before verifying if there is a relationship between the average scores for a particular characteristic and the age of the people who provided the EMG data, a non-parametric Friedman's ANOVA test was performed for the characteristic's average scores. This was done to make sure that the differences between the average scores were statistically significant overall. If this test showed no significance, then it would mean that all the average scores could have the same value and therefore it would not make sense to calculate a correlation between the average scores and the age of the people who provided the EMG data. If the Friedman's ANOVA test was significant, however, then the average scores were ordered by the age of the person whose muscles were portrayed by the sonification and Spearman's rank correlation was calculated between the characteristic's average scores and the age of the people providing the EMG data.

8.5.1 Correlation with age

Roughness

Roughness scores: significance of means	*Friedman's ANOVA Chi-Square(29)*	*significance*	*effect size*
Roughness scores	299.023	$p < 0.05$ significant	0.76 (large)

Table 8.4: Roughness scores: significance of means

The differences between the average scores are statistically significant.

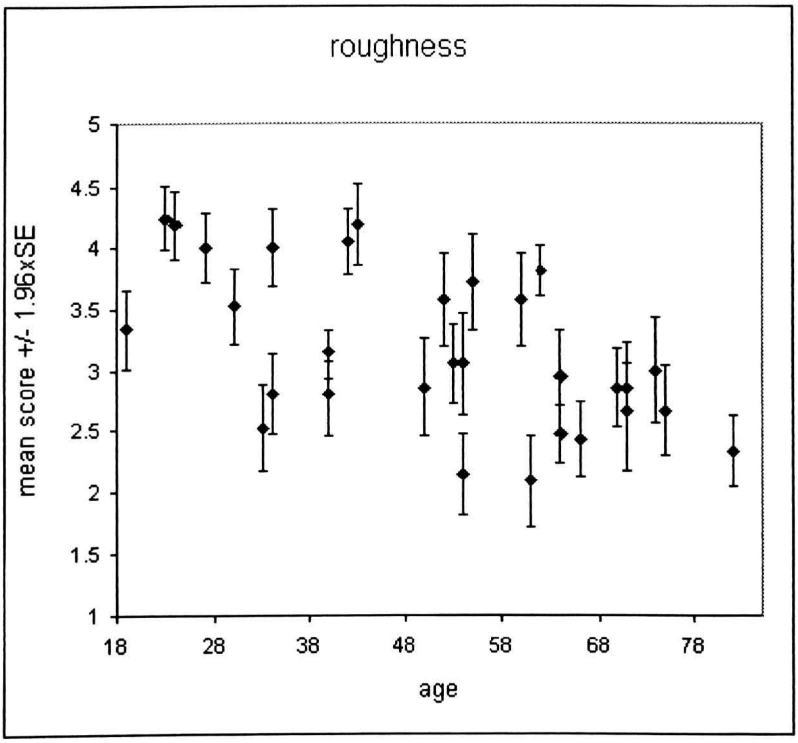


Figure 8.12: Average roughness scores ordered by age of the subjects providing the EMG data

Spearman's rank correlation

$r_s = -0.54$, p(one-tailed) < 0.01, effect size: large

A significant negative rank correlation was found between the average roughness scores for the sonifications of the subjects providing the EMG data and their age.

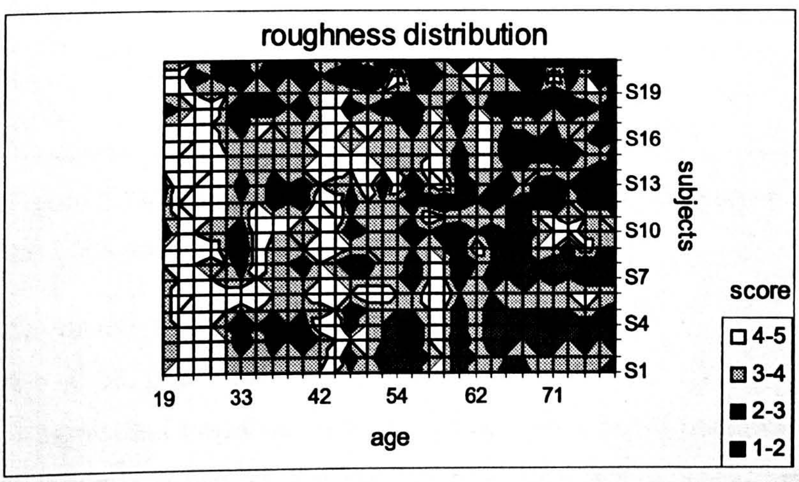


Figure 8.13: Distribution of the entire scoring for the roughness

This means that the younger the person providing the EMG data the rougher the sonification of the EMG data.

Loudness

Loudness scores: significance of means	*Friedman's ANOVA Chi-Square(29)*	*significance*	*effect size*
Roughness scores	426.941	$p < 0.05$ significant	1.20 (large)

Table 8.5: Loudness scores: significance of means

The differences between the average scores are statistically significant.

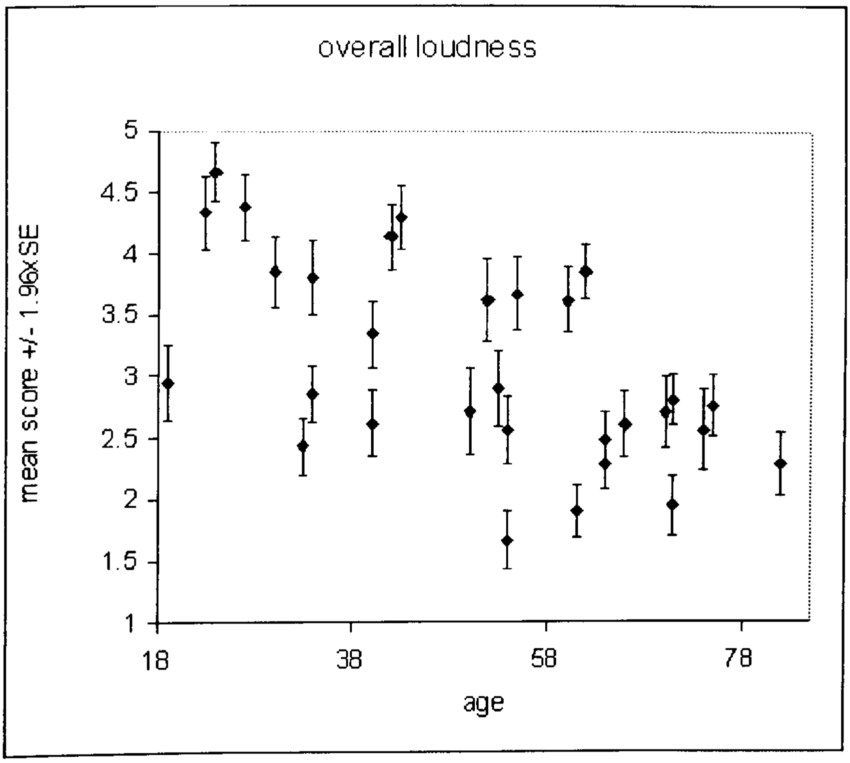


Figure 8.14: Average overall loudness scores ordered by age of the subjects providing the EMG data

Spearman's rank correlation

$r_s = -0.58$, p(one-tailed) < 0.01, effect size: large

A significant negative rank correlation was found between the average overall loudness scores for the sonifications of the subjects providing the EMG data and their age.

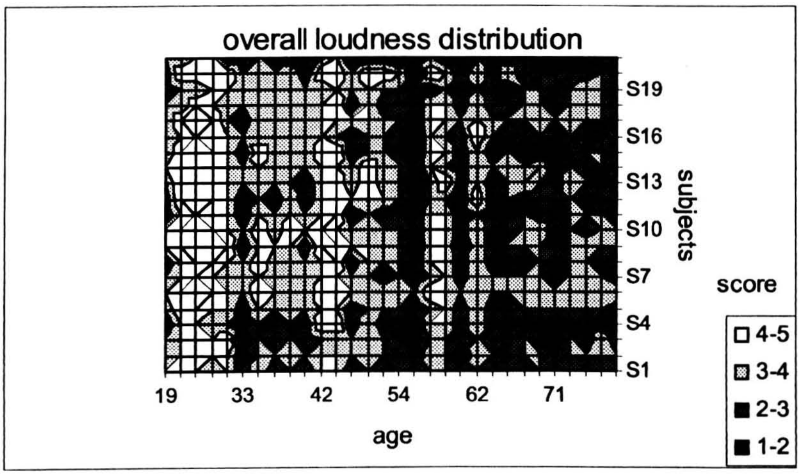


Figure 8.15: Distribution of the entire subjects' scoring for the overall loudness

This means that the younger the person providing the EMG data the louder the sonification of the EMG data.

Attack speed

Attack speed scores: significance of means	*Friedman's ANOVA Chi-Square(29)*	*significance*	*effect size*
Attack speed scores	326.152	$p < 0.05$ significant	0.94 (large)

Table 8.6: Attack speed scores: significance of means

The differences between the average scores are statistically significant.

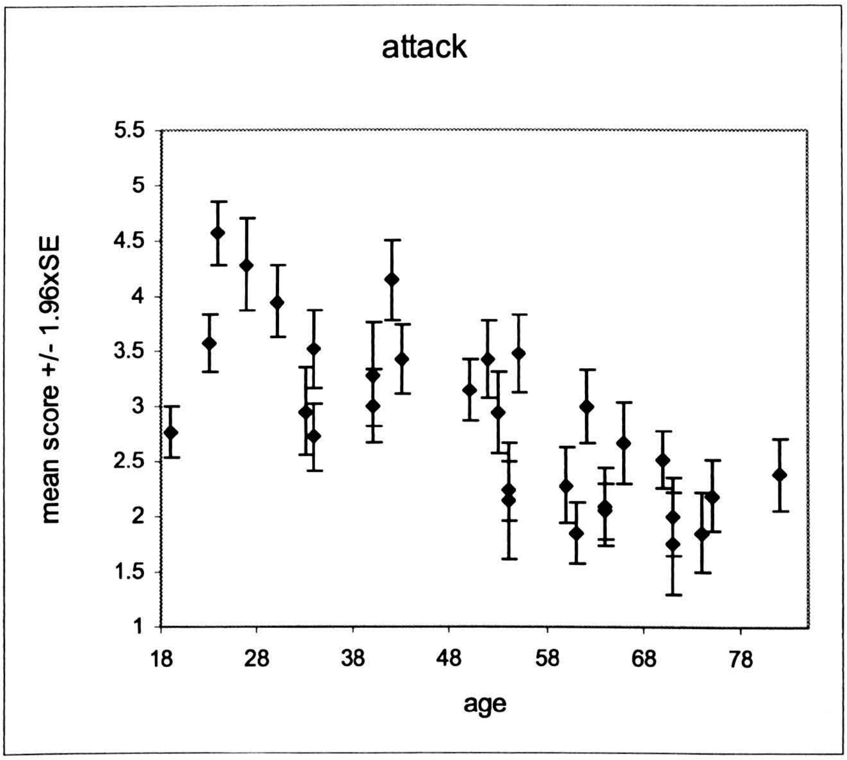


Figure 8.16: Average attack speed scores ordered by age of the subjects providing the EMG data

Spearman's rank correlation

$r_s = -0.75$, p(one-tailed) < 0.01, effect size: large

A significant negative rank correlation was found between the average attack speed scores for the sonifications of the subjects providing the EMG data and their age.

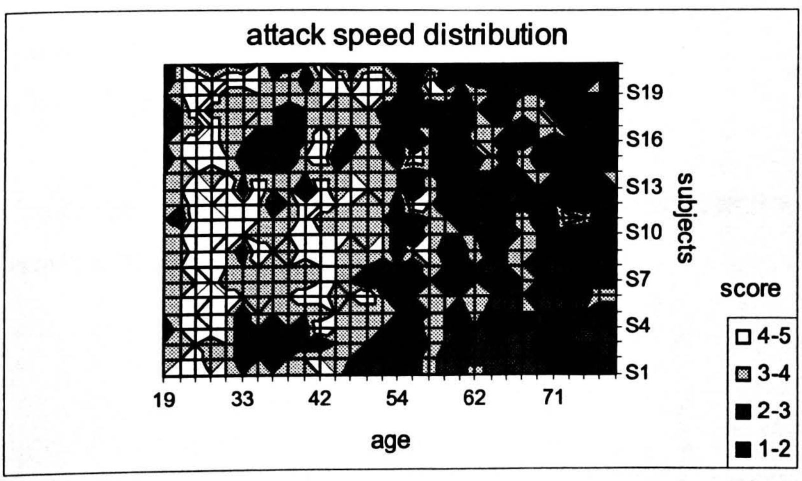


Figure 8.17: Distribution of the entire scoring for the attack speed

This means that the younger the person providing the EMG data the faster the attack of the sonification. The Spearman's correlation of this attribute with the age of the people providing the EMG data is much higher than for roughness and overall loudness. This seems to be the characteristic which best portrays the age of the person providing the EMG data.

8.5.2 Correlation between the results for roughness and both loudness and attack speed

The roughness average scores are highly correlated with the loudness average scores:

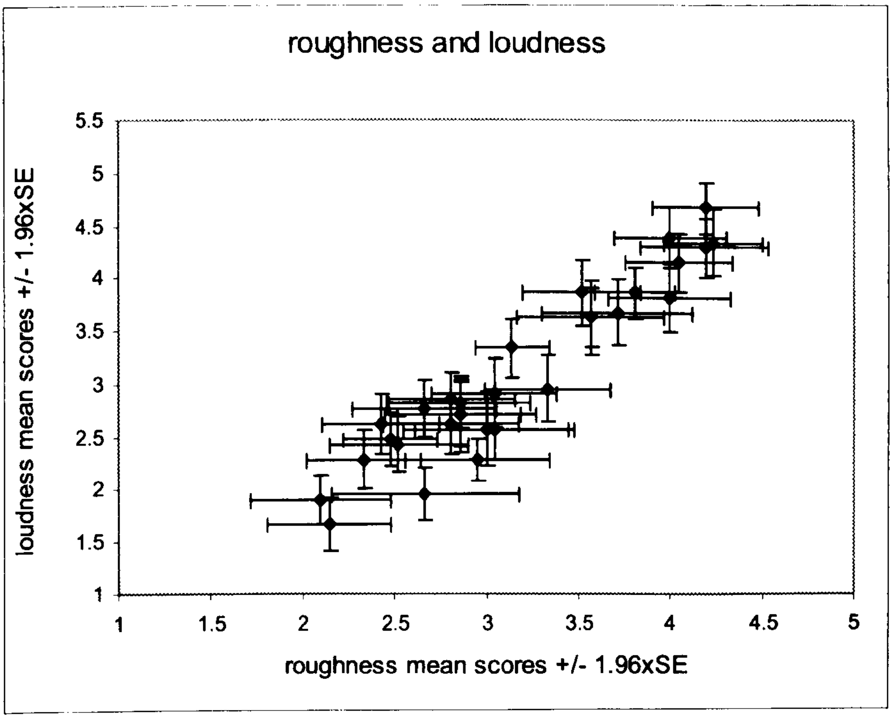


Figure 8.18: Scatter plot of roughness mean scores and overall loudness mean scores

Spearman's rank correlation

$r_s = 0.91$, p(one-tailed) < 0.01, effect size: large

The roughness average scores are also highly correlated with the attack speed average scores:

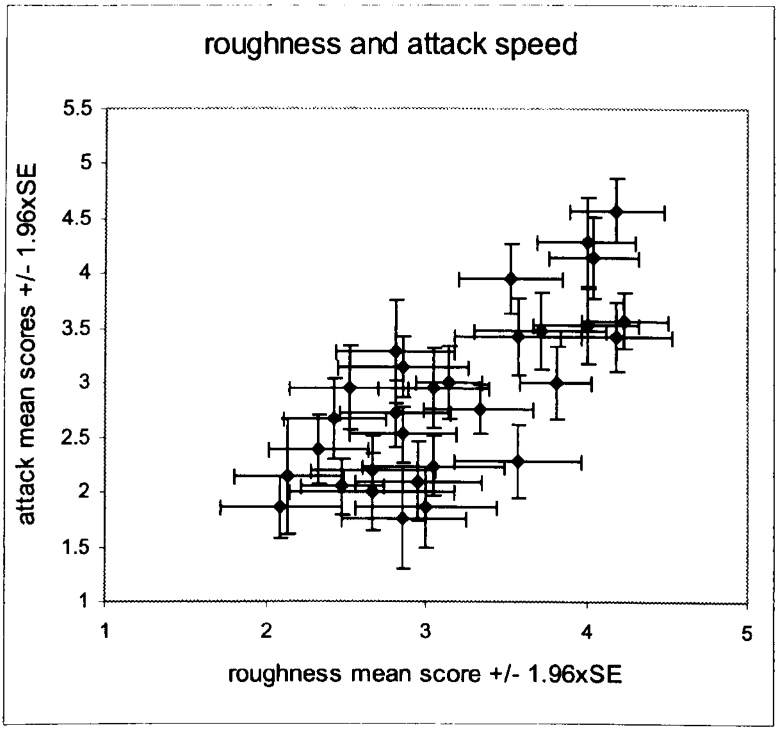


Figure 8.19: Scatter plot of roughness mean scores and attack's speed mean

Spearman's rank correlation

$r_s = 0.73$, p(one-tailed) < 0.01, effect size: large

People perceive the sound resulting from this sonification to be rough when the loudness is high and the attack speed is high. Since these two characteristics, overall loudness and attack's speed, are known to be related to age (fact confirmed by this experiment) then, on the basis of this correlation, we can expect a rough sound to belong to a young person and a less rough sound to belong to an older person (which is confirmed by the correlation between the roughness and the age).

It seems, therefore, that the descriptor 'roughness' can represent both the loudness and the attack speed of the sound simultaneously and is an example of a descriptor which could be used in the communication between a patient and a physiotherapist (both not necessarily at ease with sound parameters language). For instance, if a physiotherapist needs to tell a patient to try to make a younger and healthier muscle contraction, he/she could say "Can you please try to make the sound rougher?"

8.6 Results by groups

8.6.1 Comparison between the young and old groups

Roughness

Significance of difference in roughness mean scores between young and old groups			
Significance of means	*Wilcoxon signed rank test Z*	*significance*	*effect size*
Roughness scores	-1.923	p = 0.05 marginally significant	0.7 (medium)

Table 8.7: Significance of difference in roughness mean scores between young and old groups

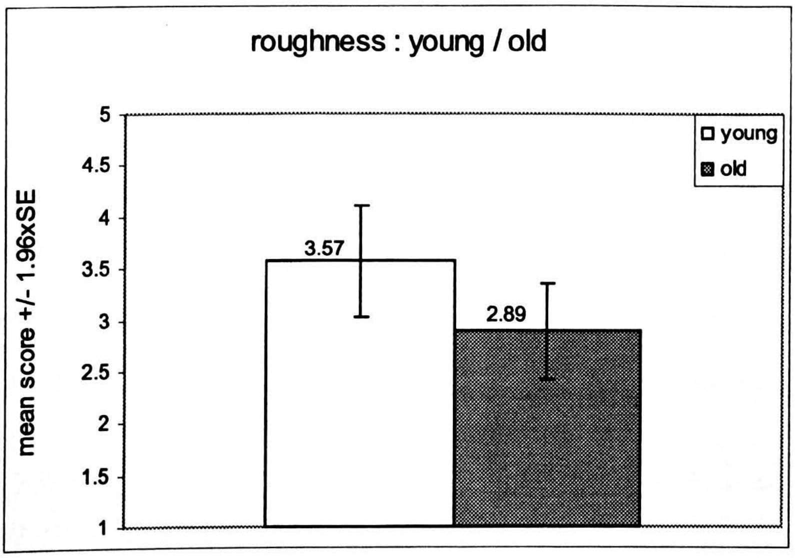


Figure 8.20: Difference in roughness means between the young and the old groups

This result can be considered marginally significant. The result is consistent with the previous result which found an inverse correlation between roughness scores and age, however, as we can see from the a-priori tests in Appendix B, a larger number of observations would confirm more clearly the significance of this result.

Loudness

Significance of difference in loudness mean scores between young and old groups			
Significance of means	*Wilcoxon signed rank test Z*	*significance*	*effect size*
Loudness scores	-2.197	$p < 0.05$ significant	0.98 (large)

Table 8.8: Significance of difference in loudness mean scores between young and old groups

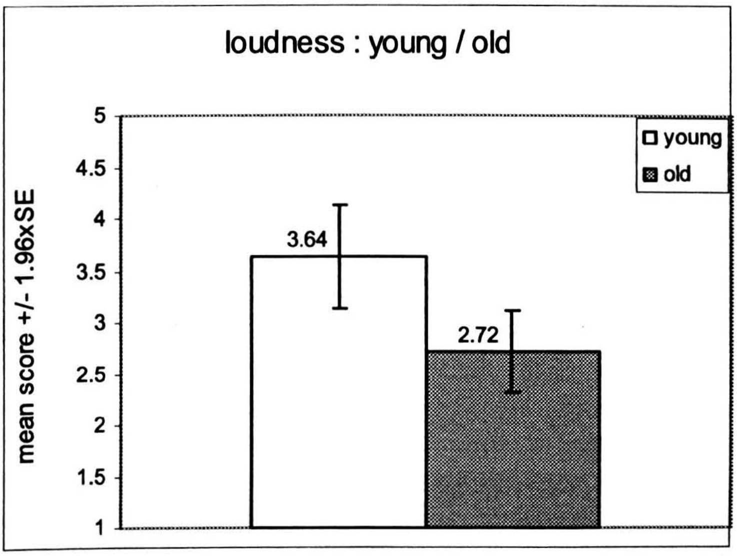


Figure 8.21: Difference in loudness means between the young and the old groups

Attack speed

Significance of difference in attack speed mean scores between young and old groups			
Significance of means	*Wilcoxon signed rank test Z*	*significance*	*effect size*
Attack's speed scores	-2.489	$p < 0.05$ significant	1.10 (large)

Table 8.9: Significance of difference in attack speed mean scores between young and old groups

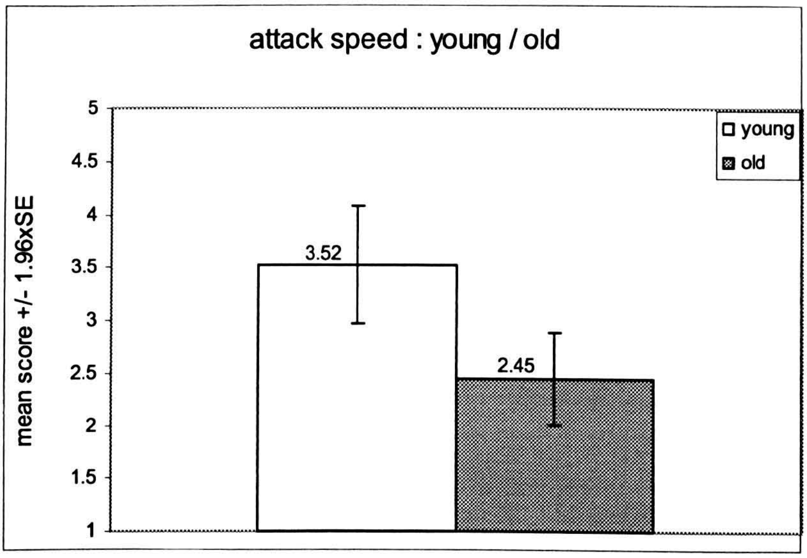


Figure 8.22: Difference in attack speed means between the young and the old groups

There is a significant difference between the mean scores for loudness and attack's speed of the young and old groups. Overall these results are consistent with the results obtained from the correlations with the age of the people providing the EMG data.

8.6.2 Comparison between the old asymptomatic and the old symptomatic group

Roughness

Significance of difference in roughness mean scores between old asymptomatic and old symptomatic groups		
Significance of means	*Wilcoxon signed rank test Z*	*significance*
Roughness scores	-1.718	p > 0.05 not significant

Table 8.10: Significance of difference in roughness mean scores between old asymptomatic and old symptomatic groups

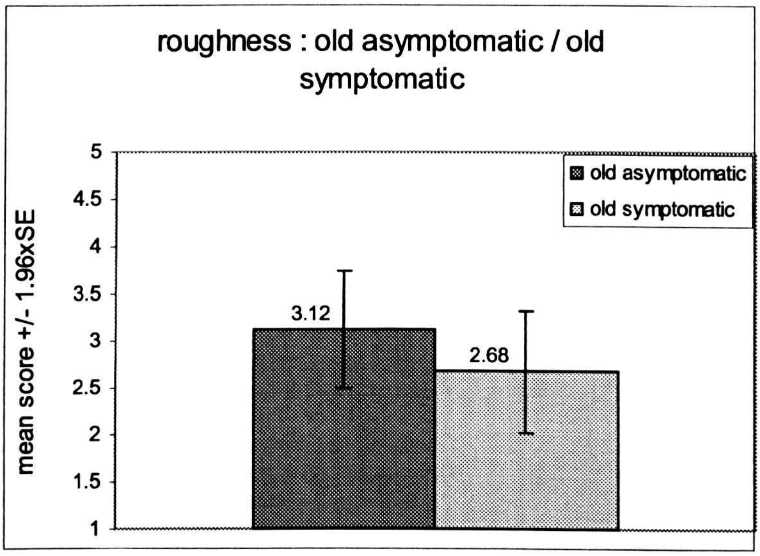


Figure 8.23: Difference in roughness means between the old asymptomatic and the old symptomatic groups

There is no significant difference in roughness between the old symptomatic and asymptomatic groups.

Loudness

Significance of difference in loudness mean scores between old asymptomatic and old symptomatic groups		
Significance of means	*Wilcoxon signed rank test Z*	*significance*
Loudness scores	-1.423	$p > 0.05$ not significant

Table 8.11: Significance of difference in loudness mean scores between old asymptomatic and old symptomatic groups

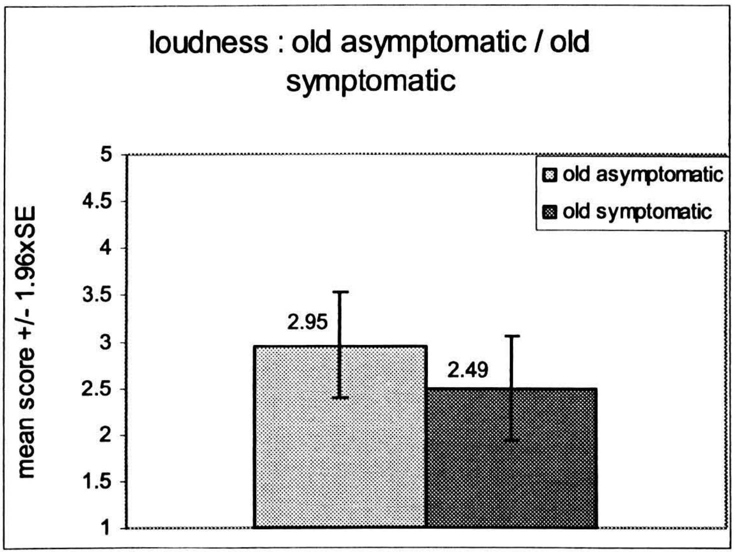


Figure 8.24: Difference in loudness means between the old asymptomatic and the old symptomatic groups

There is no significant difference in loudness between the old symptomatic and asymptomatic groups.

Attack speed

Significance of difference in attack speed mean scores between old asymptomatic and old symptomatic groups		
Significance of means	*Wilcoxon signed rank test Z*	*significance*
Attack's speed scores	-1.7193	$p > 0.05$ not significant

Table 8.12: Significance of difference in attack speed mean scores between old asymptomatic and old symptomatic groups

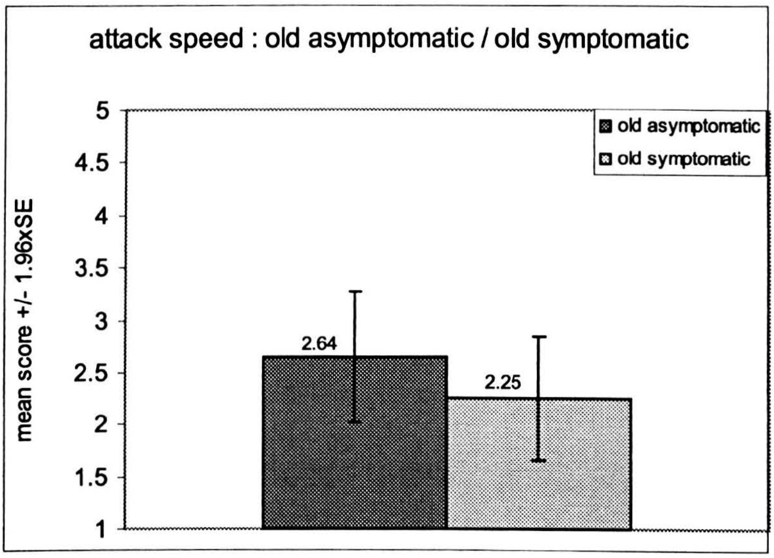


Figure 8.25: Difference in attack speed means between the old asymptomatic and the old symptomatic groups

There is no significant difference in attack speed between the old symptomatic and asymptomatic groups.

8.6.3 Summary of results for roughness, overall loudness and attack speed

There are significant correlations between average roughness, loudness and attack speed scorings and the age of the EMG subjects. When dividing the EMG subjects in two groups (young and old), a significant difference in mean is found for the attack speed and the overall loudness scoring. The differences in roughness scoring are consistent with the correlations results, but the sample size is insufficient to consider these results more than marginally significant in this test. The differences between old symptomatic and asymptomatic groups cannot be considered significant.

8.7 The last two characteristics: distinct pitches and time structure

The subjects were asked to score if they heard distinct pitches emerging in the sound and if they heard a time structure different from the expected envelope. It was noticed that the questions were difficult to interpret and that there was little variety in the data.

8.7.1 Correlation with age

Distinct pitches

Distinct pitches scores: significance of means	*Friedman's ANOVA Chi-Square(29)*	*significance*	*effect size*
Distinct pitches scores	187.220	$p < 0.05$ significant	0.51 (large)

Table 8.13: Distinct pitches scores: significance of means

The differences between the average scores are statistically significant.

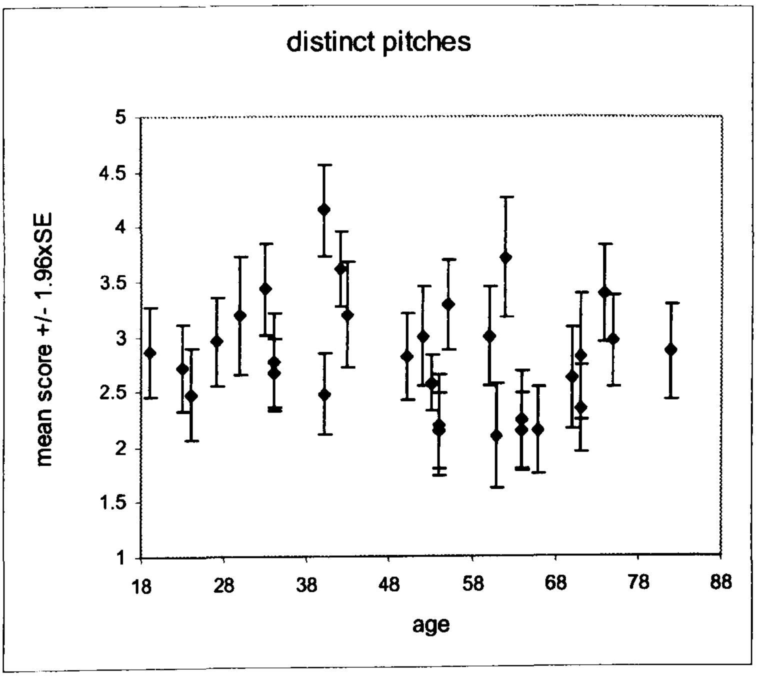


Figure 8.26: Average distinct pitches scores ordered by age of the subjects providing the EMG data

Spearman's rank correlation

$r_s = -0.176$, $p > 0.05$, not significant

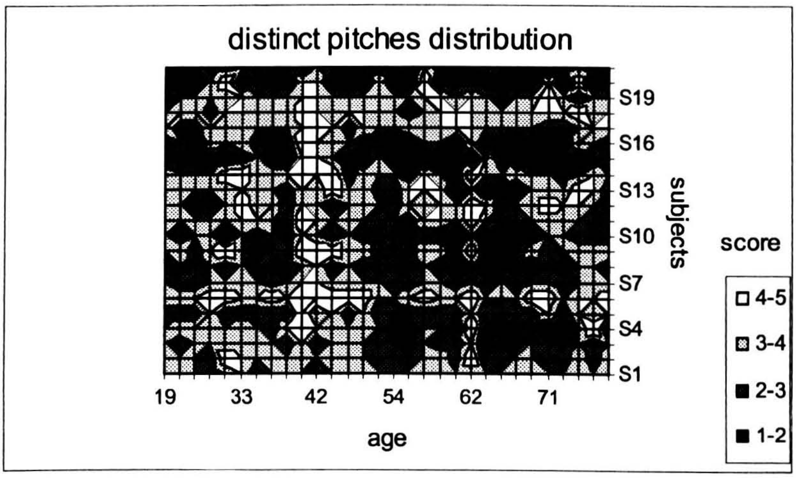


Figure 8.27: Distribution of the entire scoring for the distinct pitches

There is no significant correlation between the presence of distinct pitches and the age of the people providing the EMG data.

Time structure

Time structure scores: significance of means	*Friedman's ANOVA Chi-Square(29)*	*significance*	*effect size*
Time structure scores	166.183	$p < 0.05$ significant	0.47 (large)

Table 8.14: Time structure scores: significance of means

The differences between the average scores are statistically significant.

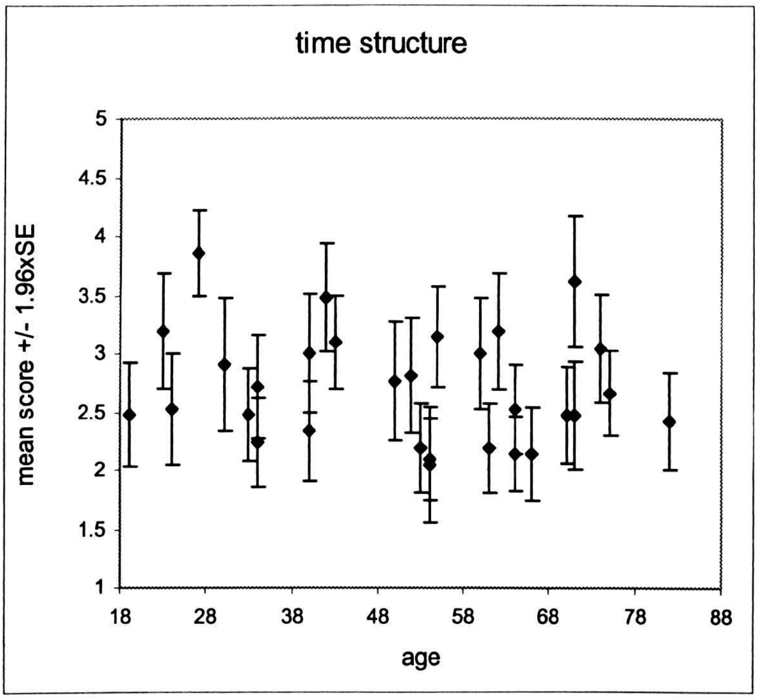


Figure 8.28: Average time structure scores ordered by age of the subjects providing the EMG data

Spearman's rank correlation

$r_s = -0.165$, $p > 0.05$, not significant

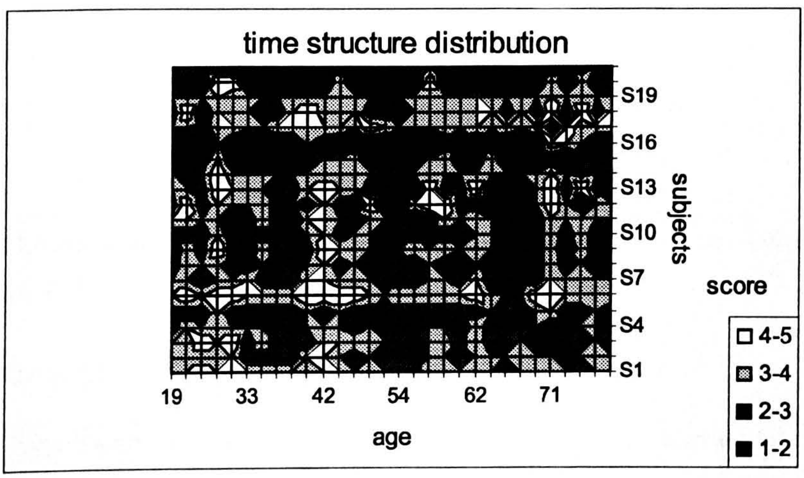


Figure 8.29: Distribution of the entire scoring for the time structure

There is no correlation between the presence of structure in time and the age of the people providing the EMG data.

8.8 Results by groups for the last two characteristics

8.8.1 Comparison between the young and old groups

Distinct pitches

Significance of difference in distinct pitches mean scores between young and old groups		
Significance of means	*Wilcoxon signed rank test Z*	*significance*
Distinct pitches scores	-1.412	$p > 0.05$ not significant

Table 8.15: Significance of difference in distinct pitches mean scores between young and old groups

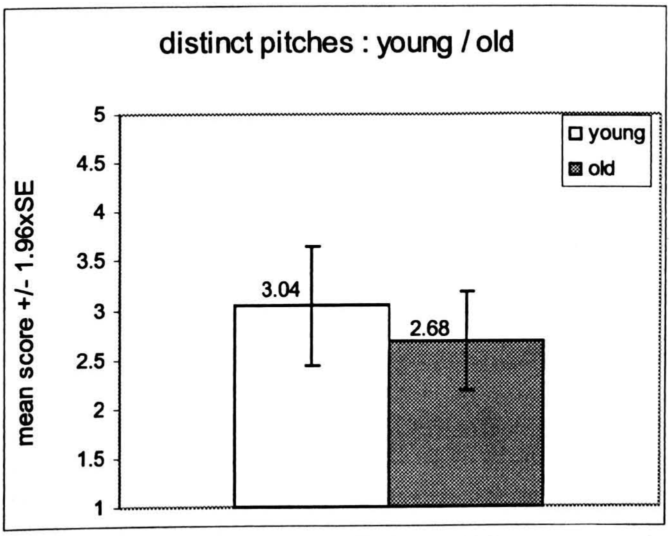


Figure 8.30: Difference in distinct pitches means between the young and the old groups

Time structure

Significance of difference in time structure mean scores between young and old groups		
Significance of means	*Wilcoxon signed rank test Z*	*significance*
Time structure scores	-1.452	$p > 0.05$ not significant

Table 8.16: Significance of difference in time structure mean scores between young and old groups

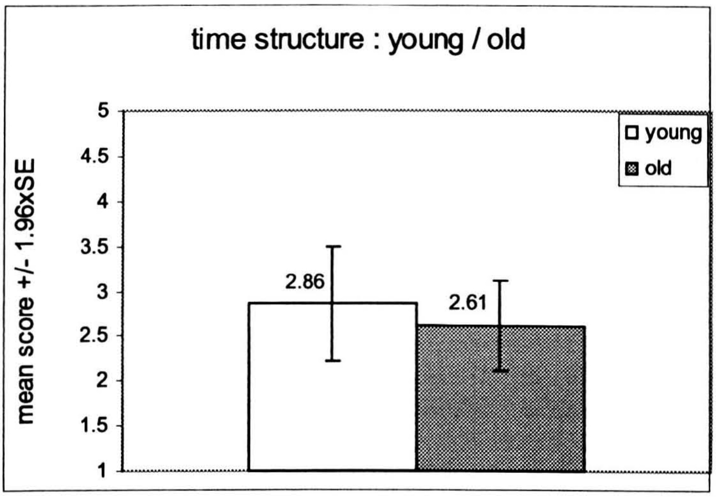


Figure 8.31: Difference in time structure means between the young and the old groups

The differences in scores for distinct pitches and structure in time between the old and young groups are not significant.

8.8.2 Comparison between the old asymptomatic and the old symptomatic group

Distinct pitches

Significance of difference in distinct pitches mean scores between old asymptomatic and old symptomatic groups		
Significance of means	*Wilcoxon signed rank test Z*	*significance*
Distinct pitches scores	-0.356	$p > 0.05$ not significant

Table 8.17: Significance of difference in distinct pitches mean scores between old asymptomatic and old symptomatic groups

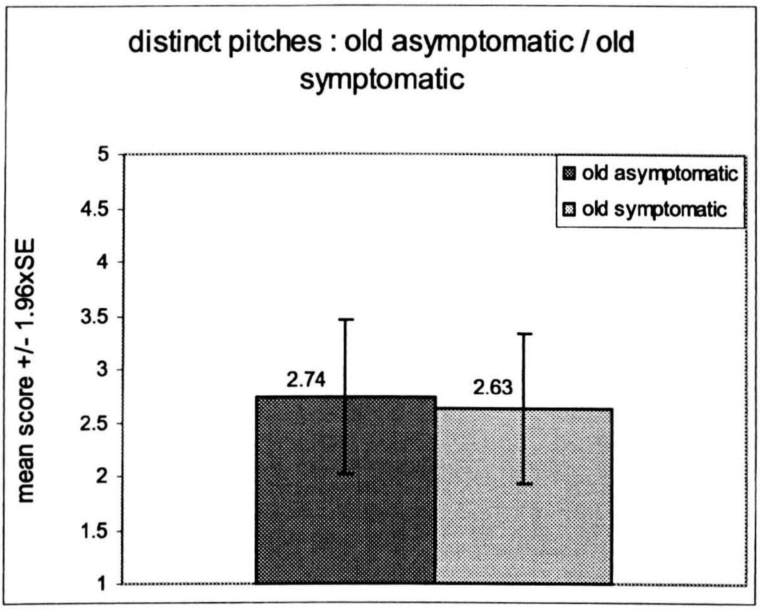


Figure 8.32: Difference in distinct pitches means between the old asymptomatic and the old symptomatic groups

Time structure

Significance of difference in time structure mean scores between old asymptomatic and old symptomatic groups		
Significance of means	*Wilcoxon signed rank test Z*	*significance*
Time structure scores	-0.059	$p > 0.05$ not significant

Table 8.18: Significance of difference in time structure mean scores between old asymptomatic and old symptomatic groups

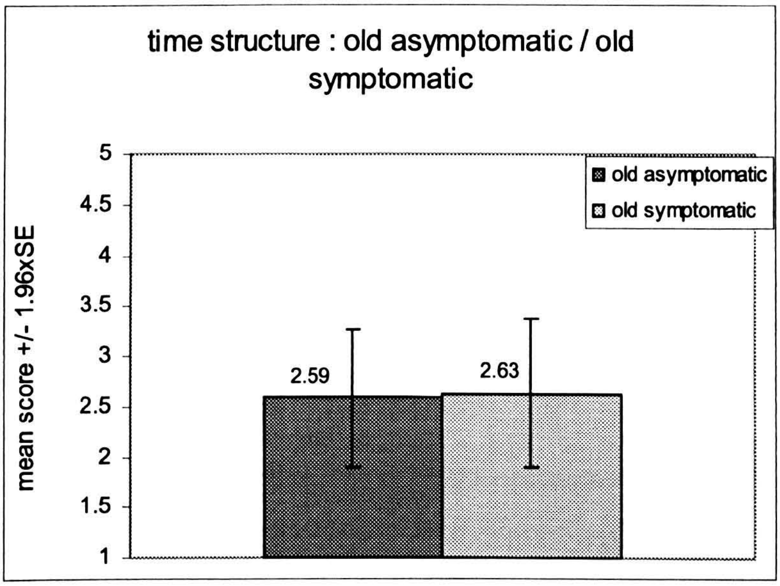


Figure 8.33: Difference in time structure means between the old asymptomatic and the old symptomatic groups

The differences in scores for distinct pitches and structure in time between the old symptomatic and asymptomatic groups are not significant.

8.8.3 Summary of results for distinct pitches and time structure

These two characteristics are not correlated with the age of the subjects providing the EMG data and do not show significant differences with respect to the groups (young/old, old asymptomatic/old symptomatic) present in this experiment.

8.9 Results' conclusions

The overall loudness of the sonifications and the speed of the sound attack are perceived to decrease with age. This confirms our hypothesis and therefore allows us to say that this type of EMG sonification does portray effectively some characteristics of the data as expected.

A significant correlation was found between the average *roughness* score of the sonifications and the age of the subjects providing the EMG data. The perception of sound roughness is highly correlated with the perception of the loudness of the sound and its attack's speed. In this type of sonification the descriptor *roughness* can be used to describe and distinguish between young people movements and old people movements.

The young and old groups are well separated looking at attack's speed and overall loudness. The young and old groups are only marginally separated with respect to roughness. No significant differences were found in roughness, attack's speed and loudness to distinguish between old symptomatic and asymptomatic groups. The presence of structure in time and distinct pitches was not found to be related to age or presence of symptoms of osteoarthritis of the knee.

8.10 Questionnaire's results

The subjects were asked to score how they found the sounds in terms of pleasantness and interest.
A score of 1 corresponded to very unpleasant and uninteresting, while a score of 5 represented very pleasant and interesting.

Is the sound pleasant?

Average score +/- 1.96*SE = **2.9** +/- 0.4

Is the sound interesting?

Average score +/- 1.96*SE = **3.7** +/- 0.4

On average, the subjects found the sound *neither pleasant nor unpleasant* and they found it *slightly interesting*.

Fatigue

In order to have a measure of how fatigued the subjects were after doing this experiment (which lasted on average around 20 minutes), the following questions were asked:

At the end of this test you found that:
You could have listened more of these sounds
You could have listened more, but not to these types of sounds
You were tired of listening to any sound

The subjects were asked to tick the appropriate answer.
20 subjects answered this question:

- 7 out of 20 (35%) said they could have listened to more of these sounds;
- 12 out of 20 (60%) said they could have listened more, but not to this type of sound;

- 1 out of 20 (5%) said they were tired of listening to any sound.

So after 20 minutes of intensive listening of these sounds without a break, the majority of the people were fatigued.

8.10.1 The audio metaphor

The sonification used in this experiment does not attempt to create an audio image that relates to how the data originated. However, it is important, for the sonification to be a good display that the sound's image is at least not in contradiction with the actual event that originated the data, because this could make the display very unclear. For example, in a visual display we would not represent apples with bananas because it would create mistakes of interpretation. To investigate this part of the problem, the subjects were asked to write what event they associated with the sounds they heard. It was interesting to see if the image, which was not predetermined during the choice of the sonification algorithm, could be consistent, or at least not in contradiction, with the representation of muscle activity. It must be noted that the subjects knew that they were being presented with sonifications of EMG data.

Did these sounds remind you of any natural sound? If yes which one?

The subjects wrote various different answers, but surprisingly many answers were similar and could be grouped under the same label. In brackets the number of people that gave the same or similar answer.

Sonification audio metaphors		
LABEL	**ANSWER**	**NUMBER OF ANSWERS**
animal sound	whale sound (2), birdsong	3
underwater sounds	divers breathing apparatus, breathing underwater, bubbles under water (3), underwater (2)	7
sea waves	waves (5), pebbles washed by sea (2)	7
musical instrument	flute, organ pipe, wind instrument	3
wind/air	wind, wind through narrow tube, air in open metal tube	3
natural event	thunder (2), earthquake	3
materials	sandpaper, dragging over wooden floor, gravel, creaking of ropes	4
mechanical sounds	factory noises, tuning sounds, flight landing and take off, crackles	4

Table 8.19: Sonification audio metaphors

The majority of these images are related to an event that is characterised by a gesture; either a human gesture, movement or a natural gesture. In particular, the majority of answers can be related to air, wind and water.

The answers related to air and wind are:

Underwater sounds 7

Musical instrument 3

Wind/air 3

Natural events 3

For a total of 16 out of 34

The answers that relate to water are:

Underwater sounds 7

Waves 7

For a total 14 out of 34

Therefore 30 answers out of 34 relate to gestures created by air, wind or water

It can be concluded that the sound portrays gesture and movement even though considerations about appropriate metaphors was not taken into account when designing the sonification.

Finally, with the following questions, the subjects were asked explicitly if they found the sounds appropriate to represent muscle activity, i.e. movement:

> *These sounds are synthesised from data produced by the activity of leg's muscles.*
>
> *Do you find this sound appropriate to represent muscle's activity, i.e. movement?*

(20 subjects answered this question)

5 out of 20 (25%) subjects said no
5 out of 20 (25%) subjects said I don't know
4 out of 20 (20%) said maybe, with some sort of adjustment
6 out of 20 (30%) said yes.

Therefore half of the subjects found the sounds appropriate (30%) or almost appropriate (20%).

A quarter (25%) of the subjects found the sounds inappropriate and another quarter (25%) of the subjects did not know and could not comment.

From this result it would appear that the sound is closer to being appropriate than to being inappropriate. Similarities can be found between these findings and the study done by Bonebright in [2001] on everyday sounds. In the experiment described in this chapter, it was found that the level of roughness of the sound is related to the age of the subject producing the EMG: the higher the roughness the younger (and normally more energetic) the subject. This type of sound was also found to be closer to being appropriate than being inappropriate in describing muscle's activation. Bonebright found that rough everyday sounds are also described as tense, a descriptor which could also relate to how energetically a muscle is activated. In both these studies, roughness seems to relate to the concept of tension.

8.10.2 Results from questionnaire

The questionnaire allowed the gathering of information about some general characteristics of this particular auditory display. In particular, the questionnaire algorithm as they are produced. Future evaluations of this sonification within an auditory biofeedback system could utilise this real-time set up.

As described in previous sections (see Section 8.2 of this Chapter), the muscle's firings are detected via electrodes attached to the skin of the patient which are then sent to the Biopac transducer which then sends them to the computer where the data are read by the Biopac software. Fortunately, the Biopac system allows routing the incoming data to analogue outputs from which the data can be sent to other interfaces. The data channels were sent, through 6 cables, to the PICO ADC converter, ADC-12, by PicoTech Ltd [Pico Technology] which connected to the PC parallel port. An object called "ADCreader" in PD allows reading the data coming through the PICO interface. This object was programmed by a student of Limerick University (Ireland) following the instructions provided by PicoTech for creating a software interface to the PICO ADC. As the current version of the Toolkit only allows the use of data that has been recorded into a file, for the video demonstration, a new simpler patch in PD was programmed that reproduced the EMG sonification algorithm described in this chapter that accepted the data incoming in real-time.

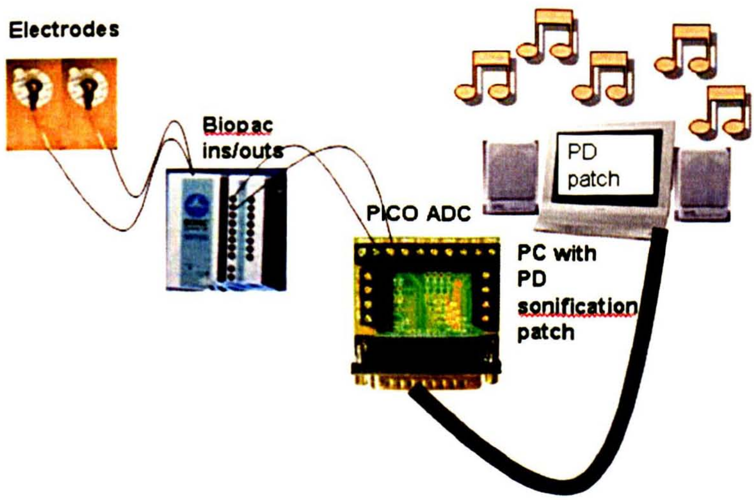


Figure 8.34: Connections for the real-time implementation of the EMG sonification (original in colour)

The physiotherapy video in the accompanying DVD clearly shows: the production of Voluntary Isometric Muscles Contractions (VIMC), the electrodes placement, the connection between the electrodes and the BIOPAC system, aimed at finding out qualitative information about how the display was perceived by the users. The display was considered neither pleasant or unpleasant and slightly interesting. This result suggests that the display should be aesthetically improved while maintaining accuracy. This would increase accessibility by analysts, physiotherapists and patients. The type of sound produced with this type of sonification seems to cause a certain degree of fatigue. After 20 minutes of intense listening of these sounds, 65% of the test subjects didn't want to listen to these sounds anymore. This type of sound was found to create in the listener's mind an *audio image* of a natural movement/gesture. This result is encouraging because it means that the overall type of sound is not in contradiction with what it tries to represent, i.e. limb gesture. 50% of the test subjects found the sound appropriate or fairly appropriate, 25% found it inappropriate and 25% didn't know. A majority of people found it appropriate.

8.11 Summary of overall results

In this chapter an experiment has been described which aimed at verifying if a particular sonification of EMG data could be a good auditory display of these data. The sonification was found to display well certain characteristics known to be present in the data. These characteristics were also found to be well described by a single timbre descriptor of the sound: *roughness*. This means that complex physiotherapist/patient communication could potentially be made easier by adopting a common simpler language based on sound descriptors.

The sonification was found to be a fairly appropriate sound metaphor of the type of data it represents. Intense listening to these sonifications was found to produce a certain degree of fatigue and the sound is not considered either pleasant or unpleasant. This display would gain by being improved aesthetically, but without losing its accuracy.

8.12 A step further: a real-time implementation of the EMG sonification

By a way of a coda to this chapter on the sonification of EMG data, a real-time implementation of this sonification is briefly described. The video found in the accompanying DVD in the folder *Real-time sonification of EMG/Video* shows this implementation in which the EMG data are fed into the sonification the connection between the electrodes to the ADC and the PC for the real-time setup.

8.13 Conclusions

The experiment described in this chapter shows that it is possible to create an effective sonification display of a multivariate large data set. This sonification uses raw EMG data and can be produced in real-time, making it a good candidate for the creation of an EMG auditory feedback. An experiment was conducted in this research that confirmed that the sonification effectively displays the relationship between the overall loudness and attack's speed of the sonification sound with the age of the person producing the EMG data. The roughness of the sonification also relates to the age of the person performing the movement. This means that the general descriptor roughness could be used by the physiotherapist to instruct the patient to produce a younger movement. A questionnaire showed that, although the sonification could be aesthetically improved, it is considered a fairly appropriate display of what it represents. The final section of this chapter described one possible implementation of this sonification as a real-time auditory feedback.

Chapter 9 – Experiment three: evaluating interaction in sonification

9.1 Introduction

The two experiments described in the previous chapters mainly aimed to verify if sonifications of multivariate large data sets could be effective in displaying information. The experiment described in this chapter aims to verify if the addition of user interaction can improve the efficiency, effectiveness and user satisfaction of such displays. More specifically, this chapter describes the role of interaction in a sonification system. The general process of designing, implementing and evaluating various interaction methods is described. The particular interaction methods designed in the context of this research are explained. The experiment that was conducted to evaluate these interaction methods is explained and its results analysed in detail.

The combination of all the results from these three experiments (see Chapter 7, 8 and 9) answers the issues posed by this thesis hypothesis.

9.2 The design, implementation and evaluation of interaction

So far, we have focused on evaluating how effective a sonification can be as a display of data sets, however, in an interactive sonification system, the sonification is not the only aspect of the system that needs evaluation and careful calibration. The second fundamental aspect that needs evaluation is the interaction. In an interactive sonification system, the user is allowed to "experience" the sonification having a certain degree of control over the way the sonification is displayed. This means that the user is allowed to "observe" the data/sonification in many ways, i.e. the user's point of view can easily and quickly change. There are many issues that need to be taken into consideration when designing the interactive aspect of a system and it is crucial to evaluate this aspect. A good interaction design enhances a system and makes it incredibly powerful (because it makes it more "human" and less "machine-like"), whereas a bad interaction design can obscure good aspects of the system and make it unusable. Highly interactive websites are good examples of how the quality of the interaction overshadows other aspects. For example, a user would not normally mind if a website has simple graphics, not very nice colours, but allows buying a plane ticket very easily. On the other hand, a beautiful website, with very pleasing graphics and colours, would not be successful if it required a very long and intricate process to make an equivalent purchase. While it is clear that the interaction aspect of a system can really influence the system's effective usefulness, it is not straightforward to state what qualities a system should have in order to allow a good interaction.

In *Interaction Design*, Preece *et al.* [2002] state that a good design should include the following characteristics: one-step actions, simplicity and elegance. While we can understand that such qualities are probably good, it is not easy to establish how to measure them. To have an overview of what is involved in designing interaction can help us find out what are the important variables to evaluate when assessing the interactive aspect of a system. The following information is summarised from [Preece *et al.*, 2002].

First of all, one has to identify the needs and requirements of the future users of the system. Then one or more alternative designs that meet the requirements are developed. These designs need to be embodied in some form so that they can be communicated and assessed. Finally the designs need to be evaluated. The results of the evaluation will often suggest changes and improvements of the original designs. These changes need to be implemented and assessed again until a particular design is judged to be satisfactory.

The design process is, therefore, a feedback loop that pushes the design to its best version.

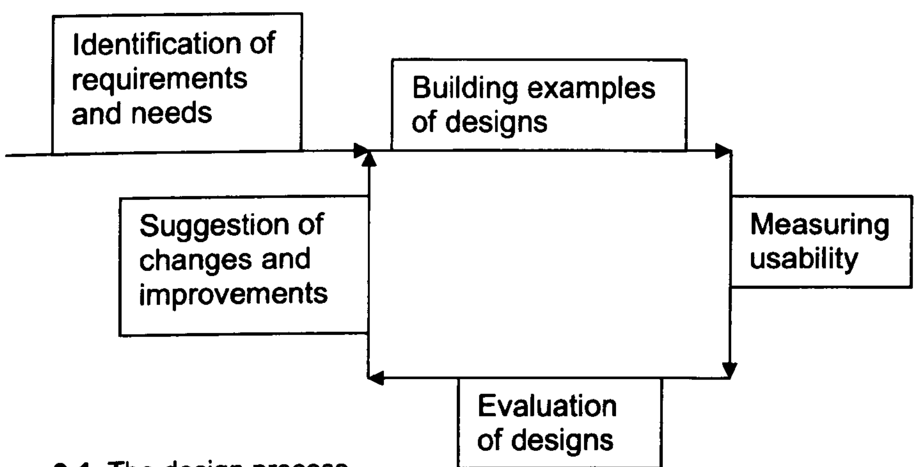


Figure 9.1: The design process

A very important step of the design process is the evaluation. All the steps of this loop are important, but if the evaluation is badly done or has biases, the design process can lose improvement power. A bad evaluation can transport the final design far from the ideal design, while a good evaluation brings the newer version of the design closer and closer to the ideal design. The evaluation step can be considered the gain of the feedback loop and continuing with this analogy: good evaluation is equivalent to a gain less than one, i.e. the distance between ideal design and real design gets smaller at each passage through the loop; on the other hand, a bad evaluation is equivalent to a gain higher than one, i.e. the distance between ideal design and real design gets bigger at each passage through the loop.

Two main concepts are important in design evaluation: user-centred design and usability. User-centred design refers to a design process that involves the users at all stages. In particular users are involved when identifying requirements and at the evaluation stage. This is important because the users can remind the designers, at all times, of their real needs. The general term usability refers to the fundamental aspect of the design that needs to be evaluated. The evaluation "measures" how "well" the system can be used if a certain design is utilised.

For a system to be "usable" it needs to have many characteristics. The following is a list of some of the most important characteristics a usable system needs to have:

- it needs to be effective, e.g. deliver the results it promises;
- it needs to be efficient, e.g. deliver quickly;
- it needs to be safe;
- it should be easy to remember how to use.

Typical measures of usability are [Shneiderman and Plaisant, 2005]:

- the number of errors made while performing the task, i.e. the fewer the errors, the higher the effectiveness of the system: measure of effectiveness;
- the time to complete a task, i.e. the shorter the time, the higher the efficiency of the system: measure of efficiency;

- user satisfaction, i.e. how much the users like the system. Data can be gathered via questionnaires, interviews and open comments.

Other usability principles are:

- *visibility*: functions of the system must be evident;
- *feedback*: there should a very short delay between an action and its result. Feedback is about sending to the user information about what action has been done and what still needs to be done. For an interaction to be effective, the feedback needs to be quick;
- *constraints*: restricting the user to only useful and correct actions;
- *mapping*: relationship between controls and their effects;
- *consistency*: similar operations use similar elements to produce similar tasks;
- *affordance*: term used to refer to an attribute of an object that allows people to know how to use it. In this context to "afford" means "to give a clue" [Preece at al; 2002]. An example could be: a door handle affords pulling.

In the evaluation stage of the design process it is important to select the most important usability principles that need to be measured. It is important as well to gather qualitative information about how the interaction is perceived (e.g. pleasant, difficult, etc.). This is important because a system needs not only to be objectively usable (efficient, effective, etc.) it also needs to be perceived as useful, pleasant, helpful, etc., or, in other words, satisfactory.

To evaluate a system means to ask a certain number of people to do a test. It is necessary to have:

1. the system in some form, i.e. a prototype of the system;
2. the test, e.g. a series of tasks the subjects need to perform, a questionnaire they need to answer, etc.

A prototype can be anything between a storyboard and a complex piece of software. These different types of prototypes can be divided into two main categories: low fidelity and high fidelity prototypes [Preece *et al.*, 2002]. An example of a low fidelity prototype is storyboarding: illustrating with a series of cartoons the chain of main actions and scenarios that happen when interacting with the system. An example of high fidelity prototype is a piece of software that reproduces the main characteristics of the system.

In both cases a prototype is a compromise between what can be produced relatively quickly and the aimed final system. In particular the speed of response is normally slow, graphical interface is sketchy and the functionality is limited. It is also possible to distinguish between:

- *Horizontal* prototyping: a prototype with a wide range of functions, but little detail;
- *Vertical* prototyping: a prototype with a lot of detail for few functions only;
- *Evolutionary prototyping*: where the prototype will evolve into the final product;
- *Throwaway prototyping*: where the prototype will be discarded when creating the final product, which will be implemented from the beginning using the knowledge gained from the user tests.

Prototypes are used for various purposes, e.g. to test out the feasibility of an idea, clarify requirements, do evaluation and some user testing, check out the compatibility of certain design directions with the rest of the system, etc.

Both low and high fidelity prototypes have advantages and disadvantages and these should be taken into account in each particular situation.

Advantages and disadvantages of different types of prototypes		
Prototype category	Advantages	Disadvantages
Low fidelity	• Lower cost • Multiple designs can be evaluated quickly • Useful communication device	• Limited error checking • Poor detailed specification • Limited utility after requirements established • Limited usefulness for usability tests • Navigational and flow limitations
High fidelity	• Higher functionality • Look and feel of final product • Can serve as a living specification • Can serve as marketing and sales tool	• More expensive to develop • Time consuming to develop • Not effective for requirements gathering

Table 9.1: Advantages and disadvantages of different types of prototypes. This table is summarised and adapted by the author from table in [Preece *et al.*, 2002; p. 246].

Once a design idea is ready for testing and a prototype has been implemented, one needs to build a test to measure the usability of the system and to evaluate how it is perceived by the users.

Depending on what type of information is needed, different evaluation paradigms can be used:

- *Informal feedback*: an informal quick feedback from users;
- *Usability testing*: measuring typical users' performance on carefully prepared tasks that are typical of those for which the system was designed. Three typical measures are: the number of errors, the time to complete task and the interaction path. Interviews and questionnaires are also used in this type of testing. The conditions under which the test is performed are controlled as much as possible. This has the advantage of producing results that are easier to interpret, but the disadvantage that what is tested is not the real situation;
- *Field studies*: in this case the evaluation is done in natural settings with the aim of increasing the understanding about what users normally do and how technology impacts on them. The data produced in this type of study are conversations, video, artefacts, etc. ;
- *Predictive evaluation*: in this case experts apply their knowledge of the typical users to predict usability problems. In this case the user is not present during the evaluation.

There are also different techniques in evaluation:

- *Observing users*: the observation is done using notes, audio, video, diary, etc.;
- *Asking the opinions of the users*: interviews and questionnaires are utilised;
- *Asking the opinions of experts*: in this case experts step through tasks role-playing typical users and identify problems;
- *Testing the performance of users*: two or more designs are compared by measuring the users' performance in both designs;
- *Modelling users' performance of tasks*: this allows predicting the efficiency of a user interface.

To prepare for evaluation it is important to check the following steps:

1. determine the goals of the evaluation;
2. explore the questions to be answered;
3. choose the appropriate evaluation paradigm and techniques;
4. identify practical issues, e.g. how many test subjects is it necessary and practical to have;
5. analyse, interpret and present the data.

9.3 Evaluating the interaction: an experiment

The aim of the experiment described in this chapter is to evaluate and compare three different methods for interacting with a sonification of data streams. A usability test was carried out using the Interactive Sonification Toolkit as a high fidelity prototype. The usability test focused on measuring the *efficiency, effectiveness* of the three interaction methods. At the end of the test, subjects were asked to answer a questionnaire (see Appendix E for the questionnaire) which gathered information about the users' perception of the interaction provided by different interaction methods. This gave some information also about *user satisfaction*.

Here there follows the description of the rationale behind the experiment design. The final goal of a software tool which uses sonification to display complex data sets is to support data analysis and exploration *efficiently* and *effectively*. A fundamental task in data analysis is to be able to identify particular structures present in the data under examination. Three separate factors can affect the efficiency and effectiveness of data analysis via sonification:

1. the specificity of the data set;
2. the sonification algorithm;
3. the interaction method.

This means that once a sonification method has been proven to be a good display for a particular data set and once the user knows what kind of data structures can be found in the data set, the efficiency and the effectiveness with which a user analyses a data set and recognises the data structures it contains depends only on how the user is allowed to navigate the auditory display. In this experiment the independent variable is the interaction method which is evaluated under three different conditions. The dependent variables, on the other hand, are the time spent to complete a task (measure of efficiency: the higher the time, the lower the efficiency) and the number of incorrect data structures' identifications made during the execution of the task (measure of effectiveness: the higher the number of incorrect answers, the lower the effectiveness).

9.3.1 Experiment description

In this experiment, the subjects were asked to navigate and listen to a sonification using three different interaction methods. Their task was to recognise which data structures were present in the data set and in which order. Before starting the test, the subjects were trained to listen to the particular sonification and to the particular kind of data sets provided in the test so that they had experience in how to analyse and recognise the structures present in the data sets. This training was given so that the experimenter did not need to consider the sonification and the type of data sets as variables in this experiment. In the test, the subjects used the Interactive Sonification Toolkit (see Chapter 6 of this thesis) developed by the author.

In this toolkit, the user is presented with a red rectangular area (or screen area) on the top half of the screen. This screen area represents the sonification from beginning to end (the beginning is mapped to the left corner and the end to the right corner of the area). The user can navigate the sonification by interacting with this screen area using two types of interface devices: the mouse and the jog wheel/shuttle interface [Contour Design].

9.3.2 The interaction methods

In a qualitative scale of interaction (see Figure 9.2) that has as its extremes non-interaction and very high interactivity, there are many middle stages where there is interaction, but control is still quite simple and evident.

Non interactive				Very interactive
Fixed sonification	Possibility of changing the fixed duration of the sonification	Quickly change the speed of the sonification while it is playing	Immediate change in direction and speed of sonification	Instant modification of many sonification parameters: speed, position, direction, mapping, range, etc.

Figure 9.2: Interaction scale

The interaction methods studied in this experiment can be considered to sit in the middle section of this scale. Furthermore, the interaction methods studied here use well known and common interface devices. These devices were chosen in this study because they are common and therefore it is highly probable that they will be used in a final implementation of an interactive sonification-based data analysis system. The fact that these interaction methods are relatively close to each other, in terms of interactivity, and easy to learn, because of their popularity, could make the results of the comparison difficult to predict and therefore interesting. For these reasons this experiment explores three interaction methods which lay in the middle in this interactivity scale.

9.3.3 The three interaction methods in detail

The three interaction methods studied in this experiment are:

1) *Low interaction method*: a section of sonification can be selected and the playback duration can be fixed before playback;
2) *Medium interaction method*: it is possible to switch between different fixed playback speeds and direction of playback;
3) *High interaction method*: immediate change in direction and speed of playback is possible.

9.3.4 Low interaction

In the first method, the subject selects a section of the sonification by right-clicking the mouse somewhere within the screen area, dragging it towards the right, and then letting go of the right button of the mouse wherever the selection ends (see Figure 9.3).

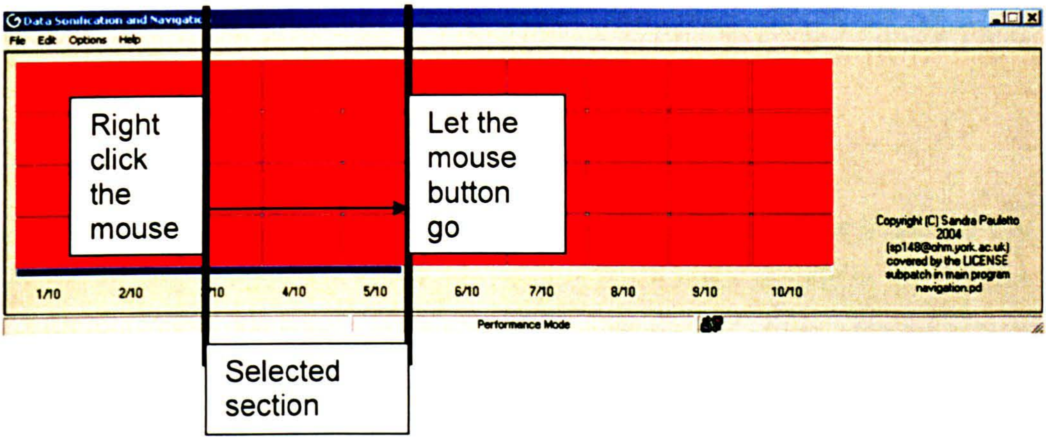


Figure 9.3: Selection of section of sonification (original in colour)

After the selection has occurred, the subject needs to enter the amount of time in which he/she wants to hear the selected section of sonification. Finally, the subject clicks the button 'play' to hear the selected section. This interaction method derives from the way in which, in various audio editing programs (e.g. Soundforge), a section of sound can be selected for manipulation. In this case the only possible manipulation is to choose the length of time in which to hear the selection (which could be considered a kind of temporal zoom).

9.3.5 Medium interaction

In the second method, the subject is asked to navigate the sonification using the jog wheel/shuttle interface (see Figure 9.4).

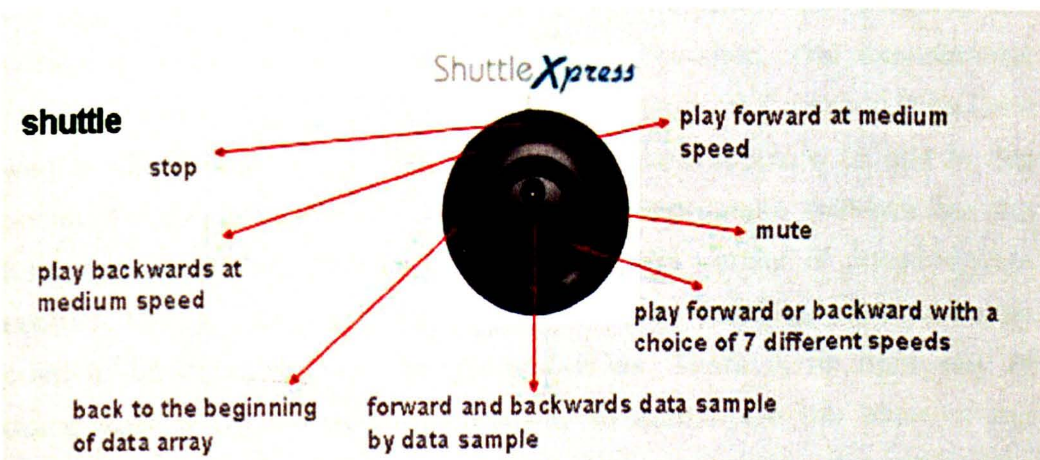


Figure 9.4: The Shuttle XPress Interface

The buttons and the wheels of the shuttle are mapped to defined presets as shown in Figure 9.4. With this interaction method, the subject can move backwards and forwards in the sonification at various constant speeds. The user can jump to the beginning of the sonification by pressing one button and can stop and change direction instantly.

9.3.6 High interaction

In the third method, as the subject left-clicks the mouse and drags it around the screen area, the sonification plays. By moving the mouse around in the screen area, the mouse is instantly mapped to the scaled values of the data set which are instantly fed into the sonification algorithm that produces the sound. The speed at which the sonification is played depends on the speed of the movement of the mouse on the screen. In this case the speed is rarely constant. One might think that the medium or high interaction methods are more efficient than the low method because they allow a more immediate user control of the playback. However this immediacy could actually produce more confusion when performing the task, i.e. the user could make more mistakes in recognising the data structures because of the speed of interaction this immediacy allows. As a result, the explorations using high and medium interaction methods could be less effective and have overall a lower usability than the low interaction method. This experiment will determine if this situation occurs.

9.3.7 The data and its sonification

The structures present in the data need to be fixed and known by the experimenter so that, when the test subjects are asked to recognise the structures present, the correct answers can be counted. The experimenter created the data sets using various algorithms in Microsoft Excel and Pure Data. It was a difficult task to decide what types of data structure to use in this experiment. Complex data sets come from both man-made systems and the natural world and they present us with an infinite variety of possible data structures. In this experiment only a few structures could be used and they needed to be controlled, i.e. completely known. There is no right way of deciding how to choose and create these structures. On the basis of the experimenter's experience in working with data sonification, five main data structures were considered to be very basic and common in data sets produced by any type of process (e.g. natural, mechanical, etc.):

1) a noisy structure;

2) a constant structure;

3) a linear structure (in particular an ascending linear ramp);

4) a discontinuous structure;

5) a periodic structure.

Each data set used in this experiment included all of these structures and each data set channel contained the same number of data samples (220500). This number of data per channel was chosen to set the following timing reference: if the data was played back at audio rate (44.1kHz), then the sonification would have lasted 5 seconds.

1) Noisy structure

The noisy structure contains random numbers produced using the algorithm INT(RAND()*65536)-32768 in Excel (i.e. integer random numbers between -32768 and 32768).

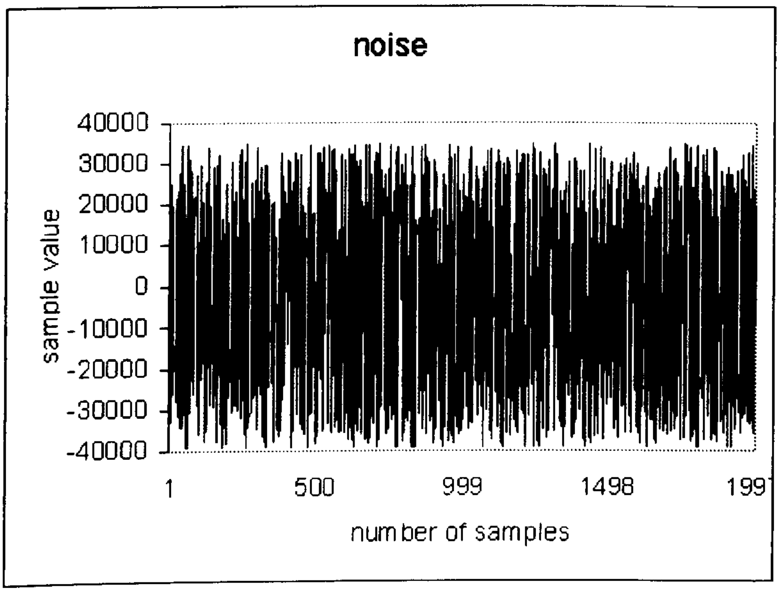


Figure 9.5: Graph of noisy data structure

2) Constant structure

It was decided that the constant value would be set at 49152, i.e. ¾ of 65536.

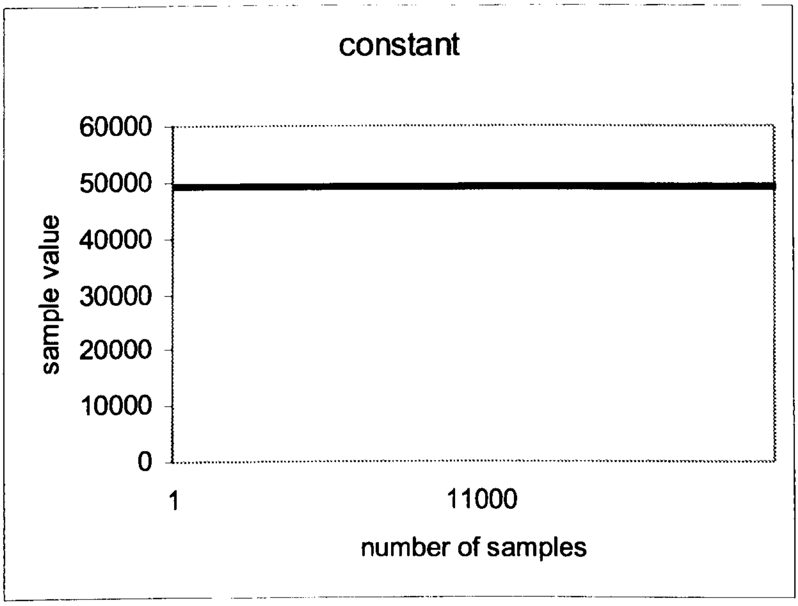


Figure 9.6: Graph of constant data structure

3) Ascending linear ramp structure

The ascending linear ramp is constructed by a sequence of values going from 0 to a maximum of 65536 by adding 1 at each new step.

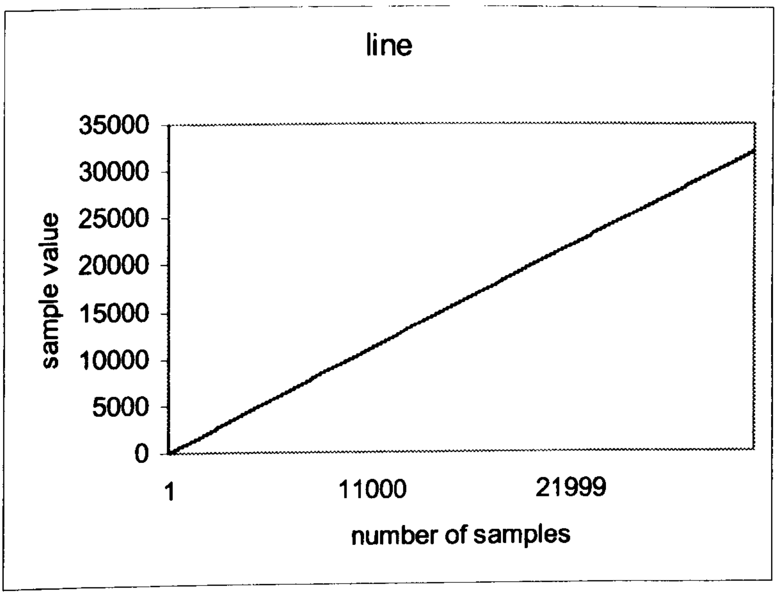


Figure 9.7: Graph of ascending linear ramp data structure

4) Discontinuous structure

The discontinuous structure was created by having always zero apart from the occasional value of 65536 at random positions.

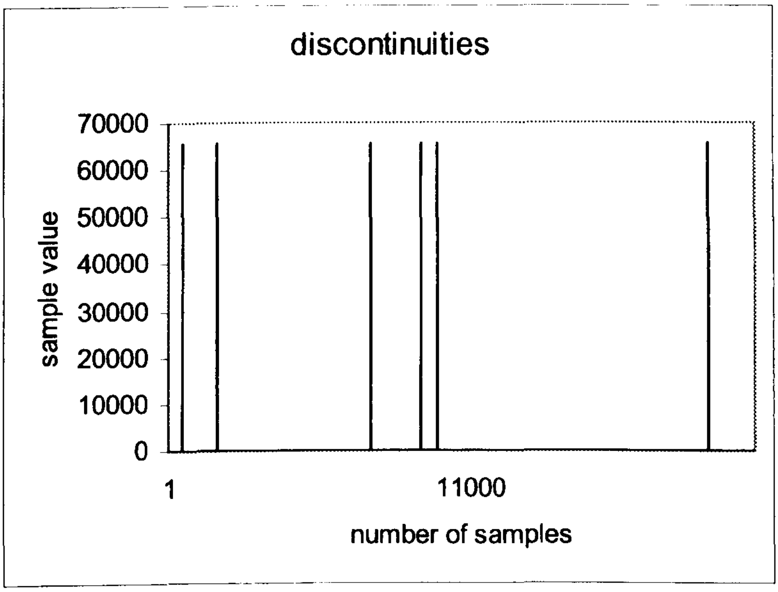


Figure 9.8: Graph of discontinuous data structure

5) Periodic structure

The periodic structure was produced by creating one period of a sine wave in an array of 256 points in PD. The data of this array were exported in a text file. The data then were imported in Excel and repeated 44100/256 times. The data (originally between -1 and 1) were then scaled between -32768 and 32768 by multiplying the original data by 32768. The rate of repetition of the sine wave is: 44100 / 256 = 172.2656Hz.

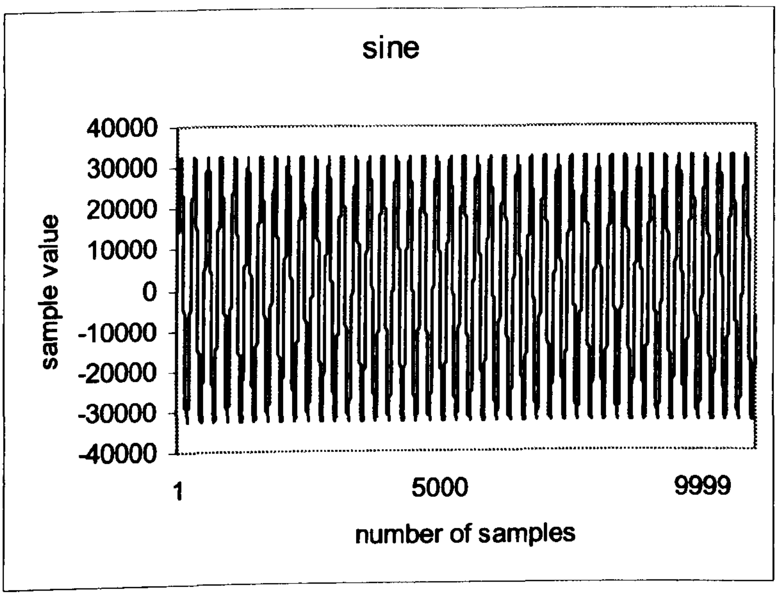


Figure 9.9: Graph of sinusoidal data structure

9.3.8 The sonification

It was decided to map the data sets to the amplitude of a sine oscillator of fundamental frequency 261.6Hz (middle C). This type of mapping is simple enough so that people can very quickly learn how the different data structures sound using this mapping.

- The noisy structure sounds like noise.
- The constant structure sounds like a sine wave of constant volume.
- The linear ramp sounds like a sine wave with an increasing volume.
- The discontinuous structure sounds like a series of clicks and silence in between clicks.
- The periodic structure sonification results in a particular case of amplitude modulation (AM). If the playing speed is low, it sounds like a vibrato, while if the speed is high two sidebands typical of AM synthesis form and therefore we hear two pitches (see Chapter 8, Section 8.4.4 of this thesis for a detailed explanation of the amplitude modulation synthesis technique).

For this experiment it is important that people can very easily recognise the data structures, given the sonification, if they are presented with a simple data set. However, in order to be able to measure the effects of different interaction methods on the identification of the structures, it is important to create data sets that need repeated listening to be understood, i.e. not so easy to understand that no one will make a mistake. Furthermore, if repeated listening was not needed, the action of navigating the sonification would not be needed and obviously would not be measurable.
The strategy used to make the datasets at the same time simple, but needing repeated listening, was to construct them in such a way so that they would challenge the subject's hearing attention all the time. Experiments show [see Bregman, 1994; pp. 207-8] that if two different sequences of words are presented to subjects, one in the left ear and the other in the right ear, and after the subjects are asked to repeat one sequence, they usually cannot report the words heard in the non-attended ear. This fact led to the idea, in this experiment, of playing different streams in the two ears simultaneously. Usually subjects would need repeated listening to switch attention from the left to the

right ear and recognise all the elements in the two streams. For this reason, it was decided that each data set should contain two channels of data, one panned to the left and one to the right, each of which containing two different sequences of the five structures mentioned above.

It was also decided that the different sections of structures should last different lengths of time (so that structures would not change simultaneously both in the left and in the right ear). Each channel had a sequence of 10 structures and two sections of data containing the same structure could not be presented one after the other. Three different data sets were constructed. In each of them the order of the structures' sections and their length was different.

First data set

Distribution of data structures in 1st data set			
Left channel	**Samples per section**	**Right channel**	**Samples per section**
constant	22050	sine	11025
line	33075	noise	22050
noise	22050	line	22050
sine	11025	constant	33075
discontinuities	11025	discontinuities	22050
sine	22050	constant	33075
noise	22050	line	22050
discontinuities	33075	sine	11025
line	22050	discontinuities	33075
constant	22050	noise	11025

Table 9.2: Distribution of data structures in 1st data set

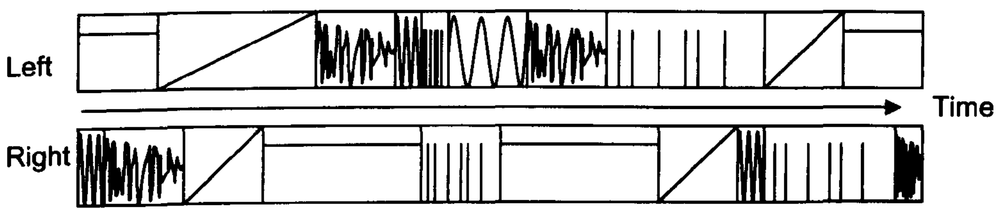


Figure 9.10: Graphical representation of 1st data set

Second data set

Distribution of data structures in 2nd data set			
Left channel	**Samples per section**	**Right channel**	**Samples per section**
sine	11025	constant	33075
noise	22050	sine	11025
discontinuities	22050	discontinuities	22050
line	33075	line	22050
constant	22050	noise	33075
line	22050	line	22050
constant	11025	noise	22050
discontinuities	33075	discontinuities	11025
noise	22050	sine	11025
sine	22050	constant	33075

Table 9.3: Distribution of data structures in 2nd data set

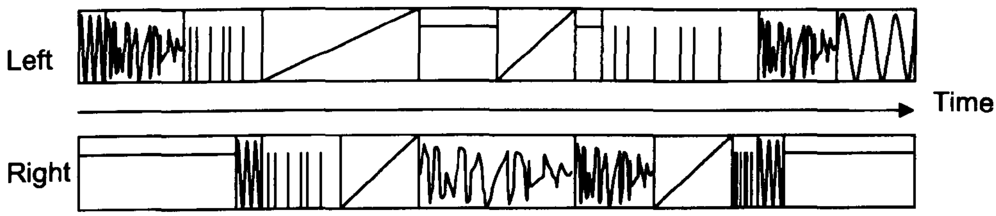


Figure 9.11: Graphical representation of 2nd data set

Third data set

Distribution of data structures in 3rd data set			
Left channel	**Samples per section**	**Right channel**	**Samples per section**
line	33075	sine	33075
discontinuities	22050	noise	11025
sine	11025	discontinuities	22050
constant	22050	line	22050
noise	11025	constant	22050
line	22050	noise	11025
sine	22050	discontinuities	33075
noise	33075	sine	22050
discontinuities	11025	constant	22050
constant	33075	line	22050

Table 9.4: Distribution of data structures in 3rd data set

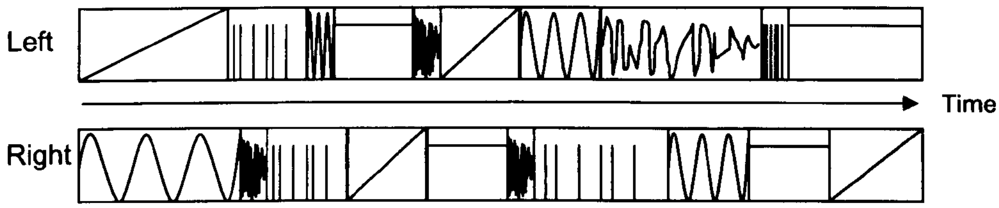


Figure 9.12: Graphical representation of 3rd data set

9.4 The Experimental Procedure

9.4.1 Experimental design

This experiment has a within-subjects (or related) design. This means that the same group of subjects do the experiment under all the conditions (the three conditions correspond to the three different interaction methods).

Various aspects of the experiment are randomised so that effects due to order of presentation are eliminated:

- in order to eliminate errors due to the order of the conditions, each subject is presented with the conditions in a different order;
- every time the interaction method, i.e. the condition, changes, the data set also changes (otherwise the subject would already know the order of the data structures);
- finally, it is important not to always assign the same data set to one interaction method because this can cause errors: for example, one interaction method could be particularly good when used in conjunction with one particular data set.

The number of possible data set/condition combinations that follow the above rules is a number that can be calculated and that tells us how many test subjects are necessary for the test. Let us call the three different interaction methods a, b, and c, and let's call the three different data sets 1, 2 and 3. The total number of permutations without repetitions can be calculated

$$n\_P\_k = \frac{n!}{(n-k)!} \qquad (9.1)$$

where:

n_P_k = number of permutations of size "k" that can be made with "n" number of objects without repetitions;

n = number of objects to be permutated;

k = size of permutation.

The total number of permutations without repetitions for the three interaction methods is

$$\frac{3!}{(3-3)!} = \frac{3!}{0!} = \frac{3*2*1}{1} = 6$$

since the number of objects to permute is 3 and the size of the permutations is 3 (remember that 0! = 1). The same is true for the number of permutations of the three data sets, i.e. there are 6 permutations.

If we pair up the 6 interaction methods permutations and the 6 data sets permutations, there are 36 pairs (6*6 = 36) of interaction methods and data sets in which neither interaction methods or data sets are repeated. Therefore the experimenter should have had 36 test subjects available to do the experiment. As it was not possible to recruit 36 subjects, the number of pairs (interaction methods/datasets) was further reduced.

These 36 pairs include the situation in which particular configurations of pairs of interaction method/data set (e.g. a1, b2, c3) are repeated, but in a different order (e.g. a1b2c3 and a1c3b2). In particular, for each particular configuration of pairs, there are 6 of such variations in order: e.g. a1b2c3, a1c3b2, b2a1c3, b2c3a1, c3a1b2, c3b2a1. Three of these variations in order were eliminated for each configuration of pairs leaving us with a total of 18 combinations which required 18 test subjects. If a, b, c, are the interaction methods and 1, 2, 3 the different data sets then the combinations used were:

Combinations of interaction method and data set					
a1	b2	c3	a2	b3	c1
b2	a1	a1	c1	c1	b3
c3	c3	b2	b3	a2	a2
a1	b3	c2	a3	b1	c2
b3	a1	a1	c2	c2	b1
c2	c2	b3	b1	a3	a3
a2	b1	c3	a3	b2	c1
b1	a2	a2	c1	c1	b2
c3	c3	b1	b2	a3	a3

Table 9.5: Combinations of interaction method and data set

9.4.2 The test subjects

Eighteen subjects were asked to participate in this experiment. Seven of these subjects had done the experiments described in Chapters 7 and 8 a few months earlier. Nine of these subjects had done only one of the two experiments described in Chapter 7 and 8. This number of subjects in a three-condition within-subjects experiment has the statistical power 80% to detect a large effect (effect higher than 0.45) with a significance level of 5%. This test has also a 60% statistical power to detect medium effect sizes (higher than 0.35) with a significance level of 5% (see a-priori power analysis tests in Appendix B of this thesis). Results with effect sizes between 0.35 and 0.45 are here considered significant if they are consistent with the results of the other tests carried out on the same people and that have an effect size higher than 0.45. The average age of the subjects was 28 years of age. It was assumed that there would not be differences in judgement due to gender, and that, given a higher education level as background, differences in cultural background could also be disregarded. The subjects were fifteen male and three females. All the participants were British apart from a Malaysian and a French person. The subjects were all researchers, students or lecturers of York University Electronics Department. Sixteen subjects specialise in Music and Audio Technology. Sixteen subjects normally work with sound, while two do sporadically. The test was carried out in a silent room (in the recording studio performance area at the University of York, UK). Good quality headphones (DT 990 Beyerdynamic) were used with a wide frequency response (5 – 35kHz). This minimised the errors that could be due to external sounds and maximised the quality of sound reproduction.

9.4.3 Description of the experiment to the subjects

The task was explained to the subjects by the experimenter and the subjects had as much time as needed to familiarise themselves with the sonifications, the data structures and the interaction methods (typically they used about 5 minutes). The experimenter made sure that the subjects could easily recognise the data structures by giving a simple training that used a simple data set example. The test subjects were asked to navigate this simple sonification example using the three interaction methods and were asked to identify the structures present in the data set example. The data set example consisted of one channel of data. This channel was sonified and sent to both left and right ears. The data set was divided into five sections, each one containing one data structure. The subjects started the experiment only when they felt confident in recognising the structures and in using the interfaces.

9.4.4 The test

The experimenter uploaded and scaled appropriately one data set (following one of the orders in the table above) and panned the sound appropriately. The subjects were presented with a sheet of paper with two tables (see Figure 9.13).

Left ear

Right ear

Figure 9.13: Test sheet

The subjects were asked to write in the tables the sequences of structures they heard in both ears just by writing the following letters to indicate the structures:

N noise

C constant

L linear ramp

D discontinuities

S sinewave

The experimenter measured how long the subject took to recognise the sequences using a hand held chronometer.

After having done the task using the three different interaction methods and the different data sets, each subject was asked to fill in a questionnaire.

In the questionnaire the subject was asked to rate from 1 to 5 the *pleasantness* of each interaction method, the *intuitiveness*, the *clarity* and the *quickness*. The questionnaire was designed to gather information about user satisfaction, and the perceived efficiency and effectiveness of the interaction methods which could be compared with the objective results produced by the test. In particular, the perception of quickness was considered to be linked with efficiency and the perception of clarity with effectiveness. Finally, intuitiveness was considered to be a measure of good affordance and close interaction feedback (qualities which should increase effectiveness). Pleasantness was considered a measure of how likeable an interaction method was and therefore of user satisfaction. Then, to gather more information about user satisfaction, the subjects were asked to select their *preferred* interaction method and to *comment* on why they chose it.

9.5 Experiment results

9.5.1 Efficiency and effectiveness results

In this experiment two main dependent variables were measured as indicators of efficiency and effectiveness:

- the length of time spent in executing the task;
- the number of incorrect answers in recognising the data structures, where the higher the number of incorrect answers, the lower the effectiveness.

For each subject, there are three 'timings' and three 'incorrect answers' results: one for each interaction method. The averages, over the number of subjects, for the timings and the number of incorrect answers, and their significance, are calculated.

9.5.2 Efficiency results

The process involved in performing a task is efficient when the task can be performed quickly. The time used to perform the task is therefore a measure of efficiency. Timings are measurements done on an interval scale (continuously in milliseconds), for this reason parametric tests, such as One-Way ANOVA, could be performed to calculate the significance of the results. However, parametric tests require that the data have certain characteristics. Specifically, they have to be normally distributed and have homogeneity of variance. To verify if the data are normally distributed, a Shapiro-Wilk test (see Chapter 5, paragraph 5.4.2 for details on the test) was performed on the test scores obtained for each interaction method. Notice that if the test is significant it means that the data deviate significantly from a normal distribution.

Shapiro-Wilk test on efficiency results			
Interaction method	**Test**	**Significance**	**Result**
LOW	D(18) = 0.941	$p > 0.05$	Normally distributed
HIGH	D(18) = 0.846	$p < 0.01$	Not normally distributed
MEDIUM	D(18) = 0.872	$p < 0.05$	Not normally distributed

Table 9.6: Shapiro-Wilk test on efficiency results

Below are the distributions of the results for all the interaction methods. The continuous line in the diagrams shows the ideal normal distribution calculated using the mean and standard deviation of the data.

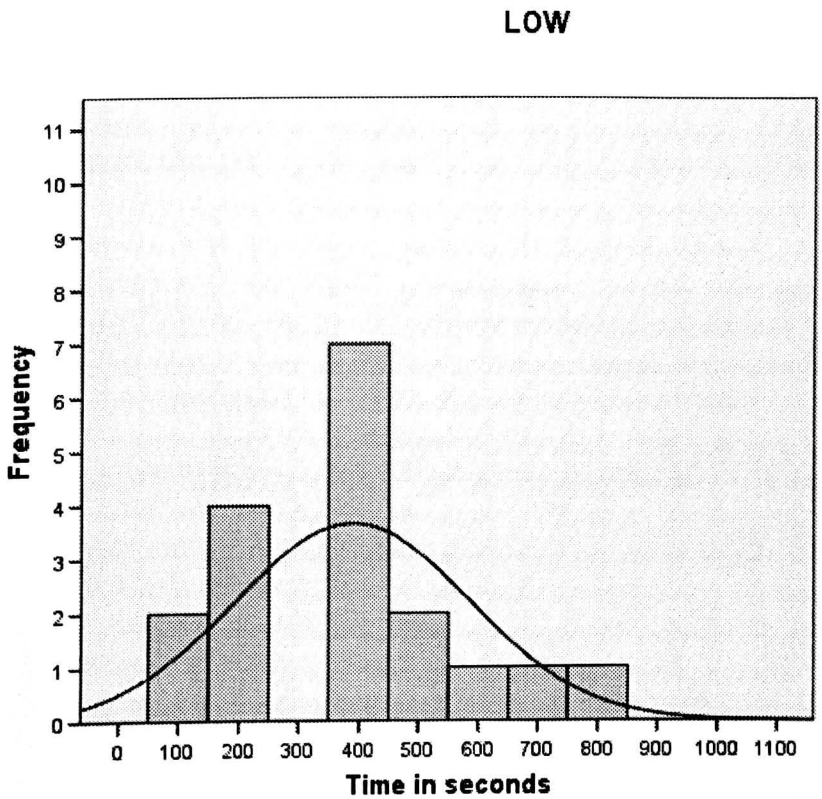


Figure 9.14: Distribution of timings for the LOW interaction method.

MEDIUM

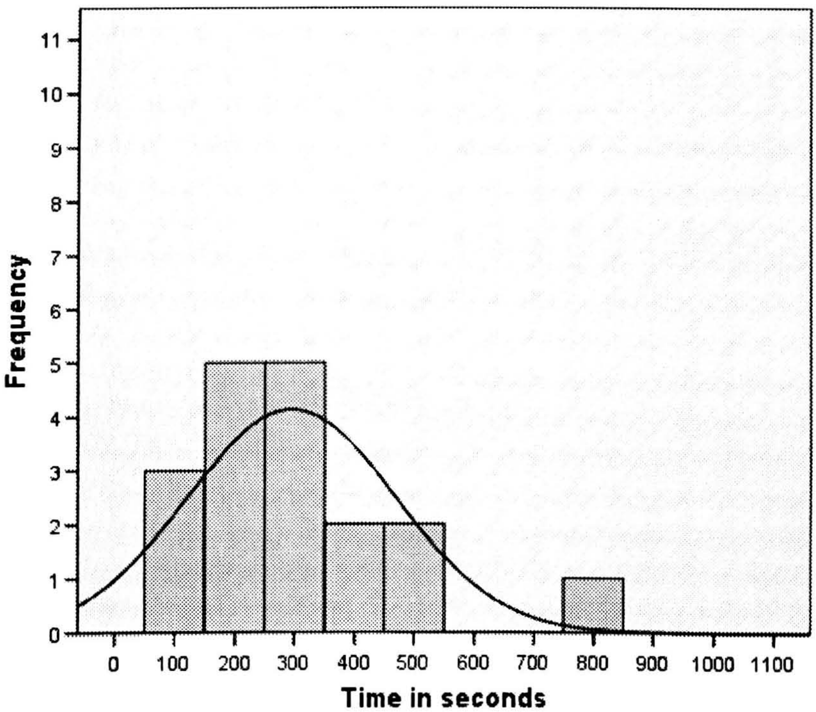


Figure 9.15: Distribution of timings for the MEDIUM interaction method

HIGH

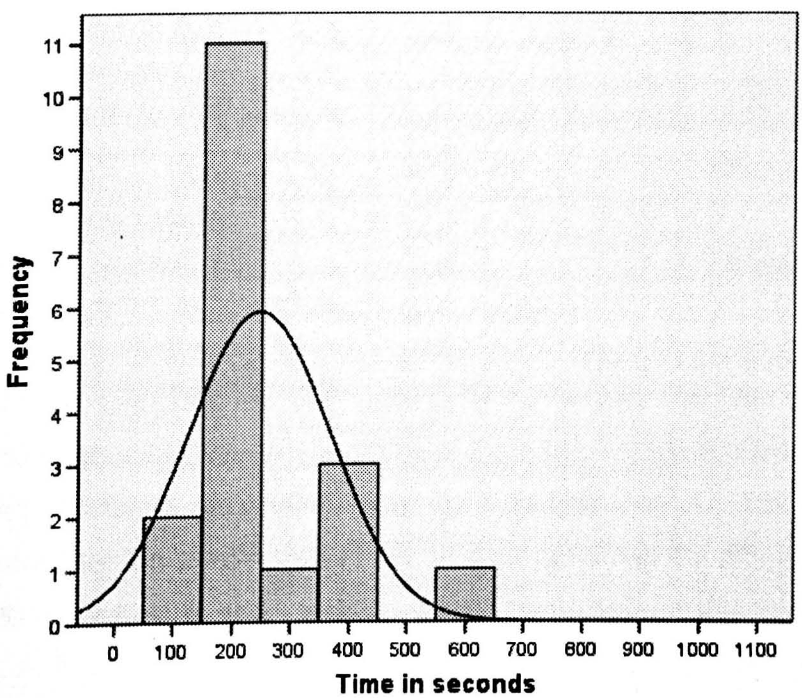


Figure 9.16: Distribution of timings for the HIGH interaction method.

From the analysis of skewness and kurtosis (see table below), it was found that the medium and high interaction methods distributions are positively skewed, i.e. they are significantly not symmetrical and the medium interaction method distribution has positive kurtosis, i.e. it has a too narrow a peak. The skewness can be due to the fact that there is a physical limit to how fast the task can be done (it cannot be done in zero or less than zero seconds!) and therefore all the scores below the mean are close to each other, while the scores over the mean can be more farther apart from each other.

Skewness and Kurtosis results for the timings distributions	**LOW**	**MEDIUM**	**HIGH**
Skewness	.528	1.607	1.282
Std. Error of Skewness	.536	.536	.536
Skewness/ Std. Error of Skewness	**0.985**	**2.998**	**2.392**
Skewness significance	**$p > 0.05$**	**$p < 0.01$**	**$p < 0.05$**
Kurtosis	.145	3.579	.976
Std. Error of Kurtosis	1.038	1.038	1.038
Kurtosis/ Std. Error of Kurtosis	**0.140**	**3.448**	**0.940**
Kurtosis significance	**$p > 0.05$**	**$p < 0.001$**	**$p > 0.05$**

Table 9.7: Skewness and kurtosis results for the timings distributions

To verify if there is homogeneity of variance in the scores, the Levene's Test (see Chapter 5, paragraph 5.4.2 for details on the test) was performed. Notice that if the test results significant it means that homogeneity of variance cannot be assumed.

Levene's test on efficiency results	**Significance**	**Result**
$F(2, 51) = 1.374$	$p > 0.05$	Variance is homogenous

Table 9.8: Levene's test on efficiency results

These tests tell us that, although the scores can be considered to have homogeneity of variance, not all of them can be considered normally distributed and therefore parametric tests, such as One-Way ANOVA, cannot be used to measure the significance of the average results in the different conditions.

Non-parametric tests such as Friedman's ANOVA (for more than two related conditions) and Wilcoxon signed rank (for two related conditions), with a Bonferroni correction applied to the significance level, will be calculated to verify the significance of the differences between the three conditions (see

Chapter 5, Section 5.4.2 for details on these statistical tests). The results for efficiency and the relative statistical tests (see Table 9.8) indicate that the Low interaction method is slower than the High interaction method. This means that exploring a sonification moving the mouse freely up and down the timeline is more efficient than selecting a section of the sonification, selecting the speed at which to hear it and then press play. The efficiency of the Medium interaction method is not significantly different from the other two methods.

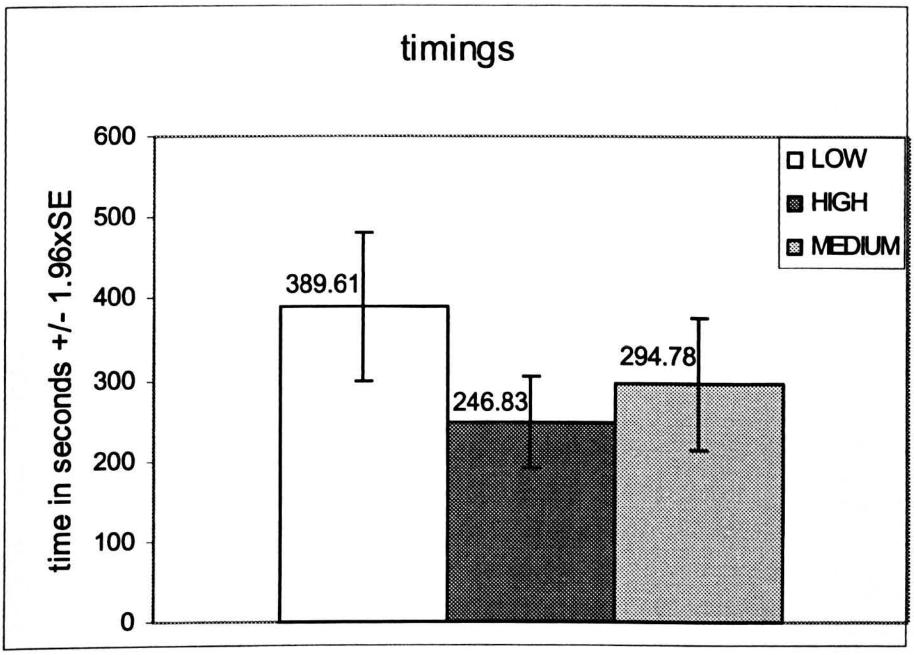


Figure 9.17: Efficiency results

Statistical tests for efficiency results		
Test: Friedman's ANOVA	**Significance**	**Effect size**
Chi-Square(2) = 8.444	p = 0.014 < 0.05	effect size = 0.35 (medium), statistical power between 60% and 80%
Post-hoc comparisons: Wilcoxon signed rank test	**Significance**	**Effect size**
Low-High: Z = -3.288	p(one-tailed) = 0.000 < 0.0167	effect size = 0.88 (large)
Low-Medium: Z = -2.069	p(one-tailed) = 0.019 > 0.0167	not significant
Medium-High Z = -1.067	p(one-tailed) = 0.149 > 0.0167	not significant

Table 9.9: Statistical tests for efficiency results

9.5.3 Effectiveness results

A method for performing a task is considered more effective than another if it minimises the chances of making errors in the task. The number of incorrect answers produced during the execution of a task is a measure of the effectiveness of the system, considering that the higher the number of incorrect answers the lower the effectiveness.

In this test, subjects were asked to perform a task under three conditions and the number of incorrect answers for each subject was recorded. As before, the Shapiro-Wilk test and Levene's test are performed on the recorded scores to verify if the data are normally distributed and have homogeneity of variance. If these conditions are satisfied parametric statistical tests can be performed on the data.

Shapiro-Wilk test on effectiveness results			
Interaction method	**Test**	**Significance**	**Result**
LOW	$D(18) = 0.887$	$p < 0.05$	Not normally distributed
HIGH	$D(18) = 0.838$	$p < 0.05$	Not normally distributed
MEDIUM	$D(18) = 0.917$	$p > 0.05$	Normally distributed

Table 9.10: Shapiro-Wilk test on effectiveness results

Below are the distributions of the results for all the interaction methods. The continuous line in the diagrams shows the ideal normal distribution calculated using the mean and standard deviation of the data.

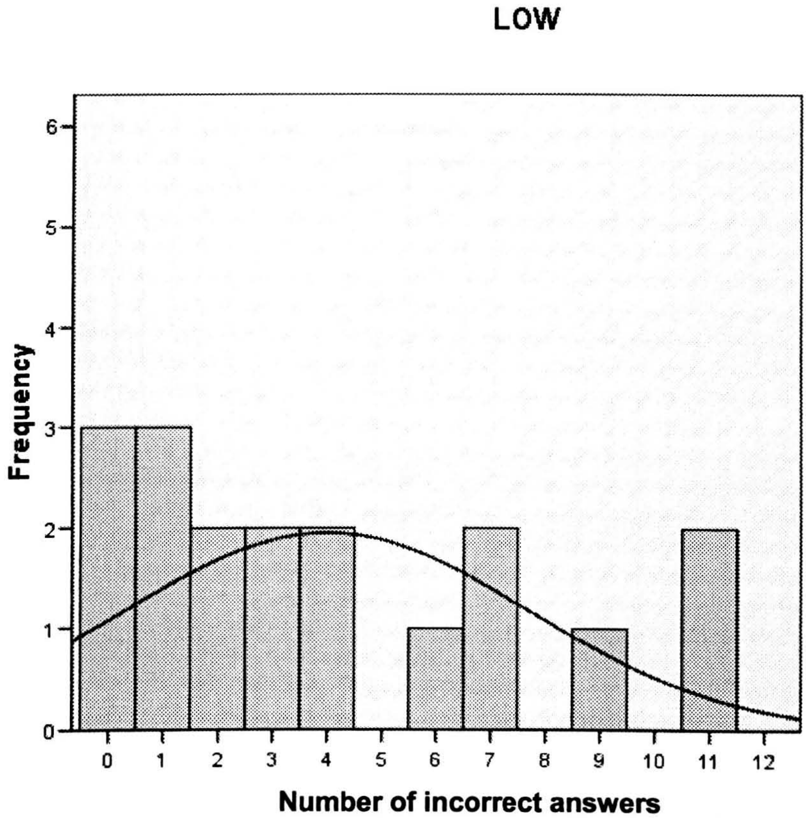


Figure 9.18: Distribution of number of incorrect answers for the LOW interaction method.

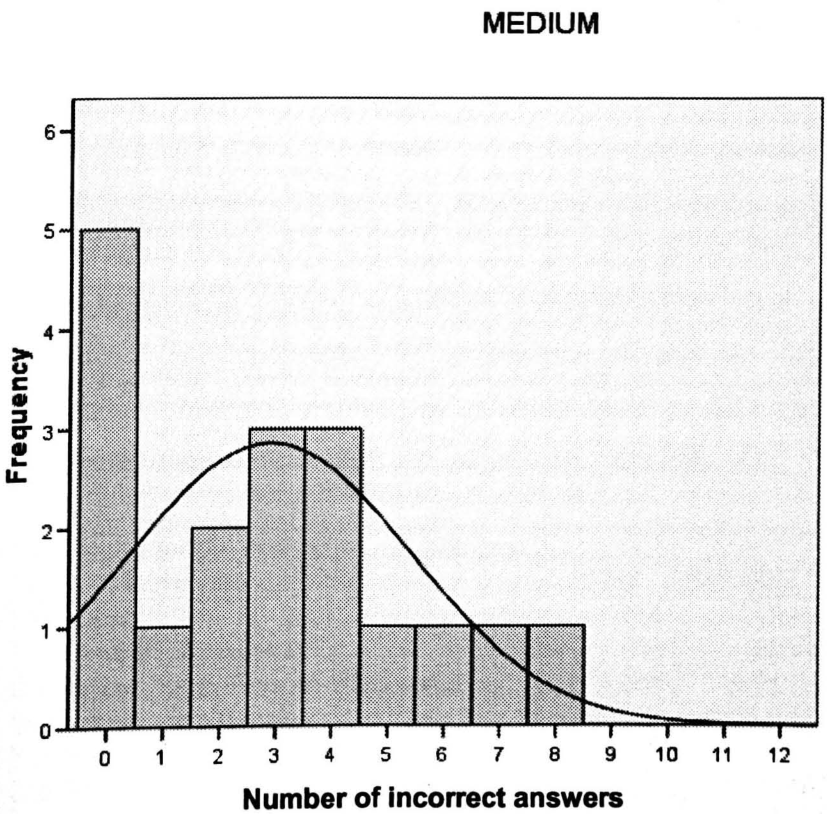


Figure 9.19: Distribution of number of incorrect answers for the MEDIUM interaction method.

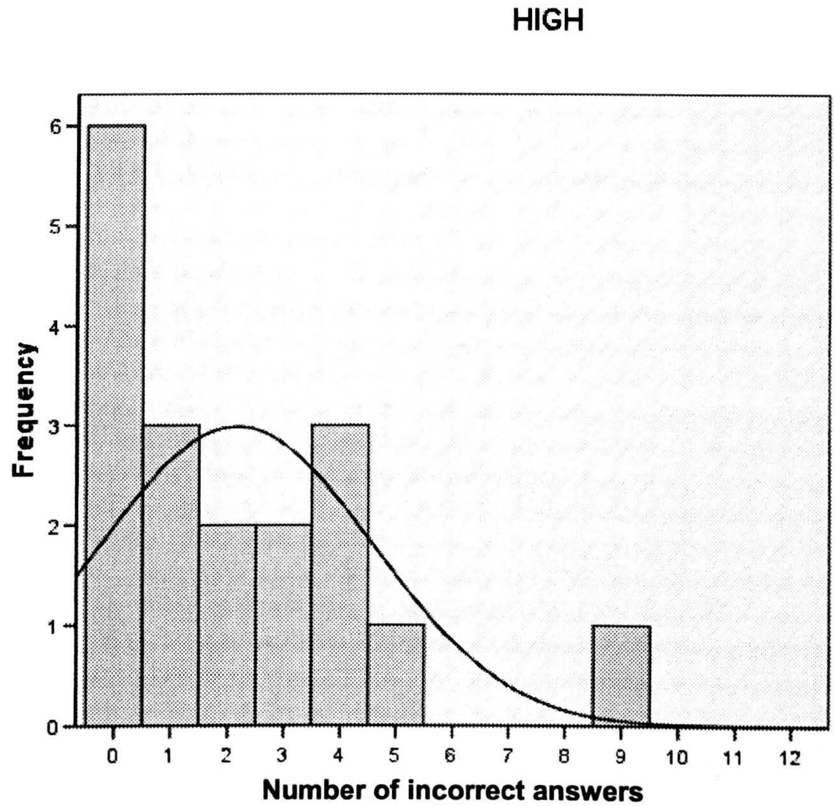


Figure 9.20: Distribution of number of incorrect answers for the HIGH interaction method.

From the analysis of the skewness and kurtosis of the data it was found that the high interaction method distribution has both significant skewness and kurtosis. The skewness can be explained by the fact that there is a lower limit to these data which is zero and by the fact that overall a large proportion of people did the task without making mistakes (otherwise the task would have been unrealistic) and therefore the distribution is skewed. The low interaction distribution does not result significantly skewed or with kurtosis, however the skewness value is quite high, and it fails the normality test.

Skewness and Kurtosis results for the incorrect answers distributions	**LOW**	**MEDIUM**	**HIGH**
Skewness	.774	.489	1.403
Std. Error of Skewness	.536	.536	.536
Skewness/ Std. Error of Skewness	**1.444**	**0.912**	**2.617**
Skewness significance	**$p > 0.05$**	**$p > 0.05$**	**$p < 0.01$**
Kurtosis	-.567	-.592	2.424
Std. Error of Kurtosis	1.038	1.038	1.038
Kurtosis/ Std. Error of Kurtosis	**-0.546**	**-0.570**	**2.335**
Kurtosis significance	**$p > 0.05$**	**$p > 0.05$**	**$p < 0.05$**

Table 9.11: Skewness and kurtosis results for the incorrect answers distributions

Levene's test on efficiency results	Significance	Result
F(2, 51) = 2.511	p > 0.05	Variance is homogenous

Table 9.12: Levene's test on effectiveness results

From these tests we can see that, although the scores present homogeneity of variance, not all of them are distributed normally and therefore non-parametric tests will be calculated to verify the significance of the average results in the three conditions. In this experiment, no significant difference in effectiveness was found between the three conditions.

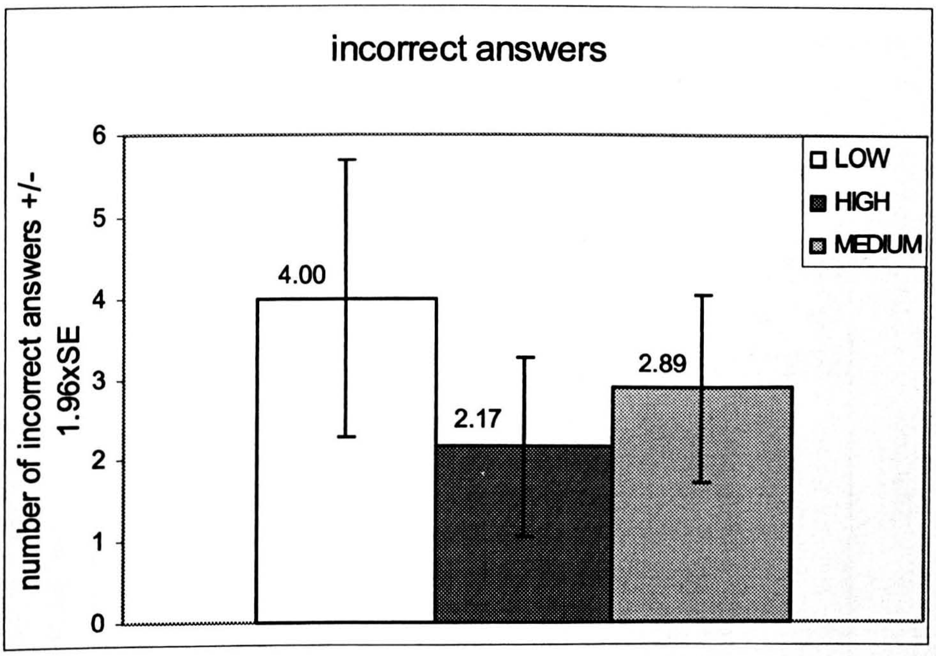


Figure 9.21: Effectiveness results

Statistical tests for effectiveness results		
Test: Friedman's ANOVA	**Significance**	**Effect size**
Chi-Square(2) = 8.035	p < 0.05	effect size = 0.26 (medium)

Table 9.13: Statistical tests for effectiveness results

Although the Friedman's ANOVA test is significant, the statistical power is too low to detect a medium effect size (see Appendix B) and to consider the results statistically significant.

9.5.4 Verification of the equivalence of the data sets used in this test

In this experiment three different, but similarly, constructed data sets were used. This was due to the fact that each subject needed to repeat the experiment three times (each one using a different interaction method) and therefore needed to analyse a different data set each time (otherwise the subject would know the answers after having done the test the first time).

The effectiveness and efficiency tests described above are based on the assumption that the three data sets used in the test are equivalent to analyse, i.e. no data set is easier to analyse than another data set. This assumption can also be verified using the test results. We can group the test scores for each data set and verify that the differences in average score between the different data sets are not significant, i.e. no data set is easier to analyse than the others.

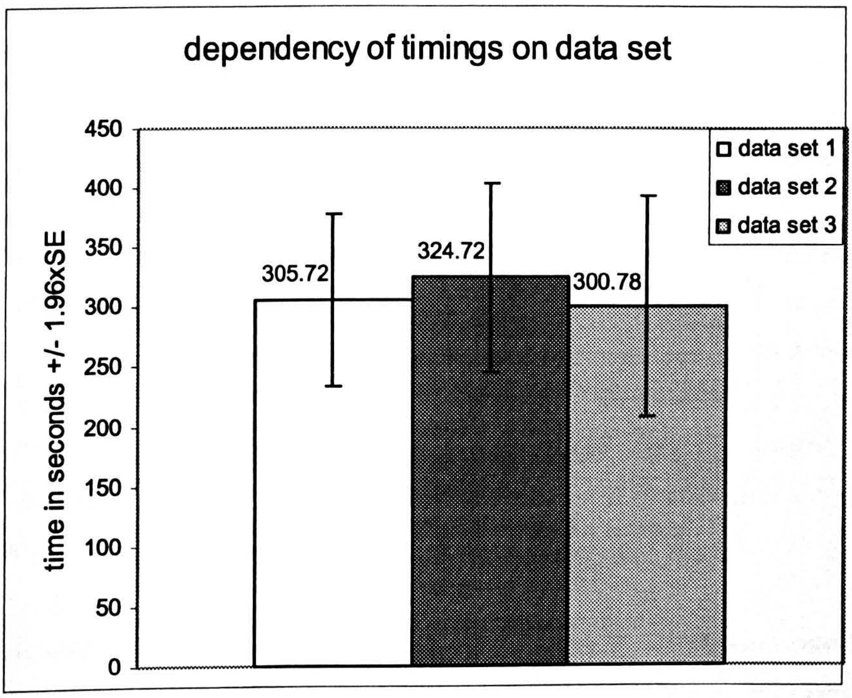


Figure 9.22: Dependency of timings on data sets

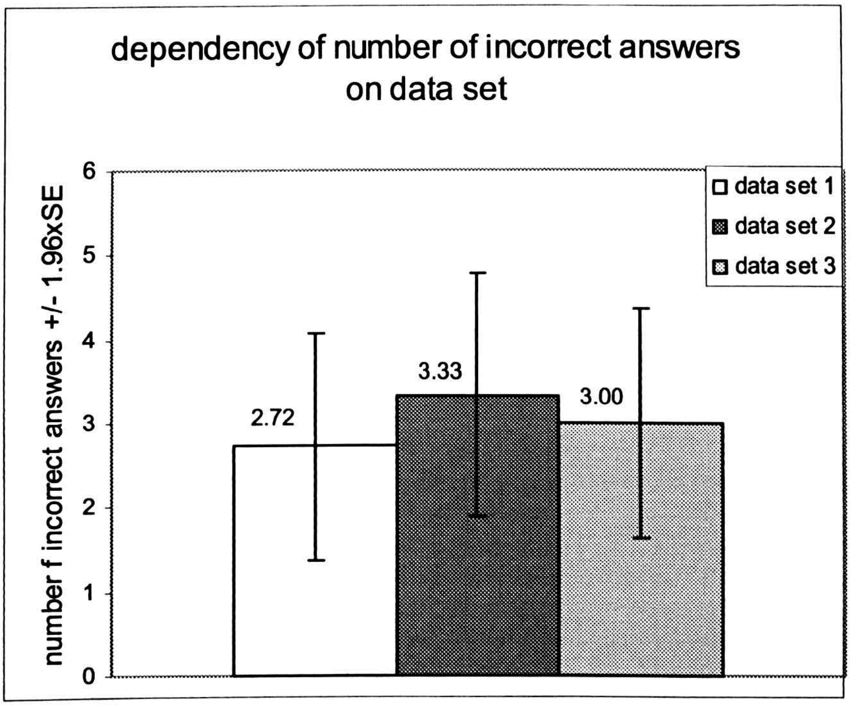


Figure 9.23: Dependency of number of incorrect answers on data sets

Statistical tests for the dependency of timings and incorrect answers on the particular data set			
Friedman's ANOVA test	Dependency of timings on data sets	**Friedman's ANOVA test**	Dependency of number of incorrect answers on data sets
Chi-Square(2) = 1.778	p > 0.05, not significant	Chi-Square(2) = 0.667	p > 0.05, not significant

Table 9.14: Statistical tests for the dependency of timings and incorrect answers on the particular data set

These tests confirm that there are no significant differences between the results obtained using the first, the second or the third data set, therefore the data sets used in this test are equivalent.

9.5.5 Effect of task's repetition on efficiency and effectiveness

From this experiment results, we can also verify if there is any significant difference in efficiency and effectiveness between the results obtained the first time the subjects did the experiment, the second and third time.

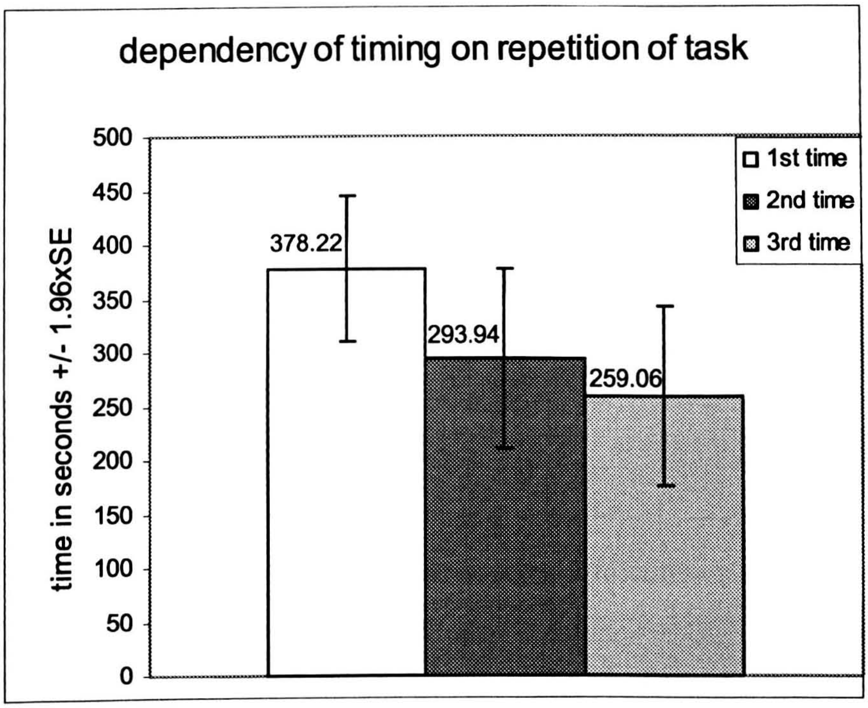


Figure 9.24: Dependency of timings on task's repetition

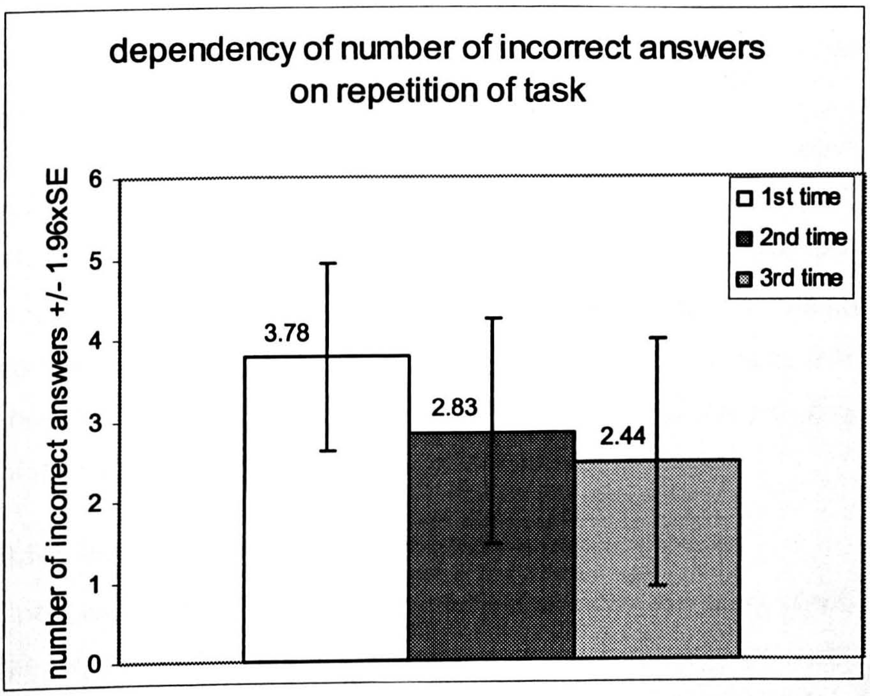


Figure 9.25: Dependency of timings on task's repetition

Statistical tests for the dependency of timings and incorrect answers on task's repetition				
Friedman's ANOVA test	Dependency of timings on task's repetition		**Friedman's ANOVA test**	Dependency of number of incorrect answers on task's repetition
Chi-Square(2) = 10.111	$p < 0.05$, effect size = 0.29 (medium)		Chi-Square(2) = 6.351	$p < 0.5$, effect size = 0.19 (small)

Table 9.15: Statistical tests for the dependency of timings and incorrect answers on task's repetition

No significant differences in efficiency and effectiveness were found between the average results obtained the first time the task was performed, the second and the third. Altough the significance value of the Friedman's ANOVA tests are below the 5% level, in both cases the effect sizes are too small to be considered relevant.

9.5.6 Questionnaire results

In the questionnaire, the subjects were asked to score each interaction method for pleasantness, intuitiveness, clarity and quickness.

Finally they were also asked to express explicitly which interaction method they preferred.

The results obtained from the questionnaire tell us how the different interaction methods were perceived subjectively. In particular, the results for quickness can tell us about the perceived efficiency of the interaction methods. The results for intuitiveness and clarity can tell us about the perceived effectiveness of the interaction methods.

9.5.7 User satisfaction: pleasantness

The Medium interaction method is found to be the most pleasant followed by the High interaction method and then the Low interaction method. This means that the subjects liked using the shuttle wheel which allows changing speed and direction of playback quickly and has a set number of constant playback speeds. To use the mouse with a direct mapping between speed of movement of the mouse and playback speed was judged to be the second most pleasant method and to select sections of the sound, select the speed of playback and press play was considered the least pleasant method of interaction.

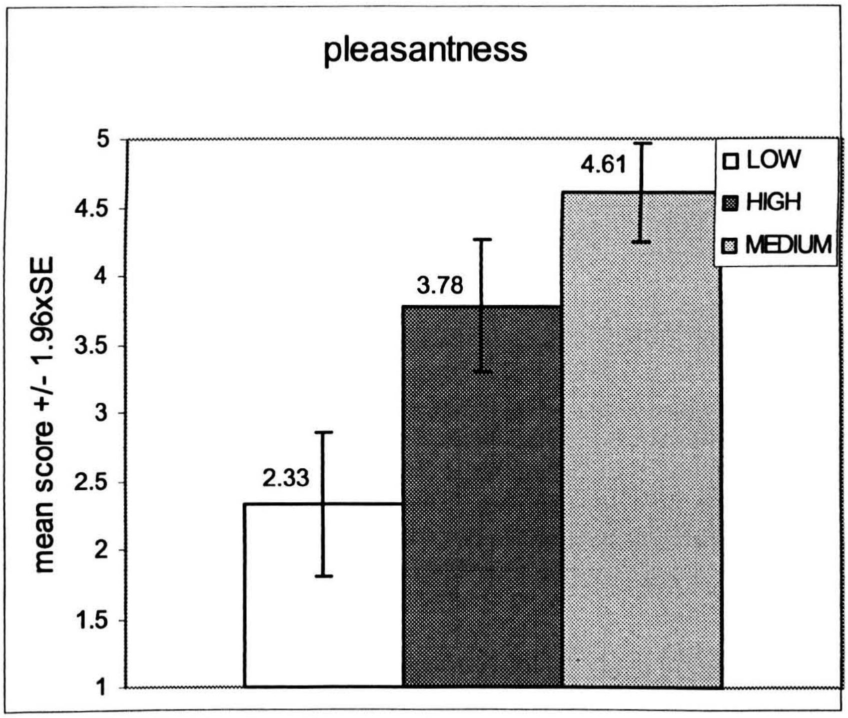


Figure 9.26: Pleasantness results

Statistical tests for the pleasantness results		
Test: Friedman's ANOVA	**Significance**	**Effect size**
Chi-Square(2) = 18.781	$p < 0.05$	effect size = 0.94 (large), statistical power > 80%
Post-hoc comparisons: Wilcoxon signed rank test	**Significance**	**Effect size**
Low-High: Z = -2.716	p(one-tailed) < 0.167	effect size = 1.32 (large)
Low-Medium: Z = -3.528	p(one-tailed) < 0.167	effect size = 2.35 (large)
Medium-High Z = -2.284	p(one-tailed) < 0.167	effect size = 0.89 (large)

Table 9.16: Statistical tests for the pleasantness results

9.5.8 Intuitiveness and clarity: perceived effectiveness

The results for intuitiveness and clarity tell us that selecting a section of audio, choose the playback speed and press play (Low interaction method) is considered significantly less intuitive and clear, than to explore the sonification using either the mouse or the shuttle interface. Differences between these last two methods are not significant.

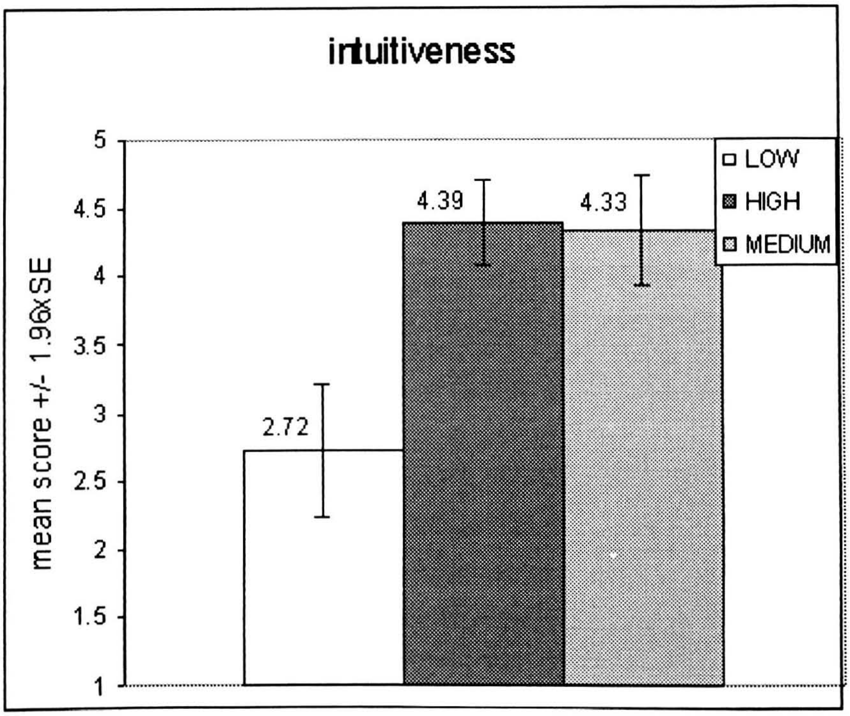


Figure 9.27: Intuitiveness results

Statistical tests for intuitiveness		
Test: Friedman's ANOVA	**Significance**	**Effect size**
Chi-Square(2) = 14.533	p < 0.05	effect size = 0.88 (large), statistical power > 80%
Post-hoc comparisons: Wilcoxon signed rank test	**Significance**	**Effect size**
Low-High: Z = -3.223	p(one-tailed) < 0.167	effect size = 1.85 (large)
Low-Medium: Z = -3.054	p(one-tailed) < 0.167	effect size = 1.68 (large)
Medium-High Z = -0.263	p > 0.05	Not significant

Table 9.17: Statistical tests for intuitiveness

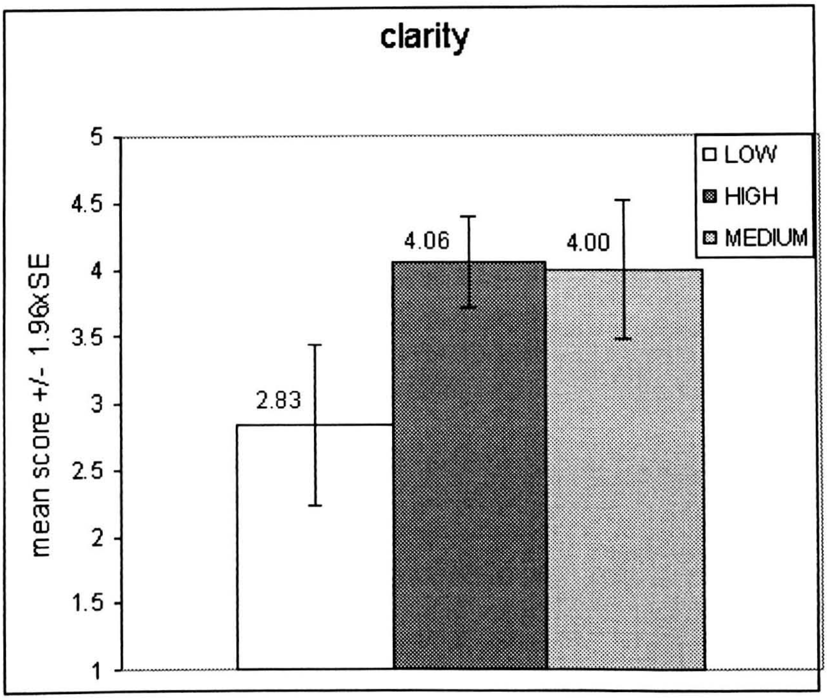


Figure 9.28: Clarity results

Statistical tests for clarity		
Test: Friedman's ANOVA	**Significance**	**Effect size**
Chi-Square(2) = 6.758	$p < 0.05$	effect size = 0.52 (large), statistical power > 80%
Post-hoc comparisons: Wilcoxon signed rank test	**Significance**	**Effect size**
Low-High: Z = -2.726	p(one-tailed) < 0.167	effect size = 1.16 (large)
Low-Medium: Z = -2.237	p(one-tailed) < 0.167	effect size = 0.96 (large)
Medium-High Z = -0.206	$p > 0.05$	Not significant

Table 9.18: Statistical tests for clarity

9.5.9 Quickness: perceived efficiency

The Low interaction method is also perceived to be significantly slower than the other interaction methods. There is no significant difference in quickness between the Medium and High interaction method.

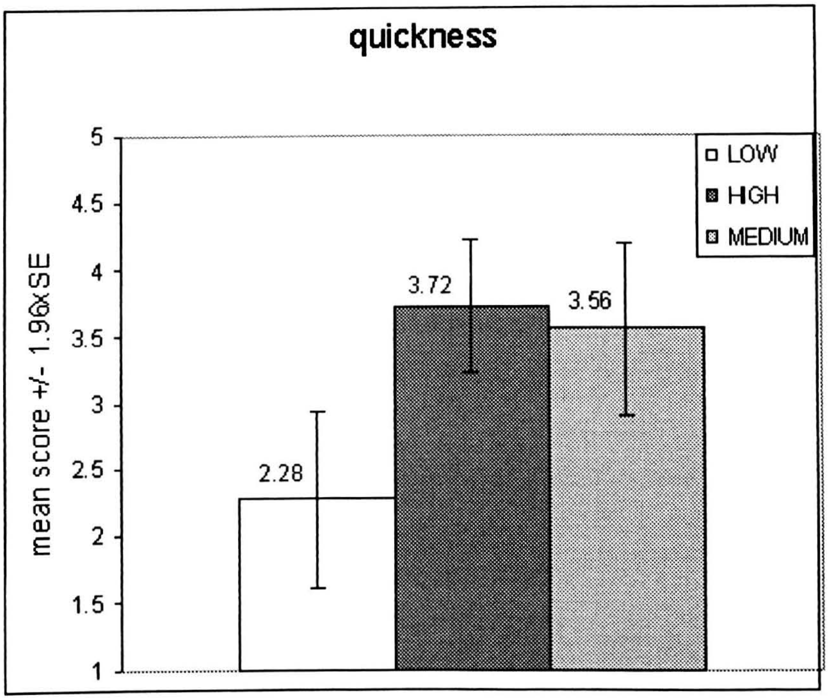


Figure 9.29: Quickness results

Statistical tests for quickness		
Test: Friedman's ANOVA	**Significance**	**Effect size**
Chi-Square(2) = 6.738	$p < 0.05$	effect size = 0.49 (large), statistical power > 80%
Post-hoc comparisons: Wilcoxon signed rank test	**Significance**	**Effect size**
Low-High: Z = -2.234	p(one-tailed) < 0.167	effect size = 1.13 (large)
Low-Medium: Z = -2.084	p(one-tailed) < 0.167	effect size = 0.89 (large)
Medium-High Z = -0.293	$p > 0.05$	Not significant

Table 9.19: Statistical tests for quickness

9.5.10 User satisfaction: preferred interaction method

The 18 test subjects were asked directly in the questionnaire which interaction method did they prefer:

- 4 out of 18 (22%) said the Low interaction method, i.e. selecting a section and a playback speed and press play;
- 6 out of 18 (33%) said the High interaction method, i.e. play the sonification by moving the mouse backward and forward in the sound;
- 8 out of 18 (45%) said the Medium interaction method, i.e. play the sonification using the shuttle interface.

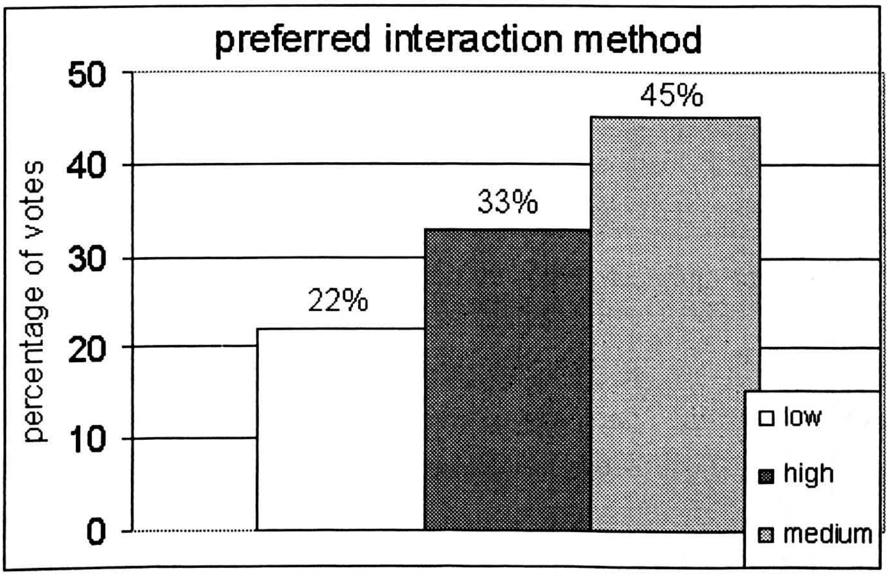


Figure 9.30: Preferred interaction method

9.5.11 User satisfaction: summary of comments

At the end of the test, each subject was asked to freely comment on the three proposed interaction methods. Below is a summary of the main points expressed in the comments. The numbers in brackets indicate the number of times the same comment was made by different people.

User satisfaction: comments	**Advantages**	**Disadvantages**
Low interaction method	• It allows better focus (2) and better understanding of a specific region of data; • It is a similar method to the one used in sound editing software, which is very familiar to me; • It allows more precision; • The temporal zoom can be very slow and therefore the sonification can be understood without need of repetition; • It allows constant speeds which is comforting.	• It is very frustrating and hard to use
Medium interaction method	• It has more functionality: the user is allowed to stop at any point, to restart with a single click and to browse slowly or quickly (2); • The shuttle can be used with the left hand and the right hand can be used to write (or do something else); • It feels professional and appropriate; • It is nice to use; • It allows a constant playback speed; • It allows fine control at low speed; • It gives the best sense of control of playback and its speed; • It is intuitive and simpler; • The user does not need to look at the screen and therefore he/she can concentrate more on the sound; • The hand can be off the interface device (hands free) without losing sense of where you are (2).	• It is difficult to match the Medium interaction and the graphics; • The playback speeds are fixed.
High interaction method	• The direct relationship between hand position and speed, and playback position and speed (3); • It allows total and quick control; • It allows for going straight to a certain point in the sound; • It is more intuitive and familiar; • Results are obtained more easily.	• None

Table 9.20: User satisfaction: comments

Two people indicated a second preferred choice:

- The High interaction: because I use the mouse all the time;
- The High interaction: it is familiar and it is fast.

Two people preferred equally two interaction methods:

- High and medium interaction methods (2 people).

One person suggested that a new interface device could combine the advantages of the interaction methods. This device should combine the medium and high interaction methods and be a combination of the shuttle interface and the mouse.

9.5.12 Summary of questionnaire results

The medium interaction method is judged to be the most pleasant and it is the preferred interaction method by the subjects. The High interaction method is the second best method and for intuitiveness, clarity and quickness it scores as well as the Medium interaction method. The Low interaction method is the worst in all the questionnaire results.

The Medium interaction method is considered appropriate, "professional and nice". It has various functionalities (different fixed speeds, back and forward playback, "return to zero" button, fine control of slow speed, etc.) and works often hands free (i.e. hands can be used to do other things while listening to the display). It is not necessary to look at the screen area to know where one is in the data at any given time. This last aspect is at times considered confusing and can become also a disadvantage. The use of fixed playback speeds is both considered comforting, but at times restrictive.

The High interaction method is considered to give total control because there is a direct relationship between interface device position and speed, and the playback position and speed. With this method, changes in position are immediate. The device used is very familiar and results can be obtained very quickly.

9.6 Summary of overall results

The clearest result of this experiment is that the Low interaction method is considered the worst under all aspects. The High and Medium interaction

methods are considered better methods to use for navigating a sonification and in particular the Medium interaction method is considered significantly more pleasant and was the preferred method of this group of subjects, i.e. it is the most satisfying of the methods.
The main result from the objective measurements of efficiency and effectiveness is that the Low interaction method is slower, i.e. less efficient, than the other methods. The Low interaction method requires the user to perform many actions (select with the mouse, enter the chosen playback speed as a number, press play, etc.) before he/she can listen to the sonification. This procedure is time consuming making this method slower. The interaction methods were not significantly different in terms of effectiveness, i.e. the analysis of the sonification can be done equally well using all the methods.
The interaction methods reveal more clear differences when we look at how they are perceived by the user during the task. The Medium and High interaction methods are perceived to be more efficient and effective than the Low interaction method as they score significantly higher for quickness, clarity and intuitiveness. Therefore, although from the objective measurements of effectiveness no difference was found between the methods, the subjects perceive the Low interaction method as significantly less effective than the other two. Finally, subjects indicated in this experiment that, even if there are no particular objective differences between the Medium and High interaction methods, they prefer and find more pleasant the Medium interaction method which uses the shuttle interface. The reason for this preference could be the fact that the Medium interaction method provides, at the same time, very quick changes in playback speeds and direction, while allowing eyes and hand free moments in which the user can concentrate solely on the sound. The Medium interaction method does not require constant activity from the user (the High interaction method does) allowing to shift the attention rapidly between different tasks such as restart the sound, change playback speed, listen to the sound and analyse the data.
From this experiment we can conclude that the addition of a relatively high level of interaction to a sonification display improves the efficiency of the analysis of the sonified data. From the point of view of users' perception, the addition of a relatively high level of interaction greatly improves the overall auditory display which is then perceived as more pleasant, clear, intuitive and quick to use. Interesting ideas for further work would be to explore in detail the reasons for

the subjects' preference of the Medium interaction method and develop and test a hybrid interface device for interactive sonification which groups the qualities of the mouse and the shuttle interface.

9.7 Conclusions

The focus of this Chapter is the interaction between the user and the sonification display. Firstly, the main issues surrounding the design and the subsequent evaluation of the interaction were described. Then, on the basis of principles derived by the literature on interaction design, an experiment, which is here described, was conducted to evaluate three different interaction methods for navigating a sonification. Finally, the quantitative and qualitative results from this experiment show that adding a relatively high level of interaction to a sonification display of a generic multivariate large data set improves the efficiency and the effectiveness of the display as well as the user satisfaction in using the display.

Chapter 10 – Summary and conclusions

10.1 Introduction

The following hypothesis was investigated within this thesis:

> *The combination of user interaction and sonification allows the display of multivariate time-based large data sets effectively and efficiently.*

The above hypothesis states that data sets containing many variables varying simultaneously (multivariate), each containing thousands of data points (large), which have been sampled in time (time-based), can be displayed effectively, i.e. their main characteristics are clearly shown, and efficiently, i.e. the information is communicated rapidly, using a display that combines sound and user interaction.

At the centre of this research are three experiments which were carried out in order to investigate the above hypothesis. The first two experiments are concerned with the sonification of real-world time-based multivariate large data sets and its effectiveness as a data display when compared with traditional displays. Specifically, in the first experiment, test subjects were presented with the audifications and the spectra of data from the field of helicopter engineering. The audifications allowed presenting the data from a thirty minute flight in two and a half seconds. Test subjects were asked to score audifications and spectra, presented in random order, in a scale from one to five for important characteristics. The analysis of the results revealed how effective the sonification display was in comparison to the visual display (spectra) in portraying the principal characteristics of the data.

In the second experiment, test subjects were presented with a sonification of EMG (electromyography) data which portrayed six channels of samples simultaneously. Test subjects were asked to score the sonifications in a scale from one to five for some relevant characteristics. The analysis of the results showed how effective the sonification display was when compared with results obtained with traditional signal processing analysis.

In both experiments, test subjects were asked to provide comments on the sonification displays. These comments provided additional qualitative information on the perceived effectiveness of these displays.

The main objective of these experiments is to provide experimental evidence that a sonification display of large multivariate time-based data sets can be effective, where the effectiveness of the display is measured in comparison to other well established displays which are considered effective.

The third experiment seeks to evaluate whether adding a high level of interaction to a sonification display of multivariate large data sets can improve the efficiency and effectiveness of the display. A prototype of an interactive sonification application, programmed by the author, the Interactive Sonification Toolkit (see Chapter 6 for the detailed description of the toolkit) was used in this experiment. Three different methods for interactively exploring a sonification display were compared. These different methods differ from one another principally in the degree of control allowed during the interaction. The data to sonify in this experiment were synthesised by the author. The test subjects were trained to use and interpret the sonification utilised in this experiment so that only the interaction method could be considered the independent variable of the experiment. The methods were compared on the basis of their objective and perceived efficiency and effectiveness. The test subjects were also asked to provide qualitative comments on the three interaction methods in order to gather some information about user satisfaction.

The specific hypothesis of this third experiment is that a sonification display that allows a higher level of interaction with the sound, and therefore with the data, is more efficient and effective in portraying the data's information than a display with a lower level of interaction.

10.2 Summary of results

In the first experiment, described in Chapter 7 of this thesis, audifications of measurements of twenty eight helicopter flight's parameters were compared to the spectra of the same data sets. The test subjects were asked to score each audification and spectrum, which were presented in a random order, on a scale from one to five for the following characteristics: presence of noise, presence of distinct frequencies, presence of discontinuities, presence of repetitive elements and signal power. Spearman's correlations were calculated to analyse the relationship between average auditory results and average visual results. Significant one-tailed correlations were found for all five characteristics. This result demonstrates that the auditory (audification) display portrays information similarly to the visual display (spectrogram) despite the fact that the audification of these data produces highly complex and noisy sounds. In the qualitative comments some subjects found that some characteristics were more difficult to score than otehrs and that the sound provided more detail about the data and in particular that noise, as a sound, presents more detail than noise as a visual element.

In the second experiment, described in Chapter 8 of this thesis, EMG data are sonified using a sound synthesis method based on amplitude modulation. The data were provided by Dr John Dixon physiologist from Teesside University. The EMG data were recorded while subjects performed a Maximal Voluntary Isometric Contraction (MVIC) with their leg. This type of data can generally be described as being noisy with a simple amplitude envelope that includes an attack, a sustained and a release part. Dr Dixon reported that previous analysis on this type of EMG data based on signal processing techniques show that the signal amplitude and the speed of the signal's attack depend on the age of the subject performing the MVIC: the higher the age, the lower the amplitude of the signal and the speed of the signal's attack. In the experiment described on this thesis, the sonification of EMG data were presented to listening subjects who were asked to score, on a scale from one to five, the overall loudness of the sonifications and the attack's speed. One-tailed significant Spearman's negative correlations, between the average scores for these characteristics and the age of the subjects performing the MVIC, were found. This result confirms that the sonification display portrays the general trend of these two characteristics as well as the signal processing techniques.

In the field of physiotherapy, the communication between physiotherapist and patient is an important issue. The physiotherapist needs to find a way to communicate to the patient various issues in a simple, non technical fashion. As part of this experiment, the author investigated if the descriptor of sound's timbre *roughness* could be used in this sonification to communicate to the patient the main characteristics of his/her muscle activity. The subjects of this experiment were asked to score on a scale from one to five the roughness of the sonification. One-tailed significant Spearman negative correlation was found between the age of the subject performing the MVIC and the average score of the sonifications.

This result tells us that the older the subject performing the MVIC action, the rougher the sound of the sonification. This result indicates a trend: when using this type of sonification the roughness of the sound can indicate how young the subject performing the action is.

This type of sonification can be produced in real-time, as the subject performs the action (a description of an example of real-time set up can be found in Chapter 8, Section 8.12). Using this real-time set up, given this sonification and given the results of this experiment, the physiotherapist should expect that when he/she asks a subject to try to produce a rougher sound, a younger, healthier movement will be produced by the subject under investigation.

The EMG data sets used in this experiment were also divided in two groups: asymptomatic subjects and symptomatic subjects with osteoarthritis of the knee (OA patients). Statistical tests were carried out to find out if the scores for the above mentioned characteristics of the sonifications were significantly different for these two groups. These differences were found not to be significant.

Finally, in this experiment, a questionnaire was submitted to the listening subjects to find out if they considered the sonification to be appropriate as auditory metaphor of muscle activity. The sonification was found to be appropriate in portraying a general gesture and therefore also a muscle movement. The sound, however, was found to be quite fatiguing suggesting that the sonification algorithm should be improved in order to create a better display.

In the third experiment, described in chapter 9 of this thesis, three methods that allow navigating the sonification were compared. These three methods differ for their allowed level of interaction and they are called: Low Interaction Method, Medium Interaction Method and High Interaction Method. Using the Low Interaction Method an analyst can select, using the mouse, a section of sonification to listen and select at which speed to hear this selection when played back. Using Medium Interaction Method, the analyst employs the jog wheel/shuttle interface [Contour Design] to navigate the sonification instantly changing between different speeds and direction (backward or forward). Using the High Interaction Method, the analyst navigates the sonification by moving the mouse over the screen area that represents the sonification. The speed of navigation directly maps the speed of movement of the mouse and direction changes instantly.

The data used in this experiment was synthesised by the author and it contained five types of data structures that are commonly found in data sets. The data structures were noise, a periodic wave form, a constant, a ramp and randomly placed discontinuities. The sonification used was simple and well understood by the listening subjects. Three data sets were built by combining in different orders the five chosen data structures. The data sets contained two channels of data each and the sequence of structures present in one channel differed from the sequence present in the second channel. The sonification of the two channels was panned so that channel one was heard in the left ear and channel two was heard in the right ear. The listening subjects were asked to navigate a data set three times: each time using a different interaction method and data set. They were then asked to write down the sequences of structures recognised in the data sets. The way the data sets were built and the panning configuration allowed to produce a test in which, even if the sonification in itself was not difficult to interpret and therefore did not constitute a variable in the experiment, the recognition of the structures required multiple navigations because it was difficult to attend simultaneously to what was heard in both ears.

Two objective measures of efficiency and effectiveness were measured: the time spent to recognise all the structures (the shorter the time the higher the efficiency of the interaction method) and the number of errors in recognising the structures (the lower the number of errors the higher the effectiveness of the interaction method). The perceived effectiveness and efficiency of the three methods were explored by asking the listening subjects to score, from one to five, each interaction method for intuitiveness, quickness and clarity. User satisfaction was explored in various ways: by asking the listeners to score on a scale from one to five the pleasantness of the interaction method, by asking which method was preferred and by collecting free comments from the test subjects.

The results show that the Low Interaction Method is just significantly less efficient than the High Interaction Method. No Significant differences could be found between Low Interaction Method, Medium Interaction Method and High Interaction Method for effectiveness. The results for the perceived characteristics are clearer. The Low Interaction Method is considered significantly less pleasant, intuitive, quick and clear than the Medium Interaction Method and High Interaction Method. There is no significant difference between the High Interaction Method and Medium Interaction Method methods for quickness, intuitiveness and clarity, while the Medium Interaction Method is considered more pleasant than the High Interaction Method. The subjects were also asked which method they preferred and the Medium Interaction Method turned out to be the preferred method.

The overall result of this experiment tell us that the two methods (Medium Interaction Method and High Interaction Method) that allow a higher level of interaction are significantly better than the method with low degree of interactivity (Low Interaction Method). Overall the Medium Interaction Method and High Interaction Method methods score similarly in both objective and perceived characteristics, however the Medium Interaction Method was thought to be significantly more pleasant than both the other two methods and it was voted the preferred method.

10.3 Discussion

This research has found that multivariate time-based large data sets can be displayed using sonification in an effective way, i.e. in a way that is at least as effective as alternative and well established displays. It was also found that the effectiveness, efficiency and user satisfaction of the display increase when the user is allowed to navigate the sonification with a high level of interaction.

The importance of the results obtained in this research is twofold. The results give us both general information about auditory displays of large data sets and about the specific data domains explored in this project. The following characteristics of this project contribute towards making its results generally applicable.

The sonification algorithms do not depend on the type of data used and can be interpreted easily. The data sets used in this research are highly complex and contain a variety of different data structures: from different types of noise, to amplitude discontinuities, from variable frequency periodicities to slow overall changes in amplitude. These data sets cannot be said to represent every possible large data set, however they are derived from real research domains and are highly varied.

The characteristics tested in the experiments are also very general as they are characteristics that can be found in many data structures. On the other hand, our interpretation of the experimental results in the context of the specific field domain makes the tested characteristics particularly meaningful for the specific domain (i.e. physiotherapy or helicopter engineering).

In the third experiment, the interaction methods which have been compared are independent of the type of sonification explored. The experimental task (recognition of data structures) is a fundamental function that any display representing data must facilitate. This experiment was made explicitly independent from the type of sonification explored so that its results are general for any auditory display.

The results of this research also have a more specific importance. They provide new knowledge in the field of helicopter flight analysis and EMG analysis in physiotherapy. The main result for the field of helicopter engineering is that through audification general characteristics of flight data can be displayed using very short sound files. This is encouraging and creates the basis for investigating further how sound can be used to solve the problem of how to make data analysis in this field more efficient and effective.

The main result in the field of EMG analysis for physiotherapy is that sound can be used to display general characteristics of EMG data in real-time. This result allows us to progress with designing more precise auditory displays for this type of data, which can provide meaningful feedback not only for the analyst and the physiotherapist, but also for the patient. Such a tool (real-time auditory feedback of EMG data) could both be a diagnostic tool and a rehabilitation tool. A research grant proposal to further investigate this particular result was submitted to EPSRC in February 2007. This new project is a follow up project from the EPSRC project "Data mining through an interactive sonic approach" which funded most of the research presented in this thesis. Collaborators for this new grant will be Dr Andy Hunt from York University, Prof. Keith Rome and Dr John Dixon from Teesside University and Sandra Pauletto from Huddersfield University.

The work presented in this thesis is highly original. Partially the originality derives from the fact that the field of auditory display is quite new and therefore few data domains have been portrayed as sound. To the best of the author's knowledge there has been no other research on the sonification of EMG data and this type of helicopter flight data. The design of the experiments (and specifically which characteristics of data have been evaluated, which displays have been compared, etc.) is highly original, and there are no examples of structured methodologies for designing experiments that evaluate the display of complex data through complex sounds. The experiments designed in this thesis, as mentioned before, are generally applicable and could be used to evaluate other methods of sonification for large data sets.

The final experiment is also highly original. There exists no standard way to evaluate the effect of using interaction in an auditory display of large data sets. The experiment described here, and its overall logic, could be used to evaluate other interaction methods. It could also be expanded to include other general data structures and tasks. This research, therefore, provides structured experimental designs that can be used in other research.

10.4 Further work

Listed here are some of the most important and immediate developments that could stem from this research.

1) More experiments could be done to explore the effectiveness of displaying large data sets through audification in comparison to using spectra. Examples of issues to investigate are:
    - which structures are better displayed with sound than with a spectrogram;
    - which are the most appropriate time and frequency ranges for the display of specific characteristics;
    - how well short audified data channels can be compared to each other when compared sequentially or simultaneously;
    - what the advantages and disadvantages are of using a multimodal display that uses audification, spectra and interactive navigation.

2) Concerning the use of amplitude modulation for the sonification of several channels, further experiments can be done to find out:
    - to extent to which the separate characteristics of the different channels can be heard;

- how many channels can be sonified together while still maintaining a reasonable resolution;
- which characteristics are better displayed with sound than visually.

3) Further work can be done in the investigation of EMG sonification to produce a real-time auditory display to use as a diagnostic tool and a rehabilitation tool. This is the focus of the new collaborative research proposal submitted to the EPSRC and outlined above. In this context different sonification algorithms will be developed and tuned, and the displays of various EMG data (coming from various muscles, subjects and patients) will be compared. We aim to develop a portable tool which will incorporate sensors, analysis and adjustable sonic feedback.

4) In the field of helicopter engineering, the analysis of the audification of different helicopter flights could be compared to discover if the sound of a standard flight, in which no problems were experienced, can be clearly defined and easily separated from the sounds of a flight where problems have occurred. If this can be done, it would then be important to investigate if there is a correlation between particular sounds in the audification of flights and specific problems which occurred during the flight.

5) Concerning interaction, a variety of different interaction methods could be evaluated, coupled with various common analysis tasks (for example comparing data streams, recognising a particular pattern in different data streams, etc.) and specific sonifications (such as audification or a sonification that uses AM synthesis).

10.5 Final summary

In conclusion, this thesis presents new research on the usability (efficiency, effectiveness and user satisfaction) of interactive non-speech auditory displays of large multivariate time-based data sets. This is a multidisciplinary subject which requires the study and understanding of issues from several disciplines, e.g. sound perception, sound synthesis, human computer interaction, etc. Chapters 2 to 4 introduce and summarise the main background knowledge and recent research from the disciplines related to this work. Chapter 5 clarifies the research approach used in this thesis and summarises the main statistical concepts used in this work. Chapters 6 to 9 present the prototype application and experimental work produced during this research.

At the heart of this work are three experiments and the combination of their results gives us new knowledge about the usability of interactive auditory displays of large data sets. Experiments one and two use real world data from helicopter flight analysis and physiotherapy (EMG data). In experiment one, the audification technique is used to portray each channel of data as a 2.5s sound file. This display portrays basic information about the data well when compared with spectra. This result tells us that using sound to analyse this type of data is a feasible option. It also tells us more generally that audification is an effective way of portraying some basic structures typically present in large data sets.

In experiment two, amplitude modulation is used to sonify six channels of EMG data. The results show that this display method is effective in displaying characteristics of the data. Other advantages of the use of this sonification technique are that the sound produced is quite appropriate in timbre for representing muscle activity and that it can be used to provide real-time auditory feedback because the original real-time data update rate can be maintained during sonification.

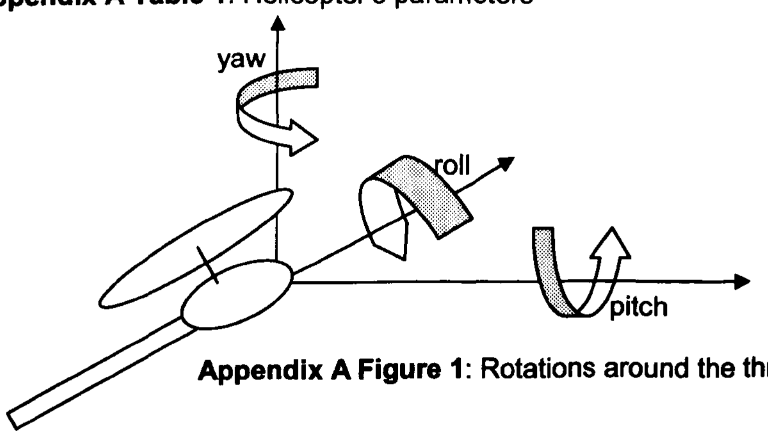
The third experiment compares three methods of interacting with a sonification of a large data set. These three methods differ from one another in the level of user control they allow during the interaction. The results show that the addition of a relatively high level of user interaction improves the efficiency, perceived effectiveness and user satisfaction (i.e. the usability) of the overall display.

In summary, we have learnt that human beings can indeed extract useful information from large multivariate time-based data sets portrayed as sound, and that a high degree of interaction with the playback of these sounds improves the effectiveness and efficiency of the display.

Appendix A - Helicopter's data set: channels and parameters

Helicopter's parameters			
channel	**parameter**	**channel**	**parameter**
1	Engine 1 Torque	15	Main rotor speed
2	Engine 2 Torque	16	Outside air temperature
3	Collective control lever	17	Pitch rate
4	Cyclic stick forward angle	18	Pitch attitude
5	Cyclic stick lateral angle	19	Altitude
6	Servo	20	Rotation attitude
7	Pedal	21	Rotation rate
8	Barometric altitude	22	Vertical acceleration
9	Air speed	23	Yaw rate
10	Velocity forward	24	Power turbo speed engine 1
11	Velocity lateral	25	Power turbo speed engine 2
12	Acceleration forward	26	Lever servo
13	Acceleration lateral	27	Stick servo forward
14	Magnetic heading	28	Stick servo lateral

Appendix A Table 1: Helicopter's parameters



Appendix A Figure 1: Rotations around the three orthogonal axes

Appendix B – A priori power analysis tests for the experiments of this thesis

Experiment one

Significance of means

28 groups, 18 subjects per group (504 subjects in total)

Effect size	Alpha (significance)	Power	Number of groups	Total number of observations
Small (f = 0.10)	0.05	0.8	28	2380
Medium (f = 0.25)	0.05	0.8	28	420
Large (f = 0.40)	0.05	0.8	28	196

Appendix B Table 1: A priori power analysis test for the significance of means in Experiment one

With 504 total number of subjects, this experiment has the power to detect large and medium effects.

Correlations

28 samples per correlation

Effect size	Alpha (significance)	Power	Sample size
Small (r = 0.1)	0.05	0.8	614
Medium (r = 0.3)	0.05	0.8	64
Large (r = 0.5)	0.05	0.8	21

Appendix B Table 2: A priori power analysis test for the correlations in Experiment one

With 28 samples, this experiment has the power to detect large effects.

Experiment two

Significance of means

30 groups, 21 subjects per group (630 total number of observations)

Effect size	Alpha (significance)	Power	Number of groups	Total number of observations
Small (f = 0.1)	0.05	0.8	30	2460
Medium (f = 0.25)	0.05	0.8	30	420
Large (f = 0.4)	0.05	0.8	30	180

Appendix B Table 3: A priori power analysis test for the significance of means in Experiment two

With 630 total number of observations, the experiment has the power to detect medium and large effects.

Correlations

30 samples per correlation

Effect size	Alpha	Power	Number of samples
Small (r = 0.1)	0.05	0.8	614
Medium (r = 0.3)	0.05	0.8	64
Large (r = 0.5)	0.05	0.8	21

Appendix B Table 4: A priori power analysis test for the correlations in Experiment two

With 30 means, the experiment has the power to detect large effects.

Comparison young/old

2 groups, 12 subjects (EMG subjects) in one group, 18 subjects (EMG subjects) in the second group

Effect size	Alpha (significance)	Power	Number of groups	Total number of observations
Small (d = 0.2)	0.05	0.8	2	620
Medium (d = 0.50)	0.05	0.8	2	102
Large (d = 0.80)	0.05	0.8	2	42
Very large (d = 0.95)	0.05	0.8	2	30

Appendix B Table 5: A priori power analysis test for the significance of means (comparison between young and old groups) in Experiment two

With a total of 30 EMG subjects, the experiment can detect only a very large effect (d > 0.95).

Comparison old symptomatic/old asymptomatic

2 groups, 18 subjects (EMG subjects) in total

Effect size	Alpha (significance)	power	Number of groups	Total number of observations
Small (d = 0.2)	0.05	0.8	2	620
Medium (d = 0.5)	0.05	0.8	2	102
Large (d = 0.8) Very large (d = 1.23)	0.05 0.05	0.8 0.8	2 2	42 18

Appendix B Table 6: A priori power analysis test for the significance of means (comparison between symptomatic and asymptomatic groups) in Experiment two

With 18 observations in total, the experiment can detect only a very large effect.

Experiment three

Significance of means

3 groups, 18 subjects per group

Effect size	Alpha (significance)	Power	Number of groups	Total number of observations
Small (f = 0.1)	0.05	0.8	3	969
Medium (f = 0.25)	0.05	0.8	3	159
Large (f = 0.4) Larger (f = 0.44)	0.05 0.05	0.8 0.8	3 3	66 54

Appendix B Table 7: A priori power analysis test for the significance of means in Experiment three

With a total of 54 observations, the experiment have the power to detect large effects: higher than f = 0.44.

Post-hoc comparisons

2 groups, 36 observations

Effect size	Alpha (significance)	power	Number of groups	Total number of observations
Small (d = 0.2)	0.0167	0.8	2	884
Medium (d = 0.5)	0.0167	0.8	2	144
Large (d = 0.8)	0.0167	0.8	2	58
Very large (d = 1.03)	0.0167	0.8	2	36

Appendix B Table 8: A priori power analysis test for the significance of means (post-hoc comparisons) in Experiment three

With 36 total observations, the experiment has the power to detect very large effects: higher than d = 1.03.

Appendix C - Dissemination

The results of this thesis have been disseminated through the following publications, workshops participations and collaborations:

Hunt, A. and Pauletto, S., (2004) *Interactive sonification in physical rehabilitation* (presentation) EMusk Conference, School of Health of Teesside University, Middlesbrough

Pauletto S. & Hunt A., (2004) *Interactive Sonification of Helicopter flight data and muscle (EMG) data*, International Workshop on Interactive Sonification, Bielefeld, Germany

Hunt, A., Hermann, T. and Pauletto, S. (2004) *Interacting with sonification systems: closing the loop*, Proc. of the IEEE Conference on Information Visualisation, pp. 879- 884

Pauletto, S. and Hunt, A., (2004) *A Toolkit for Interactive Sonification*, Proc. International Conference on Auditory Display, Sydney, Australia

Pauletto, S. and Hunt, A. (2005) *A Comparison of Audio & Visual Analysis of Complex Time-Series Data Sets*, Proc. International Conference on Auditory Display, Limerick, Ireland

Principal Investigator: Dr Andy Hunt, Research Assistant: Sandra Pauletto, (2005) *Improved data mining through an interactive sonic approach*, Final detailed Report: EPSRC Grant award: GR/S08886/01

Pauletto, S., (2005) *Interaction in sonification*, presentation at the tutorial on *Interactive Sonification – Auditory Data Mining for Uncovering Structure in High-Dimensional Data*, International Conference on Auditory Displays, Limerick, Ireland

Pauletto, S., (2005) *Interactive Sonification as a tool for data exploration*, (seminar) Neuroinformatics Department, Bielefeld University, this seminar was given at the end of the collaboration with researcher T. Hermann which was funded by the COST-287 Congas European Project [Congas] under the Short Term Scientific Missions programme

Pauletto, S. (2005) *Short Term Scientific Mission report*, [Congas]

Pauletto, S. and Hunt, A. (2006) *The sonification of EMG data*, Proc. International Conference on Auditory Display, London, England

Collaboration with Irish composer Fergal Dowling [Dowling, 2006].
Dr Dowling composed an interactive piece using the audifications of the helicopter data produced in this research (see Chapter 7). The piece titled *I'm a helicopter* is part of the portfolio of compositions submitted by Dr. Dowling for his Doctorate obtained from York University in 2006. See Appendix D for a detailed description of the piece and see the accompanying DVD to listen to a performance of the piece.

The author was invited to participate at the *Science by Ear workshop* (March 2006) at The Institute of Electronic Music and Acoustics (IEM), Graz, Austria [see http://sonenvir.at/workshop/]

Pauletto, S. & Hunt, A., (2007) *Interacting with sonifications: an evaluation*, International Conference on Auditory Display, Montreal, Canada

Appendix D – *I'm a helicopter:* musical composition by Irish composer Fergal Dowling

Here follows a description provided by the composer of the musical composition *I'm a helicopter* by Fergal Dowling (this description, with additional details, can be found in the composer doctoral thesis [Dowling, 2006]). This piece uses the audifications of helicopter's data presented in Chapter 7 of this thesis. The recording given on the accompanying DVD, was made by performer David Stalling, at the composer's studio, Dublin 2006.

I'm a Helicopter

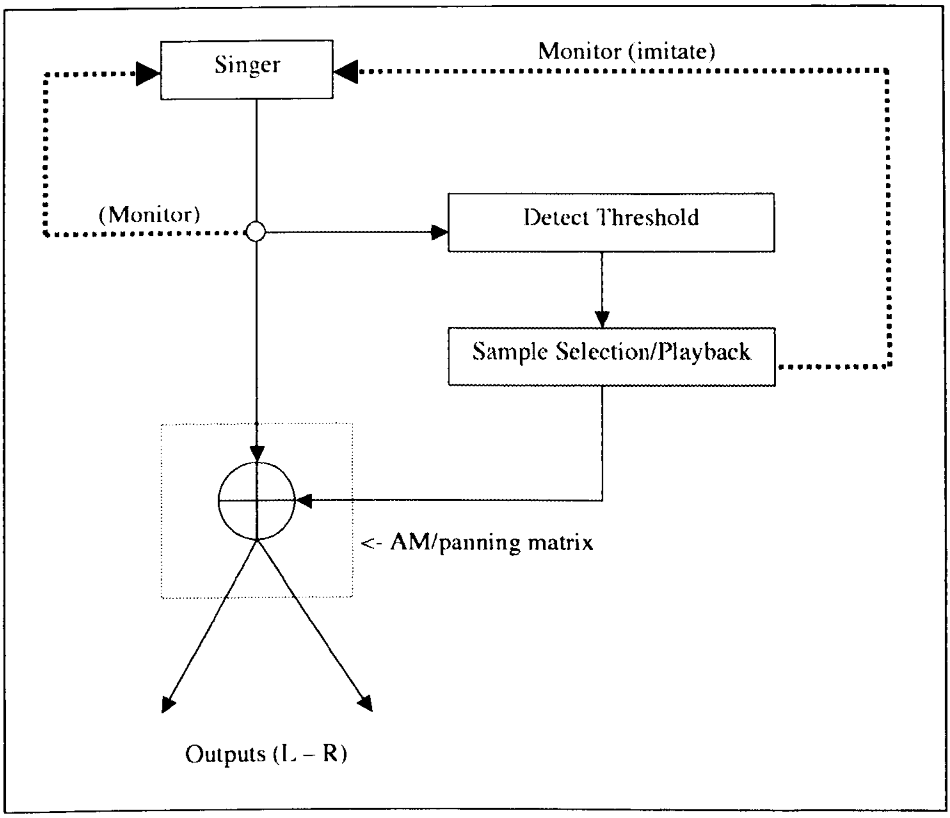
Real-time interaction for singer, microphone, computer (MaxMSP), twenty-eight samples and two loudspeakers
York, April 2005

I'm a Helicopter Operational overview

I'm a Helicopter is intended to function exclusively as a performance piece and [...] should not be presented as an installation. The performer is required to wear headphones and is unable to directly hear the full output of the system, that is to say that he cannot hear his own output in combination with the system, but can only hear the playback of the pre-recorded samples themselves. The details of the operating procedures are outlined below [...].

A singer articulates at a microphone. If sufficiently loud, his sounds trigger the playback of one of a number of samples. The samples are sonifications of data, the translation of data into the audio domain (Kramer, 1994), extracted from a helicopter flight. The samples are played back in the order given by the sample soundfile names, i.e. 2.wav–29.wav.

Each sample is replayed both to the singer's headphones and to a panning matrix. Upon hearing the sample playback, the singer must imitate the sound of the sonification. The singer's output is also routed to the panning matrix. His vocalisations are amplitude modulated by the sample/sonification playback and outputted on two channels to a stereo loudspeaker pair. The modulation is arranged so that a negative sample signal will cause the vocal output to be panned to the left channel, and positive sample signal will cause the vocalisations to be panned to the right channel. The strength of the panning effect can be varied from within the MaxMSP patch by setting the panning modulation level. The panning modulation level defaults to 1.0 and should normally remain in the range 0.5–1.0. The diagrams below should make the operational procedures of *I'm a Helicopter* evident.

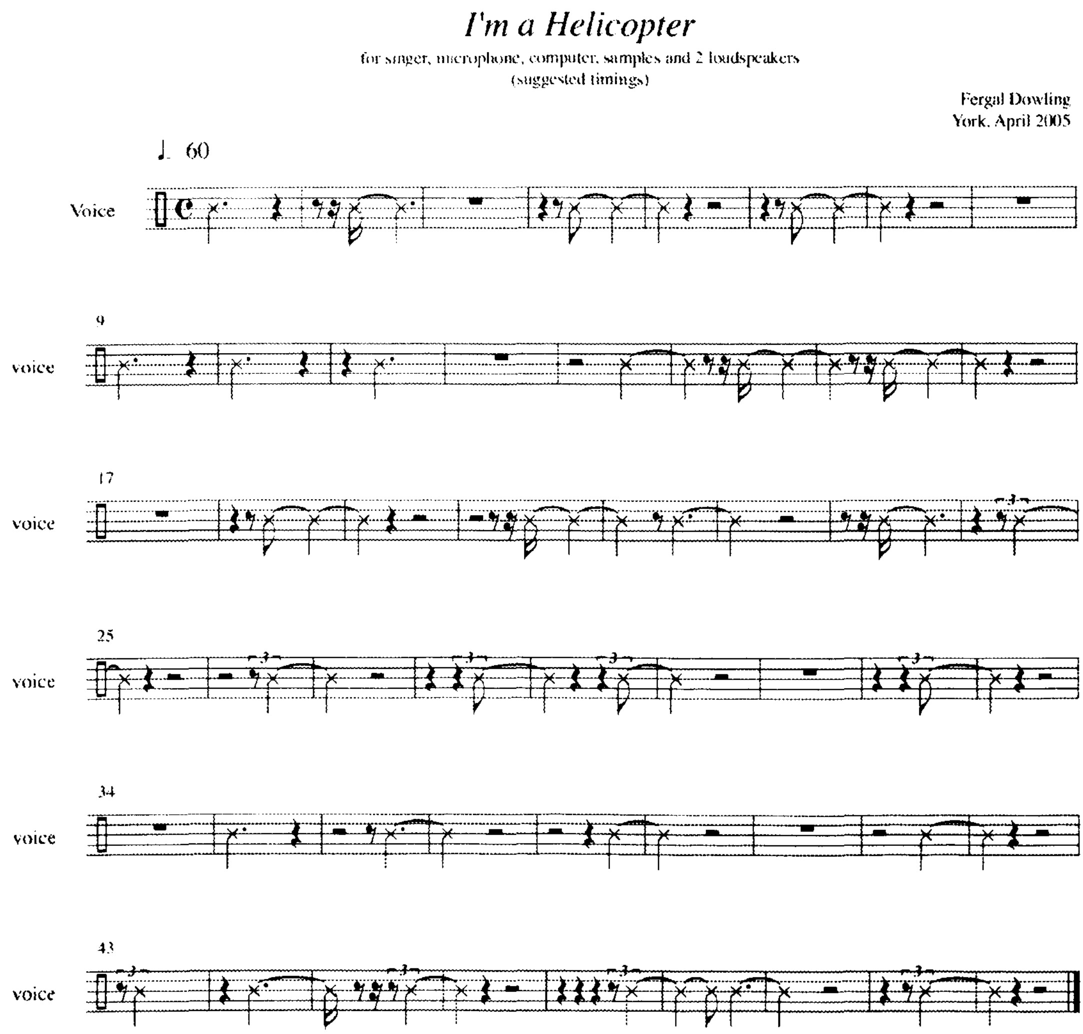


Appendix D Figure 1: *I'm a Helicopter*, schematic flowchart

Each helicopter-based sonification has a duration of 2.415 seconds. The singer may sing (imitate the sonification) for the full 2.415 seconds of its duration. If he continues for longer than 2.415 seconds he will cause the next sample to be replayed, and so on. This is possible, but whether or not it is desirable is a matter for the singer. In this manner, the singer continues until all samples have been replayed, at which point he may stop. He may stop before all samples are replayed, or he may continue after the last sample (29.wav) has been replayed. In this case he would cause the first samples in the sequence to be replayed again, and the process could continue cycling through all the samples again, and could continue ad nauseam.

A score is provided for the purpose of giving suggested timings of the first twenty-eight entries, triggering one sample at a time. The score shows only entry points (rhythms); all durations are assumed to be 2.415 seconds, however the singer is free to make articulations of any duration. Pitch, timbre and dynamic markings are necessarily absent since the performer should copy these qualities from the samples as heard. As with other members of this group of pieces, there is potential for either a restrained, controlled performance or an entirely chaotic rendition. This score may be followed or ignored. Likewise, other sample sets might be substituted.

I'm a Helicopter Score



I'm a Helicopter Sonification

The cross-synthesis panning matrix principle implemented here allows for the performer to predict, at least somewhat, the possible output resulting from his interaction with the samples.

The samples used here are based on data extracted from a helicopter flight which was sonified by Sandra Pauletto using her Sonification Toolkit [see Chapter 6 of this thesis]. The Sonification Toolkit reads data arrays and can map the data to a range of audio-musical parameters. The user can quickly experiment with various mapping possibilities and reading and rendering rates.

Appendix E – Experiments' questionnaires

Questionnaire experiment One

Dear Madam/Sir,

For statistical reasons I need to gather some general data about yourself.

Please answer the following questions:

1) What is your age?

2) Are you male or female?

3) What is your nationality?

4) What is your occupation?

5) In your occupation, do you work with sound and spectrograms of sounds?

6) Do you know of any problem with your hearing?

If yes, please, explain.

7) Do you know of any problem with your vision?

If yes, please, explain.

8) After the test, please feel free to comment on the experience

Questionnaire experiment Two

Dear Madam/Sir,

For statistical reasons I need to gather some general data about yourself.

Please answer the following questions:

1) What is your age?

2) Are you male or female?

3) What is your nationality?

4) What is your occupation?

5) In your occupation, do you work with sound?

6) Do you know of any problem with your hearing?
If yes, please, explain.

Some questions about the sounds you heard.

Did you find these sounds:

Very unpleasant						Very pleasant
Very uninteresting						Very interesting

At the end of this test you found that:

You could have listened more of these sounds	
You could have listened more, but not to this type of sounds	
You were tired of listening to any sound	

Did these sounds remind you of any natural sound? If yes which one?

These sounds are synthesised from data produced by the activity of leg's muscles. Do you find this sound appropriate to represent muscle's activity, i.e. movement? (please comment if you wish)

Questionnaire experiment Three

Dear Madam/Sir,

For statistical reasons I need to gather some general data about yourself.

Please answer the following questions:

1) What is your age?

2) Are you male or female?

3) What is your nationality?

4) What is your occupation?

5) In your occupation, do you work with sound?

6) Do you know of any problem with your hearing?
If yes, please, explain.

Please score the 3 interaction methods:

1) Low interaction method

Pleasant						Unpleasant
Intuitive						Unintuitive
Clear						Confusing
Quick						Slow

2) Medium interaction method

Pleasant						Unpleasant
Intuitive						Unintuitive
Clear						Confusing
Quick						Slow

3) High interaction method

Pleasant						Unpleasant
Intuitive						Unintuitive
Clear						Confusing
Quick						Slow

Which interaction method did you prefer?

Low interaction method	
Medium interaction method	
High interaction method	

Please tell me in a sentence why you preferred this interaction method

Any other comment

Appendix F – DVD content

Folder: Interactive Sonification Toolkit

The *readme* file explains how to install PD and the 2 versions of the Toolkit in your computer.

Once installed, to open the Toolkit double click on the *.bat files:
toolkit29.bat opens the Toolkit without graphical display;
toolkitV29gem.bat opens the version of the toolkit with the graphical display.

The two versions of the Toolkit are in folders:
Interactive Sonification Toolkit\pd\toolkitV29
Interactive Sonification Toolkit\pd\toolkitV29gem
The file *readme.pdf* contained in these folders is the Toolkit User Guide.

In folder *Interactive Sonification Toolkit\pd\bielefeldTidy* there are the PD files for the model-based sonification created at Bielefeld University in collaboration with Thomas Hermann (see Chapter 7 for description of this sonification).
This folder also contains the mathematically processed helicopter's data that are necessary to create this sonification.

In *Interactive Sonification Toolkit\pd\PICO_dll* there are the PD files for the real-time EMG sonification described in Chapter 8 that uses the PICO analogue to digital converter interface and the adcReader object in PD.

ToolkitVideo.avi is a video demonstration of an early version of the Toolkit created for the presentation at the International Conference on Auditory Display in Sidney in 2004.

Folder: Experiment 1 (helicopters)

Experiment 1 (helicopters)\heli_mtx_for_experiment contains the helicopter's data used in this research.

Experiment 1 (helicopters)\heliExperiment contains the PD files for the Experiment 1 program and the experiment questionnaire.

The *readme* file in *Experiment 1 (helicopters)* explains how to install the programs for the experiment.

To open the experiment programs (once installed) double click *heliExperimentSounds.bat* and *heliExperimentSpectra.bat*.

In Experiment 1 (helicopters)\heliExperiment\sounds are the audifications used in Experiment 1.

In Experiment 1 (helicopters)\heliExperiment\spectra are the spectra used in Experiment 1.

Folder: Experiment 2 (EMG)

The *readme* file says how to install PD program for Experiment 2.

In *Experiment 2 (EMG)\physioexperiment* are the PD files for the Experiment program, the experiment questionnaire and the program to randomise the order of presentation of the sounds.

To open the Experiment 2 program (once installed) double click the *physioexperiment.bat* file.

The .doc file *physio data sub number age and illness table* gives a table to link the EMG subject's number with his/her age.

In *Experiment 2 (EMG)\physiomtx* is the EMG data used for Experiment 2.

In *Experiment 2 (EMG)\physioexperiment\sounds* are the sounds used in the experiment.

Folder: Experiment 3 (interaction)

The *readme* file explains that Experiment 3 uses the Interactive Sonification Toolkit.

dataset1, *dataset2* and *dataset3* are the data files used in Experiment 3.

The three *.doc files are the experiment questionnaire and the sheet given to the test subject to order the structures.

Folder: Model-based sonification

The *readme* file explains how to install the program created for the model-based sonification of the helicopter's data (see Chapter 7 for the description of this sonification).

To open the program (once installed) double click *Bielefeld.bat*

Model-based sonification\video and sonification contains:
- one example of the model-based sonification created in collaboration with researcher Thomas Hermann as a .wav file;
- a very low resolution video of the graphical display of the model-based sonification;
- the installation file for the free BBFlash Player which allows to open the file *BielefedVideoNeedBBFlashBackFreePlayer* which is a much higher resolution version of the video of the graphical display of the model-based sonification;
- the *readme* file explains how to use the above mentioned files.

Folder: Real-time sonification of EMG

This folder contains the program used to create the real-time EMG sonification described in Chapter 8. The *readme* file explains how to install the program and open it using the *teesside.bat* file.

Real-time sonification of EMG\Video contains a video that shows how the real-time EMG sonification works and is set up.

Folder: Fergal Dowling composition

This folder contains Dr Fergal Dowling composition as a *.wma file. This composition used the sonifications created for Experiment 1 of this thesis. The composition is described in Appendix D and mentioned in Appendix C.

References

AGS2005 (2005) *First International Symposium on Auditory Graphs* (AGS 2005) [online] Available at: <http://sonify.psych.gatech.edu/ags2005/index.html> [Accessed 24 March 2007]

ANSI (1960) *American Standard Acoustical Terminology*, American National Standards Institute, New York

Avanzini, F., Rocchesso, D. and Serafin, S. (2004) *Friction sounds for sensorial substitution*, Proc. of the International Conference on Auditory Display, Sydney, Australia

Baier, G. and Hermann, T. (2004) *The sonification of rhythms in human electroencephalogram*, Proc. International Conference on Auditory Display, Sydney, Australia

Baier, G., Hermann, T. and Lara, O. M., (2005) *Using Sonification to detect weak cross-correlations in coupled excitable systems*, Proc. International Conference on Auditory Display, Limerick, Ireland

Baier, G., Hermann, T., Sahle, S. and Stephani, U., (2006) *Sonified epileptic rhythms*, Proc. International Conference on Auditory Display, London, England

Ballas, J. A. (1993) *Common factors in the identification of an assortment of brief everyday sounds*, Journal of Experimental Psychology, Vol 19(2), pp. 250-267

Barrass, S. and Zehner, B. (2000) *Responsive sonification of well logs*, Proc. of the International Conference on Auditory Display, Atlanta, USA

Barrass, S., (1995) *Personify: a toolkit for perceptually meaningful sonification*, Proc. Australian Computer Music Conference (ACMA)

Beamish, T., Van Den Doel, K. MacLean, K. and Fels, S. (2003) *D'groove: A haptic turntable for digital audio control*, Proc. of the International Conference on Auditory Display, Boston, USA

Beilharz, K. (2005) *Wireless gesture controllers to affect information sonification*, Proc. of the International Conference on Auditory Display, Limerick, Ireland

Ben-Tal, O., Berger, J., Cook, B., Daniels, M., Scavone, G. and Cook, P., (2002), *SONART: the sonification application research toolbox*, Proc. International Conference on Auditory Display, Kyoto

BIOPAC, *Biopac Systems, Inc.* [online] Available at: <http://www.biopac.com> [Accessed 24 March 2007]

Blattner, M. M., Sumikawa, D. A. and Greenberg, R. M. (1989) *Earcons and Icons: their structure and common design principles*, Human-Computer Interaction, Vol. 4, No. 1, pp. 11-44

Bonebright, T. (2001) *Perceptual structure of everyday sounds: A multidimensional scaling approach.* Proc. of the International Conference on Auditory Display, Helsinki, Finland

Bonebright, T. L. (2005) *A Suggested Agenda For Auditory Graph Research*, International Symposium on Auditory Graphs, Limerick, Ireland

Bonebright, T. L., Nees, M. A., Connerly, T. T. and McCain, G. R. (2001) *Testing the effectiveness of sonified graphs for education: a programmatic research project*, Proc. International Conference on Auditory Display, Helsinki, Finland

Bovermann, T., Hermann, T. and Ritter, H. (2005) *The Local Heat Exploration Model for Interactive Sonification*, Proc. International Conference on Auditory Display, Limerick, Ireland

Bovermann, T., Hermann, T. and Ritter, H. (2006) *Tangible data scanning sonification model*, Proc. of the International Conference on Auditory Display, London, UK

Brazil, E. and Fernström, M. (2006) *Investigating concurrent auditory icon recognition*, Proc. International Conference on Auditory Display, London, UK

Brazil, E., Fernstrom, M., Tzanetakis, G. and Cook, P. R. (2002) *Enhancing sonic browsing using audio information retrieval*, Proc. of the International Conference on Auditory Display, Kyoto, Japan, pp. 113–118

Bregman, A. S. (1994) *Auditory Scene Analysis*, Cambridge, Massachusetts: MIT Press

Brewster, S. (1994) Providing a structured method for integrating non-speech audio into human-computer interfaces, PhD thesis, York University, England

Brown L. M. and Brewster S. A., (2003) *Drawing by ear: Interpreting Sonified Line Graphs*, Proc. International Conference on Auditory Display, Boston, USA

Brown, A., Stevens, R. and Pettifer, S. (2006) *Audio representation of graphs: a quick look*, Proc. International Conference on Auditory Display, London, England

Bruce, J.W. and Palmer, N., (2005), *SIFT: Sonification Integrable Flexible Toolkit*, Proc. International Conference on Auditory Display, Limerick, Ireland

Burns, R. B. (2000), *Introduction to research methods*, SAGE, London

Camurri, A. and G. Volpe (Eds.), (2003) *Gesture-based communication in human-computer interactions*, Genova, Italy, vol. 2915 of *Lecture Notes in Computer Science*, pp. 369–379, Springer

Cassidy, R.J., Berger, J. and Lee, K. (2004) *Auditory display of hyperspectral colon tissue images using vocal synthesis models*, Proc. International Conference on Auditory Display, Sydney, Australia

Chion, M., Gorbman, C. and Murch, W. (1994) *Audio-Vision, Sound on screen*, Columbia University Press, New York

Cohen, J. (1988), *Statistical power analysis for the behavioural sciences*, Hillsdale, N. J.; Hove: L. Erlbaum Associates

Congas, *Congas, Gesture Controlled Audio Systems* [online] Available at: <http://www.cost287.org/> [Accessed 24 March 2007]

Contour Design, *Contour Design* [online] Available at: <http://www.contourdesign.com> [Accessed 24 March 2007]

Cook, P. R. (1999) *Music, Cognition and Computerized Sound*, Cambridge, Massachusetts: MIT Press

Cox, D. R. and Cox, M. A. A. (1994) *Multidimensional Scaling*: Chapman & Hall, London

Csikszentmihalyi, M. (2000) Beyond Boredom and Anxiety: Experiencing Flow in Work and Play, Jossey Bass Wiley

Csound, *cSounds.com* [online] Available at: <http://www.csounds.com> [Accessed 24 March 2007]

Cullen, C. and Coyle, E. (2005) *TrioSon: A Graphical User Interface for Pattern Sonification*, Proc. International Conference on Auditory Display, Limerick, Ireland

Cycling74, *Cycling74* [online] Available at: <http://www.cycling74.com> [Accessed 24 March 2007]

De Campo, A. (2007) *Toward a data sonification design space map*, Proc. of the International Conference on Auditory Display, Montreal, Canada

Di Filippo, D. and Pai, D. K. (2000) *Contact interaction with integrated audio and haptics*, Proc. of the International Conference on Auditory Display, Atlanta, USA

Dixon, J. (2004) An Electromyographic Analysis of Quadriceps Femoris in Patients with Osteoarthritis of the Knee, PhD Thesis, School of Health, Teesside University, Middlesborough, 2004

Dodge, C. and Jerse, T. A. (1997) *Computer Music, Synthesis, Composition and Performance*, Schirmer Books, New York, 1997

Dombois, F. (2001), *Using Audification in Planetary Seismology*, Proc. International Conference on Auditory Display, Helsinki, Finland

Dombois, F. (2002), Auditory Seismology - on free oscillations, focal mechanisms, explosions and synthetic seismograms, Proc. International Conference on Auditory Display, Kyoto, Japan

Dowling, F., (2006), *Composition Portfolio*, Doctoral Thesis, Music Department, University of York, UK

Dowling, F., *Fergal Dowling* [online] Available at: <http://www.fergaldowling.com/index.htm> [Accessed 24 March 2007]

Edmonds E., Candy L., Fell M., Knott R., Pauletto S., Weakley A. (2003), *Developing Interactive Art Using Visual Programming*, Proceedings of HCI International Conference, Greece, 2003

Effenberg, A. O. (2005) *Movement sonification: effects on perception and action*, IEEE Multimedia, April-June 2005, Vol. 12, No. 2, pp. 53-59

Eslambolchilar, P., Crossan, A. and Murray-Smith, R. (2004) *Model-based Target Sonification on Mobile Devices*, Proceedings of the International Workshop on Interactive Sonification, Bielefeld, Germany

Eslambolchilar, P. and Murray-Smith, R. (2006) *Model-based, Multimodal Interaction in Document Browsing*, 3rd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI'06), Washington DC

Ferguson, S., Cabrera, D., Beilharz, K. and Hong Jun Song (2006) *Using psychoacoustical models for auditory display*, Proc. International Conference on Auditory Display, London, England

Fernström, M. and McNamara, C. (1998) *After direct manipulation - direct sonification*, Proc. of the International Conference on Auditory Display, British Computer Society, Glasgow, UK

Fernström, M., Brazil, E. and Bannon, L. (2005) *HCI Design and Interactive Sonification for Fingers and Hears*, IEEE Multimedia, vol. 12, pp. 36-44

Field, A. (2005), *Discovering statistics using SPSS*, Sage Publications Ltd.

Fitch, W. T. and Kramer, G. (1994) *Sonifing the Body Electric: Superiority of an Auditory over a Visual Display in a Complex, Multivariate System*, in Kramer G. (ed) Auditory Display: Sonification, Audification, and Auditory Interface, Reading, Addison-Wesley

Fletcher, H. and Munson, W. (1933) *Loudness, its measurement and calculation*, Journal of the Acoustical Society of America, Vol. 5, pp. 82-108

Flowers, J. H., (2005) Thirteen years of reflection on auditory graphing: promises, pitfalls, and potential new directions, International Symposium on Auditory Graphs, Limerick, Ireland

Fox, J. and Carlile, J. (2005) *SoniMime: Sonification of Fine Motor Skills*, Proc. International Conference on Auditory Display, Limerick, Ireland

Gaver, W. W. (1986) *Auditory Icons: Using Sound in Computer Interfaces*, Human-Computer Interaction, Vol. 2, No. 2, pp. 167-177

GEM library, *GEM* [online] Available at: <http://gem.iem.at/> [Accessed 24 March 2007]

Ghez, C., Rikakis, T., DuBois, R. L. and Cook, P. R. (2000) *An auditory display system for aiding interjoint coordination*, Proc. International Conference on Auditory Display, Atlanta, USA

GPower software, *What is G*Power* [online] Available at: <http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/> [Accessed 24 March 2007]

Greene, J. and D'Oliveira, M. (1982), *Learning to use statistical tests in psychology*, Milton Keynes: Open University Press

Grey, J. (1977) *Timbre discrimination in musical patterns*, Journal of the Acoustical Society of America, Vol. 64, pp. 467-472

GrIPD library, *GrIPD* [online] Available at: <http://crca.ucsd.edu/~jsarlo/gripd/> [Accessed 24 March 2007]

Gröhn, M. (1992) *Sound probe: An interactive sonification tool*, Proc. of the International Conference on Auditory Display, Santa Fe, USA

Hayward, C., (1994) *Listening to the Earth Sing*, in Kramer, G. (ed), Auditory Display. Sonification, Audification and Auditory Interfaces, Addison-Wesley, Reading, USA, pp. 369-404

Hermann, T. (2002) *Sonification for Exploratory Data Analysis*, PhD Thesis in Faculty of Technology, Bielefeld University, Germany

Hermann, T. and Hunt, A., (2005) *An Introduction to Interactive Sonification*, IEEE Multimedia, April-June, Vol. 12, No. 2, pp. 20-24

Hermann, T. and Ritter, H. (1999) *Listen to your data: Model-Based Sonification for Data Analysis*, in Lasker, G. E. (ed) Advances in intelligent computing and multimedia systems, Baden-Baden, Germany, pp. 189-194

Hermann, T. and Ritter, H. (2002) *Crystallisation sonification of high dimensional datasets*, Proc. International Conference on Auditory Display, Kyoto, Japan

Hermann, T. and Ritter, H. (2004) *Sound and meaning in auditory data display*, Proceedings of the IEEE, Special Issue on Engineering and Music - Supervisory Control and Auditory Communication, April 2004, Vol. 92, no. 4, pp. 730-741

Hermann, T., Baier, G., Stephani, U. and Ritter, H. (2006) *Vocal sonification of pathologic EEG features*, Proc. International Conference on Auditory Display, London, England

Hermann, T., Hansen, M. H. and Ritter, H. (2001) *Sonification of Markov Chain Monte Carlo Simulations*, Proc. International Conference on Auditory Display, Helsinki, Finland

Hermann, T., Henning, T. and Ritter, H. (2003) *Gesture desk - an integrated multi-modal workplace for interactive sonification*, in International Gesture Workshop, Genova, Italy

Hermann, T., Meinicke, P. and Ritter, H. (2000) *Principal Curve Sonification*, Proc. International Conference on Auditory Display, Atlanta, USA

Hermann, T., Meinicke, P., Bekel, H., Ritter, H., Muller, H. M. and Weiss, S. (2002) *Sonifications for EEG data analysis*, Proc. International Conference on Auditory Display, Kyoto, Japan

Hermann, T., Nölker, C. and Ritter, H. (2001) *Hand postures for sonification control*, in Gesture and Sign Language in Human-Computer Interaction, Proc. Int. Gesture Workshop GW2001, Ipke Wachsmuth and Timo Sowa, Eds. 2002, pp. 307–316, Springer

Hinterberger, T. and Baier, G. (2005) *Parametric Orchestral Sonification of EEG in Real Time*, IEEE Multimedia, April-June 2005, Vol. 12, No. 2, pp. 70-79

History of Max, *Une brève histoire de MAX* [online] Available at: <http://freesoftware.ircam.fr/article.php3?id_article=7> [Accessed 24 March 2007]

Holmes, J. (2005) *Interacting with an information space using sound: accuracy and patterns*, Proc. of the International Conference on Auditory Display, Limerick, Ireland

Houtsma, A. J. M., Rossing, T. D. and Wagenaars, W. M. (1987) *Auditory demonstrations*, CD booklet, prepared at the Institute for Perception Research, Eindhoven, The Netherlands, supported by the Acoustical Society of America

Howard, D. and Angus, J. (1996) *Acoustics and Psychoacoustics*, Music Technology Series, Oxford: Focal Press

Hunt, A. (2000) *Radical User Interfaces for Real-time Musical Control*, Ph.D. thesis, University of York, UK

Hunt, A. D., Paradis, M. and Wanderley, M. (2003) *The importance of parameter mapping in electronic instrument design*, Journal of New Music Research, vol. 32, no. 4, pp. 429–440, December, special issue on New Musical Performance and Interaction

Hunt, A., Hermann, T. and Pauletto, S. (2004) *Interacting with sonification systems: closing the loop*, Proc. of the IEEE Conference on Information Visualisation, pp. 879- 884

Hutchins, E. L., Hollan, J. D. and Norman, D. (1986) *Direct manipulation interfaces*, in *User-Centred System Design*, Norman, D. and Draper, S. (Eds.) Hillsdale, NJ, USA: Lawrence Erlbaum Associates, pp. 87-124

Interactive Sonification, *Interactive Sonification Homepage* [online] Available at: <http://www.interactive-sonification.org> [Accessed 24 March 2007]

International Conference on Auditory Display, *ICAD International Community for Auditory Display* [online] Available at: <http://www.icad.org> [Accessed 24 March 2007]

IRCAM, *IRCAM* [online] Available at: <http://www.ircam.fr> [Accessed 24 March 2007]

ISO 9241-11 (1998), Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability, International Standard Organisation

Janata P. and Childs, E., (2004) *Martketbuzz: Sonification of Real-Time Financial Data*, Proc. International Conference on Auditory Display, Sydney, Australia

JMax, *JMax 4.1.0 released* [online] Available at: <http://sourceforge.net/projects/jmax/> [Accessed 24 March 2007]

Joseph, A. and Lodha, S. K., (2002) *MUSART: Musical Audio Transfer Function Real-Time Toolkit*, Proc. International Conference on Auditory Display, Kyoto, Japan

Jovanov, E., Starcevic, D., Wegner, K., Karron, D. and Radivojevic, V. (1998) *Acoustic rendering as support for sustained attention during biomedical processing*, Proc. of the International Conference on Auditory Display, British Computer Society, Glasgow, UK

Kazem, M. L. N., Noyes, J. M. and Lieven, N. J. (2003) *Design considerations for a background auditory display to aid pilot situation awareness*, Proc. International Conference on Auditory Display, Boston, USA

Keller, J., Prather, E.E., Boynton, W.V., Enos, H.L., Jones, L.V., Pompea, S.M. Slater, T.F. and Quinn, M. (2003) *Educational testing of auditory display regarding seasonal variation of polar ice caps on Mars*, Proc. International Conference on Auditory Display, Boston, USA

Kildal, J. and Brewster, S., (2006) *Providing a size-independent overview of non-visual tables*, Proc. International Conference on Auditory Display, London, England

Kildal, J. and Brewster, S. (2006) *Explore the Matrix: Browsing Numerical Data Tables Using Sound*, Proc. International Conference on Auditory Display, London, England

Kopra Andy, *Digital Sound Synthesis*, [online] Available at: <http://homepage.mac.com/andykopra/pdm/lectures/digital_sound_synthesis.html> [Accessed 24 March 2007]

Kramer, G. (1994) *Some organizing principles for representing data with sound*, in Kramer G. (ed.) "Auditory Display: Sonification, Audification, and Auditory Interface", Addison-Wesley, Reading, MA

Kramer, G. (ed), (1994) Auditory Displays. Sonification, Audification and Auditory Interfaces, Addison-Wesley

Kramer, G., Walker, B. N., Bonebright, T., Cook, P., Flowers, J., Miner, N., *et al.* (1999), *The Sonification Report: Status of the Field and Research Agenda*, Report prepared for the National Science Foundation by members of the International Community for Auditory Display. Santa Fe, NM: International Community for Auditory Display (ICAD)

Krishnan S., Rangayyan R. M., Bell G. D., Frank C. B., (2001) *Auditory display of knee-joint vibration signals*, Journal of the Acoustical Society of America, 110(6), December 2001, pp. 3292-3304

LabVIEW, *LabVIEW: 20 Years of Innovation* [online] Available at: <http://www.ni.com/labview/> [Accessed 24 March 2007]

Lee, K., Sell, G. and Berger, J. (2005) *Sonification using digital waveguides and 2- and 3- dimensional digital waveguide mesh*, Proc. International Conference on Auditory Display, Limerick, Ireland

LeVeau, B. and Andersson, G. B. J. (1992) *Output Forms: Data Analysis and Applications*, in 'Selected Topics in Surface Electromyography for Use in the Occupational Setting: Expert Perspectives', Soderberg G. L., Ed.: U. S. Department of Health and Human Services, Public Health Service, Centers for Diseases Control, National Institute for Occupational Safety and Health, 1992, pp. 69-102

Levitin, D. J. (1999), *Experimental design in psychoacoustic research,* in Music, Cognition and Computerized Sound, Cook. P. R. (Ed.), Cambridge, Mass.: MIT Press, pp. 299-328

Lodha, S. K., Beahan, J., Heppe, T., Joseph, A. and Zane-Ulman, B. (1997) *MUSE: a musical data sonification toolkit*, Proc. International Conference on Auditory Display, Palo Alto, USA

Lokki, T. and Gröhn, M. (2005) *Navigation with auditory cues in a virtual environment*, IEEE Multimedia, vol. 12, pp. 80-86

Malandrino, D., Mea, D., Negro, A., Palmieri, G. and Scarano, V. (2003) *NeMoS: Network monitoring with sound*, Proc. International Conference on Auditory Display, Boston, USA

Mansur, D. L., Blattner, M. M. and Joy, I. J. (1985) *Sound Graphs: A Numerical Data Analysis Method for the Blind*, Journal of Medical Systems, vol. 9, no. 3, pp. 163-174

Matlab, *The MathWorks* [online] Available at: <http://www.mathworks.co.uk/> [Accessed 24 March 2007]

Menéndez, R. G. and Bernard, J. E. (2001) *Flight Simulation in Synthetic Environments*, IEEE on Aerospace and Electronic Systems Magazine, vol. 16, pp. 19-23

Miele, J.A., (2003) *Smith-Kettlewell display kit: A sonification toolkit for Matlab*, Proc. International Conference on Auditory Display, Boston, USA

Milczynski, M., Hermann, T., Bovermann, T. and Ritter, H. (2006) *A malleable device with applications to sonification-based data exploration*, Proc. of the International Conference on Auditory Display, London, UK

Nees, M. A. and Walker, B. N., (2006) *Relative intensity of auditory context for auditory graph design*, Proc. International Conference on Auditory Display, London, UK

Neo/NST, *The Graphical Simulation ToolkitNeo/NST* [online] Available at: <http://www.techfak.uni-bielefeld.de/ags/ni/projects/neo/neo_e.html> [Accessed 24 March 2007]

Nesbitt K. V. and Barrass S., (2004) *Finding Trading Patterns in Stock Market Data*, IEEE Computer Graphics and Applications, September/October 2004, pp. 45-55

Neuhoff, J. (ed) (2004) *Ecological Psychoacoustics*, New York: Academic Press

Novartis Foundation, *The Novartis Foundation* [online] Available at: <http://www.novartisfound.org.uk> [Accessed 24 March 2007]

Open Sound Control, *Home Page* [online] Available at: <http://cnmat.berkeley.edu/OpenSoundControl/> [Accessed 24 March 2007]

Palomäki, H. (2006) *Meanings conveyed by the simple auditory rhythm*, Proc. International Conference on Auditory Display, London, UK

Pauletto, S. and Hunt, A., (2004) *A Toolkit for Interactive Sonification*, Proc. International Conference on Auditory Display, Sydney, Australia

Pauletto, S. and Hunt, A. (2005) *A Comparison of Audio & Visual Analysis of Complex Time-Series Data Sets*, Proc. International Conference on Auditory Display, Limerick, Ireland

Pauletto, S. and Hunt, A. (2006) *The sonification of EMG data*, Proc. International Conference on Auditory Display, London, England

Pauletto, S. & Hunt, A., (2007) *Interacting with sonifications: an evaluation*, Proc. International Conference on Auditory Display, Montreal, Canada

Peres, S. C. and Lane, D. M. (2003) *Sonification of statistical graphs*, Proc. International Conference on Auditory Display, Boston, USA

Pico Technology, *PC Oscilloscope and Data Acquisition Products* [online] Available at: <http://www.picotech.com> [Accessed 24 March 2007]

Pierce, J. (1999) *Hearing in time and space*, in Cook, P. (ed.) Music, cognition and computerized sound, MIT Press, pp. 89-103

Preece, J., Rogers, Y. and Sharp, H. (2002), *Interaction Design. Beyond human-computer interaction*, John Wiley & Sons, Inc.

Puckette, M. (1988) *The Patcher*, Proc. of the International Computer Music Conference, pp. 420-429

Puckette, M. and Lippe, C. (1992) *Score Following In Practice*, Proc. of the Society for Electroacoustic Music in the United States, San Jose, USA

Pure Data programming environment, *PURE DATA* [online] Available at: <http://pd.iem.at/> [Accessed 24 March 2007]

Rath, M. and Rocchesso, D. (2005) *Continuous sonic feedback from a rolling ball*, IEEE Multimedia, vol. 12, pp. 60-69

Rizzi, S. A., Sullivan, B. M. and Sandridge, C. A. (2003) *A three-dimensional virtual simulator for aircraft flyover presentation*, Proc. International Conference on Auditory Display, Boston, USA

Saue, S. (2000) *A model for interaction in exploratory sonification displays*, Proc. of the International Conference on Auditory Display, Atlanta, US

School of Health and Social Care of Teesside University, *School of Health and Social Care* [online] Available at: <http://www.tees.ac.uk/schools/SOH/index.cfm> [Accessed 24 March 2007]

Schouten, J. F. (1940) *The perception of pitch*, Philips Technical Review, Vol. 5, p. 286

Shneiderman, B. (1983) *Direct Manipulation: A step beyond programming languages*, IEEE Computer, vol. 16, Issue 8, pp. 57-69

Shneiderman, B. and Plaisant, C. (2005) *Designing the User Interface*, (Fourth edition), Addison Wesley

SKDTools, *Smith-Kettlewell Display Tools (SDTools) An Auditory/Tactile Data Representation Toolbox For Matlab* [online] Available at: <http://www.ski.org/rerc/SKDtools/index.html> [Accessed 24 March 2007]

Smalltalk, *Smalltalk.org* [online] Available at: <http://www.smalltalk.org/main/> [Accessed 24 March 2007]

Smith, D. R. and Walker, B. N. (2002) *Tick-marks, axes, and labels: The effects of adding context to auditory graphs* Proc. of the International Conference on Auditory Display pp. 362-367, Kyoto, Japan

Snyder, B. (2000) *Music and Memory*, MIT Press

Sonification Lab, GT Sonification Lab, School of Psychology, Georgia Institute of Technology [online] Available at: <http://sonify.psych.gatech.edu/research/index.html> [Accessed 24 March 2007]

Sonnenschein, D. (2001) Sound design. The expressive power of music, voice and sound effects in Cinema. Saline, Michigan, USA: Mc. Naughton and Gunn Inc.

SPSS software, *Welcome to SPSS.com. Enabling the Predictive Enterprise* [online] Available at: <http://www.spss.com> [Accessed 24 March 2007]

Stevens, S. S. (1935) *The relation of pitch to intensity*, Journal of the Acoustical Society of America, Vol. 6, pp. 150-54

Stevens, S. S., Volkmann, J. and Newman, E. B. (1937) *A Scale for the Measurement of the Psychological Magnitude of Pitch*, Journal of the Acoustical Society of America, Vol. 8, pp. 185-190

Stockman, T. (2004) *The design and evaluation of auditory access to spreadsheets*, Proc. International Conference on Auditory Display, Sydney, Australia

Stockman, T., Hind, G. and Frauenberger, C. (2005) *Interactive sonification of spreadsheets*, Proc. of the International Conference on Auditory Display, Limerick, Ireland

Stokes M. and Blythe M. (2001) Muscle Sounds in physiology, sport science and clinical investigation, applications and history of mechanomyography. Medintel, Oxford, 2001

Streiner, D. L. (1996) Maintaining standards: differences between the standard deviation and standard error, and when to use each, Canadian Journal of Psychiatry, Vol. 41, pp. 498-502

Sturm, B. L. (2002) *Water Music: Sonification of Ocean Buoy Spectral Data*, Proc. International Conference on Auditory Display, Kyoto, Japan

Sturm, B. L. (2003) *Ocean buoy spectral data sonification: research update*, Proc. International Conference on Auditory Display, Boston, USA

Super Collider, *Super Collider* [online] Available at: <http://www.dewdrop-world.net/sc3/index.php> [Accessed 24 March 2007]

Supercollider, *Supercollider* [online] Available at: <http://www.audiosynth.com/> [Accessed 24 March 2007]

Thomas Hermann, *Dr. Thomas Hermann – Welcome to my Homepage* [online] Available at: <http://www.techfak.uni-bielefeld.de/~thermann/> [Accessed 24 March 2007]

Valenzuela, M. L., Sansalone, M. J., Streett, W. B. and Krumhansl, C. L. (1997) *Use of Sound for the Interpretation of Impact-Echo Signals*, Proc. International Conference on Auditory Display, Palo Alto, USA

Van Den Doel, K. (2004) *Physically-based models for liquid sounds*, Proc. of the International Conference on Auditory Display, Sydney, Australia

Walker, A., and Brewster, S. (2001) Sitting too close to the screen can be bad for your ears: A study of audio-visual location discrepancy detection under different visual projections Proc. of the International Conference on Auditory Display (pp. 86-89), Helsinki, Finland

Walker, B. N. (2000) *Magnitude estimation of conceptual data dimensions for use in sonification.* Unpublished Ph.D. Dissertation, Rice University, Houston, TX

Walker, B. N. (2002) *Magnitude estimation of conceptual data dimensions for use in sonification* Journal of Experimental Psychology: Applied, 8(4), 211-221

Walker, B. N. and Cothran, J.T. (2003) *Sonification sandbox: A graphical toolkit for auditory graphs*, Proc. International Conference on Auditory Display, Boston, USA

Walker, B. N. and Kramer, G. (2004) *Ecological Psychoacoustics and Auditory Displays: Hearing, Grouping and Meaning Making*, in Neuhoff, J. (ed) Ecological Psychoacoustics, New York: Academic Press, pp. 150-175

Walker, B. N. and Kramer, G., (2005), *Sonification design and metaphors: Comments on Walker and Kramer, ICAD 1996* ACM Transactions on Applied Perception, 2(4), 413-417

Walker, B. N. and Lane, D. M. (2001) *Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted listeners,* Proc. of the International Conference on Auditory Display (pp. 90-94), Helsinki, Finland

Walker, B. N. and Mauney (2004) *Individual differences, cognitive abilities, and the interpretation of auditory graphs*, Proc. International Conference on Auditory Display, Sydney, Australia

Walker, B. N. and Nees, M. A., (2005) Brief training for performance of a point estimation sonification task, Proceedings of the International Conference on Auditory Display (ICAD2005), Limerick, Ireland

Walker, B. N., & Kramer, G. (1996b). *Mappings and metaphors in auditory displays: an experimental assessment* Proc. of the International Conference on Auditory Display, Palo Alto, USA

Walker, B. N., Kramer, G. and Lane, D. M. (2000) *Psychophysical scaling of sonification mappings*, Proc. International Conference on Auditory Display, Atlanta, USA

Westland, *Westland* [online] Available at: <http://www.whl.co.uk/index.cfm> [Accessed 24 March 2007]

Williamson, J. and Murray-Smith, R. (2005) *Sonification of probabilistic feedback through granular synthesis*, IEEE Multimedia (April-June 2005) Vol. 12(2) pp. 45-52

Wilson, C. M. and Lodha, S. K., (1996) *Listen: a data sonification toolkit*, Proc. International Conference on Auditory Display, Palo Alto, USA

Winberg, F. and Hellström, S. O. (2001) *Qualitative aspects of auditory direct manipulation: a case study of the towers of Hanoi*," Proc. of the International Conference on Auditory Display, Helsinki, Finland

Zexy library, *PURE DATA downloadables* [online] Available at: <http://iem.at/~zmoelnig/pd/download/index.html#zexy> [Accessed 24 March 2007]

Zhao, H., Smith, B. K., Norman, K., Plaisant, C. and Shneiderman, B., (2005) *Interactive Sonification of Choropleth Maps*, IEEE Multimedia, April-June 2005, Vol. 12, No. 2, pp. 26-35

Zwicker, E. and Fastl, H. (1999) *Psychoacoustics: facts and models*, Berlin; New York: Springer