Developing an Interactive Overview for Non-Visual Exploration of Tabular Numerical Information

Johan Kildal

Submitted for the Degree of Doctor of Philosophy

Department of Computing Science, University of Glasgow

September 2008

Abstract

This thesis investigates the problem of obtaining overview information from complex tabular numerical data sets non-visually. Blind and visually impaired people need to access and analyse numerical data, both in education and in professional occupations. Obtaining an overview is a necessary first step in data analysis, for which current non-visual data accessibility methods offer little support.

This thesis describes a new interactive parametric sonification technique called High-Density Sonification (HDS), which facilitates the process of extracting overview information from the data easily and efficiently by rendering multiple data points as single auditory events. Beyond obtaining an overview of the data, experimental studies showed that the capabilities of human auditory perception and cognition to extract meaning from HDS representations could be used to reliably estimate relative arithmetic mean values within large tabular data sets.

Following a user-centred design methodology, HDS was implemented as the primary form of overview information display in a multimodal interface called TableVis. This interface supports the active process of interactive data exploration non-visually, making use of proprioception to maintain contextual information during exploration (non-visual focus+context), vibrotactile data annotations (EMA-Tactons) that can be used as external memory aids to prevent high mental workload levels, and speech synthesis to access detailed information on demand.

A series of empirical studies was conducted to quantify the performance attained in the exploration of tabular data sets for overview information using TableVis. This was done by comparing HDS with the main current non-visual accessibility technique (speech synthesis), and by quantifying the effect of different sizes of data sets on user performance, which showed that HDS resulted in better performance than speech, and that this performance was not heavily dependent on the size of the data set. In addition, levels of subjective workload during exploration tasks using TableVis were investigated, resulting in the proposal of EMA-Tactons, vibrotactile annotations that the user can add to the data in order to prevent working memory saturation in the most demanding data exploration scenarios. An experimental evaluation found that EMA-Tactons significantly reduced mental workload in data exploration tasks.

Thus, the work described in this thesis provides a basis for the interactive non-visual exploration of a broad range of sizes of numerical data tables by offering techniques to extract overview information quickly, performing perceptual estimations of data descriptors (relative arithmetic mean) and managing demands on mental workload through vibrotactile data annotations, while seamlessly linking with explorations at different levels of detail and preserving spatial data representation metaphors to support collaboration with sighted users.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Stephen Brewster, for his help, encouragement and invaluable advice over the course of this research project. I would also like to thank my second supervisor, Dr Helen Purchase. Thanks are also due to Professor Stephen McAdams at McGill University, Montreal. I am also grateful to my colleagues in the Multimodal Interaction Group at the Department of Computing Science, University of Glasgow. In particular, I want to thank Dr. Andrew Crossan, Dr. David McGookin and Dr. Steven Wall, for many interesting research related discussions over the course of these years. I would also like to acknowledge my examiners, Professor Roderick Murray-Smith and Dr. Robert Stevens, for their constructive feedback.

Particular thanks go to the students at the Royal National College for the Blind in Hereford, the Royal Blind School in Edinburgh, and Uddingston Grammar School in Glasgow, who participated in the requirements capture process and experimental studies for this thesis. Thanks are also due to the staff of these institutions and of the Royal National Institute for Blind People. I would also like to acknowledge the Ken Browning scholarship fund, which supported my placement at McGill University.

Additionally, I want mention the support received in the final stages from my colleagues at Nokia Research Center in Finland. In particular, I want to thank Martin Schrader and Jari Kangas, Jyri Huopaniemi, Vuokko Lantz and my colleagues in the IC and MMI teams.

Last but not least, I cannot overstate the importance of the support received from the people that have been on my side during the years that took to complete this piece of work. My biggest thanks go to Dr. Louisa Gairn, whose care and understanding have seen no limits. I am also hugely grateful to her parents, Margaret and Jim. Similarly, my parents, Johan and Axun, have been, as always before, of inestimable help and support, and for that I cannot be thankful enough. I also want to most dearly thank my three sisters (Karen, Liv and Anne) for so much care and worry. Finally, I must not forget to mention the names of Frode Laursen and Katrin Kirm, and to also extend my sincere thanks to my dear friends, whose friendship I have always carried with me while away from home.

This research was supported by EPSRC Grant GR/S86150/01 as part of the MultiVis Project.

Dedication

To my father and friend Johan, who is an example of integrity and courage that I am always safe in trying to follow. *Eskerrik asko, Aitatxo*.

Declaration

The material presented in this thesis is the result of my own research carried out at the Department of Computing Science at the University of Glasgow working under the supervision of Professor Stephen Brewster and Doctor Helen Purchase.

The work contained in Chapter 4 has been published as Kildal & Brewster at CHI2006 [82]. The work contained in Chapter 5 has been published at CHI2007 [85]. The work contained in Chapter 6 has been published in the *International Journal on Disability and Human Development* [81] and in ICAD2006 [83]. Work contained in Chapter 7 has been published at INTERACT 2007 [84].

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

81. Kildal, J. and S.A. Brewster, *Exploratory Strategies and Procedures to Obtain Non-Visual Overviews Using TableVis.* International Journal on Disability and Human Development, 2006. **5**(3): p. 285-294.
82. Kildal, J. and S.A. Brewster. *Non-visual Overviews of Complex Data Sets*. in *Extended Abstracts of CHI 2006*. 2006. Montreal, Quèbec, Canada: ACM Press.
83. Kildal, J. and S.A. Brewster. *Providing a Size-Independent Overview of Non-Visual Tables*. in *12th International Conference on Auditory Display (ICAD2006)*. 2006. Queen Mary, University of London.
84. Kildal, J. and S.A. Brewster. *EMA-Tactons: Vibrotactile External Memory Aids in an Auditory Display*. in *IFIP Interact 2007*. 2007. Rio de Janeiro, Brasil.
85. Kildal, J. and S.A. Brewster, *Interactive Generation of Overview Information Using Speech*, in *CHI '07 extended abstracts on Human factors in computing systems*. 2007, ACM Press: San Jose, CA, USA. p. 2489-2494.

Chapter 1 Introduction

1.1 Motivation of the Thesis

The World Health Organization reports that in 2002 the number of people with visual impairment worldwide was in excess of 161 million, of whom about 37 million were blind; in Europe alone, the population of blind and visually-impaired people was estimated at 15.5 million [124]. The United Nations Convention on the Rights of Disabled Persons (2006) sets out the obligation for governments to ensure that people with disabilities can enjoy real equality in society. As part of this requirement, UN member states are requested to take measures to ensure accessibility of the physical environment and information and communications technology [1]. In line with this, and other legislation such as the UK's Disability Discrimination Act (1995), governments, employers and educators are now obligated to provide equality of opportunity to disabled people, including those who are blind and visually impaired.

Advances in information technology in the last decades have contributed enormously to making information more accessible for users who are blind or visually impaired. The advent of screen reading technology, the generalisation of personal computers for work, study and leisure, and the democratisation of information availability though the internet, has drawn a new scenario in which information is, by and large, far more accessible in non-visual forms. However, in spite of this technological revolution and the accessibility policies now in force, there still exist significant barriers for blind and partially-sighted users to access many of the most common types of information. One example of particular concern in this thesis is access to numerical data sets. In the case of sighted users, highly-specialised data visualisations (designed to exploit human vision and visual information processing) provide quick and easy access to overview information and help with the performance of various aspects of data analysis, such as identifying trends, patterns and singular features that might be of interest. Understanding and analysing information using visualisations such as graphs, bar charts and 3-dimensional (3D) plots is a common task for sighted people. The skills needed are learned early in school and then used throughout life. More importantly, the basic skills needed for creating and interpreting graphs are necessary for many aspects of education and employment. So ubiquitous has the use of such visualisation techniques become that an estimated 2.2 trillion graphs were published in 1994 alone (not including graphs displayed on television or computer terminals) [76].

Recent studies have found that subjects such as business management and accounting, information technology and computing science, and social and applied sciences are popular choices amongst blind and visually impaired students in UK higher education [60]. However, blind people have very restricted access to information presented in the data visualisations commonly utilised in the study of these subjects. While the majority of blind and visually impaired computer users utilise computers for word processing, email and web access, a significant percentage also regularly deal with spreadsheets and databases – more than one third of working-age blind adults, according to a 2007 survey [42]. The lack of accessibility tools to support obtaining quick overviews of data sets non-visually has serious implications for people who are blind or visually impaired, limiting their freedom to decide about their studies or professional careers – a freedom which is now protected as a human right by the United Nations, European Union and national governments [169]. Not only do such barriers prevent blind and visually impaired people from fulfilling their potential and contributing to society in their chosen field of study, lack of access to employment also affects their sense of personal identity, status and self-respect [152]. A report commissioned by the Royal National Institute for Blind People found that 75% of blind and partially sighted people in the UK were unemployed in 2002 [126]. While several socio-political factors including workplace discrimination have been hypothesised as contributing to this statistic, research also suggests that educational attainment, supported by appropriate and accessible technologies, is a key factor influencing employment for blind and visually impaired people [117]. In order to ensure blind and visually impaired people have equality of educational opportunity, as well as the freedom to pursue a career in a scientific or financial discipline, new technologies need to be developed to make numerical data accessible non-visually.

Various computer-based accessibility technologies for visually impaired users already exist, such as screen readers, refreshable Braille displays, pin arrays and haptic interfaces such as SensAble Technology's PHANTOM force-feedback device [137]. However, these are subject to a number of limitations for exploring numerical data sets. While such technologies provide access to computer based information non-visually, they do so by presenting information in full detail and often sequentially, with little or no support to obtain overview information quickly and easily at the beginning of the exploration of a new data set. As Shneiderman [138] asserts, any data exploration begins by obtaining an overview, to be followed by zooming and filtering operations, with detailed data available on demand by the user, in order to avoid saturation of the working memory.

In addition to the potential benefits the development of such technologies would bring to blind and visually impaired people, such research can also contribute to interface improvements for all users. As Newell [112] has postulated, the "*extra-ordinary needs [of disabled people] are only exaggerated ordinary needs*", thus interfaces designed for disabled people, including the blind and visually impaired, may also have applications for other users, either in the general populace or in specific circumstances such as fatigue or "high workload environments", where the number of tasks required and quantity of data the user must assimilate exceeds the "input and output bandwidths" available to her. Besides the need to develop non-visual techniques to browse data sets so that blind and visually impaired users can access them efficiently, these techniques can also be useful for sighted users under certain circumstances. For instance, situations in which the user's visual channel is busy with some monitoring or controlling task and accessing other visual information could lead to making mistakes, or even be dangerous. Mobile and handheld devices with small screens that are often accessed while walking on the street could also benefit from non-visual data representations. Unlike other media, a radio program cannot display visual data representations to its audience, and doing so with speech only may not be the most efficient way. Hence, these kinds of media could potentially benefit from non-visual data representations, for example to summarise information about weather or the economy. In addition, some data sets, like temporal data and large collections of scientific data, are easier to deal with when presented on auditory displays as Kramer and Walker illustrate with several examples from the literature [86, 164].

Therefore, visualisations alone are not always the best choice for data perceptualisation, and representations in modalities other than visual can be of as much interest for the sighted user as for the visually impaired.

As Vanderheiden [157] observes, it is often the case that information presented through visual display "*is not itself visual in nature, but is simply displayed visually for sighted users*". Instead of producing what Roberts [129] calls "equivalence designs", it is interesting to consider whether certain types of data can be represented through techniques which do not directly translate a visual metaphor (such as a graph or diagram) into another modality since, "*by focusing on an equivalent design, the temptation – and perhaps the consequence – is to recreate the design itself rather than representing the underlying phenomena's aspects*". In this sense, there is, strictly speaking, nothing essentially spatial about a data set except for the fact that to present the raw data visually on some print medium a certain amount of physical space is required. Information visualisation techniques are designed to lay out data in the physical space so that mathematical or logical relations between data points become apparent. Those relations, however, are generally abstract and as such the spatial attributes are intrinsic to the visual representation and not to the data themselves. This thesis investigates the problem of obtaining overview information from complex numerical data sets non-visually, with a focus on tabular data structures which form an intermediate stage between raw data and data representations such as graphs, as it will be justified in Section 4.2.1. As such, non-visual representation techniques are investigated, which exploit the properties of the sensory channels involved (auditory, proprioceptive and cutaneous, and the interactions between them). In doing so, however, the essential spatial attributes of data in visual representations are preserved in order to provide a "common language" across different representations, and thus facilitate collaboration. Although primarily directed to the blind and visually impaired community, the author hopes that some of the findings can benefit sighted users as well.

1.2 Aims of the Thesis

Thesis Statement: *Obtaining an overview is a necessary first step in data analysis, for which current non-visual data accessibility methods offer little support. High-Density Sonification, in combination with proprioception and vibrotactile cues, forms a non-visual multimodal interaction technique that permits a user to extract overview information from tabular data sets efficiently and effectively.*

This thesis investigates the problem of obtaining overview information from complex numerical data sets non-visually. In an attempt to reconcile the spatial metaphors employed in visual data representations with intrinsically non-visual forms of data representation, the work in this thesis focuses on the particular case of tabular numerical data structures, a middle ground between the raw data and the data perceptualisations constructed from them. This problem is addressed by focusing on the following thesis research questions:

RQ-1. How can overview information from tabular numerical data sets be obtained non-visually?
RQ-2. What level of performance can be achieved with the techniques proposed in this thesis?

RQ1 is addressed by combining a literature review with a user-centred design methodology, which results in the proposal of High-Density Sonification (HDS), a novel technique to extract overview information from tabular data interactively. HDS is implemented in TableVis, a multimodal interface in which the user maintains contextual information through proprioception and where vibrotactile cues can be used as external memory aids to annotate data and prevent high mental workload levels.

To address RQ2, a series of empirical studies is conducted, quantifying the performance attained in the exploration of tabular data sets for overview information. This is done by comparing HDS with the main current non-visual accessibility technique (speech synthesis), and by quantifying the effect of different sizes of data sets on user performance. In addition, experimental studies are conducted to discover the capabilities of the human auditory system to extract meaning from auditory representations generated by HDS. Finally, levels of subjective workload during exploration tasks using TableVis are investigated, resulting in the proposal of EMA-Tactons, vibrotactile annotations that the user can add to the data in order to prevent working memory saturation in the most demanding data exploration scenarios.

This work is novel in that it tackles the problem of obtaining overview information from complex numerical data sets efficiently, by taking an intrinsically non-visual approach in the solution proposed, while preserving spatial metaphors that are necessary for collaboration with users that utilize visual representations of the same data. A new data sonification technique is proposed (High-Density Sonification), which is also novel for permitting the representation of subsets of information in single auditory events, which can be compared against each other, revealing overview information and even enabling the user to perform basic statistical analysis of the data by exploiting properties of the human auditory system. In TableVis, the proposed multimodal interactive technique, a further contribution is the concept of EMA Tactons, a form of intrinsically non-visual data annotation which, through structured vibrotactile messages, support the user's working memory during a cognitively-demanding exploration.

1.3 Terminology

The English language offers a rich vocabulary to signify concepts related to accessing information in order to extract knowledge. "Explore", "browse", "analyse", "search", "seek", "navigate", "examine", "look for", and "data mine" are just a few examples, some of which are more commonly used than others in HCI, all with overlapping meanings and each with different connotations. In the literature relevant to this thesis, some of these terms are found to be used interchangeably and possibly subject to interpretation, influenced by previous uses of the terms. To make an informed decision about the term or terms to be used in this thesis, the *Oxford English Dictionary* (*OED*) [154] has been taken as the reference source for English language, and without reproducing verbatim its lengthy entries, the meanings of some of these terms are explained below and commented on in the context of this thesis:

- **Browse**. Its original meaning was to "*feed on the leaves and shoots of trees and bushes; to crop the shoots or tender parts of rough plants for food: said of goats, deer, cattle*". Regarding this first meaning, Cove & Walsh [36] comment that the term "*…carries the connotation of selecting worthwhile and useful information*" (p.31). From the late nineteenth century the term started to be used, by association, in the context of looking through collections of documents or even a single document. In any case, the meaning of browsing involved casually looking around in order to discover items of interest while sampling, especially without knowledge of what to look for beforehand, but without a clear objective or any plan as such.
- **Explore**. To explore implies a stronger sense of purposeful search: "*to investigate, seek to ascertain or find out*" (*e.g.*, to explore a possibility); "*to look into closely, scrutinise*" (*e.g.*, to explore a problem); most recently, it came to mean "*to search into or examine (a country, a place, etc.) by going through it; to go into or range over for the purpose of discovery*". Exploring involves having some kind of planned objective or particular aim for learning about what a system contains, and it leaves room for having a strategy to accomplish that objective.
- **Analyse**. According to the OED, to analyse is "*to take to pieces; to separate, distinguish, or ascertain the elements of anything complex, as a material collection, chemical compound, light, sound, a miscellaneous list, account or statement, a sentence, phrase, word, conception, feeling, action, process, etc.*". This separation of anything complex into its constituent parts, for the purpose of an examination of each separately involves taking a very close up look, retrieving every detail from that system.
- **Seek.** Its main meaning is synonymous of "look for": "*To go in search or quest of; to try to find, look for (either a particular object, person, thing, or place whose whereabouts are unknown, or an indefinite object suitable for a particular purpose)*". It is a purposeful search for something specific known or expected to be somewhere in the system.
- **Search. "***To go about (a country or place) in order to find, or to ascertain the presence or absence of, some person or thing; to explore in quest of some object*". Again, the user looks for something specific, the existence of which in the system searched in know or expected.

The most comprehensive of all the verbs listed above seems to be "*to explore"*, which can encompass both an overview-related approach and a more targeted one to specific information. For this reason, "*data exploration*" will be hereafter used to describe the complete process of approaching a data set and extracting the information relevant to the user. Among the other terms listed above, the most relevant one for the focus of this thesis on the stage of obtaining an overview is "to browse". Examples of uses of "browsing" in the literature that are relevant to the work presented in the thesis are included in the next chapter.

Before ending this section on terminology, it is worth also looking up the term "overview" in the *OED*, as a starting point to define it for this thesis:

- **Overview.** After its original, obsolete meaning of "*inspection, overseeing, supervision*", its 20th century meaning is the one relevant here: "*A general survey; a comprehensive review of facts or ideas; a concise statement or outline of a subject. Also: a broad or overall view of a subject*".

By association, it can be applied to any collection of information as referring to a review or summary of the main features contained within it. The challenge is to define which of these features belong to this general survey and at what point the features are too detailed to be considered part of an overview. This thesis does not provide a general answer to this question, but it does offer insights into the nature of overview and detail in exploring tabular numerical data sets in Chapter 4.

1.4 Thesis Structure

Chapter 2, *Literature Review*, reviews the literature on browsing for overview information. An overview of accessibility technologies for blind and visually impaired users is presented, followed by a discussion of previous research on non-visual techniques for the exploration of numerical data.

Chapter 3, *Requirements Capture*, outlines the user-centred approach taken in the design of an interface to enable blind and visually impaired users to access numerical data non-visually. Feedback from focus groups, interviews and classroom observations is reported. Based on this, a set of requirements for consideration in the design of the solution is presented.

Chapter 4, *High-Density Sonification and its Implementation in TableVis*, starts by discussing the scope of the data sets considered in this thesis. High-Density Sonification is then defined, both in its general form and in its implementation in TableVis. The design process for TableVis is justified, and the results of a pilot study for the evaluation of the interface are presented.

Chapter 5, *Quantitative and Qualitative Evaluations of TableVis and High-Density Sonification*, compares performance in data exploration tasks carried out using High-Density Sonification and speech synthesis. Qualitative data collected from the pilot study are presented, providing an analysis of participants' exploratory strategies and procedures.

Chapter 6, *The Effect of Table Size in the use of TableVis*, presents an experimental study in which the main aim was to measure the effect of table size on user performance and subjective workload. The chapter concludes with a discussion of the experimental results, which provides evidence of the low impact of table size on these two metrics.

Chapter 7, *EMA Tactons* describes data exploration scenarios in which users' working memory reaches saturation. EMA Tactons are designed as external memory aids that the user can add as annotations to the data to relieve cognitive demands. An iterative design and experimental evaluation process is described, which highlights some of the challenges of designing multimodal interaction techniques.

Chapter 8, *Understanding Perceptual and Cognitive Aspects of High-Density Sonification*, reports an in-depth experimental study investigating the reliability with which relative arithmetic means can be estimated by accessing a collection of numbers in a single auditory event with High-Density Sonification. A series of experiments is reported, with which the reliability of this technique is quantified and discussed in terms of data configurations.

Chapter 9, *Discussion & Conclusions*, summarises the work presented in this thesis, and relates the findings back to the research questions outlined in Chapter 1. It also sets out the findings in terms of a set of guidelines. The limitations of this research are identified, and suggestions for future work based on the issues raised by the thesis are proposed.

Chapter 2 Literature Review

The previous chapter introduced the main research focus of this thesis, namely the investigation of non-visual techniques to obtain overview information of tabular numerical data. In order to design, implement and evaluate such techniques, it is important to understand the state of the art in this field of research and to know what accessibility solutions are already available as well as the support they offer for exploring numerical data.

This chapter aims to provide answers to the first thesis research question, defined in Chapter 1: "*How can overview information from tabular numerical data sets be obtained non-visually?*". The chapter therefore begins by discussing the importance of obtaining an overview as part of a general data browsing strategy. The main current accessibility tools available for blind and visually impaired computer users are then reviewed, and the chapter concludes with a review and discussion of previous research on representing numerical data non-visually.

2.1 Browsing for Overview Information in Data Explorations

2.1.1 The importance of obtaining an overview

A fundamental premise in this thesis is that obtaining an overview is an essential early stage in the exploration of any collection of information. This may not have to be done explicitly if the user is already familiar with the information being explored and therefore already in possession of the necessary overview information, but as a general case the user has to be able to obtain such overview information quickly and easily, as it is not practically possible to undertake detailed analysis without setting the scene first. The literature review in this section supports this idea, as it is widely proposed, that obtaining an overview should be the first step in any data exploration. Probably the most influential of such statements in HCI was proposed by Shneiderman in his *Visual Information Seeking Mantra* [138], where this process is expressed as: "*overview first, zoom and filter, then details on demand*", where the function of each steps is summarised as:

- *Overview* (gain an overview of the entire collection);
- *Zoom* (zoom in on items of interest);
- *Filter* (filter out uninteresting items);
- *Details-on-demand* (select an item or group and get details when needed).

Shneiderman's influential statement was preceded in nearly three decades by Bertin [14] who, working in the field of visual representation of information using graphs, stated that general graph-reading operations were characterised by their level or scope, beginning with *Overall* (looking for overall structure), *Intermediate* (identifying trends); and finally *Elementary* (looking for specific, perhaps quantitative, information). Related to this, and also preceding Shneiderman, is the concept of exploratory data analysis of scientific results developed by Tukey [153], who argued that before confirming features of a data set through statistical analysis, it is first essential to produce data overviews (particularly through visualizations such as graphs) which can simply and easily display general trends and highlight anomalies or unexpected results.

Although Shneiderman formulated his mantra for visual information, this principle has since been extended to other modalities, including hearing, where it is asserted that any data exploration should always start by retrieving overview information. Zhao *et al.* [173] proposed an Auditory Information-Seeking Principle (AISP): "*gist, navigate, filter, and details-on-demand*", suggesting that if information-seeking in the auditory mode follows the same pattern as in visual information-seeking, then the collaboration between visual users and auditory users might become easier. In more detail, Zhao defines each one of the stages in auditory information-seeking as follows: *Gist:* Quick grasp of the overall data trends and patterns from a short auditory message; *Navigate*: Fly through the data collection and closely examine portions of interest; *Filter*: Seek data items satisfying certain criteria; *Details-on-demand*: Obtain details of groups or an individual item for comparison. In the overview stage in AISP (which Zhao called "*gist*"), some characteristics of the approach chosen in this thesis are mentioned: an overview should be obtained *quickly*; in the case of an auditory overview, the auditory message should be *short*; finally, the overall information about a data set should inform about *overall trends and patterns* in the data.

Continuing in the auditory domain, Pérez-Quiñones *et al.* [119] proposed another adapted version of Shneiderman's Visual Information Seeking Mantra, for the particular context of exploring web pages using speech: *Situate* (help the user identify his or her location within the information space); *Navigate* (refers to the user's movement within the information space or control of the interface); *Query (*enable the user to issue specific information requests); *Details-on-demand* (refers to user's task where s/he requests more details on specific items). The "situate" stage is an important concept for this thesis, where the user needs to obtain an overview of an unknown collection of information, which depends upon being able to place data features within the context of the whole information space. For the exploration of large spatial data sets using data sonification, Saue [132] makes a simpler two-stage distinction between *orientation* (finding interesting regions) and *analysis* (investigation of these regions). In the exploration stage, Saue distinguishes four levels of detail the user focuses on: *Global* (the whole data set); *Intermediate* (one part of the data set); *Local* (a few data points); *Point* (a single data point).

In all the above models of data exploration processes, the authors propose an initial overview of the data is required, before gradually focusing on the areas that are identified as of interest. This thesis adheres to this consensus by adopting this principle as a foundational guideline to provide non-visual access to tabular data. The requirements capture process that is reported in Chapter 3 provides further confirmation of the importance of this principle in explorations of numerical data sets by visually impaired users. Based on the premise that being able to obtain overview information is a prerequisite to being able to complete a data exploration efficiently, a lack of the appropriate tools and techniques to obtain an overview can be seen as a barrier to information accessibility. This barrier is probably less of a problem for small data sets, or for data collections with which a user is already familiar; in the first case because the selection of relevant data has already taken place, and in the second because the user is already in possession of the overview information. However, once the data sets are of a larger size and higher complexity, and particularly when the user is not familiar with the data, not being able to obtain an overview can be a real barrier to complete the exploration.

In the literature, the concept of *browsing* is closely related to the initial stage of data exploration to obtain overview information. The process of browsing a data set has received the attention of researchers, who have theorised about its purpose and structure. Relevant conclusions from reviewing this part of the literature are collected in the next sub-section.

2.1.2 Browsing data sets

In the literature, browsing is identified as a process which takes place in the early stages of the exploration and provides overview information about the data set. Browsing takes place when a user needs to obtain an overview because she does not know what the data set contains. Thus, an initial overview will highlight the main features of the data and help to distinguish any anomalies or "outliers", allowing the user to make decisions about the data, and about how to proceed with a more detailed analysis. According to Cove & Walsh [36], browsing "*is related to searching where the initial search criteria are only partly defined (…). It is a purposeful activity occasioned by a felt information need or interest. In addition, because recognition is easier than recall (or not knowing), a further way of describing browsing is to say it is the art of not knowing what one wants until one finds it*" (p.31). By this definition, browsing can be considered as a fairly common activity which humans use to explore the array of information available in the physical world, and it makes sense for browsing to be supported for man-made sets of information, as Marchionini [96] suggests: "*Browsing is a natural and effective approach to many types of information-seeking problems. It is natural because it coordinates human physical, emotive, and cognitive resources in the same way that humans monitor the physical world and search for physical objects. It can be effective because the environment and particularly human-created environments are generally organized and highly redundant – especially information environments that are designed according to organizational principles*" (p.100). However, Marchionini also notes, browsing for information retrieval can be subject to high attention demands and possible information overloads, depending on the nature of the task and the user's individual differences.

In an attempt to understand and provide structure to the process of browsing a data set, Kwasnik [90] proposed that browsing is a complex process that involves several functional components. These functional components may occur in any order and number of iterations during the browsing process. This means that browsing is not a passive activity. As Kwasnik says, the user "*is in charge of the direction, pace, and depth of the search*".

- *Orientation*. Answers the question "where am I?", and implies knowledge of the structure of the data set;
- *Comparison*. Occurs throughout the entire browsing process. Values in the data set are compared to each other;

- *Place marking.* Consists of defining "mental landmarks" with potentially interesting pieces of information for possible later re-consideration;
- *Identification.* Confirmation or rejection of the mental landmarks identified during place marking, *i.e.* "Am I still interested in this?"
- *Transition.* Movement from one value to another, in anticipation of a goal. It can also be movement away after identification and rejection, or after success or exhaustion of information;
- *Resolution of anomalies.* People tend to give structure to the data as they browse, even if they do not need to, ignoring odd values that would spoil the structure that the user was expecting to find. It can be seen as a consequence of the wish to find the target.

This set of components will be revisited throughout this thesis, reflecting on the important fact that, as Kwasnik says, these components do not take place sequentially, but can appear in any order and any number of times throughout the iterative processes and sub-processes found during data exploration for overview information.

2.2 Review of current accessibility tools for data exploration by blind and visually impaired users

A wide variety of accessibility tools has been developed to enable blind and visually impaired people to access information. This section reviews the accessibility tools and methods currently available and others in development with the intention of discerning the positive and negative attributes of each with respect to accessing numerical data, and in particular their suitability to provide overview information.

2.2.1 Screen readers, speech synthesis and refreshable Braille displays

Screen readers are currently the most widely-used accessibility tools for blind and visually-impaired computer users. Offering output in both Braille and speech, screen readers gather the content of the computer screen from the computer's operating system, converting this into computer-synthesised speech and/or Braille characters on a refreshable Braille display. The screen reader not only renders the text content of the visual display but also describes elements of the graphical user interface itself, such as scroll bars and windows. Screen reader (SR) applications are available for Microsoft Windows, Apple Mackintosh and Linux systems, however Freedom Scientific's JAWS for Windows [135], GW Micro's Windows-Eyes [62] and Dolphin's SuperNova [40] are three of the most widely-used versions for English speakers worldwide. As reported by Stockman [147], most SR users demonstrate a preference for speech output, utilising either speech-only mode or Braille + speech as the output of the application, rather than Braille alone. Stockman suggests this can be attributed both to the expense of acquiring the hardware components necessary for Braille output and the lack of fluency in reading Braille amongst many blind and visually impaired users. Indeed, Braille literacy was reported to be as low as 10% among the blind and visually impaired population of the United States in the 1980s [2], declining to 9.45% in 1994 [133]. Braille reading rates average 104 words per minute for experienced adult users [52], a rate much slower in comparison to speech output (which enables rates of up to 400-600 words per minute [157]) and to average reading speeds of sighted readers (280 wpm for high school students) [150].

Screen readers are affordable, widely available, can be used with many documents that can be displayed on a computer screen and, particularly when in speech mode, provide fast access to the content of a document. However, while screen readers providing speech or Braille output offer good support for accessing text contained in web pages and word-processing documents, they are less effective in enabling users to access numerical data such as those contained in spreadsheets and other data sets [127] [55]. One of the main limitations of these existing tools is that they are mentally demanding and ill-suited for carrying out the functional components of browsing, as defined above. The screen reader only provides a sequential read-out of the data contained in the spreadsheet. In order to obtain an overview of an entire numerical data set, the user has to access each of the values in the set sequentially, compare each new value with all the previous ones, remembering, at least approximately, both value and position for all the data points. Even using small data sets, the user's short term memory becomes saturated before the browsing task can be completed, since the number of units of information a person can maintain in their short-term memory at any one time is limited to between 5 and 9 [103]. As Ramloll *et al.* [123] observe for the case of tabular numerical data sets, when accessing them using speech synthesis the user's navigational moves produce an overload of speech feedback and an overall picture of the data structure cannot be obtained. A similar conclusion could be expected when accessing numerical data sets using Braille displays [123]. Because of this, as Stockman [147] observes, "*independent examination of spreadsheet data remains cumbersome and cognitively demanding for most speech-based SR users*", thus blind and visually impaired users often require sighted assistance when dealing with numerical data sets.

There are many other accessibility tools based on speech technology, as reviewed by Freitas & Kouroupetroglou [55]. These range from simple speech output devices for everyday needs, such as talking thermometers, clocks and calculators, to handheld electronic text reading devices (also known as digital talking book players) which convert electronic text files to speech, where the files are stored on removable flash memory cards or downloaded from personal computers or the Internet). The most advanced versions of these devices, such as the Dolphin Easyreader [39] include the ability for the user to bookmark and annotate points of interest within the text, which can then be "jumped" to on demand. Automatic reading devices for accessibility to printed material are also available in both desktop and handheld versions. These are normally "standalone" devices which do not need to be connected to a computer. They scan printed documents to capture an image of the text which is then processed by an optical character recognition application and rendered into synthesized speech. They are designed for scanning normal continuous text and as such do not handle data tables well.

2.2.2 Screen Magnifiers

Screen magnification software interfaces with a computer's graphical output to present enlarged screen content. Users require some functional vision in order to be able to utilize this accessibility tool. Large data sets (such as large spreadsheets) are difficult to navigate, so the user is unable to obtain an overview and can get "lost" in the data [127]. A similar problem is encountered with the small display screens on mobile telephones and other mobile devices [19]. Much of the difficulty is due to the lack of context for the user to locate the information under focus, which makes it difficult to implement the functional components of browsing presented in the previous section.

2.2.3 Braille print-outs, raised paper and embossing film

Data tables can be printed out using Braille transcription software (such as MegaDots [45], Duxbury Braille Translator [44] or Braille Maker for Windows [3]) which converts text documents, spreadsheets and other common computer documents into Braille code, and a Braille embosser or printer which prints the Braille onto special paper. Tactile versions of graphs can also be rendered in Braille as raised paper diagrams. Braille print-outs are portable, and they allow two-handed exploration of the document, whereby users can "place-mark" points of interest while retaining a sense of the overall "shape" of the document – also an important "affordance" of paper for sighted readers noted by Sellen & Harper [136]. However, Braille print-outs are less appropriate for large data sets, as it is often not possible to fit all the data onto one sheet of paper. For instance, large graphs or data tables may need to be printed over multiple pages or split into smaller sections. This complicates the user's two-handed exploration of the data sets and creates problems for obtaining an overview of the data and keeping detail information in context. Another important disadvantage is that such print-outs are inherently non-dynamic, and are thus less appropriate for users needing to work with numerous or dynamically-created data sets.

A further drawback is that, while Braille print-outs themselves are relatively cheap, the cost of purchasing the Braille embosser hardware can be prohibitive for individuals. The production of tables, charts and graphs can also be difficult for those unfamiliar with the transcription software, such that the RNIB advise inexperienced Braille transcribers to seek professional assistance when reproducing graphical or complex information [128]. This need to rely on professional transcription and print services, or to gain access to an embosser separate to one's home or place of work, is impractical for those dealing with numerical data sets on a regular basis.

Microcapsule paper (also known as "Minolta paper", or "swell paper") and plastic embossing film (also known as German film) are commonly used to produce tactile graphics, and to enable blind and visually impaired students to produce diagrams and maps as part of the school curriculum. Plastic embossing film is used by placing it on a rubber mat and drawing on it with a ballpoint pen or special embossing tools (known as spur wheels) to create a raised line, which can be read by touch. Graphs can also be produced using embossed graph paper (showing gridlines) with a cork geometry base (into which pins can be pushed) and tactile geometry tools (See Figure 1). This can be quite a laborious process which often requires sighted assistance, and graphs produced in this manner cannot easily be stored for later use. In addition, the pins and tools needed for this method are sharp and can cause injury.

Microcapsule paper has microcapsules of alcohol embedded in the paper which burst when exposed to heat and make the surface of the paper swell up. Graphs can be printed or photocopied onto the paper, which is then passed through a "fuser" which activates the marked area with heat and thus produces the raised image. The actual graphing of the information produced using microcapsule paper is not modified from the visual form, and as such a bar graph or a line graph will have the same form as presented on a computer screen or on any print media, where the relative size of the bars represent relative numerical values, and the variations in the slope of a line graph represent variations in the trends of the magnitude represented. Some adaptation of the visual version of a graph for printing on swell paper can take place. For instance, colour coding on bar graphs is converted into different textures that fingers can discriminate, and text in the title, labels on the axes, scaling, legends etc. are converted into Braille.

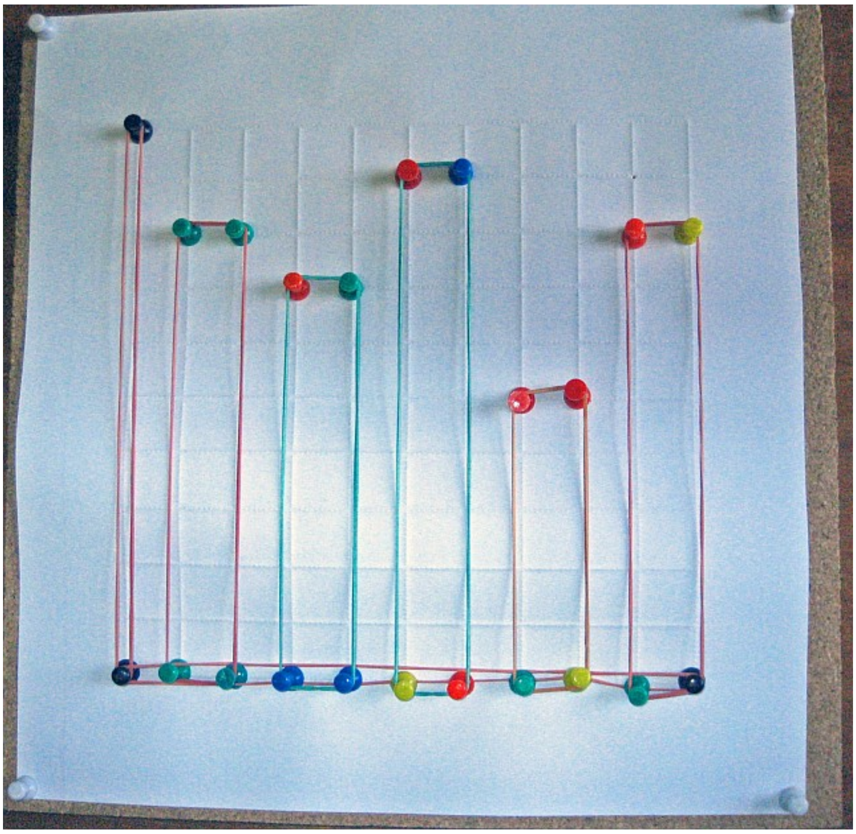


Figure 1. Bar chart constructed with embossed graph paper, pins and rubber bands. From [99].

One advantage of tactile graphs is that they can be explored with both hands – a natural way of exploring the physical environment and a very rich source of information. As a result, an overview of the information in print can be obtained. Once more, one of the biggest drawbacks of this method is that it is inherently non-dynamic. In supporting Shneiderman's information seeking mantra, explicit zooming and filtering to explore the graph requires the production of new print-outs, which is quite impractical. Details-on-demand are difficult to obtain, as detail information is written in Braille, which takes up a lot of space on the paper. Often, Braille key-characters (rather than full-text labels) are written on the axis of the graphs, and these key-characters must then be looked up in a list with the full Braille text. Figure 2 shows an example of a bar chart printed on raised paper, with Braille text.

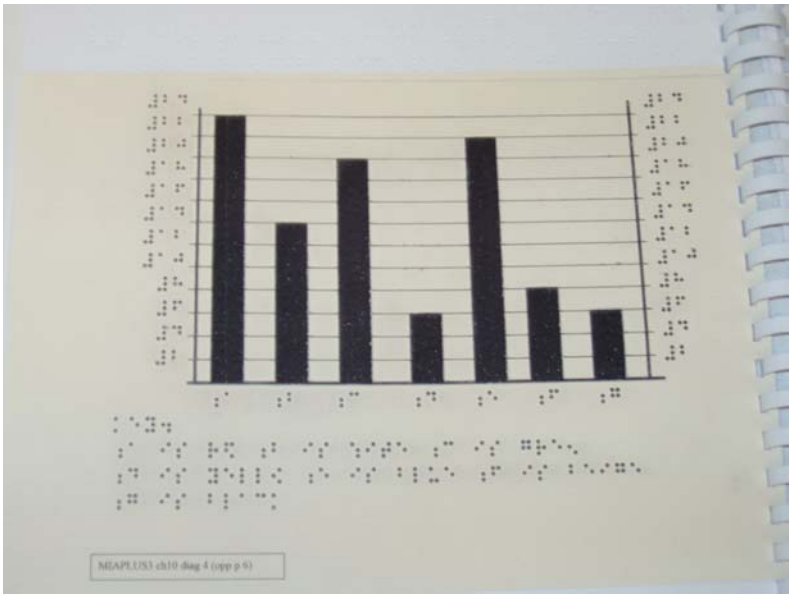


Figure 2. Example of a bar chart printed on raised paper. Under the graph, a list of the labels of the bars can be seen, each label related to the corresponding bar by a key-character that the user can look up. From [101].

2.2.4 Haptic technologies

Accessibility devices using haptic technologies include tactile devices, which appeal to the cutaneous senses of the skin, utilizing pressure, vibration or temperature to stimulate the skin, and force-feedback technologies which stimulate the kinaesthetic sense by presenting computer-controlled forces to create an illusion of contact with a rigid surface, perceived through the muscles, tendons and joints.

Based on the widespread use of raised paper tactile images, Parkes developed NOMAD [115] and Tactile Audio Graphics kit [116] to enable blind and visually impaired users to explore traditional tactile diagrams and maps augmented with audio information, with the aim of improving graphicacy[1] and mobility. The Royal National College for the Blind's "Talking Tactile Tablet" (T3) is a similar tool, comprising a touch-sensitive device resembling a graphics tablet, which provides instant audio feedback from tactile images [92]. The device connects to the user's computer, and is activated by the user touching a tactile diagram overlay which is placed on the surface of the tablet. When a user presses on various parts of the tactile diagram they hear appropriate descriptive audio feedback. These methods combine the affordances of raised paper with speech synthesis to enable the exploration of static content contained on customised CD-ROMs (atlases, encyclopaedias, and topics relating to the school curriculum). None of these tools currently offer access to dynamically generated content, as they must be programmed to relate to pre-determined data sets which are commercially supplied.

Extending the metaphor of raised paper, but addressing the issue of rendering dynamic content, some researchers are developing tactile graphics tablets which utilise electro-rheological or magneto-rheological displays, whereby a thin layer of fluid or gel reactive to electricity or magnetism is incorporated into a graphics tablet-like base. When exposed to small electrical or magnetic fields on the surface of the tablet, for instance

[1] Capacity required to generate and interpret information in the form of graphics.

corresponding to the lines of a diagram, the fluid changes into a solid state, producing a refreshable tactile graphic display which can be used in conjunction with a user's personal computer [56]. Such technologies are still in their infancy, however, and are not commercially available.

A more widespread haptic technology are pin array displays. Pin arrays are accessibility devices in which an array of small pins stimulates the user's fingertip. They are most commonly used for presenting Braille text in dynamic Braille displays, a line of 40 or 80 cells each with 6 or 8 pins which move up and down to represent the dots of a Braille cell. Pin arrays are capable of presenting fine cues for surface textures, edges and lines for graphical information. These have been adapted to enable users to navigate through a computer screen, and thus explore larger tactile images, by mounting the pin array on a mouse or graphics tablet. The user explores the image displayed on the computer by navigating with the mouse or moving their hand across the tablet, and the pin array dynamically presents the information to the user's fingertip. Only a small fraction of the whole image is presented on the pin array display at any one time, the size of which depends on the number of pins in the array. Examples of commercially available pin arrays include Freedom Scientific's Focus Braille Displays [134] and VirTouch's VTPlayer mouse [159] (see Figure 3).

Figure 3. Close up of a pin array from VirTouch's VTPlayer mouse [159].

Force-feedback systems are another form of haptic technology adapted for use by blind and visually impaired users. Although originally developed for computer-aided design and industrial applications, these also offer applications for blind and visually impaired users by exploiting the kinaesthetic senses. SensAble Technology's PHANTOM Haptic Device is one of the more widely available desktop force feedback devices [137]. The device is operated by moving a pen-stylus in a 3D workspace. The stylus is affixed to a mechanical arm powered by electrical motors which generate forces to simulate the user's interaction with virtual objects. The device can simulate kinaesthetic but not cutaneous sensations, and the single point of contact of the stylus means that there is a significantly reduced "bandwidth" of sensation in comparison to the body's own wide array of haptic receptors [172]. A related problem is that the stylus can "slip off" the object being explored and it can be difficult for blind and visually impaired users to locate the object in the 3D space. Researchers have sought to improve this through "virtual fixing", utilising orthogonal forces to guide the user's stylus back to the object, somewhat like a magnet [12, 57]. Research has also resulted in useful guidelines about how to present information haptically to minimise slip-off problems. For instance, if the user is expected to follow a linear trajectory on a virtual surface with a PHANTOM device it is easier to follow the line if this is "engraved" in the virtual surface than if it is "embossed" [172]. Currently the PHANTOM and similar devices are expensive for individuals to purchase (although some devices cost less than a Braille display), and this, combined with the limited interaction method means force-feedback is not widely used as an accessibility tool. Such devices are also non-portable, and susceptible to mechanical failure, potentially requiring expensive maintenance.

Another form of tactile display currently being researched is vibrotactile representation. Brewster and Brown [22] proposed Tactons, or tactile icons, which are structured, abstract tactile messages used to communicate complex concepts to users non-visually. Tactons exploit the basic parameters of cutaneous perception, such as waveform and rhythmic patterns, to encode information. Brown *et al.* [26] found users could recognize different Tactons with an accuracy rate of 71%. Brewster & King [20] found simple Tactons could be successfully used to encode information for a vibrotactile progress bar for blind and visually impaired computer users.

2.2.5 Auditory Displays using non-speech sounds

Auditory displays utilising non-speech sounds form the basis for another accessibility tool that is currently not widely implemented. The screen reader JAWS for Windows v.5 (2003), includes a facility for non-speech sound which can be customised to provide feedback to users about the interface itself, such as indicating upper or lower case text, paragraph formatting, and the focus of the browser window [147]. A broad range of sonification techniques for non-visual access to numerical data are being researched, which will be discussed in Section 2.4. However, few of these have been implemented in commercial interfaces. Frauenberger *et al.* [54] note this gap between research and practice. Surveying views on the use of audio in the user interface amongst 86 HCI interface designers, they discovered that only speech synthesis was considered as an audio interface, and non-speech sounds were generally viewed as potentially annoying for users and not considered a “serious alternative” to the visual channel.

Opposing this generalised view, the next section includes a discussion of the research being conducted in the field of the auditory display and the various non-speech-sound-based techniques developed to access information. It also reviews research on accessibility techniques utilising haptic modalities.

2.3 Current Research on Non-Visual Presentation of Numerical Information

Researchers in HCI and related disciplines have approached the problem of accessing numerical data by means other than visual representations, generating research questions and knowledge that are part of an actively ongoing process. Often motivated by the need to provide accessibility solutions for the visually impaired population, many of these techniques make use of accessibility tools and technologies reviewed in the previous section. Attention is turned in this section to reviewing the most relevant contributions from ongoing research, in order to learn about their strengths and limitations, and the support that they offer to obtain an overview of numerical data.

2.3.1 Haptic Access to Numerical Information

Data Representations Using Force-Feedback

Research has mostly approached the problem of representing data in haptic modalities by using equivalence representations of the data, *i.e.* by accessing data visualisations rather than specifically developed haptic representations. Haptic graphs generated using force-feedback haptic devices offer a way of presenting visual graphs to the kinaesthetic sense dynamically, that is, the computer which generates the haptic representation can update the view presented to the user. Furthermore, the user can modify the haptic graph by applying forces to it, and changes in the underlying data can be recorded by the computer in real time. The most notable device used for this is SensAble Technology's PHANTOM device, described in Section 2.2.4. Graphs are generated dynamically from numerical data contained, for example, in a spreadsheet. Then, the haptic graphs are accessed in a virtual 3D space in which all the features of a visual graph can be felt. Context is maintained through proprioception within the working space, aided by virtual walls and frames which delimit the work space and lead the stylus towards the haptic graphs by providing physical constraints to the movement of the hand.

Some of the most extensive work in this area has been carried out by Yu & Brewster, whose research has covered several types of graphs, including line graphs and bar graphs, explored using PHANTOM force-feedback devices. Their early studies found that retaining a sense of context when exploring line graphs with the PHANTOM was difficult for users and it was hard to find the maximum and minimum values on the graph [172]. A later iteration of the interface was developed and evaluated where virtual line graphs were explored as "grooves" to enable the user to retain the pointer on the line [171]. Columns in bar graphs were also represented as vertical grooves. By exploring graphs with the PHANTOM, users were able to detect trends in the data and make comparisons between different variables more accurately than by exploring the same graphs using raised paper. Haptic feedback was found to be useful for assisting users to navigate the data representation. However, the time taken to extract information was significantly longer and the perceived workload was significantly higher than with a raised paper version. This was attributed to the single point of contact interface provided by the PHANTOM, which led to a reduction in perceptual "bandwidth" (in comparison to two-handed exploration of raised paper) and thus a less time-efficient way of exploring the data. They found that introducing speech and non-speech audio to provide abstract and detailed information on the data series was an effective way to improve performance.

In a similar vein, Fritz *et al.* [57] developed a "Haptic Graphing" technique using the PHANTOM device to allow blind and visually impaired users to explore the plot of two- and three-dimensional functions, and discrete data points in space. Context was provided through the use of a haptic grid, with thin "walls" generating very small forces, intended to enable the user to feel the grid without confusing this with the data structure itself. Synthesised speech provided coordinates on demand. Unfortunately no evaluation of this interface has been reported. Van Scoy *et al.* [155] developed a similar technique to enable users to feel the graph produced by a mathematical equation, by following the curve of the function displayed as a "groove" in a virtual haptic surface using a PHANTOM stylus. In a further iteration of the design, musical notes were played to provide further information about the curve of the graph, with pitch indicating the relative numeric value of the position at which the stylus was pointing [156]. No formal evaluation of this interface has been reported either.

McGookin & Brewster [98] developed GraphBuilder to support the creation, manipulation and collaborative access to bar graphs by blind and visually impaired users, also enabling remote collaboration. The system utilised the PHANTOM device to enable users to browse and modify bar graphs, supplemented by speech and non-speech sound feedback. When a bar is selected, moving the stylus up or down modifies the height of the bar, and pitch-mapped musical notes are played to indicate the number of units the bar has increased or decreased. Providing audio feedback to the haptic interface was found to be beneficial in enabling users to keep track of the interaction and review the accuracy of the completed task. On the basis of this evaluation, McGookin & Brewster proposed a guideline for designing haptic interfaces for the visually impaired, that all "touchable objects" in a haptic interface should provide audio feedback. In a separate study [100], they investigated concurrent presentation of two line graphs in haptic and auditory modalities, but found that participants were distracted by the different modalities and were thus unable to compare graphs presented in this way. However, auditory feedback did assist the users in retaining context, while use of the force-feedback device gave users a sense of control over the sonification of the data.

As will be argued later in the thesis, a 2-dimensional (2D) data table can be understood as a collection of as many data series as rows or columns the table contains. Each data series (each row or column in the table) can be represented as a graph. Visually, 3-dimensional (3D) graphs (also known as *stereogram*s, resembling mountainous landscapes) are often produced to represent all the data in a table. Wall & Brewster [167] constructed a haptic table-browsing interface that converted the data in the table into a virtual 3D surface, which could then be explored (non-visually) using a PHANTOM force-feedback device as well as sonification of values in the cells. They provided external memory aids called "beacons" to enable users to mark values of interest. These showed potential usefulness, although evaluations and further work are needed in this direction. This interface permitted participants to freely explore the whole 3D surface, without assistance in constraining the exploration to a single data series in the data set.

Data Representations on Tactile displays

A similar approach of retrieving the visual data representation (accessing visual graphs) has been taken by research work that has incorporated tactile displays, specifically pin-array displays. Tactile displays can have similar limitations of narrow focus when the size of the display is small, as in the case of pin arrays found on devices like the VT-Player mouse. The tactile display acts as a small window through which small fractions of a graph can be accessed. Thus, the trajectory of the hand that follows the length of a visual graph to show it on a tactile display can perceive information about the characteristics of that graph via proprioception. The information presented in the display at any one time is, however, difficult to interpret. The fraction of information displayed on the graph is too small to provide any contextual information in itself, and it is therefore difficult to provide overview information. Wall & Brewster [168] investigated improving access to graphs using small tactile displays by combining the display of information with the exploration of the workspace on an absolute pointing device (a graphics tablet). Thus, the dominant hand controlled the navigation by selecting a specific area of the graph to be displayed, and this movement of the hand provided proprioceptive information that helped keep the navigation in context. At the same time, the non-dominant hand could feel the information on a narrow area around the point under focus. Detailed information (such as absolute values corresponding to the fraction of graph under focus) was available on demand at any time in the form of speech. However, in terms of facilitating means of obtaining a quick overview the limitations were similar to those posed by explorations with force-feedback devices.

2.3.2 Auditory Displays to Present Numerical Information

In the auditory modality, sonification techniques have been used successfully to present numerical information non-visually. Put simply, sonification is "*the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation*" [87]. Sound offers a large number of attributes which can be perceived by humans, including pitch, loudness, timbre, time (duration of notes and rhythm) and direction (location in 3D space) [110]. By utilising different combinations of these attributes to encode or represent information, it is possible to develop multi-dimensional auditory displays, capable of representing large or complex numerical data sets. For example, Bly [16] showed that six-dimensional information could be encoded in non-speech sounds, while Ramloll & Brewster [122] and Brown *et al.* [25] found that it was possible to browse multiple data series in parallel when these were presented in 3D space around the listener. The multidimensionality of the auditory modality has also led to the development of research into sonification of other kinds of data including dynamic real-time financial data [74], images [53] and even mapping the movements of fish in a public aquarium [163].

The following types of sonification have been distinguished in the literature:

- *Audification*: the direct translation of data into a form audible by humans, commonly applied to geophysical or astrophysical waveform data [41, 86].
- *Parameter Mapping*: features of a data set are mapped to pitch, timbre or other acoustic attributes [66], as in auditory graphs, discussed below.
- *Auditory icons*: Originally developed by Gaver [58], auditory icons are "*everyday sounds meant to convey information about events in the computer by analogy with everyday events*" (p.3), and are often used to relay information about the computer's operating system to users.
- *Earcons*: First proposed by Blattner [15], Earcons are symbolic sounds composed of short, rhythmic sequences of pitches with variable timbre, loudness etc. which can be used to communicate complex information about computer objects, operations or interactions.
- *Model-based sonification (MBS)*: Proposed by Hermann and colleagues [67, 70], MBS draws on real-world interaction with objects, where applying forces or manipulating them creates a sound which conveys information about the properties of the object to the listener. In this way, complex data sets could be transformed into a virtual object, where "*data more or less directly becomes the sounding instrument, which is examined, excited, or played by the listener*" (p.190) [67].

In line with the Auditory Information Seeking Principle (AISP) proposed by Zhao *et al.* [173], some recent sonification research has focused on the need to obtain an overview of complex data sets. A number of authors have argued for the importance of this approach in research on auditory display. Harrar & Stockman [63] have commented that the "*presenting overviews is a very under used concept when it comes to auditory displays and even more so when applied to auditory graphs*" (p.299), and suggest that more work is needed on the optimum way of sonifying graphs so that users can obtain a useful overview. Stevens [144] notes that screen reading technology is useful for extracting fine detail but insufficient for obtaining an overview, and proposes the development of an "*auditory glance*" to enable a rapid overview of a data set. Brown [23] concentrated on the problem of obtaining an auditory "glance" of abstract graphs such as molecular structures and flow charts. Obtaining overview information from other types of data collections with complex structures and relations, like computer program code, has also been approached via sonification, attempting to make use of the high pitch and temporal resolution of human audition to identify patterns in the code [11, 49, 158]. Nickerson & Stockman [113] also researched new approaches to obtaining overview information from collections of items, by exploring the objects on a computer desktop presented in a 3D sound environment.

Edwards *et al.* stressed the importance of the user being in control when browsing data through non-visual interfaces, in order to avoid information overload [46]. Stevens [144] also emphasises the importance of control of information flow in accessibility interfaces, proposing that this should be incorporated into the design principles for such interfaces in order to ensure blind and visually impaired users can engage in active data exploration (rather than passive listening). Hunt & Hermann [70] have extended this to emphasise the importance of interaction, to reflect properties of real-world interaction between humans and the objects in their environments. Interactive sonification has been defined as "*the use of sound within a tightly closed human-computer interface where the auditory signal provides information about the data under analysis, or about the interaction itself, which is useful for refining the activity*" [66] (p.20).

One line of sonification research has investigated ways of using musical sounds to provide auditory representations of data visualisations and pictorial diagrams. Alty & Rigas [125] investigated the use of music to communicate diagrams to blind users. They continued this line of research with AUDIOGRAPH, and found that context was crucial for exploring complex diagrams with non-speech sounds [6]. Kennel [78] developed a similar "diagram reading" application, to allow audio-tactile exploration of graphs and diagrams, where diagrams are displayed on a touch-sensitive glass panel connected to a computer and are explored by the fingertips. Pressing on part of the diagram elicits an "auditory symbol", or alternatively a speech description, to indicate the kind of graphic element encountered. Kennel found that, after training, users were able to read and accurately describe a simple diagram within three minutes. Cohen *et al.* [33] developed the PLUMB system to enable blind students to actively explore diagrams using continuous tones with varied pitch and loudness to guide the user around the vertices and edges of the diagram, and synthesised speech feedback to describe elements of the diagram. As yet, only limited work has been undertaken to identify how other forms of diagram such as pie and bar charts can be sonified [53].

A key focus of research utilising parameter mapping techniques is the development of auditory graphs. Since Mansur proposed his *Sound Graphs* as a method to sonify line graphs [95], data sonification has been used extensively and successfully to perceptualise series of numbers non-visually, as an alternative to the construction of data visualisations such as line graphs and bar charts. To construct auditory graphs, discrete data points are mapped to properties of sound, and rendering them sequentially conveys a clear sense of the shape of a graph, conveying relative rather than absolute values, just as with visual graphs. While a number of different sound parameters can be utilised to represent data values (as will be discussed in more depth in Section 0), experience and research have shown that pitch is the most accurate parameter of sound to be used for the mapping of numerical data, due to its high resolution [50]. Flowers *et al.* [51] found a high correlation between participants' ability to make judgments on data which was presented via auditory and visual graphs. Edwards et al. [46] developed "Soundgraph" to facilitate access to and creation of simple line graphs for blind and visually impaired users, where pitch is mapped to the height of the curve in a line graph. The user is able to move backwards and forwards along the curve to locate points of interest, details of which can be obtained through a speech read-out of the x and y coordinates. The program also permitted users to create their own graphs by entering a function, drawing the graph, or importing data from a spreadsheet or other data source. Brown *et al.* [25] and Ramloll & Brewster [122] worked on the problem of exploring two and three data series respectively, finding in both cases that concurrent presentation of two audio graphs is an efficient way of finding intersection points. Several authors have collected much of the current knowledge on auditory graphs in the form of guidelines or graph creation toolkits. Thus, Brown *et al.* [28] and Flowers [50] released sets of guidelines, highlighting which techniques work and which ones do not. Additionally, Walker & Cothran [162] developed "Sonification Sandbox", a sonification toolkit enabling users to generate auditory graphs using sound-parameter mapping, permitting the implementation of a proven successful technique as well as flexibility to experiment with new ones. A design principle that enjoys general consensus is that designing auditory graphs should exploit human high temporal and pitch discrimination resolutions. Thus, for single data series (or single-variable $y=f(x)$ mathematical functions) that would be presented visually as a line graph or bar chart on a Cartesian reference frame, it has become the norm to map values on the x axis to time and the corresponding values on the y axis to pitch of sound.

As discussed above, a number of authors have focused on the access of two or three data series concurrently. Data tables constitute a generalisation of this problem, since a data table can be seen as a collection of graphs (see also the discussion in Section 2.3.1). To obtain overview information from larger collections of data series (2D data tables), Ramloll *et al.* [123] observed that using only non-speech sounds was preferred over using a combination of speech and non-speech sounds, which produced higher mental workload (when the extra level of information provided by the speech was not strictly necessary). However, even utilising only non-speech sounds, extending the single-graph sonification metaphor to collections of graphs resulted in complex and demanding exploration tasks. In an attempt to reduce the perceived complexity in exploring collections of data series (of 2D numerical data tables), Kildal & Brewster [80] investigated accelerating the rendering of the data points in each data series to rapid successions of sounds. Instead of traversing the data points in each auditory graph and retrieving fine detail, the traversing of each graph that made up the complete set occurred automatically at a fast speed, sonifying a data point every ~60ms. Experimental evaluations showed that much of the detail about the inflections in the melody line of the auditory graphs was not perceived by the users when they were rendered rapidly, and only the overall trend, general shape of the data series and the presence and approximate location of outliers were perceived clearly. In other words, with fast rendering of audio graphs, fine detail was filtered out, while overview information became apparent. Since fast rendering resulted in playing the whole series of data points in a graph in a short time (the graphs with 10 data points used in that study where played in 0.6 seconds), each rapid auditory graph was perceived as a single (yet complex) auditory event conveying a relatively simple amount of information. With this simplification of the information perceived, the study showed that it was possible for the participants to access sequentially all the rows in a table (each row being sonified as a rapid auditory graph) and compare the information retrieved from all of them, providing an overview of the whole data table. The same process of sonifying a table as a collection of rapid auditory graphs can be applied to the rows or to the columns in the table. The effect rendering speed has on filtering out detail and revealing overview information will be taken as one of the bases for the solution proposed and investigated in this thesis.

Given the fact that tables are the underlying paradigm for generic software applications as popular as spreadsheets and databases, it is not surprising that there is specific research going on about how to access these applications more easily in a non-visual way. In this direction, Stockman and colleagues [147, 148] are developing sonification tools to enhance one of the most popular screen readers, JAWS, in order to improve accessibility for visually impaired users. Their approach utilises various mechanisms to support interactive sonification of spreadsheet data, including automatic sonification of rows and columns and manual sonification utilising a sonification cursor, which when activated would sonify the values in the cells as the user moved the cursor across the spreadsheet using the keyboard. Maximum and minimum numeric values were automatically calibrated to the available frequency range for sonification, but users could set boundaries around selected areas in the spreadsheet, in which case the system recalibrates the frequency range to the maximum and minimum values within that area, creating an "auditory lens". Users could also mark areas of interest within the spreadsheet by using "bookmarks" for up to 4 cell ranges. Most users found both automatic sonification and manual sonification fairly straightforward to understand and utilise for exploration. Manual sonification was found to be particularly useful for exploring numerical data, where the intention was to detect general characteristics of the data, outliers or patterns, rather than precise values. The bookmarks mechanism was found to be useful in this instance, as it allowed the user to mark a point of interest without specifying cell coordinates, for easy retrieval for further analysis.

Other work on sonification of tabular data includes research conducted by Smith *et al.* [140] in the area of automated information processing, who used various parameters of sound to augment iconographic visualisations of large data sets distributed on a surface in quasi-tabular form, and help with the identification of patterns. In another example of somewhat irregular tabular structures, Zhao *et al.* [21, 94] investigated non-visual accessibility to geo-referenced data (data presented on a geographic map of the United States, divided into 51 states), assimilating the map with a quasi-tabular structure. The fact that the map was not a regular table was a source of confusion for the participants in the evaluations. Zhao reported that the best strategy with that interface was to be already very familiar with the geography shown in the map (the names of the states and where they are supposed to be) and to find some key states using the speech synthesiser, to create a mental image of the map. This conclusion suggests that the method utilised in this study cannot be generalised to other cases, as knowing the specific layout of the particular map being explored seems to be a prerequisite.

2.4 Conclusions

This chapter has presented a review of the literature of the main fields that are relevant to the research work presented in this thesis, with the aim of providing answers to RQ-1, which asks "*How can overview information from tabular numerical data sets be obtained non-visually?*". This review began with a discussion about the importance of obtaining an overview as part of a general data exploration strategy, and how the action of browsing has been studied as central to obtaining overview information from a collection of information. Following this, the main current accessibility tools available for blind and visually impaired computer users were considered, highlighting the main benefits and limitations in each case. In the final section of this chapter, current ongoing research techniques to access numerical information non-visually were reviewed, paying special attention to their support of the process to obtain overview information.

This review revealed a consensus in both the data exploration and accessibility research literature about the need to obtain an overview as the first step to analyse or simply explore any set of information. The process of browsing information, linked to extracting an overview was considered, and the discussions on its importance and structure made by various authors were included. Among these, the *functional components of browsing* were reported in more detail, for the guiding importance that they have in some of the coming chapters in this thesis.

When reviewing both current accessibility tools and technologies available for visually impaired users as well as new accessibility techniques for numerical information that are under development, the need to provide both context and overview information was a concern which appeared repeatedly. Some of the technologies and techniques had *a priori* more limitations in enabling users to access overview information in a quick and easy way. One conclusion that emerged was that the properties of the auditory system can be utilised to provide overview information from collections of data rendered in sound. The high temporal resolution and fine pitch discrimination capabilities that the auditory system offers have been used successfully for this purpose and could provide the key to obtaining at least part of the synoptic properties of human vision, at the time of obtaining a summarising "glance" of a collection of information. This is further supported by the fact that many researchers working on haptic techniques for data representation have also found it useful to implement some form of data sonification, to assist users in obtaining an overview.

Building on the knowledge and awareness of the current research work presented in this chapter, the next chapter captures the main requirements for non-visual overviews of numerical data sets from the end users themselves, which in turn forms the basis for the design and implementation of TableVis and the High-Density Sonification method developed by this thesis.

Chapter 3 Requirements Capture

3.1 Introduction

While there are several scenarios in which sighted users could potentially benefit from non-visual access to tabular data sets (for example, in scientific research, industrial applications, and many data-mining tasks where vision is limited or where an overview can more easily be conveyed through use of an auditory display), the current thesis concentrates primarily on responding to the needs of the blind and visually-impaired community. As discussed in Chapter 1, blind and visually impaired users often encounter barriers to accessibility of numerical data, which can lead to them being hindered in, or even prevented from, progressing in mathematical and scientific disciplines, whether at school, in higher education or during their professional careers. Chapter 2 reviewed a range of approaches to non-visual access to numerical data, and found that many currently available accessibility technologies were not ideal for accessing overviews of numerical data.

In order to apply a user-centred approach to the design of techniques that can tackle this need, representatives of the end-user population and the main stakeholders were actively involved in the process of requirements capture for this thesis. The information obtained through this process thus provides further answers to the first research question posed by this thesis: "*How can overview information from tabular numerical data sets be obtained non-visually?*" (RQ-1). This chapter reports the feedback gathered from this requirements capture, and presents the resultant requirements which need to be taken into account in the process of decision-making for the design on a prototype solution. The design process itself, together with the results of initial evaluations, is described in Chapter 4.

The main aims of the requirements capture process were to establish:

- whether members of the community of blind and visually impaired computer users need to, or wish to, access tabular numerical data in their everyday activities, particularly in their education;
- tools and methods currently used to access such data, and strategies employed by users;
- user opinions on current accessibility technologies, including benefits and difficulties;
- key requirements for accessibility to numerical data sets and user opinions on how to improve tools to access such data.

In order to answer these questions, it was essential for the research to be carried out with the assistance of and advice from blind and visually impaired users, as well as those who are involved in their education. For this purpose, a network of contacts with institutions for the blind and visually impaired was developed, where ideas could be discussed and feedback received, and which enabled the testing and evaluation of prototypes with key stakeholders for an iterative process of improved designs.

This chapter reports the feedback obtained from stakeholders through focus groups, interviews and classroom observations held with blind and visually impaired students. Further requirements identified from the literature and through informal discussion with stakeholders are then discussed. The chapter concludes with a summary of the main requirements identified, which informed the design of TableVis (reported in Chapter 4).

3.2 Stakeholders

Two groups of people were identified to be the main stakeholders in this research work:

- *Blind and visually-impaired computer users*. Users with many different levels of computing skills were involved in the requirements capture process, from individuals that used basic web-browsing and word-processing techniques to students of disciplines in which the computer played a central role, and professionals in computing-related fields.
- *Educators of children and young people with visual impairments*. Through classroom observations and dialogue with teachers, views and experiences of educators of blind and visually impaired people were obtained.

The process of requirements capture was conducted in collaboration with three British institutions for people with visual impairments: the Royal Blind School of Edinburgh, the Scottish branch of the Royal National Institute for Blind People (RNIB) in Edinburgh, and the Royal National College for the Blind (RNC) in Hereford. Several techniques were employed to find out about users' and professionals' opinions and needs:

- Focus groups and interviews were conducted with blind and visually impaired students to discuss issues about using data visualisations in general, and data tables in particular, and to gather information about the kinds of accessibility technologies used;
- Three hours of lessons with blind and visually impaired students were observed: two classes with a single student, reviewing graph understanding and mathematical subjects; and one group tutorial on algebra.

In addition, a number of informal discussions and presentations were conducted with professionals from the RNIB and RNC.

3.3 Focus Groups and Interviews

Focus groups and interviews were conducted with students from the Royal National College for the Blind in Hereford. These involved eight blind participants, two in individual interviews (Participants 1 and 2), and six in focus groups (Participants 3-8). They were aged 18 to early 30s, but were mainly in their late teens and early twenties. Three were male and five were female. Two of the participants had jobs and the other six were students, four of whom were studying music technology. Although all the participants were totally blind, some of them had had partial vision in their earlier years. The focus groups were prepared beforehand following guidelines from Morgan [108], Krueger & Casey [88] and Kuniavsky [89]. They were designed to explore the following aspects of using graphs and tables:

- *Experiences of accessing graphs and tables in everyday life:*
  Participants' experiences of using graphs and tables; in which subjects they encountered them at school or college; whether they encounter them outside educational contexts, in their everyday lives; their appreciation of the usefulness of graphs and tables, and whether these help or hinder their understanding of the topics that use them to present information, and why; problems encountered and participants' feelings about this.
- *Tools used to access graphs and tables, and benefits or drawbacks of these:*
  A description of the tools participants use to access graphs and tables; how they use those tools; the strategies they follow and whether they are dependent on the tasks; which tools they consider to be best or worst, and why; identification of any tools that they have not used but they have heard of, or that they can conceive would be useful.

The following information was found from the focus groups, regarding ways and means of accessing graphs and tables in their everyday lives

3.3.1 Where and when graphs and tables are encountered

Graphs and tables are mainly used at school, in subjects such as maths, science and geography, although they are also encountered in further and higher education subjects such as psychology and music technology, and in professional settings. Tables were reported to be mainly encountered in web pages, where information is commonly structured using tabular structures, for page formatting. They were also encountered in word processing documents and spreadsheets. One participant still occasionally used tables in her job to structure information, but remarked that she was glad this was only an occasional task. Some students accessed information made available to them in Excel spreadsheets, and most of the group had used Excel at some point. Graphs had been encountered at school as part of the national curriculum, however none of the users utilised graphs to access information in their jobs, current studies or leisure time.

3.3.2 How graphs and tables are normally accessed

Graphs and tables were most commonly accessed through screen readers, which read out table coordinates and cell values sequentially. None of the participants had explored large tables using this method. Braille print-outs and embossing film were also used for exploring, transcribing and constructing tables and, more commonly graphs, in order to meet the requirements of the national curriculum.

3.3.3 General views and experiences of using graphs and tables

The group concurred that visually-impaired people generally do not enjoy working with graphs and tables, and blind or visually impaired students will often avoid dealing with information visualisations where possible, even in exam situations. The group consensus was that this aversion to graphs and tables was mainly due to a lack of accessibility. However, participants remarked that blind and visually impaired people do have an understanding about what graphs and tables are for, and can work out the tasks that were set for them with graphs and tables, albeit with difficulties: *"…if there was a way […] that [tables and graphs] really could be* *accessible, then perhaps in the future more visually-impaired people wouldn't be so frightened of them."* (Participant 4)

3.3.4 Perceived usefulness of graphs and tables

The group agreed that if graphs and tables were easier for blind and visually impaired people to access, they would possibly use them more frequently and perceive them to be more useful. Those who utilised spreadsheets in further education or in a professional setting tended to perceive tables and graphs as more useful than those who had last used data visualisations at school: *"If graphs were accessible though, I do think blind and visually-impaired people would use them just as much as sighted people."* (Participant 6) One participant suggested that accessing the data in tabular form made sense, but not accessing visual representations: *"…Excel, that's all right as long as there are no charts and graphs and things involved. I found it very useful when I used it in the past."* (Participant 8)

3.3.5 Problems encountered and barriers to accessibility

A key frustration for all the participants was the difficulty of obtaining an overview when presented with a new table or new graph. The group agreed that getting a sense of the boundaries and layout, as well as the overall content of the graph or table, was a necessary first step to a more detailed exploration. A contributing factor was the linearity of presentation when exploring a table or graph using a screen reader.

> *"When somebody fully sighted is looking at a screen, they can see it as a whole instantly, but somebody visually impaired has got to navigate actually bit by bit all around the screen, which is why it's so hard to get a complete picture on the screen."* (Participant 4)
>
> *"The best way to learn anything is to get a look at the whole picture first."* (Participant 1)
>
> *"When you go into Excel, if you don't know what the table's about, you must explore it and orient yourself around the table. Because if you don't do that, you won't get the full picture of it, you won't get all the information from that table that you need, all the information that the people who gave it to you intended you to have."* (Participant 8)
>
> *"I find in databases in Excel I have to sort of try and visualise the screen. Knowing what kind of area you are working it helps to give you some idea of how the screen is laid out."* (Participant 7)

Another significant problem for all the participants was the need for sighted help in accessing and working with numerical data, particularly in graph exploration and construction. Participants remarked that graphs and charts are not accessible with screen readers such as JAWS, and that creating a graph from an Excel table is often not possible without assistance. The steps can be learned, but it is not possible to check if the result is correct. Participants commented that it is important to be able to review what has just been done in order to make sure that there are no mistakes: *"With a Brailler, you can see exactly what you've written straight away… You can* *check to see if you have any mistakes."* (Participant 1). Sighted help was also sometimes necessary when constructing or exploring graphs using raised paper or embossing film: *"…we had visual support assistance in our maths lesson …*they *would see me doing the graph and because our graphs had to reflect something that we had been taught to reflect, they would check that we were doing it OK."* (Participant 8).

While most participants had successfully used raised paper graphs augmented with elastic bands as part of the school curriculum, they commented that this solution had some limitations as the bands could only be used for straight lines, and were prone to coming out of position. An added drawback was that the graphs could not be stored for later use, unlike electronic versions rendered in Excel.

Participants also indicated that retaining a sense of context when exploring graphs is important: *"A graph that isn't very well done is a graph without any squares that you can't work out where anything is on it."* (Participant 1)

3.3.6 Tools used to access graphs and tables: benefits and drawbacks

The two most common tools used to access graphs and tables were screen readers (which read out sequential values from the computer screen) and tactile outputs such as Braille or embossing film (also utilised in the construction or transcription of graphs and tables). Preferences between these two categories varied across the group, and seemed partly to depend upon whether participants had been educated in a special or mainstream school, as different educational settings sometimes differ in the tools offered to blind and visually-impaired students. In the case of this group, one participant who had attended mainstream school was more comfortable with a screen reader, while another who had attended a specialist school had received more training in Braille and thus was more comfortable with a Braille transcription of tables and graphs. None of the users mentioned using non-speech sonification or force-feedback haptic devices as accessibility tools.

Screen readers

A regular user of Excel said that accessing it with a screen reader is not necessarily easy as this means table values are read out sequentially, and a lot of keyboard tabbing is required to move the cursor across the table. Others in the group pointed out that obtaining an overview of a numerical data table with a screen reader can be very time-consuming, and that it is extremely difficult to remember the numbers read out so as to look for a particular value or type of data within the table.

> *"I use tables in Excel. That's on a computer with JAWS, a screen reader. I don't necessarily find that very easy because you have to navigate all the columns. There's a lot of tabbing involved and cursors up and down."* (Participant 3)
>
> *"[With a screen reader] you just have to go over it and over it and over it, as many times as you can, to try to get yourself an overall picture of it [the data table]. That might take a while."* (Participant 4)

> On the possibility of retrieving a particular value from a table accessed with a screen reader: *"…that's assuming you're going to be able to take in the table in your head, and remember it."* (Participant 7)

Braille and other tactile methods

Several participants showed a preference for Braille tables, as long as these were "simple". It was suggested that displaying data tables in Braille allows the user to get a holistic view of the table, *"almost equivalent to somebody seeing it on a screen"* (Participant 4). With larger or more complex data sets, Braille is not always practical or available, and it is often difficult to get the whole table embossed onto a single sheet: *"Sometimes I've found that when embossing a table, some of it will be on one line and some on the next, because the table's too big for the embosser to cope with. And something like that can be very hard, and confusing."* (Participant 4)

The group as a whole agreed that graphs and tables are often difficult to read in Braille and other tactile visualisations (including swell paper and embossing film), and thus making sense of them can be very time consuming and frustrating. In exams blind and visually impaired students tended to avoid questions requiring use of these tactile visualisations, because, while they were capable of answering them, the answers required a very long time that was better invested answering other questions. One advantage of Braille, however, is that when creating graphs from data sets, it enables the user to review what has just been done in order to ensure there are no mistakes.

3.3.7 Strategies employed when using accessibility tools

Users described how they used two hands to navigate Braille tables such as bus and train timetables: "*I have used train timetables and things in Braille, where you would have a column and you would follow your hand. Your right hand would take the lead and your left hand would follow it.*" (Participant 3) This was felt to be a useful strategy as it enables the user to keep track of the position of data in the table.

When exploring data tables using a screen reader, some participants said they used the strategy of repeated listening in an attempt to obtain an overview. This was felt to be inadequate as with the sequential read-out it was very difficult to get an overview of the data, and also difficult to retain a sense of context or remember points of interest.

3.3.8 Summary

The feedback from blind and visually impaired participants in the focus groups and interviews conducted at the RNC emphasised the difficulty of obtaining overviews of data tables and graphs. Participants also talked about the need to retain a sense of context during explorations, often supported by two-handed exploration of raised paper graphs and tables. Participants commented that retaining context was difficult when using screen readers to access data tables (due to sequential read-out of information) and tactile methods such as raised paper to access graphs (due to factors such as interference from grid lines and poor layout of the representation). This feedback accords with the review of current accessibility tools used to access numerical data discussed in Section 2.2 above. Further key points highlighted by participants included their experiences of needing sighted assistance when exploring or constructing data representations. This, together with the perceived complexity of the data due to the lack of overview and loss of context, contributed to the aversion many participants felt for graphs and data tables. Despite this, participants felt that tables and graphs were useful and that they would make more and better use of them if appropriate accessibility techniques were made available. This feedback helped to define a set of key requirements, which are presented in Section 3.6.

3.4 Classroom Observations

In addition to the focus groups and interviews described above, a visit was made to The Royal Blind School of Edinburgh, to observe a mathematics class and gather feedback from school pupils and teachers on accessibility requirements in support of the school curriculum. Two students and one teacher participated in this requirements capture. The first student, who was blind since birth, was studying foundation-level mathematics, while the second student who was partially sighted, and utilised large print and a magnifier, was studying Higher Grade mathematics.

3.4.1 Requirements of the school curriculum

The mathematics curriculum for blind and visually impaired pupils is identical to the curriculum in mainstream schools. The course texts are the same (*Mathematics in Action*), although the teacher said she sometimes modified the examples based on her judgment. She commented that she felt that each student was different, and that the curriculum and teaching methods used for each student had to be modified in a separate way. The examinations used are set by the Scottish Qualifications Authority and are modified, though the modifications are sometimes contentious, and may impede the student. For example, the task “Draw by eye a best fit trend line” is modified to “Draw a best fit trend line, or explain how to do so”. The sorts of questions asked by the mathematical texts and the examination board are precise. Although the teacher did ask the student to describe the general trend of the graph, most questions were far more specific, for example, “What were the sales of coffee for January?”.

3.4.2 Ways of accessing numerical data and difficulties encountered

Browsing graphical data was mainly achieved by using raised paper diagrams of charts and graphs. Bar graphs were printed on raised paper, with Braille used to denote scales and legends. In order to fit all the information onto the graph on one sheet of paper, a key for the X axis was used. This was sometimes confusing for the students as they needed to cross-refer to the key, located below the graph itself. Line graphs were also produced on raised paper, and browsing these was difficult since the embossed grid lines frequently interfered with the interpretation of the graph line. As a result, the graph line had to be enhanced using rubber bands, which were used to highlight the line, kept in place with drawing pins. One of the students commented that this improved access to the graph but it was not an ideal solution as there were some practical difficulties with stretching the rubber bands and keeping them in place. At foundation level mathematics, pupils do not have to create graphs but do have to demonstrate understanding of the coordinate system in a graph by inserting pins into given coordinates and then working out the coordinates of another point to determine the shape of a graph or a geometric diagram. This also incorporates the use of negative coordinates and negative coordinate systems. The teacher commented that in such situations a map pin was used to denote the origin of the space, as shown in [98].

In order to access and modify these data visualisations the students required sighted help. The teacher occasionally had to move the students' fingers to assist them in exploring graphs, resulting in a more cooperative exploration process than would normally be expected for sighted pupils. The teacher commented that she needed to be able to control, direct and observe the students to ensure they are picking up the material, and due to the individual differences between students she also had to adopt a more improvised teaching strategy.

It was observed that the students always worked from the origin of the graph line for exploration, using two hands. To find the appropriate points on a graph the students used one hand to move along the Y axis and then the other finger to move from there along to the X-axis. When trying to find a given point in a coordinate space, the student followed the X axis with one finger and then the Y axis with the other to the given coordinates. The fingers were then moved together to find the intersection.

3.4.3 Summary

A key observation on both the browsing and construction of graphs using raised paper is that students do not obtain a useful overview of the data. Instead, the data representation often appears complex to the student and the use of multiple graphical elements (grid lines, separate keys and so on) is often confusing. The use of pins and rubber bands to enhance the graph line, while an improvement on raised paper alone, is not an ideal solution. In order to retain a sense of context when exploring the graphs, the students used two hands and found it helpful to work from the origin of the axes to follow each data line or bar. A further important point is that the use of these methods does not support blind and visually impaired students' independence, as the students often have to rely upon the teacher's assistance in exploring data representations.

These observations further confirm the requirements obtained from the focus group discussions and interviews conducted at the RNC in Hereford, that blind and visually impaired users need to be able to explore data independently, obtain an overview of data, and retain a sense of context, with support for two-handed exploration where possible. This information contributes to the key requirements defined in Section 3.6.

3.5 Additional Requirements

In addition to the requirements captured directly from end-users through the focus groups, interviews and observations described in Sections 3.3 and 3.4, further requirements for tools for blind and partially sighted users have been identified from the literature, and through informal discussions with blind and visually impaired people, and professionals from RNIB and RNC.

3.5.1 Ability to collaborate with sighted colleagues

A key factor in enabling blind and visually impaired people to pursue studies and careers in disciplines which make use of numerical data sets (such as business and accountancy, science and engineering, and social sciences) is that they should be able to collaborate with sighted colleagues. This is also an important element of the normal school experience, where many tasks are focused on group work. The classroom observations reported in Section 3.4 showed that blind and visually impaired students often require intervention from the teacher to guide them in the browsing of numerical data representations. While these observations were conducted in a specialist school, it can be inferred that blind and visually impaired students being educated in a mainstream setting would encounter significant difficulties in collaborating with sighted classmates in the browsing or construction of graphs. Such lack of equality in access to curriculum materials and involvement in group tasks has been shown to result in social exclusion and frustration for blind and visually impaired pupils [77] [31].

Similar barriers to collaboration can be expected in the case of visually-presented tabular numerical data. Here, data are spatially laid out, thus collaboration focusing on such data sets is likely to refer to space-related terminology, such as “up-down”, “left-right”, “centre”, “top-left corner”, and so on. The spatial layout of tables as represented in spreadsheets and other documents is not always obvious for blind and visually impaired people, as discovered in the focus groups, where the discussion between participants revealed some confusion as to the layout of rows and columns within tables, due to the sequential read-out provided by screen readers. Thus, another key requirement is that the spatial arrangement of information should be preserved in data representations, in order to facilitate a common language as a basis for collaboration with sighted peers.

3.5.2 Controlling the exploration and accessing details on demand

As discussed in Section 2.3, a number of authors working in the field of non-visual data accessibility tools have stressed the importance of the user being in control when browsing data, in order to avoid information overload [46]. Related to this is the need for any non-visual data exploration technique to enable the user to access details on demand, in line with Shneiderman’s information seeking mantra: “*overview first, zoom and filter, then details on demand*” [138]. Since many interactions with datasets require both a view of the overall trends and patterns within the data, but also may require pinpointing exact values, accessibility techniques need to support the user in focusing on points of interest and retrieving exact values within data sets.

3.5.3 Ability to deal with real-world data

In order to achieve equality in their choice of career and study options, blind and visually impaired people need to be able to access different types of data (such as sales statistics for business, population demographics for social science, experimental results for physical sciences, and so on). Thus, accessibility techniques should be flexible enough to present real world data. It should be possible to accommodate a variety of data sets, data meanings and features in the data within the same interface and access them with the same techniques, so that there is no need to use heavily specialised, data and task-dependent solutions. This is in line with recent work by Walker & Nees [166], who contend that real-world data should be used in the study of data sonification to help ensure that the techniques developed can be used outside the laboratory.

3.5.4 Simplicity and affordability

Related to the above point about specialised or data-dependent solutions is the need for techniques to be as simple and easy to use as possible. The focus groups and classroom observations showed that even apparently simple accessibility tools such as raised paper can lead to confusion and complexity in their representation of numerical data sets, particularly when data visualisations are transferred directly into an equivalent haptic representation (as in the case of raised paper line graphs), or when values are read out sequentially and without a sense of context (as in the case of screen readers). Novel accessibility techniques should aim to avoid confusion and overload of the working memory.

Furthermore, it was noted in Section 2.2 that accessibility hardware such as refreshable Braille displays, Braille embossers, and force-feedback devices can be expensive or not widely available. If possible, techniques should utilise technology which is affordable and widely available, so that users can easily acquire or set up a similar system.

3.6 Conclusions and Summary of Requirements

This chapter has reported the requirements capture carried out as part of the user-centred approach to the design and implementation of TableVis, providing further information in answer to RQ-1, "*How can overview information from tabular numerical data sets be obtained non-visually?*".

The requirements capture involved a series of focus groups, interviews and classroom observations carried out with blind and visually impaired students based in Hereford and in Edinburgh. The focus groups and interviews were conducted to gather feedback on blind and visually impaired people's views and experiences of accessing numerical data sets, while the classroom observations were conducted in order to gather information on students' real-world experiences, and to obtain views of education professionals.

While many of the blind and visually impaired participants said they did not find it easy to access numerical data using current accessibility tools, and as a result sometimes avoided tasks involving graphs and tables, most said that they believed graphs and tables to be useful and, if better accessibility techniques were available, they thought that blind and visually impaired people would work with numerical data and data representations more frequently. In addition to highlighting scenarios in which graphs and tables are commonly encountered, the focus groups and classroom observations also provided feedback on the benefits and drawbacks of current accessibility tools, and some general requirements for non-visual access to numerical data sets. Having analysed the feedback obtained from the requirements capture as described above, it was possible to assemble a set of requirements for the design of techniques for non-visual access to numerical data. These can be summarised as follows:

I. *The user needs to be able to obtain a quick overview of the data*. Several participants noted the usefulness of being able to obtain an instant global view of a data set (as with visualisations available for sighted people) and commented that they believed this was the best way to commence exploration of the data. This feedback also accords with Shneiderman's information seeking mantra, which stresses the importance of "overview first" before proceeding to a more detailed exploration, as discussed in Section 2.1 [138].

II. *The user needs to have control over the data exploration process and to retrieve details on demand*. This will help to ensure the user is not overloaded with information, and also that they can zoom and filter areas of interest within the data and retrieve detailed information, as key stages in the information seeking process defined above.

III. *The interface should enable the user to retain a sense of context to avoid "getting lost" in the data*. A frequent problem for blind and visually impaired users, particularly those using screen readers, was that it was difficult to retain a sense of context when navigating electronically through a large data set. In the classroom observations, the students preferred exploring graphs by tracing the data series from the point of origin, thus retaining a sense of context within graph exploration. The focus group participants suggested that where possible, the extent of tables and their boundaries should be made obvious.

IV. *The interface should not overload the user's working memory*. Some participants commented that when exploring tabular numerical data using a screen reader the volume of data can easily become overwhelming. In order to identify trends or points of interest within the table users had to listen repeatedly to the sequential read-out of values, and this was considered both time consuming and mentally demanding.

V. *Two-handed exploration may be preferable for tactile interaction*. Focus group participants commented that they explored Braille tables such as bus timetables with two hands, which enabled them to retain a sense of context when exploring simple tables. Classroom observations also showed that students tend to explore tactile visualisations using two hands. It would make sense to retain this facility if designing an interface which utilises tactile interaction.

VI. *Data need to fit on one "page" or work space, and the spatial layout of the data should be preserved*. A frequent problem with Braille print-outs of graphs and tables is that the table may not fit onto one sheet of paper. In order to enable the user to retain control of the document and to avoid confusion, it is better to ensure that one workspace can encompass the entire data set. In addition, in order to ensure a common language for collaboration with sighted colleagues, the spatial layout of the data should be retained.

VII. *Users need to be able to access real world data*. The solution must be flexible enough to be applied to different kinds of numerical data sets of varying levels of complexity. An implicit requirement is that tables can be explored according to the logical structures that they represent, namely by rows and by columns.

VIII. *The solution should be as simple to use and affordable as possible*. It should be easy for the user to acquire, set up and use a similar system without having to purchase expensive task-dependent devices.

IX. *The entire data exploration should be possible without sighted help*. The need for sighted assistance in exploring data representations is a source of frustration and a barrier to equality for blind and visually impaired people. The design of novel techniques for non-visual access to numerical data should ensure truly independent access for blind and visually impaired users.

These requirements were used to inform the development of the non-visual exploration techniques proposed in this thesis. The next chapter will report the design, implementation and pilot evaluation of these techniques.

Chapter 4 High-Density Sonification (HDS) and its Implementation in TableVis

4.1 Introduction

The first three chapters have laid the groundwork for the main contributions in this thesis, driven by the thesis research questions formulated in Section 1.2: "*How can overview information from tabular numerical data sets be obtained non-visually?*" (RQ-1) and "*What level of performance can be achieved with the non-visual techniques proposed in this thesis?*" (RQ-2).

In a quick recapitulation, Chapter 1 discussed the motivation and the relevance of these questions in the context of accessibility to complex numerical information non-visually, with particular focus on the stage of gathering overview information. Chapter 3 surveyed topics from the literature that are relevant to the investigation of this research: the role of an overview in data exploration, current accessibility technologies and their support for gathering overview information, and research on non-visual access to numerical information and support for overviews. Chapter 3 reported the process of user-centred capture of requirements conducted as part of this thesis and defined a set of requirements to inform the design of the solution investigated in this thesis.

The information from these initial chapters led to a consideration of the approaches, existing exploratory techniques and sensory modalities that could be best employed towards answering the first research question: how to obtain overview information from tabular numerical information non-visually. The outcome was the proposal of an interface to explore tabular numerical data sets non-visually, called *TableVis.* The main technique implemented in this interface is a novel data sonification technique called *High-Density Sonification (HDS)*, which is described and justified in Section 4.3. Section 4.4 then describes how HDS was implemented in TableVis, in combination with other input and output techniques, such as speech synthesis and a tangible user interface. Finally, a pilot evaluation of the interface is reported at the end of the chapter. Before all this, the chapter starts by defining the scope of the data sets (data structures and types of data) considered in this thesis. While this thesis has from the beginning stressed its focus on tabular numerical data sets, this is a broad definition. Thus, in the next section the focus of this thesis on data tables is further discussed, defining the scope of data sets that will be considered in the rest of the thesis.

4.2 Data Sets Considered in this Thesis

Much of the discussion so far has focused on numerical data sets, without further specifying their structure or characteristics, although they are key to this thesis. This focus on tables was already mentioned in Chapter 1, and the thesis research questions formulated in Section 1.2. This section seeks to provide further justification about the importance of studying tabular data structures, and to define which tabular structures will be considered in the scope of this thesis.

4.2.1 Why numerical data tables?

The reason for studying tables relates to the aim for studying non-visual representations of numerical data sets in general, rather than accessing specific visualisations constructed from those data. In that sense, while data tables can be considered a form of data visualisation (because they reveal certain relations among the data when laid out on paper or on screen), tables are logical structures on which data can be arranged. In fact, as Card *et al.* [30] argue, tables (plus the metadata that describes and gives sense to the values contained in them) are a first step to rationalise any set of raw data before creating a visualisation of it, as represented in Figure 4.

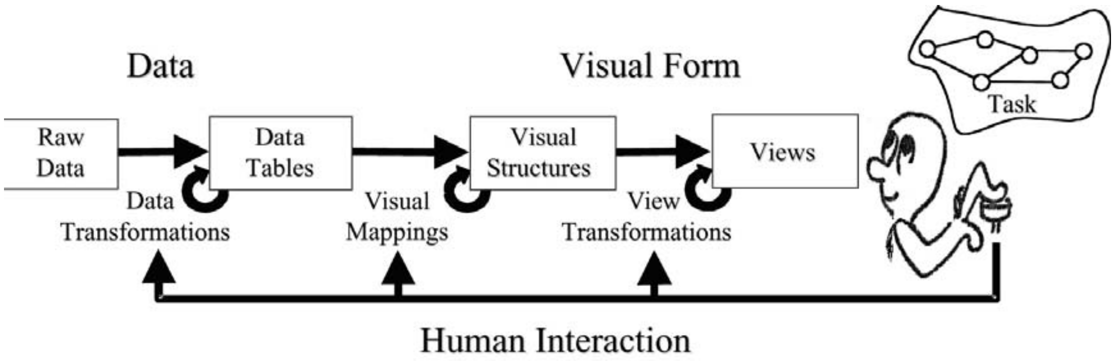


Figure 4. Reference model for the creations of visualisations from raw data, as proposed by Card *et al.* [30]. Data tables are the result of the first transformations applied to the raw data in a process called *normalisation*. Data tables are formed from raw data before any mapping to sensory parameters takes place, and are therefore independent of the sensory channel that will access them.

In this model, data tables are created from raw data as the result of a data normalisation process, but they are independent of the sensory channel with which the data will ultimately be accessed. As seen in Figure 4, it is only after the tabular structures are formed that data are mapped to visual parameters. This independence of the sensory channel permits the mapping of data to parameters of any sensory channel, or even to several channels. This idea is stressed by Card *et al.* [30] after they present their model: "*…while we are emphasizing visualization, the general case is for perceptualization. It is just as possible to design systems for information sonification or tactilization of data as for multiple perceptualisations*" (p.7).

4.2.2 Which numerical data tables?

While Card *et al.* [30] describe a detailed process to normalise raw data into tables, there are actually many structures that can be called tabular. Lopresti & Nagy [94] provide a non-exhaustive, yet comprehensive, taxonomy of 2D tables containing a variety of types of data (textual, numerical, graphic), as well as operations and transformation that can be done with a table, and they propose a table-related terminology in a paper that, itself, is written as a big and complex table. Yesilada *et al.* [170] also report a survey of 2D tables, finding 35 different groups of tabular structures based on data type, levels of table nesting, visibility and design of cell boundaries, and labelling. They also provide a classification of table reading and browsing styles, as well as presentation and content styles supported by tables. Their work concentrates on providing improved access to tables using screen readers, by using a large combination of keystrokes.

As derived from these two references, many data structures can be referred to as data tables. In the scope of this thesis, however, only certain kinds of tables are considered. The data structures considered in this thesis are matrices, or what in computing science are referred to as 2D arrays. Such data structures consist of collections of one-dimensional arrays, which are data series with the same number of elements. Thus, any one element in a multidimensional array can be accessed using pairs of integers, or indices that provide the coordinates or address of an element in the 2D array. A characteristic of these structures is their regularity: they are a lattice with m x n number of cells, where m is the number of columns and n is the number of rows. Despite regularity in the structure being adopted as a requirement for a table to be studied in this thesis, the limitations introduced by this decision are not many. In fact, a spreadsheet is an example of one such table. The results of the process of normalisation of raw data into tabular structures described by Card *et al.* [30] (from which to construct data perceptualisations) are also examples of the tables considered in this thesis.

Regarding the type of numerical information stored in the tables that will be considered in the scope of this thesis, no limitations are introduced *a priori* regarding the range of values, sign or number of decimals. In summary, any collection of numerical data that can be arranged in a 2D array or in a matrix (be it a collection of data series, samples from a mathematical function of the type $z=f(x,y)$ or other experimentally recorded sets of data from a system that fit logically into a 2D table, to mention some obvious cases), is considered in the scope of this thesis. The only exception would be data tables where one or more of the cells are empty. Such cases are left for future work. The techniques investigated in this thesis are intended to cater for data with any meaning which fit into a 2D-table. As discussed in section 4.4, there is ongoing research regarding how the customisation of sonification techniques to data with specific types of meaning affects the success of data explorations. This will also be taken into account at the time of extracting conclusions from this thesis.

4.3 High-Density Sonification (HDS)

In Section 2.3.2, it was shown that a large body of work exists around the use of auditory graphs, where relative values in a graph are presented by mapping them to some parameter or parameters of sound. To recall some aspects from that review, auditory graphs can successfully present detailed information about a graph, and to some extent about two graphs presented simultaneously [24], providing relative values of the data. For larger collections of graphs (the data points of which would constitute a tabular structure), the amount of information conveyed to the user in this way (the number, relative location and relevance of such features, plus the time it can take to traverse all the graphs) can soon grow to levels that mentally overload the user. In a study with rapid auditory graphs [80] (where each graph was sonified in its entirety automatically, at a fast pace of ~60ms per value), Kildal & Brewster found that the higher speed with which data were rendered made it more difficult to notice all the inflexions in a data series, although this provided the benefit of revealing overall characteristics (general trends and patterns) more clearly and with less effort. This simplified representation of the data, combined with the short time that it took to render one graph, made it possible to present several graphs in a sequence and extract knowledge from the whole set. In that study, users tended to perceive each rapid auditory graph as a single (although complex) auditory event, which could easily be compared against the auditory event originating from another rapid auditory graph, so that users could extract conclusions about the overall comparison of graphs. In other words, increasing the speed of rendering auditory graphs was a method to hide detail and let overview information surface, generating manageable levels of mental workload. That study was conducted using tables with 10 rows and 10 columns. Unpublished qualitative pilot evaluations by the same authors [79] suggested that with larger table sizes this data representation could again become excessively complex. To summarise, rapid auditory graphs showed that higher speeds of rendering the information affected the level of detail revealed, facilitating users to obtain overview information of larger data sets than with others forms of presentation of auditory graphs. This technique, however, was still only useful for relatively small and/or simple data sets. Harrar & Stockman [63] have also looked at the effect that modifying speed of presentation of single auditory graphs has on obtaining overview information from them. In their study, three presentation speeds were compared for graphs with 12 data points: "Slow" (14 seconds), "Medium" (7 seconds) and "Fast" (3 seconds). They determined that the medium presentation speed supported the most accurate graph comprehension, for graphs rendered in discrete sounds. The fastest speed tested (around 250ms per sonified data point) was slower than the 60ms per data point tried by Kildal & Brewster [80]. Combining the results from both studies, it seems that speeds of rendering auditory graphs of half to one second per sound are best to obtain clarity about the overview of a single auditory graph (a one-dimensional array of numbers). Increasing the speed makes it more difficult to obtain clear descriptions of a graph. However, at very high speeds (around 60ms per sound) the information that can be extracted from a single graph is so condensed that the resulting audio rendering becomes a representation of the graph a as whole, and can be compared against similar representations of other graphs, thus providing a technique for the exploration of collections of graphs (two dimensional arrays of numbers). This thesis studies the highest end of rendering speeds and their suitability to explore such two-dimensional arrays or tabular numerical data sets.

Taking into account this effect of high data rendering speeds on filtering out detail and revealing very general overview information, a logical progression to obtain overview information from larger data sets was to further increase this data rendering speed. A consequence of increasing the speed beyond the 60ms per sound tested by Kildal and Brewster [80] was that, eventually, the duration of the gap between the sounds representing the data points on a graph was so short that the auditory event generated with the rendering of a complete data series was perceived as a chord. This change in the nature of the auditory representation (from a series of sounds to a chord) would result in the loss of information about the order of the values in the data series. Additionally, the complexity in the combined frequency spectrum of the resulting chord would be expected to lead to the masking of some of the frequencies [106]. This leads to the question of whether numerical information conveyed in this manner could be usefully analysed by a listener.

This new form of sonifying series of numbers perceptually simultaneously rather than sequentially is the basis for one of the main contributions of this thesis. With the name *High-Density Sonification (HDS),* this sonification technique is formally defined in the next section. Following a general definition, its actual implementation in TableVis is discussed. Then, a pilot evaluation of this implementation is reported. The implications of HDS in its implementation for TableVis and the capabilities that it offers as a technique for discovery of knowledge from tabular data sets are investigated in much of the research work reported in the remaining chapters of this thesis.

4.3.1 General Definition of High-Density Sonification (HDS)

In its most general form HDS is defined independently of its actual implementation:

> *High-Density Sonification (HDS) is the simultaneous (or perceptually simultaneous) sonification of all or some of the data points in a data set, following the same parametric sonification strategy for all the data points. The resulting single auditory event is a representation of that data set (or sub-set), which can be compared against the representations of other data sets (or sub-sets) that are generated using the same data-to-sound mapping strategy.*

This definition of HDS encompasses all the possible instances of the technique by specifying its defining characteristics without describing the actual form of its implementation. Hence, this definition leaves freedom for designers to decide on the best implementation for each case, and for researchers to investigate aspects of usability and user performance, as well as perceptual and cognitive properties of different implementations.

It is important to highlight that a key characteristic of HDS is to sonify data points from a data set simultaneously or perceptually simultaneously. These data points can be all the data points in the data set (in which case the complete data set would be represented in a single auditory event) or just a subset; the definition does not, however, specify the structure of the data set or the way in which subsets are formed. Another statement in the definition is that the same parametric sonification strategy is applied to each data point in the subset, but it does not provide any further description about which aspects of the data are mapped to which parameters of sound, nor how this should be done. Finally, the definition states that auditory events generated using the same parametric sonification strategy over different data subsets can be compared against each other. The aspects that are not determined in this definition are left for design decisions to be made in each implementation. The next section discusses the implementation of HDS proposed for TableVis.

4.4 Implementation of HDS in TableVis

The general definition of HDS given above encompasses all the possible instances of this technique by specifying only the defining aspects of it, which can take a variety of forms in different implementations. This section discusses the implementation of HDS made for TableVis, in the context of its integration in an interface to interactively explore tabular numerical information non-visually.

4.4.1 Grouping Data to Render with HDS

The definition of HDS does not limit its use to any particular data structure, type or configuration. Data tables are not mentioned in the definition, which means that it can be used with other data structures as well. Similarly, the definition does not stipulate which data points in a data set are simultaneously sonified. In practice, the set of data that is sonified with HDS into a single auditory event will depend on the meaning of the data being explored and the aims of the exploration. A common scenario could be that the user wants to sonify data subsets in which the data points have certain common characteristics, and compare them against other subsets that share the same characteristics but vary with respect to one or more parameters, the effect of which the user is trying to assess. In the case of 2D tabular numerical data, an obvious criterion for generating data subsets which can be meaningfully sonified with HDS (although not the only one) is to combine data cells that share the same value in one of the coordinate axes. By doing this in a 2D-table, data can be grouped by rows and by columns. For each one of these ways of grouping cells, the whole data table can be represented with the auditory events generated using HDS. Thus, all the sonified rows contain a representation of all the cells in the table. Likewise, all the sonified columns also represent all the cells, but with a different set of sound events. These two different sets of auditory events, each conveying the whole data set but generated by grouping the cells differently, can be seen as two *complementary views* of the same data set. It is somewhat like looking at the same object from different angles: each view reveals certain characteristics of it, and putting both views together can offer an overview description of the whole.

For a 2D-table, grouping and sonifying the cells with the same value on one of the axes is called the canonical HDS of the table, and the patterns for grouping cells are graphically represented on the top half of Figure 5: grouping cells by rows and by columns, obtained by grouping together all the cells with the same coordinate value on the vertical and on the horizontal axis respectively. The definition of the canonical HDS can be extended to tables with more dimensions. For instance, with a 3D-table constructed around an XYZ orthogonal frame of reference, the canonical HDS is obtained by grouping cells that share the value in one of the axes, each one of those subsets being a 2D-table. Thus a collection of 2D-tables can be created along one axis, which together contains all the cells in the 3D-table. Each 2D-table is represented as a single auditory event, providing an array of sounds that represent the whole data set. Repeating the same process for each axis provides three complementary views of the same whole data set. The bottom half of Figure 5 illustrates graphically the way in which subsets of data are generated for canonical HDS in 3D-tables.

The canonical HDS for tabular data structures with any number of dimensions can be formally defined as follows:

> *Considering a data table with* n *orthogonal dimensions, its canonical HDS consists of* n *complementary views. Each view is generated along one of the axes, rendering in a single auditory event all the cells that share the same value on that axis. As a result, all the auditory events in the same view originate from data subsets with the same size and shape. Also within a view, all the cells from the data set belong to one and only one auditory event (i.e., all the cells are represented in the set of auditory events of each view, and each cell is represented only once in each view).*

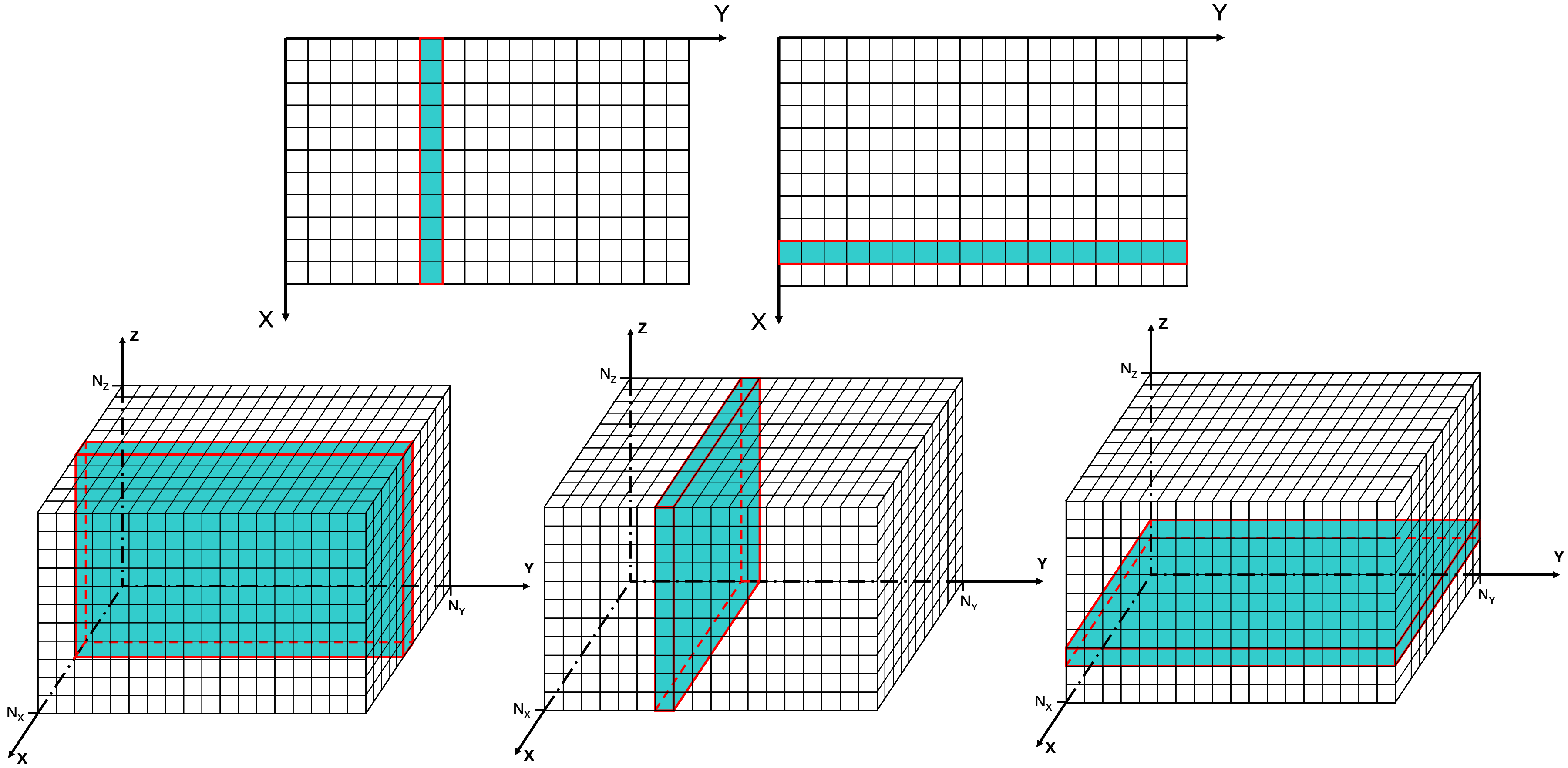


Figure 5. Data subset patterns for canonical HDS of 2D-tables (above) and 3D-tables (below). The canonical sonification of 2D-tables generates two complementary views of the data set, whereas for 3D-tables it generates three complementary views.

It should be highlighted once more that the canonical HDS of a tabular data set is by no means the only possible form of HDS. To give only a few examples, data points that share values in two or more axes (instead of just in one) could be grouped together, resulting in smaller data subsets than with the canonical HDS. Data spanning over certain ranges on one or more axes, or data subsets that are not arranged in parallel to any of the axes could be other examples. Grouping data points in concentric round shapes (*e.g.* concentric spheres in 3D-tables, or concentric circles in 2D-tables) that expand around a certain point of interest could be a particularly useful form of grouping data for HDS that is different from the canonical. Ultimately, the preferred pattern in each case will depend upon aspects such as the meaning of the data, previous knowledge about the features in the data and the aims of the exploration.

In the scope of this thesis, which focuses on 2D tabular data structures, the implementation of HDS in TableVis was chosen to be in the canonical form, due to it being a well defined form with properties that are suited to 2D-tables and which can be easily extended to tables with larger numbers of dimensions.

4.4.2 Parametric Sonification

To fully define the implementation of HDS in TableVis, the mapping of numerical information to sound needed to be decided upon, since it is not determined in the general definition of HDS. Pitch was selected to be the single parameter of sound used in HDS to map numerical information, in the implementation for TableVis. This decision is grounded on previous research on auditory graphs, as discussed in the literature review (Section 2.3.2) and in the description of the process that led to the definition of HDS, described earlier in this chapter (Section 4.3). One of the main reasons for selecting this parameter of sound is that it provides higher resolution than others parameters such as loudness, spatial location or timbre [50]. An additional advantage of using this auditory display dimension is that it is quite robust against influences from the listening environment [164]. Some researchers assert that the meaning of the data should be taken into account at the time of deciding about the parameter of sound for the mapping of the data, and that this should be done based on metaphors. Thus, Walker and Kramer [160] studied experimentally which of four parameters of sound (loudness, pitch, tempo or onset) was best suited to map data informing about temperature, pressure, size and rate. They found that some of these combinations were more meaningful than others, although not always conforming to the predictions of the sonification designers. Extending the argument of choosing mappings based on metaphors, Neuhoff [111] proposed the use of ecological sound sources and events that evoke mental models (such as the speed of steps), as opposed to low level acoustic parameters (like pitch and loudness). As discussed in Section 4.2.2, these considerations are taken into account for the general discussion of the results in the thesis, although different choices of parametric sonification are not discussed in this thesis. Instead, user performance achieved with pitch as the parameter to access numerical information, as well as its implications at perceptual and cognitive level, is studied in detail in this thesis.

By using pitch to convey quantity, only relative values can normally be conveyed to a user. Users in possession of absolute pitch could potentially be able to derive the absolute values of the numbers that were mapped to pitch, provided that information about the range of values being represented was known. However, the prevalence of absolute pitch in the general population is very low, estimated at 1 in 10,000 [7]. Thus, in general it is not feasible to represent absolute values in pitch. Mapping data to pitch has however proved to be very effective at representing relative values, as the work on auditory graphs referenced in the literature review demonstrates (Section 2.3.2). Most people have no difficulty in distinguishing which of two sounds are higher in pitch, and can readily interpret this difference as the two numeric values relative to each other. As Moore [107] reports, several studies on pitch discrimination of pure tones show that listeners can detect differences in frequencies (Hz) of less than 0.5%, throughout the frequency range. Only a small minority of people have a deficit in processing such fine-grained pitch differences – a condition known as amusia, thought to affect up to 4% of the general population [139].

A parametric mapping is not fully defined until polarity and scaling are also chosen. Walker [161] asserts that, as for the sound parameter chosen to represent the data, both polarity and scaling should also be chosen taking the meaning of the data into account. Polarity indicates how the sound parameters vary with the change of data, *i.e.* whether an increase in the sound parameter should indicate an increase or decrease in the data dimension [63]. *Positive mapping polarity* would take place when the sound parameter changed in the same way as the data dimension, and *negative mapping polarity* would happen when the opposite was true [161]. Ongoing research has provided some evidence that the preferred polarity for a population of listeners may not be the same for different types of data mapped to the same parameter of sound [164]. Walker & Lane [165] studied preferred data-to-auditory-display mappings, polarities, and psychophysical scaling functions relating data values, for blind and visually impaired listeners. The results of this study indicated that for the majority of data-to-display mappings there were strong similarities and general agreement between populations with regard to the magnitudes of the slopes of the scaling functions and the polarities. In the case of mapping values to frequency, as utilised in this thesis, a difference in preference between populations was found for data that represented number of dollars (positive polarity for sighted and negative for blind users). For the implementation of TableVis, a positive mapping polarity was chosen, *i.e.* mapping values to pitch of sound in such a way that higher frequencies represented higher numerical values and *vice versa*. The reasons for choosing positive over negative mapping polarity were, firstly, that this is the option implemented in most of the previous work on auditory graphs HDS derives from, and which proved an efficient way of representing abstract series of numbers [24, 28, 80, 95]. In addition, for data with specific meanings (temperature, pressure, money, etc), Walker & Lane [165] found that, with the exception of mapping dollars to frequency mentioned above, both sighted and blind groups of participants in their study preferred positive mapping polarity. Further support comes from research conducted in the 1960s by Roffler and others regarding the localisation of tonal stimuli in the vertical plane. According to Roffler & Butler [130], listeners reported sounds with higher pitch as originating at higher positions on the vertical plane than sounds with lower pitch. In the same study, they found that this phenomenon was also observed in congenitally blind persons, and in young children who were likely to be unaware of the use of the words “high” and “low” in describing differently pitched sounds. A recent study by Rusconi *et al.* confirmed this perceptual phenomenon [131].

The other aspect to consider in the design of a parametric sonification refers to the scaling function. Scaling refers to how much change in a data dimension is represented by a given change in the display dimension [166]. In the case of using pitch of sound to map values, as in this implementation, if a doubling of frequency results in a perceived doubling of data dimension represented, then the slope of the graph, or scaling function, is 1.0. If a doubling of frequency yields less than a doubling in the data dimension, then the slope of the line is less than 1.0 [165]. It is, therefore, a subjective scale used for relative magnitude estimation [145]. In TableVis, choosing a single parameter and polarity for every data set meant that the scaling factor was already determined, regardless of the type of data in the table. However, since the data exploration tasks investigated in this thesis (intended to extract overview information), required making comparisons of relative values but not magnitude estimation, the scaling function in each exploration was not considered. This is, however, an interesting and important aspect to investigate as future work, to facilitate comprehension of the data at higher levels of detail.

Other design decisions for the implementation of HDS in TableVis refer to the resolution that the display offered. This was determined by the ratio of the frequency range utilised in the display to the range of numerical values in the data set. As the interface was intended to access any 2D tabular data set so as to obtain overview information from it, regardless of the data stored in it, the implementation was designed to maximise this resolution in every case by always mapping the range of numerical values in the set to the full extent of the pitch space utilised (approximately between 62Hz -2600Hz, as discussed below).

The implementation of the auditory display was programmed in C++, controlling directly the MIDI synthesiser of the soundcard in a computer. A piano MIDI instrument was used in this implementation to produce the actual sound. The rich range of harmonic frequencies of this instrument can provide high *toneness* (*i.e.* clarity of pitch perception) [71]. The fast attack characteristic of the piano notes provides accuracy on the onset, which is well suited for fast and brief rendering of the notes.

The general algorithm that was used to map values to pitch was the following:

$$M = \left(\frac{x - S_{\min}}{S_{\max} - S_{\min}} \right) \cdot \left(M_{\max} - M_{\min} \right) + M_{\min}$$

Equation 1

where

- x: Numerical value that is mapped to pitch of sound;
- M: MIDI number (pitch) to generate the sound representing the numerical value;
- S_{min}: Lowest numerical value in the data set;
- S_{max}: Highest numerical value in the data set;
- M_{min}: MIDI number corresponding to the lowest end of the pitch space utilised in the display;
- M_{max}: MIDI number corresponding to the highest end of the pitch space utilised in the display;

With this formula, the range of numerical values in the data table (from S_{min} to S_{max}) was mapped to the pitch space ranging between MIDI numbers M_{min} and M_{max}. One advantage of this mapping strategy was that it would permit an easy identification of outliers since the outlier value would be rendered with a very different pitch than the rest of the values in the data set. This would ensure that the majority of values in the data set would be rendered in a narrow band of frequencies, providing low resolution. This is exactly what happens with visual graphing of data: in the presence of outliers, these appear as spikes emerging from an otherwise rather flat curve, also providing low display resolution. This can often be the first approach to obtaining an overview of the data. Once the user removes the outliers, the new range of values is rescaled to make use of the complete range of pitches, providing maximum resolution in that display.

While MIDI uses integer numbers to generate sounds in a 12-tone equally-tempered (12-TET) scale (*i.e.* the chromatic scale used in modern western music), the algorithm presented above (Equation 1) returns real numbers, with decimal point digits. In other words, it provides an exact mapping to a continuum pitch space rather than to the discrete 12-TET scale. Using MIDI pitch bending, TableVis was implemented to render sounds in the exact pitch, as returned by the algorithm. Additionally, a rounding option was implemented, with which the values obtained from the algorithm were rounded downwards to the nearest note in the 12-TET scale. Except in the study reported in Chapter 8, the research studies reported in this thesis were conducted using the 12-TET scale pitches only. This decision was made on the basis that sonifications would result in musical sound events that were more familiar-sounding for the majority of participants. While the dissonances produced within the 12-TET scale are commonplace in contemporary music (including film music and other non-avant-garde sources), chords generated with a continuum pitch space are much rarer and could potentially lead to difficulty in listening. The study reported in Chapter 8 that utilises the full frequency range to map sounds suggests, however, that accurate knowledge can be extracted from data with exact mappings too, and opens a direction of future research.

Following the design guidelines developed by Brown *et al.* [28] for the audio presentation of graphs and tables, the pitch space used to display the information was defined to be between MIDI values 35 and 100 (approximately 62Hz – 2600Hz). For effectiveness in the auditory display, Walker & Kramer [164] recommend using the range 200-5000Hz for sonification with pure tones. For complex sounds like the piano instrument used in this thesis, toneness is highest when its fundamental lies around 300Hz, and remains high in the range 80-800Hz [71]. The range selected for the implementation in TableVis is well centred with this range, and makes no use of high and low frequency sounds well outside this range, where toneness becomes lower.

4.5 Interaction in TableVis

The previous section was devoted to the definition of HDS and the description and justification of the design decisions in the process of implementing HDS for TableVis. This interface was intended to enable users to interactively explore tabular numerical data sets so as to obtain overview information from them, whilst providing a bridge to exploration techniques for accessing information at higher levels of detail. This section describes and justifies the design of the interaction by which users control the explorations.

4.5.1 Interaction Design Premises

Several of the user requirements summarised at the end of Chapter 3 and some guidelines derived from the literature review in Chapter 2 provide key concepts for the design described in this section. One conclusion extracted from the literature review was that it is important to provide means of generating data sonifications interactively [70] and that providing control over the interaction helps avoid information overload [46]. This was also a requirement that was derived from the focus groups reported in Chapter 3. Linking these aspects in the context of information browsing, Marchionini [96] states: "*…information must be assessed and examined during browsing. The information seeker's challenge is to make rapid judgements of relevance, and the designer's challenge is to provide flexible display facilities for examination and assessment.*" He adds: "*…browsing may be more of a direct manipulation strategy than the formal methods are because it involves cognitive, perceptual, and motor systems concurrently and continually*" (p.101). Referring to interaction techniques implemented in auditory interfaces, Fernström *et al.* [48] suggest that user interaction can be viewed not as discrete events (*e.g.* clicking a button) which generate processes without further user control, but as "*a continuous flow, such as a pen stroke, where we continuously move a pencil on a surface, relying on our learned gesture through proprioception, as well as haptic, visual, and auditory feedback*" (p.37). This view, they suggest, is increasingly important due to the availability of input devices such as pens and cameras which can detect complex human actions. This continual involvement of cognitive, perceptual and motor systems forms the basis for the interaction method for TableVis, and the means by which users control the exploration process.

Another requirement collected form the target end users was that the interface should make it possible for the user to retain a sense of context to avoid "getting lost" in the data. In the visual domain, Bederson & Hollan [10] put this formally by saying: "*One challenge in navigating through any large data space is maintaining a sense of relationship between what you are looking at and where it is with respect to the rest of the data (i.e. balancing local detail and global context)*" (p.20). In a similar line, Walker [166] argues strongly for the presentation of context information in the design of auditory graphs, asserting that additional contextual marks shown in visual graphs, like scale divisions on the coordinate axes, should also be provided with the presentation of auditory graphs. Also relating to the need to preserve context during the exploration, an additional requirement collected from the focus groups stated that the data being accessed should fit on one page or work space.

Blind and visually impaired participants in the focus groups also reported that using two hands helps retain context information in data explorations. It was thus determined that a good way of presenting the context of the whole data set was one that could be explored tactually while not interfering with the auditory representation of the information subset in focus. This was achieved with the use of a physical, tangible device on which to present the data. A WACOM graphics tablet [35] was chosen for this, as described in the next subsection. Once the tabular information was made available on the surface of the tablet, users could access data by directly pointing at them and retrieving them in sound.

4.5.2 Support for Focus+Context

A graphics tablet is a tangible input device consisting of a flat rectangular surface on which the user can control the position of the mouse pointer by moving a wireless electronic pen over the surface. Unlike an ordinary computer mouse, which is a relative positioning pointing device (*i.e.* the position of the mouse on the desk does not determine the coordinates of the pointer on the visual display), a graphics tablet can be configured to function as an absolute positioning pointing device. In other words, there is a one-to-one-relationship between the coordinates of the pointer on the computer screen and the rectangular working area of the graphics tablet. As a consequence, by pointing at the same physical position on the active area of the tablet, the computer always receives the same pair of coordinates for the pointer. In TableVis, the tablet was augmented by adding a physical, tangible frame around the active area, delimiting it and making the borders easily identifiable using both hands, as shown in Figure 6. As reported in Chapter 3, classroom observations showed that when visually impaired students explored tactile graphs, one of the hands (the non-dominant hand) provided reference to known locations in the exploration area (*e.g.* the origin of coordinate axes, in the case of a Cartesian graph), while the other (dominant) hand accessed information on the graph. The use of a tangibly-enhanced graphics tablet was intended to enable the users to take a similar approach to data exploration. The interface was further augmented by the addition of a small tangible dot on the middle of each side of the frame, providing additional navigation cues. Using these tangible references, and simply by relying on their senses of proprioception (the sense of the position of the limbs) and kinesthesis (the sense of the movement of the limbs) [59], users could tell how far apart one hand was from the other, and judge approximately where on the tablet the pen had been placed. Similarly, they could also change the position of the pen by jumping directly to specific areas, such as the centre of the rectangle or any of the corners.

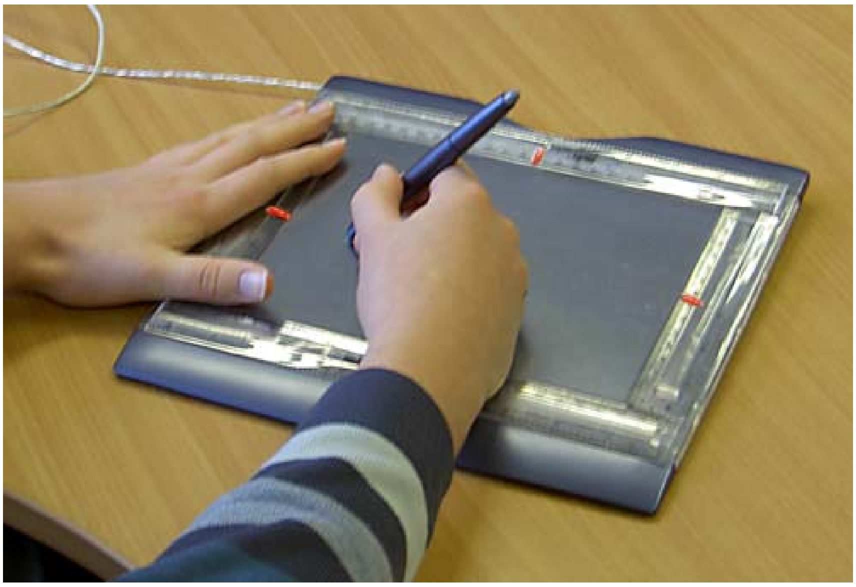
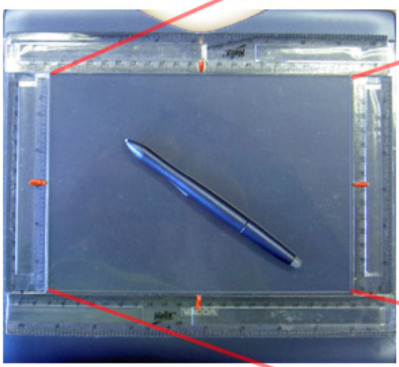


Figure 6. Graphics tablet augmented with a tangible frame delimiting the working area. Red tangible dots can be seen indicating the midpoint of every side of the rectangle. In the image, a user feels the tangible frame with the left hand while moving the electronic pen over the working area with the right hand.

Making use of this non-visual navigation setup, the data table to be explored (whichever its number of rows and columns) could be scaled and presented on the working area of the tablet, in such a way that the data table completely filled the working area of the tablet, as represented in Figure 7. Presenting the information in this way preserved the spatial relations of a printed or visually displayed table, in order to facilitate the use of a common language between blind and sighted people collaborating using the same data.



82990	81252	78202	80073	78202	80629	80255
77327	80137	76129	78329	78172	80083	79115
72362	75306	74039	73644	72402	74290	73509
77424	72057	74503	73522	72173	75398	73257
74293	76649	73002	76224	74151	75046	74220
74206	76178	75438	74967	76843	76949	77625
79122	81579	80229	81620	77252	77768	81480
80447	78475	78504	82942	79679	81552	81867
77977	76932	78888	81039	80918	80666	81894
75282	76220	77252	75390	76861	76280	76456
78360	80530	78793	82061	78002	82644	82537
81930	82683	78879	80409	82399	82789	82653
80493	79647	78670	77294	80160	80124	78071
77126	72833	75770	73530	76117	72447	72588
70609	71454	69448	71382	67003	71334	69089
71935	68943	73637	70015	72224	70106	73368
74309	74096	74880	70932	74154	74302	71128
74457	74277	71373	72871	73069	71640	73715
78268	74478	76823	78169	77840	76497	77895
81803	83956	83704	83566	79876	83833	79497
83424	85723	85555	82359	82403	85446	83026
85210	85985	85159	86189	83203	83936	81797
79231	82386	81356	80109	81761	79581	83029
80642	76131	78330	78139	80482	79480	75770

Figure 7. The data table to be explored is presented on the active area of the tablet, scaled to fill it completely.

Having described the basic setup for data explorations with TableVis, it should be highlighted that the following three *invariants* are essential for a user to be able to maintain context during data browsing:

I. *The complete data set is always on display*. The complete data set fills the active area of the tablet and that the physical, tangible borders of this active area correspond to the boundaries of the data set;

II. *Information remains stationary in space*. Each cell in the table always occupies the same physical location on the tablet.

III. *The active area on the tablet has a fixed size*. The user knows the dimensions of the active area of the tablet.

A user can make use of these three invariants to maintain contextual information throughout the exploration, and to be able to point directly at different areas on the data set, without having to access information sequentially. Because the complete data set is presented on the tablet, there is no need to scroll, and since the location of the cells remains static, once an area of interest has been identified on the data set, the user can be confident that the data will always be located at the same physical position. By feeling the physical frames around the working area of the tablet, with the use of the senses of proprioception and kinesthesis, a user can move the pen quickly and relocate it directly on specific areas of the data set. To summarise, as users access specific pieces of data on the tablet using the electronic pen (area of the data in focus), they can keep track of its position in relation to the physical boundaries of the tablet (that is, in relation to the boundaries of the data set), maintaining contextual information at every moment.

4.5.3 Generating Data Sonifications Interactively

The exploration environment designed for TableVis as described above was intended to enable users to explore tabular data interactively and to obtain overview information using HDS. In the case of the 2D-tables studied in this thesis, using the canonical implementation of HDS offers two complementary views of the data set: one exploring the data by rows and the other one exploring by columns (as discussed in Section 4.4.1). The implementation of TableVis allowed the user to retrieve these two complementary views by implementing two navigation modes between which the user could switch at any time. These navigation modes are referred to as the *rows* and *columns* modes. In each of these modes, the complete row or column being pointed at with the pen was rendered in HDS. These two modes facilitated quick overviews of the complete data set by retrieving both complementary views of the data: drawing a line across the whole table (horizontally in columns mode or vertically in rows mode), a sonification of all the data in the table was interactively generated, as the pointer moved over different rows or columns. The sonification of a row or column was rendered in a single sound event as soon as the user entered the area on the tablet corresponding to that row or column. The user could choose to repeat the sonification of any particular row or column by tapping on the same position on the tablet that generated that chord. The movement and trajectories of the pen were decided by the user, who could move back and forth or lift the pen and place it down on a different area, in which case the rows or columns in between would not be sonified. By retrieving complete “slices” of data in this way, with each one rendered as a chord, a user could easily compare them against each other (a detailed study about the actual information that can be obtained from such comparisons will be reported in Chapter 8). Switching between rows and columns modes was achieved by pressing one of the buttons on the electronic pen of the graphics tablet, which is normally operated with the index finger (most models include two buttons on the pen, placed next to one another.

An additional exploration mode, the *cells* mode, was also implemented in TableVis. In this sonification mode, only the cell pointed at with the pen was sonified through parametric mapping to pitch, utilising the same parametric sonification as described in Section 0. As the pen moved over the tablet, every time it entered a new cell, the piano sound corresponding to that cell was played. By tapping with the pen on the same position, the sonification was repeated. This mode was intended for “freehand” browsing of the exploration area, without any constraints of direction or speed. The user had full control over where to go and at what pace to sonify individual cells. Some obvious limitations of this exploration mode, as compared to the rows and columns modes, were that the user could not traverse all the cells in the table systematically, and even traversing a single row or column in a straight line seemed a difficult task. Still, in the initial pilot evaluation of the interface prototype (reported in Section 4.6) the cells mode was also made available to the users, so as to record their first impressions about it. In such case, changing the exploration mode with the button on the pen, instead of alternating between the rows and columns modes, followed the pattern “cells-rows-cells-columns-cells-rows…”. In every case, once the sonification mode was changed, a synthesised voice confirmed the new selection, speaking the word “cells”, “rows” or “columns”, depending on the mode that was selected.

4.5.4 Accessing Higher Levels of Detail

The research questions in this thesis are focused on obtaining overview information. However, as discussed in Chapter 2, obtaining an overview is often an initial stage in the complete exploration process of a data set, summarised by Shneiderman [138] as "*overview first, zoom and filter, then details on demand*". It was therefore a requirement that the techniques proposed to obtain overview information could bridge with other existing techniques that could be utilised to extract more detailed information. Two interaction techniques were provided in the interface to achieve this, which are described in this section: control over the speed of rendering sonified data and access to information in full detail using synthesised speech.

The discussion in Section 4.3 justified the proposal of HDS as a technique to obtain overview information that derived from established auditory graph designs, where data rendering speed had been increased. It was observed that rendering such graphs at high speeds (at which it was difficult to obtain detailed descriptions of the graphs) facilitated the extraction of a very general overview from them, which was easy to compare with the overview information obtained in similar ways from other graphs. This approach was taken to the extreme of rendering a series of numbers at a speed so high that all sounds were perceived to be simultaneous. It was a design decision not to implement HDS as actual chords and to render the events as very fast successions. In this way, the only difference between the generation of HDS and that of auditory graphs would be the speed of rendering sounds, although perceptually and cognitively the result was very different. Speed for rendering sounds is a parameter that can take any value from a continuum. As such, informal tests showed that being able to control speed of rendering in real time within that continuum could provide a useful way of controlling the trade-off between ease of extracting very general overview information from the whole data set and finer detail in the features from specific areas of the data set. This thesis does not study in depth the transition of levels of detail using this technique. However, this functionality was made available in the initial evaluations. In the first prototype implementation of the interface, the scrolling wheel of a mouse was used as the control for varying the speed of rendering sounds, which by default was always set at maximum to render as HDS. Turning the wheel in the direction away from the user increased the sonification speed by shortening the duration of each sound. Turning the wheel towards the user had the opposite effect. The mouse was placed on one side of the tablet (on the side of the non-dominant hand of the user). It was disabled as a pointing device, and thus could not control the position of the pointer; only the wheel was in use. More elegant solutions were implemented following the initial pilot evaluation, as described at the end of this chapter.

The other technique provided to access detailed information was speech synthesis, as an implementation of the "details-on-demand" in Shneiderman's mantra. As discussed in Section 0, mapping data values to pitch is not useful for accessing absolute values. While exploring a data set, a user may at any time want to know the absolute value of a particular cell in the table or the exact readings corresponding to a cell, row or column, on the scales of the coordinate axes. To overcome this limitation of non-speech sound, on-demand speech was provided in the TableVis prototype. Details in speech were obtained by pressing one of the buttons on the pen (the other button was allocated to change the sonification mode, as explained above). When the rows mode was selected, the label corresponding to that row was read out in speech; the same was true for the columns mode. In the cells mode, both row and column labels were read out, followed by the numeric value in the selected cell.

The user was thus able to obtain fully detailed information about coordinates and numeric value for a particular position on demand during the data exploration.

Pressing the front button on the pen activated a message giving the following information, depending on the sonification mode:

- *Rows mode.* Label of the current row;
- *Columns mode.* Label of the current column;
- *Cells mode.* Labels of the current row and column, plus value in the cell;

For example, in the cells mode, in a table with information about average monthly rainfall in different years, the speech information would be something like: "*March, 1995. Value: 327*". If with the same data set the user was exploring in the columns mode to compare the monthly average rainfall over a range of years and found a particularly rainy month, requesting details for that column would result in the speech synthesiser saying "*March*", for an example. The labels of rows and columns are part of the metadata of the data set. So is the general description of the data set. A function was added to the prototype to request a description of the meaning of the data in the data set, to be read by the speech synthesiser on request. This is probably the most basic overview information about a data set, which gives a meaning to the data. Although this function is essential, it was not utilised in any of the evaluations and studies in this thesis, because the meaning of the data was made available to the participants by the experimenter, in the introduction to each study. Therefore, a design decision about the best way of making this information available on demand was not made. Still, it is an important aspect to consider in future developments.

Speech output is compatible with non-speech sound in multimodal interfaces, and the combination of both (each providing a different kind of information) can be preferred by users of assistive technologies, as research suggests [109]. In TableVis, both sonified relative numerical values and spoken metadata and absolute values can take place simultaneously if the user decides to keep browsing data while listening to the information in speech. To ensure maximum responsiveness in the interaction, if a user requests speech details in a cell before the speech details for another cell are finished, the previous speech message will be truncated before the new one begins.

The following section reports an evaluation of the prototype implemented as described in this chapter. From the outcome of that evaluation, improvements to the design were added.

4.5.5 Design limitations

HDS in its implementation in TableVis described in this chapter, has been presented as a technique that can cater for any kind of data set within the broad subset considered in this thesis. However, when making design decisions to construct this particular implementation, some limitations were unavoidably introduced as to how a user can approach the exploration of a data set. One main limitation is the availability of only two directions of exploration that were made available: explorations by rows and by columns. While in theory this cannonical implementation of HDS can offer enough complementary views of the data set to be able to extract the main features in it, it is clear that other modes of exploration might facilitate the exploration of specific data sets. Some of the possible alternative direction of exploration have been mentioned earlier in this chapter, like "expanding wave" explorations centred on a particularpoint of the data, or explorations in directions that are not orthogonal with any of the coordinate axes.

The implementation of the cells mode can also limit the explorations strategies that a user would like to conduct. It was taken as a premise the fact that a freehand approach to exploring the data set cell-by cell can be useful to get a gist about the data, but not as a method for non-visual systematic exploration. Still, it is cenceivable that alternative implementations of this exploratory mode could provide with tools that might prove useful in many instances. One such implementation would invilve accessing clusters of data around a particular cell simultaneously. With one such implementation, a user would be able to sample the data more efficiently, covering a broader proportion of the data point in only a few strokes, depending on the size of the region covered around the position pointed at. It is clear that many other variants of these implementations could be deviced, which would be likely to prove useful in particular instances.

One further consideration is necessary to be made in this discussion about the limitations introduced by the particular implementation design decision presented in this theses. While the form in which data are made available to the user for exploration are presented as neutral and capable of accommodating many exploratory strategies (sometimes subject to the availability of a variety of exploration tools, as discussed in prevously in this section), in reality this form of presentation does impose some decisions aleady made on behalf of the user. For example, when deciding to present information evenly distributed over the active area of the tangible interface, a decision is implicitely made regarding the scale employed on the coordinate axes. For this implementation of TableVis, the scale is chosen to be a uniform one. Some users might want to use other types of scales that are meaningful for the exploration of particular data sets. An example from visual graphs would be the use of logarithmic scales. Such other scales cannot be used with TavleVis with the implementation presented here. Future work should be conducted to investigate ways of overcoming these design limitations.

4.6 Prototype Evaluation of TableVis

The prototype of TableVis was evaluated at the Royal National College for the Blind (RNC) in Hereford (UK). This qualitative evaluation was designed as a think-aloud session, in which the prototype was described to each participant and tasks were set for them to try out each aspect of the functionality available. Then, some additional tasks were set for the participants to explore data sets while think aloud was encouraged.

Throughout the evaluation session, information about various aspects of the prototype was collected, specifically: the participant's preferences, views on usability, usefulness, areas for improvement and browsing strategies.

4.6.1 Participants

Five participants, two females and three males, with ages ranging from 16 to 35, took part in the evaluation. All of the participants were students attending the RNC. Four of the participants were totally blind (three of them congenitally), and the fifth was partially sighted. Each evaluation took approximately one hour.

4.6.2 Data sets

Five data sets were prepared based on real or realistic data, a decision grounded on both the requirements capture discussed in Section 3.5.3 and suggestions for good practice in the study of data sonification derived from the literature [166]. The participants were asked to browse these data sets and carry out some tasks set by the experimenter. The data sets and browsing tasks are summarised in Table 1

Data Set	Topic	Table Size (rows x columns)	Type of Data	Tasks
1	UK population statistics	16 x 15	Real data about the population in the UK between 1991 and 2005 divided in 16 age groups covering population ranging from 0 to 79 years old. (Source of the data: the U.S. Census Bureau [29])	Which age group has the most people? What is the evolution of the oldest age group? What is the evolution of births?
2	Average monthly rainfall in Kenya (mm/m2)	12 x 11	Realistic data about the average monthly rainfall (mm/m2) in Kenya, between the years 1990-2000 (based on typical rainfall data in Nairobi [9])	Can you see any overall trends? Is rainfall the same all year round? Is there any rainy season? How many? When? Has there been any occasional storm? When?
3	Usage of 50 Laboratory Computers in a week	24 x 7	Realistic data about the usage of 50 computers in a hypothetical computer laboratory in one week	Can you see any overall trends? When is the busiest time of the week? Do people stay working late? What happens at night? What happens in the weekend?
4	China population statistics	16 x 16	Real data about the population of China between 1990 and 2005 divided into 16 age groups ranging from 0 to 79 years old. (Source of the data: the U.S. Census Bureau [29])	Which age group has the most people? What about old people? What about births?
5	Average monthly rainfall in Wales (mm /m2)	12 x 11	Realistic data about the average monthly rainfall (mm / m2) in Wales, between the years 1970 and 1980 (based on typical rainfall data in Cardiff [8])	Can you see any overall trends? Is rainfall the same all year round? Is there any rainy season? How many? When? Has there been any occasional storm? When?

Table 1. Summary of the data sets utilised to evaluate qualitatively the first prototype of TableVis

4.6.3 Procedure

Training for each participant was carried out using data set 1. The type of data (in this case population statistics for the UK) was described to the participant. The whole table was played in both rows and columns modes, all the functionality was described and the participant then tried out the interface. The user was then asked to explore the data in the different modes, to obtain an overview of the entire data set (exploring by cells, rows and columns). Finally the user was asked to provide answers to questions which would demonstrate an understanding of the data (see Table 1).

Following the training session, further data sets (2-5 in Table 1) were presented to each participant. In each case the type of data set was firstly described to the participant, who was then asked to explore the data using the functionality provided, and to complete some tasks to establish the capacity for TableVis to provide an overview of general trends as well as more detailed information on demand. Following completion of each task, the participants were asked to comment on their approach to obtaining the information and on the user experience.

Depending on the level of skill that the participants acquired using the TableVis prototype during the evaluation session, and also depending on the number and extent of the comments that they wished to make after each task, a varying number of the tables was shown to different participants. All the participants completed 3 or 4 data sets out of the total of 5 available. The experimenter prioritised obtaining participants' comments over completion of all the tasks within the one hour session.

Since a controlled experiment was not carried out, the format of each evaluation session was kept flexible. Comments from users were continuously recorded. The qualitative feedback collected in this study is summarised and discussed in the following section.

4.6.4 Qualitative feedback

One of the participants had some difficulty in distinguishing the terms "row" and "column" (the same issue was reported in the focus groups, where some participants mixed up the concepts, as discussed in Section 3.5.1). Although the experimenter had to clarify these terms several times during the evaluation, the participant's performance carrying out exploratory tasks was as good as the other participants'.

The parameter mapping technique implemented in TableVis appeared to work well for all the participants, who found it easy to understand and use: "*Telling the value listening to the sound is quite easy*" (Participant 2); "*It's easy to tell from the sound*" (Participant 1). This was felt to be useful for comparing relative values and detecting general trends within the data: "*Even if you're not absolutely spot on by figure, you should have a really good rough idea*" (Participant 4). Indeed, a number of the participants commented that that TableVis was good for obtaining a quick overview: "*It could be useful for quick overview information – that's what I like about it. It could also be useful for getting detailed information without reading the numbers and stuff*" (Participant 5); "*Often you need that kind of tool if you're doing analysis and stuff. It gives you a quick overview*" (Participant 3). This was felt to be an improvement over the sequential read-out of values using a screen reader: "*Often you just want an overview. It's very time consuming having things read out with speech*" (Participant 5). The ability to obtain details on demand in speech was also found to be useful: "*It's very useful because as well as being able to click and have whatever particular information you want, you know, you can also just get a general overview by the sounds*" (Participant 2).

Control over the exploration was of key importance to most of the participants. The pen was felt to offer a high degree of control over the exploration. Control over the exploration also enabled the participants to retain a sense of context within the data set: "*I think you need to do it yourself because, you know, you can take control of it. If you're listening to it, sometimes you're listening to something and you don't quite hear it all at the same time*" (Participant 2); "*I keep track of where I am*" (Participant 1); "*It's very good. Without that you'd never find your way around it [the data set]*" (Participant 3).

One of the difficulties encountered by participants was keeping in a straight line when navigating with the pen in cells mode. One participant commented that they found it difficult to know when they had entered a new column. However, when asked whether that navigation mode should be removed from the interface, another participant commented that "*It's better to have plenty of options and choose the ones you want*" (Participant 4). It was also observed that the use of the mouse wheel to control sonification speed was not an ideal setup because users expected the mouse to be used for navigation. While most of the participants said they liked using the tablet interface, it was observed that some participants had difficulty using the buttons on the pen (the front button for speech details and the back button to change mode). Participants sometimes pressed the wrong button, changing exploration mode accidentally which led to some confusion. Despite this, participants commented that they liked using the pen as it offered control over the exploration. When asked about the potential for more widespread use of TableVis, participants said they felt that a finished product could prove useful for blind and visually impaired users, and some commented that they would be keen to try out future developments of the interface.

4.6.5 Discussion

The feedback described above shows that users valued having control over the exploration, both in obtaining overviews and accessing details on demand, supported by the functionality for switching between modes and altering the speed of sonification. As discussed in Section 2.3, this need for control in multimodal techniques for data exploration has also been stressed in the literature [46] [144] [70].

The feedback also indicates that users found it easy to obtain quick overviews of the data, and that active exploration of the data sets using the tablet and pen helped users to retain a sense of context and enabled them to return to areas of interest for further analysis (the zooming and filtering stages of Shneiderman's Information Seeking Mantra [138]).

A few practical issues were highlighted by the pilot evaluation regarding the hardware setup. Firstly, having two buttons on the pen with two different functions led to confusion for some users, who often accidentally selected the wrong function. Secondly, using the mouse wheel to alter sonification was sometimes confusing, as users expected the mouse to act as a navigation device (although once this difficulty was overcome, the functionality provided by the wheel was shown to be useful). It was decided that these issues would be addressed in a further design iteration, while still retaining support for the key requirements defined in Chapter 3.

In order to address the problem with the pen buttons, where some users found it difficult to differentiate one from the other, it was decided that both pen buttons should provide the same functionality, of details on demand. This left two other functions to be implemented: navigation mode and speed of sonification. To change navigation mode it was felt that a click or push-button metaphor made sense, while speed of sonification required more continuous control (which was the motivation for using the mouse wheel, as a "scrolling" tool). However, as derived from the requirements capture, it was necessary to keep the interaction method as simple as possible. An input device incorporating a rotary knob and push-button function (Griffin Technology's PowerMate [61]) was chosen as this could accommodate both controls within an easy to use interface (shown in Figure 8).

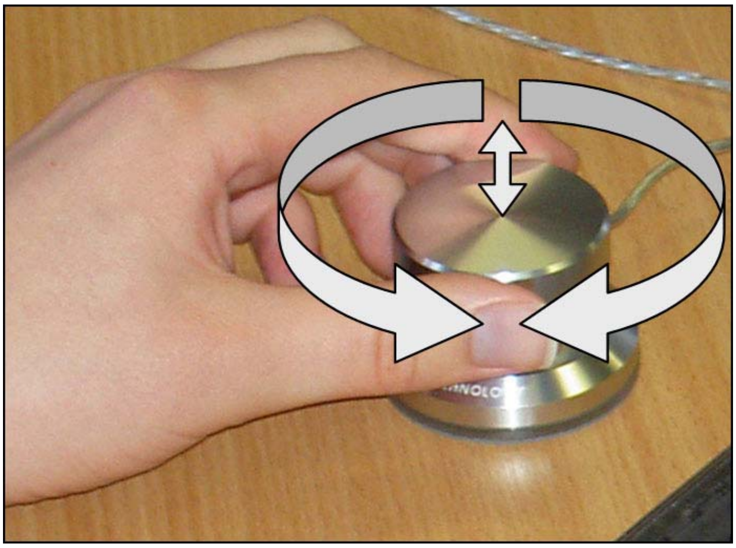


Figure 8. Input device (Griffin Technology's PowerMate) with rotary knob and push-button functions, used in TableVis to control the speed of the sonification and change the sonification mode.

Speed of sonification was controlled by turning the knob clockwise to increase speed or anticlockwise to reduce. This choice was modelled on a "hi-fi" volume dial metaphor, which informal tests showed to be a familiar and easily understandable interaction method. In order to change exploration mode, the user simply pressed the knob as a push button. The device was placed on the side of the non-dominant hand for easy access during the interaction. These changes were judged to have solved the difficulties with the hardware setup as found from user feedback.

4.7 Conclusions

This chapter began by defining what is understood by data tables within the scope of this thesis. It then introduced and justified the concept of High-Density Sonification (HDS), a novel parametric sonification technique to enable obtaining quick overviews of collections of data series, such as the tabular numerical data sets studied in this thesis. The implementation of HDS in TableVis, an interface to explore tabular numerical data sets non-visually, was then described and justified. A pilot evaluation of the interface was then reported, which showed that HDS as implemented in TableVis could be used to obtain overview information from tabular numerical data sets. It also validated the usability of the interface, while highlighting some issues which were then addressed by a further design iteration. As such, this chapter has proposed a technique to tackle the exploration need stated in the first research question: "*How can overview information from tabular numerical data sets be obtained non-visually?*"

The chapter has provided a general definition of HDS, which can be implemented in a variety of auditory displays. This offers considerable scope for the technique to be adopted and evaluated by other designers. In TableVis, canonical HDS is implemented, which for 2D tables results in the grouping and sonifying of cells with the same value on one of the axes.

The next chapter reports a study in which the implemented prototype of TableVis was fully evaluated, both quantitatively and qualitatively, with blind and visually impaired participants.

Chapter 5 TableVis and HDS: Quantitative and Qualitative Evaluations

5.1 Introduction

Chapter 4 described the design and implementation of the TableVis prototype, and reported the first evaluation of High-Density Sonification using TableVis in comparison to keyboard navigation, demonstrating that TableVis outperformed the keyboard in terms of reducing mental demand, effort, frustration and annoyance, and improved perceived performance level.

This chapter reports a study in which the implemented prototype of TableVis was fully evaluated, both quantitatively and qualitatively, with participants from the main target population of blind and visually impaired users. This study was designed to contribute towards answering both research questions defined for this thesis in Chapter 1: "*How can overview information from tabular numerical data sets be obtained non-visually?*" (RQ-1) and "*What level of performance can be achieved with the techniques proposed in this thesis?*" (RQ-2).

As a contribution towards Research Question 2, a quantitative evaluation of the prototype was carried out where performance (effectiveness and efficiency) with TableVis and HDS was compared against performance obtained with the current dominant accessibility technology (a screen reader utilising synthesised speech). Participants' subjective workloads were also compared between the two conditions.

A qualitative study was also conducted with the aim of contributing information in answer to Research Question 1. This study aimed to determine which exploratory strategies and procedures were utilised by analysing the exploration traces recorded from the eight blind and visually impaired participants during the experiment, as well as the feedback provided by all the blind and visually impaired participants who evaluated the interface with HDS. For the purposes of this study, an *exploratory strategy* is defined as the action plan that a user devises to carry out the exploration of a data set in search for information. *Exploratory procedures* are the ways in which this strategy or action plan is implemented (the ways in which actual interaction with the system takes place). This qualitative analysis of exploratory strategies and procedures led to a better understanding of the challenges and cognitive processes involved in the exploration of tabular data for overview information.

5.2 Experimental Study

The results from the pilot evaluation of TableVis reported in the previous chapter showed the benefit of using direct pointing (with a graphics tablet and the navigation modes implemented in TableVis) for exploration of tabular numerical data. The way in which data from the tables were conveyed (either as synthesised speech or as non-speech sounds generated using HDS) was the only parameter that changed between the two versions of TableVis that were used for the experiment reported in this chapter.

5.2.1 Design of the experiment

A two-condition, within-subjects, repeated-measures experiment was conducted to quantify the differences in effectiveness, efficiency and subjective workload when using each form of information presentation in overview exploration tasks:

- *Condition-I: non-speech sound condition*. Explore numerical data tables for overview information using TableVis, retrieving information in sound generated with HDS;
- *Condition-II: speech condition*. Explore numerical data tables for overview information using TableVis, retrieving information in synthesised speech;

Condition-I (HDS) included all the functionality of the TableVis prototype described in the previous chapter, with the exception of the cells mode. This mode was not made available in order to conduct the evaluation only in the rows and columns modes, *i.e.* those most suited to explore for overview information. The default sonification speed was set to the maximum, so that rows and columns were perceived as chords, but the participants could modify this setting, if desired, at any time during the experiment.

In Condition-II (speech), the functionality available in the interface was the same as in the non-speech sound condition, except that the value in each cell was conveyed in speech instead of being mapped to pitch of sound. In the rows mode, the cells in the selected row were read from left to right (top to bottom in the columns mode) at a speech rate that was set by each participant before the experiment, to match the speech rate they normally used with screen reader applications. In this way, the interface in this condition offered a compromise between speed and intelligibility that was optimised for each participant. To further optimise the use of the interface in the speech condition, users could turn the rotary knob during the experiment to select the number of digits of precision with which values were spoken. Thus, any number of most significant digits (MSD[2]) could be used, bearing in mind that the smaller the number of significant digits, the less accurate the information was, and less time was required to render data in speech.

The aim for providing this functionality was that the participants could retrieve data with the minimum necessary accuracy to obtain overview information. The number of significant digits that are required to obtain overview information at a particular level of detail varies depending on the data in the table. In the example shown in Figure 9 (left), retrieving the first MSD is enough to observe that values grow both rightwards and downwards. In Figure 9 (right), on the contrary, the first MSD alone would show a table with the same value – three million – in all the cells. However, presenting also the second MSD would reveal that the values in this example follow a similar trend as in the example on the left.

[2] The MSD in a number is the leftmost, non-zero digit in that number. It is the digit with the greatest value in the number. For example, in the number 4,326,436.394, the most significant digit is the 4 at the left end.

3,425,263	4,579,389	5,320,147
6,568.453	7,679,128	9,474,596
9,141,853	10,002,730	11,256,026

3,125,263	3,279,389	3,320,147
3,468.453	3,579,128	3,674,596
3,741,853	3,802,730	3,956,026

Figure 9. *Left:* Example of a data table in which the information in the first MSD is enough to identify the trend that the values in the table follow. *Right:* Example of a data table in which the first two MSDs are required to extract information about the trends that the values in the table follow.

A training session was provided at the beginning for each participant, to make sure that all the concepts involved were understood and to enable the participants to familiarise themselves with the interface and functionality used in the experiment. In the case of the non-speech sound condition, after the training session participants understood the relation existing between numerical values in the cells and pitch of sound used to represent those values. They also understood that accessing a full row or column using HDS was to hear all the sounds corresponding to all the values in that row or column, played as a chord. Thus, the row or column with the highest values in it would be the row or column that produced the chord that they perceived with the overall highest pitch. They were instructed to trust their first impression about the perceived overall pitch of a chord. The underlying hypothesis was that in a set of rows or columns rendered with HDS, the row or column with the highest *Perceived Overall Pitch* (POP) would also result in the highest arithmetic mean. Throughout the thesis, evidence supporting the correctness of this hypothesis is collected. In Chapter 8, an extensive study is reported in which the validity of this arithmetic mean estimation technique is analysed and quantified.

In order to keep each experimental session within one hour duration, a maximum of 2 minutes was allowed to explore each table and answer the overview exploration question (informal pilot testing suggested that this time was appropriate to perform the tasks). Each overview exploration task started with one of two possible types of question:

- *Question type I.* Which time of the day / day of the week gets more / less visits?
- *Question type II*. In which quadrant[3] of the table are values highest / lowest?

The first question type required exploring the table in one direction only, giving a precise answer (a particular row or column). The correct answer for this type of question was judged based on the arithmetic means of the rows and the column. With the second question type, the table had to be explored in both directions (horizontal and vertical), and only one of four areas had to be identified. In the case of this type of question, the correct answer was the quadrant containing the absolute highest or lowest value in a cell, depending on the question.

[3] Top-left, bottom-left, top-right or bottom-right.

For each condition, workload was measured using a modified NASA-TLX test [64]. NASA TLX scales allow for participants to rate on a set of unmarked scales the subjective workload that they experienced during an experiment. A table of the attributes used to define workload in the modified TLX scales used in this evaluation is shown in Table 2. Following Brewster [21], the NASA-TLX scales were modified for the studies in this thesis with the addition of two attributes: *annoyance experienced* and *overall preference*. Badly designed auditory displays have often been associated with annoyance, as a recent study by Frauenberger *et al.* also reported [54]. Therefore, it was judged important to measure participants' annoyance regarding the auditory displays developed in this thesis. The overall preference attribute was also incorporated as a metric to provide a relative overall subjective rating of the conditions compared in the studies.

Attribute	**End points**	**Description**
Mental Demand	Low/High	How much mental, visual and auditory activity was required? (*e.g.* thinking, deciding, calculating, looking, listening, scanning, searching)
Physical Demand	Low/High	How much physical activity was required? (*e.g.* pushing, pulling, turning, controlling)
Time Pressure	Low/High	How much time pressure did you feel because of the rate at which things occurred? (*e.g.* slow, leisurely, rapid, frantic)
Effort Expended	Low/High	How hard did you work (mentally and physically) to accomplish your level of performance?
Performance Level Achieved	Poor/Good	How successful do you think you were in doing the task set by the experimenter? How satisfied were you with your performance?
Frustration Experienced	Low/High	How much frustration did you experience? (*e.g.* were you relaxed, content, stressed, irritated, discouraged)
Annoyance Experienced	Low/High	How annoying did you find the browsing technique?
Overall Preference	Low/High	Rate your preference for the two browsing techniques. Which one made the task the easiest?

Table 2. Table showing the attributes and descriptions of modified NASA TLX workload questionnaires used in the experiment. Modified from Hart and Staveland [64].

The interactions of all the participants during the experiment were recorded, including all the traces of the movement of the pen on the tablet and all the actions performed, such as changes in exploratory mode and requests for detailed information in speech. Everything was time stamped, so that a timeline of the explorations could be reconstructed, including the speed of movement of the pen at every point. Additionally, after both experimental conditions had been completed, participants were given the opportunity to provide feedback on their experience, in an informal discussion. The qualitative evaluation of these data will be discussed in detail in Section 5.3.

5.2.2 Hypotheses

The following hypotheses were formulated for the quantitative experiment:

i. The time required to complete the exploration tasks will be significantly lower (*i.e.* efficiency will be significantly higher) with HDS than with synthesised speech;

ii. The accuracy, measured in percentage of correct answers (*i.e.*, the effectiveness) will be significantly higher with HDS than with synthesised speech;

iii. The subjective overall workload (calculated from modified NASA-TLX) will be significantly lower with HDS than with synthesised speech.

5.2.3 Data sets

Fictitious yet realistic data representing the number of visitors to an internationally popular website were produced for the experiment (see Appendix D). All the data sets had the same metadata: 7 columns (days of the week) and 24 rows (hours of the day). This metadata could accommodate results that were equally likely around the clock (the website could be at its busiest any time of the day or of the night, and any day of the week). A total of 24 tables (12 in each condition) were presented, with 6 questions of each type in each condition. The data sets in each condition were presented in random order to the participants.

5.2.4 Participants

Sixteen participants took part in the experiment: 8 totally blind participants (5 of them congenitally blind) and the remaining 8 visually impaired. They were recruited among the students and members of staff from The Royal National College for the Blind (Hereford), and students from the Uddingston Grammar School (Glasgow). 10 were male and 6 were female, with ages ranging between 16 and 55 years old. All the participants that took part in this study used screen readers to access information in computers, except two partially sighted students who used screen magnifying software and another two who used large fonts and special colour schemes.

The group recruited was heterogeneous regarding their type of visual impairment, skills and previous knowledge of relevant basic concepts. As a consequence, training sessions had different durations in order to bring all the participants to a similar level of understanding of the tool and tasks. In the cases in which a substantial amount of the experiment time had been spent in providing training and only one condition could fit into the experiment, only the non-speech sound condition was conducted. From a total of 16 participants recruited for this study, only 6 participants completed both conditions (4 of them totally blind, 2 congenitally). Another 6 completed only the non-speech sound condition (of these, 2 were totally blind, one of them congenitally), and the remaining 4 participants could not complete any of the conditions (due to hearing impairments or a lack of familiarity with basic concepts about tabular data structures). Among the 6 participants who completed both conditions, 3 of them started with Condition I and the other 3 with Condition II.

5.2.5 Results

The differences in efficiency and effectiveness measured in this experiment are shown in Figure 10 and Figure 11 respectively. In both figures, the results for the non-speech sound condition are the average of the scores obtained by 12 participants, while the results for the speech condition are the average of the results of only 6 of those participants. Since the populations in both conditions were not the same, two-sample (between groups) T-tests were performed. To test the third hypothesis regarding subjective workload, however, only the participants who used both conditions completed the modified NASA-TLX questionnaire, thus it was possible in this case to conduct two-sample (within groups) T-tests.

The results of the T-test showed that the average time to complete the task in the speech condition (M=39.1 sec., SD=13) was significantly longer than when HDS was used (M=70.5 sec., SD=29.4), (T_6=-2.496, p=0.047<0.05). These results prove the first hypothesis.

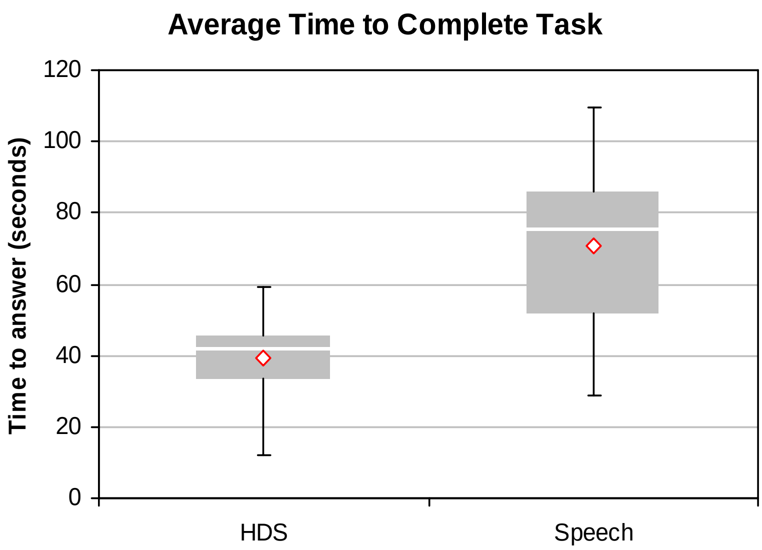


Figure 10. Efficiency in task completion for each condition, shown as average time to complete task, in seconds.

The results of the T-tests (assuming equal variances) showed a highly significant increase in the accuracy of the answers (T_{16}=3.036, p=0.007<0.01) when HDS was used (M=80.6%, SD=18.8) as opposed to the case in which speech had been used (M=59.7%, SD=17.8).

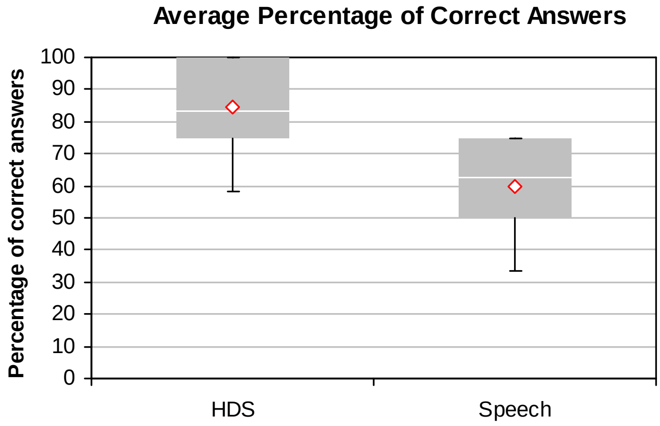


Figure 11. Effectiveness in task completion for each condition, shown as average percentage of correct answers.

Regarding the results obtained for the overall subjective workload (Figure 12), the same participants rated the modified NASA-TLX indexes in both conditions, and a within-groups analysis was performed. The T-tests (T_6=-4.833, p=0.003<0.01) showed clearly that the subjective overall workload was significantly higher when exploring with speech (M=6.3%, SD=3.6) than when using HDS (M=10.8%, SD=4.1).

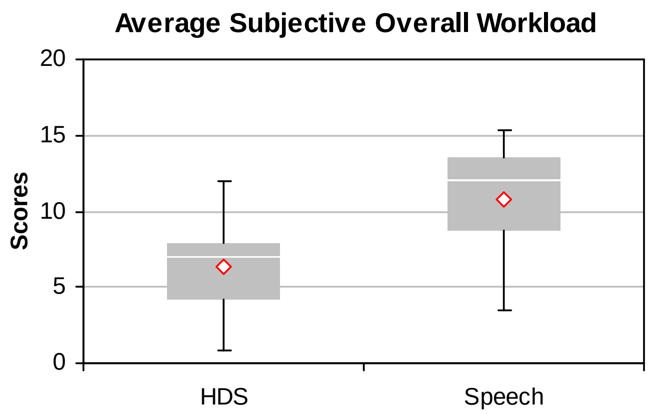


Figure 12. For each condition, average scores of subjective overall workload, calculated from the modified NASA-TLX ratings.

Looking at the average scores of each modified NASA-TLX category (Figure 13), and performing significance tests on these results showed that the significance of the overall metric originated in most of the categories. Thus, *mental demand* (T_6=3.897, p=0.008<0.01), *time pressure* (T_6=3.813, p=0.009<0.01) and *effort expended* (T_6=4.043, p=0.007<0.01) were significantly higher in the speech condition, at a level of significance of 99%. Additionally, *performance level achieved* was rated to be significantly lower (significance level of 95%) in the speech condition (T_6=3.195, p=0.019<0.05). *Annoyance* also observed a significant increase in the speech condition (T_6=2.554, p=0.043<0.05). However, in the case of *Overall Preference*, no significant difference was recorded.

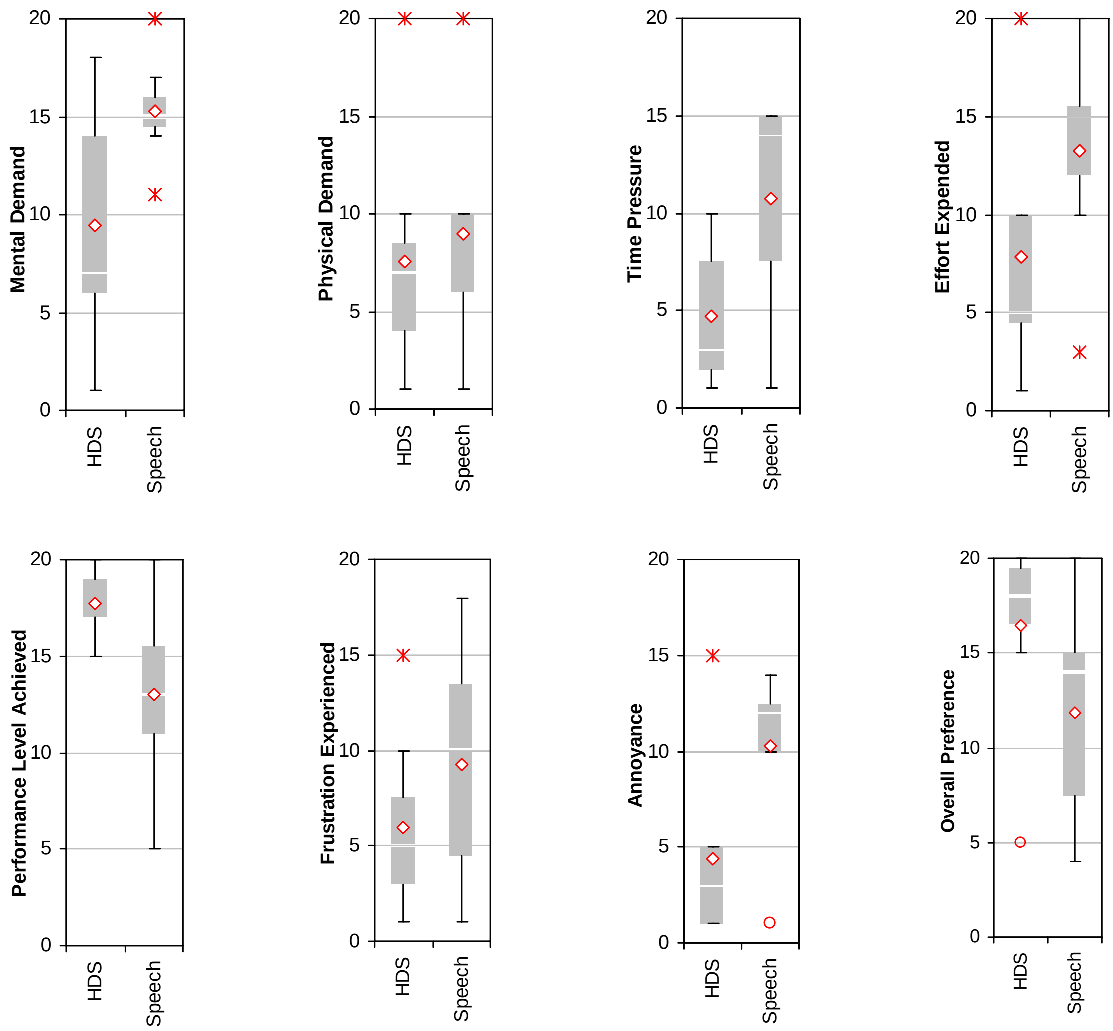


Figure 13. Mean values for the categories in the modified NASA-TLX workload data, for explorations using HDS and speech.

5.2.6 Discussion

The results presented in Section 5.2.5 show that exploring tabular data sets to obtain overview information is more successfully done using non-speech sound generated with HDS than when information is rendered in speech, as current screen-reading technologies do. Statistical analyses of the differences show that they are significant in all three metrics considered in this study. This significance is particularly high (99%) with the measures of effectiveness and subjective workload.

In the case of the measure of efficiency, with a between groups analysis of the results, the test showed that explorations were statistically significantly faster when data were explored using HDS. From the analysis of the recorded exploration traces (see Section 5.3 for a full discussion), it was observed that while HDS enabled the users to retrieve the complete set of cells in the table with each full scan, by using speech alone, a subset of the cells were retrieved before the answer was given. In other words, with speech users did not wait for the interface to speak the values in all the cells before they inferred an answer based on just a sample of all the values in the table. If participants were thorough enough to make sure that all the cells were visited before answering the questions, it can be inferred that the efficiency in the explorations with speech would have fallen much more steeply.

Some participants explained in informal discussions at the end of each session, that in the speech condition they tried to compromise and give estimates of the answers instead of making sure that the answers were correct. This behaviour is reflected in the results reported above for the modified NASA-TLX categories (Figure 13). Time pressure experienced in the speech condition was significantly higher, and compromising in this way was done to keep exploration time within reasonable bounds. Additionally, mental demand was also significantly higher in the speech condition due to the large amount of discrete units of information. Again, the users' strategy of retrieving only a sample of the values aimed at keeping mental demand within manageable levels. Participants realised that this approach had an impact on the accuracy of their results, as shown by the significantly lower scores for Performance Level Achieved in this condition. Indeed, the actual performance was such that the average number of errors in the answers was more than doubled in the speech condition, with below 60% of correct answers, in contrast with the average 80% correct answers obtained with HDS. These results are further evidence of the cognitive demands of exploring large data sets using speech, and the benefits offered by TableVis. Annoyance was also significantly higher with speech than with HDS. However, no significant difference in overall preference was derived from the questionnaires. This might be interpreted as due to that fact that all the participants were skilled regular users of speech-based screen readers.

5.3 Analysis of Exploratory Strategies & Procedures

As described in Section 5.2.1 the interactions of all the participants were recorded during the experiment, including all the traces of the movement of the pen on the tablet and all the actions performed, such as changes in exploratory mode and requests for detailed information in speech. Everything was time stamped, so that a timeline of the explorations could be reconstructed, including the speed of movement of the pen at every point. All these data, together with informal feedback collected at the end of each experimental session, were analysed with the aim of gaining further insight into the challenges of exploring data tables in search for specific kinds of overview information. This qualitative analysis was expected to provide information about the exploratory strategies and procedures devised by the users to perform the tasks with the tools available in TableVis. Such information would be valuable feedback for validation and refinement of the design, within the iterative, user-centred design procedure described in Chapter 3.

This qualitative analysis was performed with data from the first eight blind and visually impaired participants who completed the explorations in the non-speech sound condition. Four of these participants were totally blind (two of them congenitally blind). The other four were partially sighted, and they all required the use of screen readers to access information in computers.

As mentioned in the introduction to this chapter, an *exploratory strategy* is the action plan that a user devises to carry out the exploration of a data set in search for information. *Exploratory procedures* are the ways in which this strategy or action plan is implemented (the ways in which actual interaction with the system takes place). Exploring tabular data sets non-visually is a complex task that requires interpreting the mathematical relations intrinsic to a tabular structure in the context of the metadata of the particular table being explored. In TableVis, different auditory views of the same data object are obtained via interactive exploration, and combined to infer properties of that object. Users plan and implement exploratory strategies and procedures to carry out these active explorations. The complexity of tabular data explorations and the degree of flexibility that users have when they interact with the interface (users are not forced to follow a determined sequence of actions), permit them to exercise their inventiveness and, to a certain extent, devise their own ways of approaching the problem. The practices that resulted from this free approach taken by eight participants are discussed in the following sub-sections.

5.3.1 Stages of an exploration

In the bigger structure of the exploration process of a single table, three different stages could be distinguished. In the context of the tasks performed by the participants in this study, and based on the feedback obtained from them, these three stages can be rationalised as follows:

I. *Interpret the question in terms of tabular data structure, and devise a strategy for exploration*. The first step requires the user to understand the question and decide on the best strategy to explore the data in order to find an answer to that question. The question, if it is formulated in terms of the metadata (*e.g.* "which day of the week…?", "which time of the day…?"), has to be translated and reformulated into terms of tabular structure (*e.g.* "which column…?", "which row…?"). Additionally, the properties described in the question have to be translated into terms of the mapping of the data (*e.g.* "which day of the week is the website busiest?" would be translated as "which column produces the sound with the highest perceived overall pitch?"). It is important to note that the other type of question (*e.g.* "in which quadrant are the highest values?") is already phrased in terms of both tabular structure and information mapping. Once these conversions are done, a strategy is planned. Completing this stage requires a good understanding of the mathematical relationships present in tabular data structures.

II. *Interactive data exploration*. The user performs the actions planned in the previous stage, repeating and refining searches until all the information required in terms of tabular data structure and features in the data have been collected. This stage includes practically all the interaction with the interface. Interactively, all the information is retrieved in sound.

III. *Combine all the information retrieved through interaction with the system, and obtain the answer to the question*. Having obtained information in one or two navigation modes (depending on the task), the user still needs to obtain the answer for the question in terms of the meta-data. If the answer is a particular row or column, then the user still may have to interact with the system to request details in speech at the final position, obtaining the answer to the question. On the contrary, if the task was to identify a region in the table that required exploring the data set in both navigation modes, no further interaction with the interface is normally required, and the user has to perform the mental exercise of inferring the answer by superimposing both complementary views of the data (obtained in different directions of exploration) and identifying a particular area in the table (intersection of those complementary views) containing features that were being looked for.

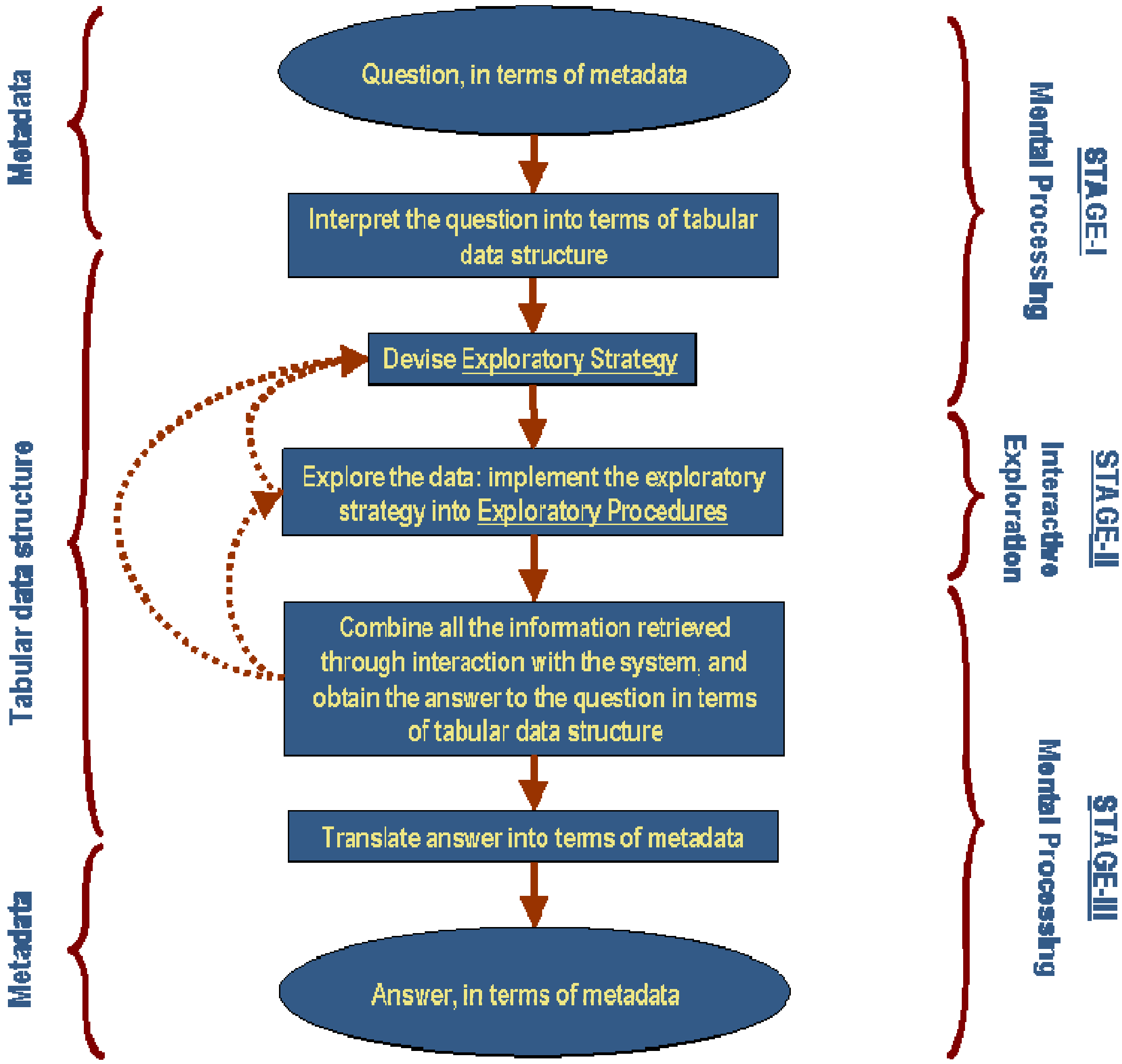


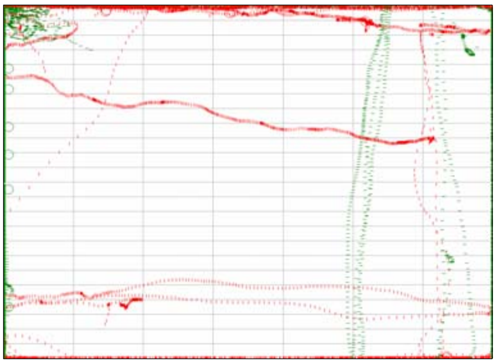
Figure 14. Steps in the exploration of a numerical data table, grouped in three stages (right) and by type of information processed in each step (left).

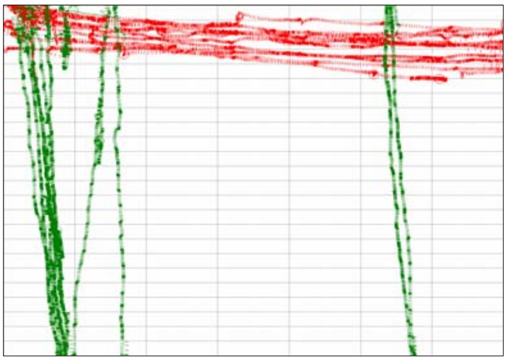
Figure 14 provides a graphic representation of the three stages of exploration. Stages I and III involve mental processing of information, but no active exploration. All the direct interaction with the system happens during stage II. The dotted arrows represent feedback loops between stages I, II and III that in a general case can take place. According to this description, and the division on the left side of the figure, this iterative process would deal with information in terms of tabular structure (rows, columns, left-right, up-down) and not with the higher level description of the data and its meaning.

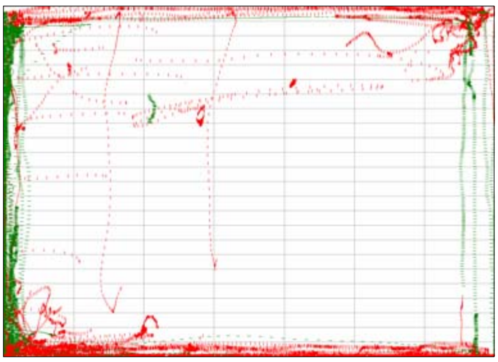
5.3.2 Exploratory strategies and procedures

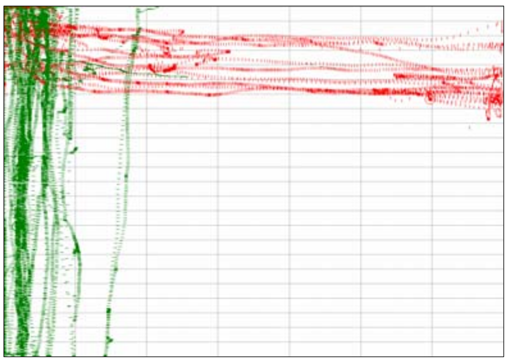
The general structure of a typical exploration with TableVis, described in the previous section, states that the interaction with the system only occurs in the second stage, in which the exploratory strategy (devised in the first stage) is implemented. The interface was designed to facilitate certain exploratory procedures (like scanning a table by rows and by columns, getting details where required), but the actual way of making use of these was left to the users' discretion. The interface gave, in fact, enough freedom to make use of the functionality in a variety of ways. Observing the way in which the functionality provided was used by blind and visually impaired users, and the success attained with different exploratory strategies and procedures was seen as invaluable information to validate the design and obtain qualitative feedback that could lead to the refinement of the design.

As mentioned above, the qualitative analysis was carried out based on two sources of information: feedback provided by users at the end of the experiment and analysis of the recorded data, describing all aspects of the interaction of the users with the interface. Figure 15 (*a* to *h*) shows the traces of the pen on the tablet, for each one of the first eight participants that completed the experiment in the non-speech sound condition (using HDS). Each Figure includes the traces of all 12 explorations of a participant, superimposed. Green traces (horizontal lines) were generated in rows mode. Red traces (vertical lines) were generated in columns mode. Marks printed close together in a trace line indicate that the pen was being moved more slowly on the tablet (the more spaced the lines, the faster the pen moved). Circles indicate requests for details in speech (label of current row or column, depending on navigation mode). Beyond simply printing them, as reproduced here, these traces were also analyzed one by one as animations that reproduced the users' interactions in each table exploration.


a


b


c


d

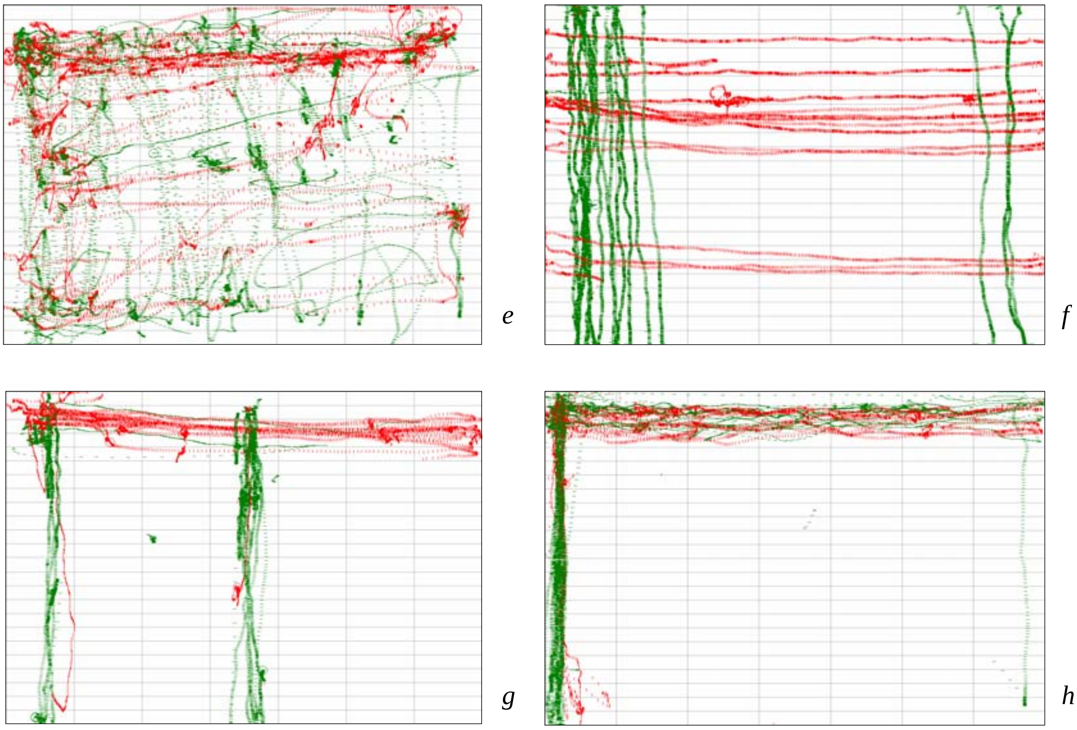


Figure 15. Traces of all 12 explorations by 8 participants (4 blind and 4 partially sighted) in the HDS mode. Green traces (small horizontal lines) were generated in rows mode, and red traces (small vertical lines) in columns mode. Circles indicate requests for details in speech (label of current row or column).

Participants *a*, *c*, *e* and *g* (those in the left column of Figure 15) were totally blind, blindness being congenital for participants *a* and *e*. The other four participants (column on the right of the figure) were partially sighted, but required the help of accessibility tools to use computers. Participants *b* and *d* used screen readers, while participants *f* and *h* used screen magnifiers and high-contrast colour schemes.

The observation of the traces, supplemented by informal feedback from participants, showed that the exploratory strategies devised to interact with the system and extract the information required (Stage-II in Figure 14) were those anticipated by the design of the interface. A typical exploratory strategy in the interaction with the system would go as:

- If a single row or column had to be found, select the appropriate navigation mode (rows or columns) and compare the data until the row or column with the required characteristics was found. Finish obtaining speech details on demand (to complete the last step in Stage-III, Figure 14);
- If a quadrant was to be found, scan the complete table in one of the modes (*e.g.* rows mode) and select one half of the table (*e.g.* the top or the bottom half). Then change to the other exploration mode (*e.g.* columns mode) and scan again the complete table, selecting again one half of it (*e.g.* the left or the right half). Mentally, combine both halves found, and take the intersection quadrant as the solution.

Within this main strategy, which was common to all the participants, differences could be found in the precise way in which this plan was implemented, *i.e.* in the exploratory procedures. On initial visual inspection of the 8 collections of traces in Figure 15, it can be observed that there are some common elements between several of them (like the abundance of lines that are close to the horizontal and the vertical directions), and also differences (like the use of the middle area of the tablet). In order to be able to conduct a systematic analysis of the analogies and differences, a set of characteristics in the exploratory procedures were identified and discussed separately (in these discussions, participants are referred to by the letter assigned to each one of them in Figure 14).

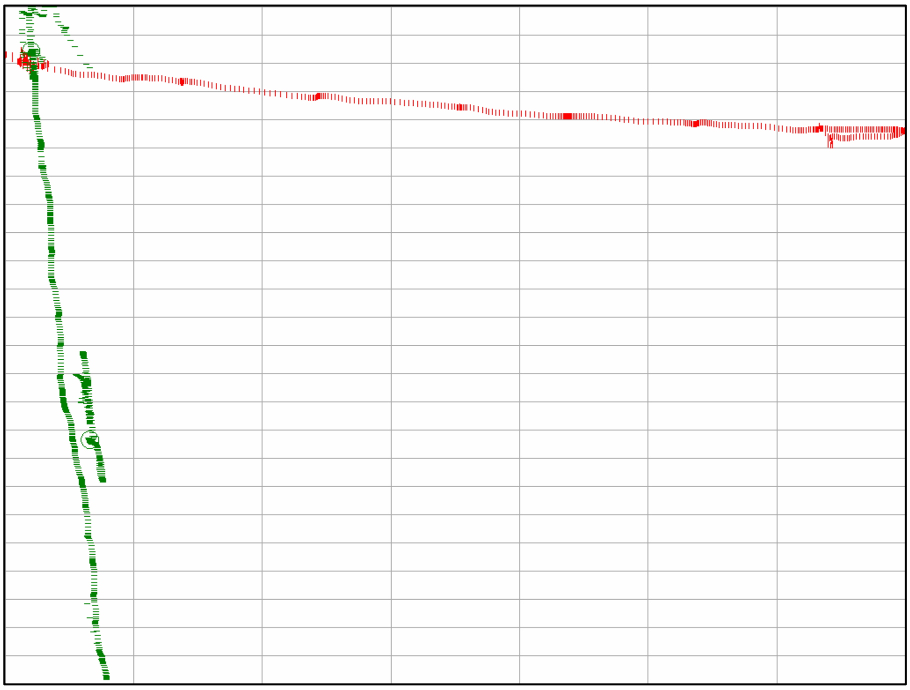


Figure 16. Example of the singular use of variable speed of scan that participant *b* implemented, in contrast to constant speeds implemented by every other participant. The dark lumps along the trace lines show slow-down positions that the participant used when hearing every new chord, in order to process it mentally.

Use of Speed of Scan

The speed at which the pen is moved on the tablet determines the number of data sonifications per unit of time that are generated. Thus, faster speeds present more information per unit of time and are more appropriate to obtain very general information (such as trends, patterns, approximate location of outliers), while more detailed characteristics about each row or column are less easily perceived. The opposite is true for slower speeds. Except for participant *b*, all the other participants developed a procedure that involved scanning the table by moving the pen at an approximately constant speed, to get an initial overview of the data. Although the speed was constant, it varied substantially between participants. Participant *a* made use, on average, of some of the fastest hand movement speeds, whereas participant *f* consistently used a slow, uniform speed. Participant *e* made use of the broadest range of uniform speeds between scans, making some of the traverses at fast speed and some others much more slowly. Participant *b* showed a different approach entirely to the use of speed. This participant scanned the tables slowly, stopping for a short time after hearing each one of the sonifications (after entering each new row or column) to analyze it mentally. In the example of a table exploration shown in Figure 16, these speed variations can be observed in the frequent variations in the density of the traces.

Number of Traverses

In most cases, fully traversing the table only once was enough to establish the approximate location of the data that contained interesting features for that particular exploration. After the first full traverse, participants normally went back to the approximate area in which they had identified the interesting data, confirming the finding (Figure 17 is an example of one such exploration in search of a particular row). If two or more interesting locations had been identified, they were quickly compared, with the user trying to ignore all the rest of the sonifications that still were being generated as the pen moved. In some cases, this appeared to be very difficult to accomplish, due to the distraction of all the unwanted sounds generated between the two that were being compared. This problem is discussed at length in Chapter 7.

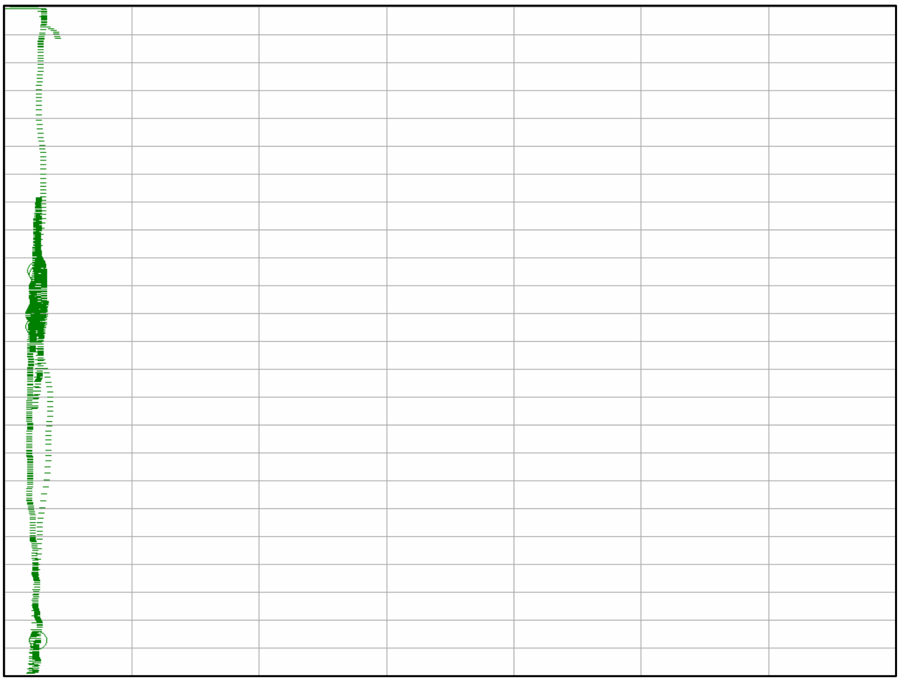


Figure 17. Typical example of a search for a row, belonging to participant *d*. Starting at the top, an initial full scan determines that the central area contains the most interesting data. Then the pen is taken back to that area and the position is fine-tuned by overshooting a couple of times, before obtaining the label of the row in speech.

Most participants performed approximately the same number of traverses in all the tables they explored. Participant *b* rarely required more than one full traverse, while participant *h* rarely performed less than three. All the other participants showed the behaviour just described, *i.e.* performing a full scan and then confirming the findings locally. As an exception in this classification, participant *e* exhibited great variability in the number of full traverses performed in each navigation mode, ranging from one to seven. It was also observed that, in general, scans performed at faster speeds of pen movement required more full traverses of the data set, but the whole process did not necessarily take a longer time to complete.

Direction of Scan

In both rows and columns navigation modes, most scans started near the top-left corner of the tables, for the majority of the participants. Clear examples of this approach are participants *d* and *h* in the summary of all their explorations shown in Figure 14. In those figures, it can be noticed that all the traces emerge from the top left corner (the animated visualisation of the traces confirmed this point), and most of the other participants show the same tendency too. An explanation for this choice may lie in the meaning or meta-data of the particular sets being explored. In these, the top-left corner is the natural origin for both axes (days of the week and time of the day, which both advance from the top-left corner right and down respectively). This might suggest that context was maintained during the exploration regarding the meaning of the data being explored. If this was true, the division between thinking about metadata and thinking about elements of tabular structure would not be as clear-cut as proposed in Figure 14. An alternative explanation for this tendency towards starting most of the scans at the top-left corner would be to follow the habit of reading rightwards and downwards, as with printed text. This arrangement of information is also present in Braille documents. Further investigation is required to confirm any of these hypotheses. A similar study in which metadata suggested an origin on a different corner could provide further information.

Figure 18. Example of table exploration in rows and columns modes, in which each scan does not start at the top left corner. Instead, the pen is never lifted and it follows the borders (notice the traces going down, right and up, tightly along the borders), showing efficiency of movement and good understanding of the metaphors in the interface. The traces belong to participant *a* (congenitally blind).

Coming back to the analysis of the directions of scan, several cases were observed in which, when the task required combining both rows and columns navigation modes, participants did not go back to the top left corner for every new scan, but very efficiently drew lines that followed the borders of the tablet. These participants (particularly participants *a* and *c*) started a new traverse where the previous one had ended, changing the navigation mode so that navigation could be resumed in a perpendicular direction, but without having to lift the pen and relocate it. The traces of the table exploration shown in Figure 18 are one example of this.

Use of Tangible, Physical Borders

By design, the rulers were intended to make the edges of the working area tangible to support obtaining context information (as discussed in Section 4.5.2). By observing the hands of the users, the way in which the tables were scanned, and the ability participants displayed to move or jump directly to revisit areas of interest, it can be inferred that the rulers were, indeed, used for this purpose by all the participants. The tangible dots in the middle points of each side of the working area were reported to be very useful by several participants, particularly to discriminate between left-right or top-bottom half in the "quadrant" questions in which high or low clusters appeared near the centre of the table.

Additionally, and while not explicitly intended from design, some participants also chose to use the rulers as guides to traverse the tables horizontally and vertically. Participant *a*, for instance, made heavy use of the top and left physical borders to slide the pen along the rulers, making virtually every exploration in this way (in Figure 14, *a*, it can be noticed that traces are densely and neatly packed along the edges, and relatively few traces are drawn elsewhere). Participant *c* also used the tangible borders as guides, mainly the left and bottom border, but also the top border.

No other participants used the physical borders as guides for the pen, except very occasionally. Two considerations can be made here. The first one is that it could be expected that participants with total blindness, particularly those who had a congenital impairment, might find more benefit in using the rulers as guides to traverse the surface of the tablet. In fact, both participants *a* and *c* were totally blind. Furthermore, participant *a*, who used the borders as guides more consistently, was congenitally blind. However, the traces of participants *e* and *g*, who also were totally blind (congenitally in the case of participant *e*), show that explorations could also be performed successfully without any support from external physical guides. The second consideration refers again to participant *a*, who made use of the highest speeds for scanning the data, and also used the tangible frame as guide for the pen most consistently, suggesting the possibility that using the frame in this way is a technique that may help obtain higher efficiency during exploration. If confirmed by further research, this would be an example of functionality not intended by design, but observed as worth supporting in the future by, for instance, making sure that the tangible border is solid enough to guide the pen leaning against it.

Shapes of trajectories

In the rows and columns modes used in this study, the pen did not need to be dragged horizontally or vertically, nor in straight lines to scan a table, as long as it travelled between the left and the right edges, or between the top and bottom. Still, straight lines were the most common shape drawn on the tablet during the data explorations. In addition to the two participants who used the physical borders of the tablet as guides (participants *a* and *c*, who drew mostly straight horizontal and vertical lines along the borders), the rest of the participants, with the exception of participant *e*, drew lines that were not only remarkably straight but also very close to being horizontal and vertical. Partially sighted participants *f* and *h* are particularly good examples, as is participant *g* among the blind participants who did not use the physical borders as guides. Although it could not have been expected that a user kept the pen within a single row or column confidently without some additional help (independent use of the interface was a design requirement for TableVis, as reported in Chapter 3), the traces presented here show good discrimination between horizontal and vertical, and the ability to stay approximately in a straight line, for participants with a variety of visual impairments. In contrast to all the rest, participant *e* used a particularly free style of exploration, in which very fast scans were drawn, vaguely following horizontal and vertical lines. Still, the performance was just as successful, as the sonifications were identical when crossing the tablet with a straight, inclined or winding line.

These observations serve as confirmation that the design is successful in freeing the users from the need to draw neat trajectories in fast scans, but it also shows that by default straight horizontal and vertical lines are chosen and executed very successfully.

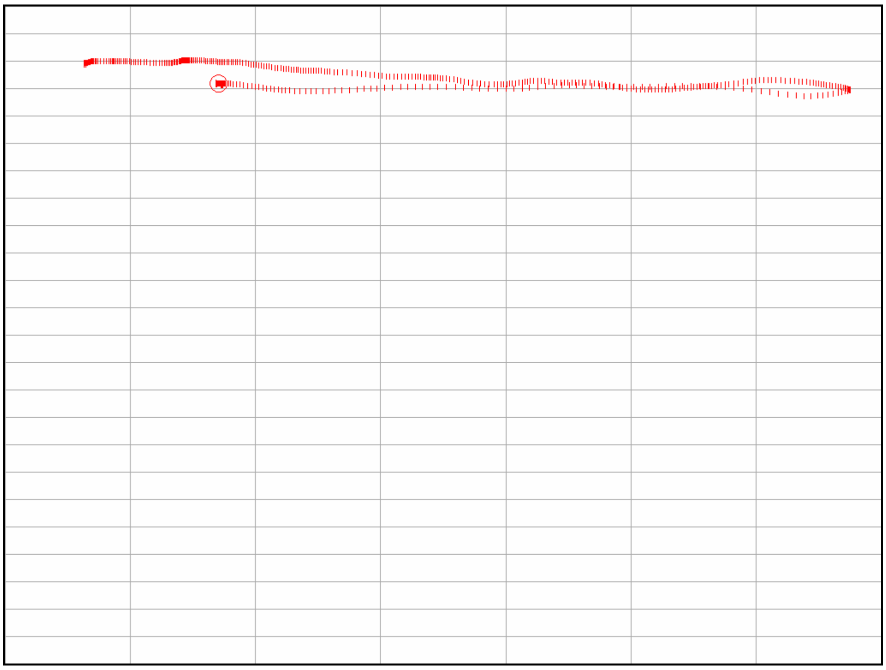


Figure 19. Typical example of the use of speech details-on-demand. Searching the columns in the table, this user started en the left and traversed the table horizontally. Then returned quickly to the position that had been identified as having the targeted properties, and details were requested to give the answer (Tuesday).

Use of Speech Details on Demand

With very few exceptions, this functionality of speech details on demand was used as it had been intended, that is, only at the end of the exploration and only if detailed information was required in the answer. A typical example is shown in Figure 19. Strictly speaking, details in speech should not be needed when choosing one of the quadrants, as proprioception should provide enough detail. This was, by and large, the case in the explorations analysed here, although occasionally participants decided to obtain the labels of rows and columns to discriminate quadrants (yet another hint that metadata might not always be completely absent in Stage-II of the exploration, as suggested in Figure 14).

A good use of this functionality was also made when a participant detected a position during the initial scan that could potentially be the answer. Then, the participant would request details in speech (the text of the label of that row or column) before continuing traversing the table. If at the end of the traverse no other part of the table had been found that could be a more likely candidate for the answer, the answer could already be given without having to go back to that position to get detailed information, because the user could still remember it. This technique was mainly used by participant *b*, as part of the generally differing strategy that this participant followed, making heavier use of working memory to remember pitches and labels. While this participant completed the task just as successfully in this study, there could be occasions when the additional use of working memory being made would lead to difficulties in completing the task, as will be discussed below and in Chapter 7.

5.3.3 Some common problems

The observation of these exploratory procedures also revealed some common situations that appeared to be confusing for some participants, and which are worth discussing here.

Difficulty to request details on demand

During the first prototype evaluations reported in Chapter 4, it was observed that users who were blind and visually impaired, and who did not normally use pens to write, did not always find it easy to hold and move the graphics tablet pen with precision. In this study, this problem became apparent for some participants when speech details were requested, by clicking on one of the buttons on the stylus. Often, once the pen was in position, the user had to search for the button on the pen, which sometimes resulted in the pen slipping away to a neighbouring row or column, and giving the wrong answer. From the analysis of the traces, it emerged that 12% of the explorations searching for a particular row or column resulted in giving the wrong answer because of the pen slipping from the correct position when clicking the button. A similar problem was encountered by McGookin & Brewster [98], who found that blind and visually impaired participants using the PHANTOM stylus to explore haptic graphs also had difficulty keeping the stylus in place while clicking the button to retrieve details on demand, resulting in what they called "off-by-one" answers. If the percentage of pen-slips was reduced, the average accuracy in solving the tasks as quantified in this study would increase. Once more, practice could potentially be the solution for this, and it is possible that if pen-based interfaces become common in accessibility tools, this percentage of errors may be reduce to levels similar to those obtained by sighted participants who are experienced in using a pen (a 2% error rate, as found in the experiment reported in Chapter 6).

Comparison of Non-Adjacent Rows or Columns

A different kind of issue was observed in situations in which a participant had identified two or more possible candidate rows or columns for the answer, which were not adjacent, and comparisons had to be performed.

Some participants used the fingers of one hand to mark the locations of points of interest and then tried to jump directly between fingers to compare only the sounds that had been pre-selected. Because those positions were not adjacent on the table, moving the pen between them would also produce sonifications of the rows or columns that were in between. This situation would benefit from some form of external memory aid that could be used to mark interesting positions that could be easily found again, and which also permitted filtering out the information that had not been selected to perform comparisons between the selected positions without interferences from other sonifications that got triggered in between. In Chapter 7, this problem is analysed in detail, and solutions are proposed and evaluated.

5.4 Conclusions

Exploring numerical data tables non-visually to obtain overview information that can answer high-level questions about the data is a complex task. TableVis provides support to complete tasks of this kind through interactive sonification of the data. Analyzing the way in which users with visual impairments approach this problem and make use of the functionality available in TableVis provides valuable insight into the practices that should be supported and the issues that must be tackled from an interaction design point of view. This approach also provides a better understanding of the problem of data exploration from the perspective of users with visual impairments.

The results of the quantitative evaluation reported in this chapter have contributed to answering RQ-2, "*What level of performance can be achieved with the techniques proposed in this thesis?*" by demonstrating that exploring with HDS is more effective, more efficient and produces lower subjective workload than exploring with speech. In addition, results from modified NASA-TLX questionnaires showed that mental demand, time pressure, effort expended and annoyance were significantly lower with HDS than in the speech condition, and participants perceived that their performance level achieved was higher with HDS.

The qualitative evaluation, carried out by observing and examining the exploratory strategies and procedures used by participants with visual impairments, provided further information in answer to RQ-1, "*How can overview information from tabular numerical data sets be obtained non-visually?*". Three distinct stages were observed in the process to complete the exploratory tasks (see also Figure 14):

I. Interpret the question in terms of tabular data structure and devise strategy for exploration;
II. Scan the data in one or to directions (depending on the strategy devised in I);
III. Combine all the information retrieved through interaction with the system and obtain the answer to the question.

The exploratory strategies were approximately the same for all the participants, although differences existed in the way in which those strategies were implemented, *i.e.* in the exploratory procedures. Further investigation is suggested to obtain evidence about which practices are the best and which cognitive aspects of the exploration should be better supported.

The conclusion is that this variety of uses should be preserved.

- *Use of speed of scan* (speed of the hand while scanning data); the use of a constant speed (chosen by most participants) offered better results than constantly varying speeds. The range of actual speeds varied considerably, without apparent correlation with performance.
- *Number of traverses*. Most participants traversed the table only once (in one or both directions, depending on the task) to obtain a good estimate of the answer, that was normally confirmed with further local searching. Thus, a single traverse was normally enough to obtain an initial overview.
- *Direction of scan*. Most searches started (and re-started) at the top left (This observation motivates further questions and hypotheses about possible reasons behind it, to be addressed in future work).
- *Use of tangible, physical borders*. From a design point of view, these borders were only provided to offer a tangible frame of reference for the non-dominant hand to provide context information during data browsing. In addition to this, some users utilised them consistently to draw fast, straight scans in both horizontal and vertical directions. Not all tangible borders can be used as support for the pen, but after this study, preserving such function can be recommended to designers.
- *Shape of lines*. While drawing the pen in straight lines was not required from a design point of view (in fact, it was a requirement that this had not to be done), most users scanned non-visually in remarkably straight and well-levelled horizontal and vertical lines, even without using support on the tangible borders.
- *Use of speech details-on-demand*. This functionality was generally used as intended, only when strictly required, what again was indication of the validity of HDS as a technique to obtain overview information.

It should be noted that, in the speech condition, participants were not using a commercial screen reading application but a custom-built interface. In this interface, all the benefits regarding context and navigation in TableVis were made available (focus+context through proprioception, direct access to any area of the data set, etc.) but information was retrieved as speech rather than non-speech sound. This was intended to provide a direct comparison between both forms of rendering information (speech and non-speech sound) however, the results cannot be interpreted as a direct comparison between TableVis and commercial screen reading applications.

Among the recommendations for improvement of the design of the interface, some indications about the need for some kind of external memory aids when dealing with particular types of data sets were observed. A detailed study regarding this issue is reported in Chapter 7.

The next chapter reports an experimental study in which the main aim was to measure the effect of the size of the tables explored using TableVis in user performance and subjective workload.

Chapter 6 Effect of Table Size in the Use of TableVis

6.1 Introduction

This chapter presents an experimental study in which the main aim was to measure the effect of the size of the tables explored using TableVis in user performance and subjective workload. An additional goal from this study was to collect performance data from sighted-blindfolded participants that could be directly compared against the performance data obtained from blind and visually impaired participants in the previous experiment, in order to obtain measurements of comparative performance between both populations.

In the study described in Chapter 5, TableVis was evaluated using data tables with a size of 24 rows and 7 columns. The results from that evaluation were analysed quantitatively and qualitatively, generating results for tables with that particular size. In practice, data tables explored with TableVis can have different sizes, and this fact motivated the question of how table size affects user performance and subjective experience in the main three metrics that were being measured in the quantitative part of the previous study:

- *Efficiency*. Time to complete exploratory task;
- *Effectiveness*. Percentage of correct answer to the exploratory tasks;
- *Subjective experience*. Modified NASA-TLX subjective workload.

It seemed likely that scores in these three metrics would be affected in such a way that performance would be lower and subjective workload higher if table size was bigger, since there would be more information to explore (requiring more exploration time), and the user's limited mental resources (namely the user's working memory) would have to deal with a larger set of data, posing higher demands on these resources and resulting in increased perceived mental workload and reduced accuracy in the user's judgements. Beyond this general hypothesis, it was difficult to predict a model that would describe the effect of table size on the three metrics, given the numerous factors that could affect such a model. On one hand, if table size affected these metrics because the amount of information to be explored changed with size, the effect should be a function of the number of cells in the table. However, since exploring for overview information with TableVis is done by retrieving rows and columns (rather than cells), the way in which cells are distributed in the table (the shape of the table) might have an effect too. Furthermore, it could be expected that for two tables with the same number of cells (same amount of information) but different shapes, the performance and subjective experience of the explorations could be different. To give an extreme example, it would be very different for a user to explore with TableVis a square 20x20 table and a stretched rectangular 2x200 table, both containing 400 cells. More specific questions would also need to be considered, as follows: does the rate at which these metrics change depend on the size and/or the shape of the table? Are there sizes beyond which exploring tables with TableVis becomes impractical? In what way do the characteristics of the data in the tables (e.g. degree of variability in the data) influence performance and usability? Are there any other aspects in relation to the amount of data in the tables that can also affect the three metrics?

Answering all these questions and inferring detailed models to describe these influences would require extensive experimental research, well beyond what could be undertaken within this thesis. The study reported in this chapter takes the first step in this line of research, by comparing performance and subjective workload measured in the exploration of tables with three different sizes, thereby covering a broad yet limited range of sizes.

The main group of participants in this study was a population of sighted-blindfolded participants. In addition, the experiment was also conducted with a small group of visually impaired participants, whose results were compared to those of the main group, in order to obtain a reference regarding the validity of the results. Since the group of visually impaired participants that could be recruited for this study was so small that it would not permit statistical testing of the results, the whole study was designed so that one of the conditions was identical to the non-speech sound condition from the previous study (reported in Chapter 5), which had been conducted with blind and visually impaired users. By proceeding in this way, the results from both groups could be compared statistically, informing about the comparability of results between populations, relevant not only to this study but to the whole of this thesis.

The aim of testing the interface with different table sizes was to find out in what way performance and subjective workload are affected by the size of the data set explored, and additionally to quantify that effect within a limited range of table sizes. An additional goal was to provide performance data from sighted, blindfolded participants that could be directly compared against the performance data from blind and visually impaired participants collected in the previous experiment. The experiments in this chapter thus attempt to answer RQ-2 as defined in Chapter 1: "*What level of performance can be achieved with the techniques proposed in this thesis?*". To achieve this, the experimental study reported in this chapter investigates how sensitive scores for effectiveness, efficiency and subjective workload are to variations in table size, and whether there are significant differences in performance between sighted and visually impaired users.

6.2 Experimental Study

6.2.1 Design of the Experiment

A one factor, repeated measures experiment was designed. The only independent variable in the experiment was the size of the table. Three conditions were compared: small tables, medium tables and large tables (the characteristics of each size are described in the next section). The order in which each condition was completed was counterbalanced between participants, creating 6 groups of participants (Table 3). The 18 sighted-blindfolded participants that took part in this study were distributed in the 6 groups, allocating 3 participants to each group.

	1st	2nd	3rd
Group-I	S	M	L
Group-II	S	L	M
Group-III	M	S	L
Group-IV	M	L	S
Group-V	L	S	M
Group-visually impaired	L	M	S

Table 3. Order in which the three conditions (S: small tables; M: medium-sized tables; L: large tables) were presented to each of the 6 groups of participants.

The experiment set-up was identical to the non-speech sound condition in the previous chapter. Before the experiment, training was provided in the three conditions. Participants practiced exploring data tables without blindfolds, until all the concepts were well understood, including the meaning of the questions and the kinds of answers that were expected, and how to use the system. Participants were informed that they would have up to 120 seconds to explore each table and give an answer to the question, and that they should give their answers as soon as they knew them. Each question was spoken in synthesised speech when each new table was presented, and the computer automatically began timing the exploration. The participant gave the answer by saying it loudly, at which time the experimenter stopped the clock and manually recorded the answer. The computer logged the time as well as the participant's exploration traces. Data sets within each condition were presented in random order. At the end of each condition, participants filled in a modified NASA-TLX questionnaire [65].

The data gathered from the medium-sized tables were used to compare the performance of sighted participants in this study against the performance of blind and visually impaired participants in the previous study. This was attained by carefully selecting the data sets and exploration task, as described in the next section, to be able to compare both collections of results directly.

6.2.2 Data Sets and Tasks

Three different table sizes were chosen for this experiment: a small size with 28 cells (4 rows x 7 columns), a medium size with 168 cells (24 rows x 7 columns), and a large size with 744 cells (24 rows x 31 columns). The medium size was the same (in size and shape) as in the evaluation of TableVis described in Chapter 5. From that study, scores in the three metrics considered were already available for a population of blind and partially-sighted users. By making the tasks and actual data for the medium size tables in the present study also identical to those used in the previous study, the results from both populations (sighted-blindfolded and visually impaired) could be compared.

The particular combinations of numbers of rows and columns used in the three sizes (*i.e.* the shapes of the tables) were selected in order for the tables to accommodate information with similar metadata for all three sizes. As in the previous study, the data produced were realistic information about the number of visits to an international imaginary website (visitors are equally likely at any time of the day), over a period of time. Different sizes combined different timescales in both axes, as described in Table 4. A visual representation of these relative sizes is provided in Figure 20, where the three table sizes appear superimposed. Additionally, Figure 21 represents on a linear scale the three relative sizes in terms of number of cells (top) and in terms of the total number of rows plus the total number of columns, R+C (bottom). This representation also indicates the multiplicity factors between the three sizes, by pairs (*e.g.* a large table has 26.57 times the number of cells and 5 times the number of R+C of a small size).

		Rows	**Columns**	**Cells**
	Small	**4** night --- (00:00-06:00) morning --- (06:00-12:00) afternoon --- (12:00-18:00) evening --- (18:00-00:00)	**7** Days in a week	**28**
Table Size	**Medium**	**24** Hours in a day	**7** Days in a week	**168**
	Large	**24** Hours in a day	**31** Days in a month	**744**

Table 4. Summary of the three table sizes, showing the number of rows, columns and cells, as well as the time scales used for each axis in each case.

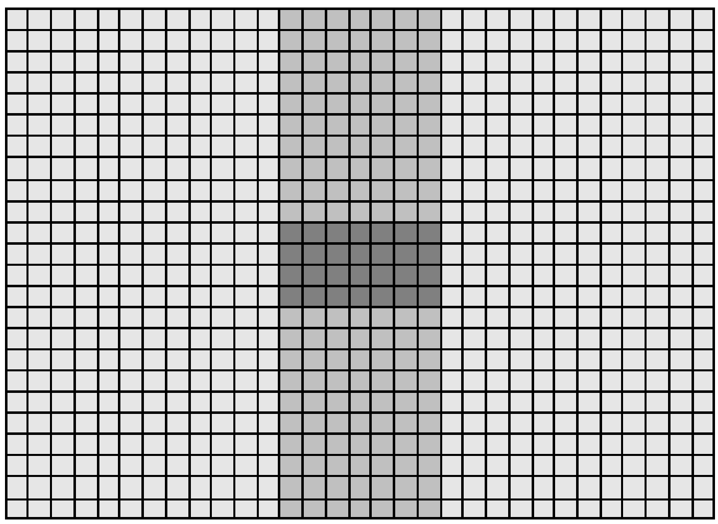


Figure 20. Visual representation of the relative sizes of the three tables selected for this study, created by super-imposing the small table (darkest shade) on the medium-sized table (medium shade), and both on the large table (lightest shade).

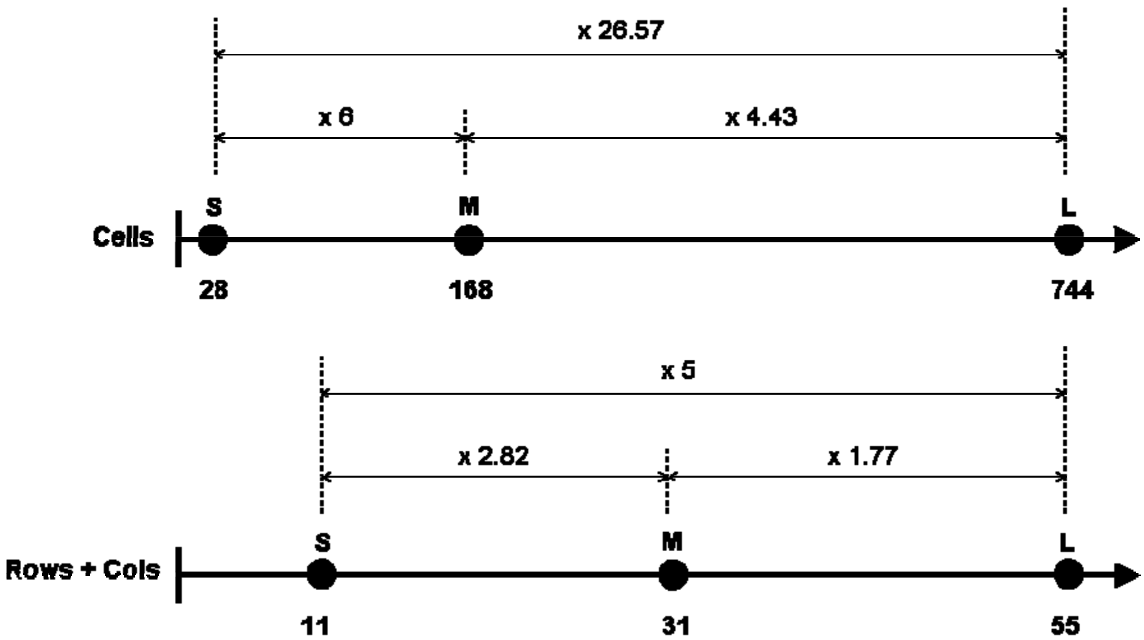


Figure 21. Representation of ratios between table sizes. *Top:* comparison of the number of cells in the three sizes, indicating pair wise ratios. *Bottom:* comparison of the total number of rows plus the total number of columns (R+C) in the three sizes, including pair wise ratios. (S: small; M: medium; L: large).

Once the tables with the three sizes defined for this study were rescaled and presented on the actual graphics tablet used for the experiment (a Wacom Graphire3 A5, G-630 [31]), the physical dimensions of a cell in each table size were as shown, in natural scale, in Figure 22. These cell size differences will become relevant in the discussion of the results (Section 0).

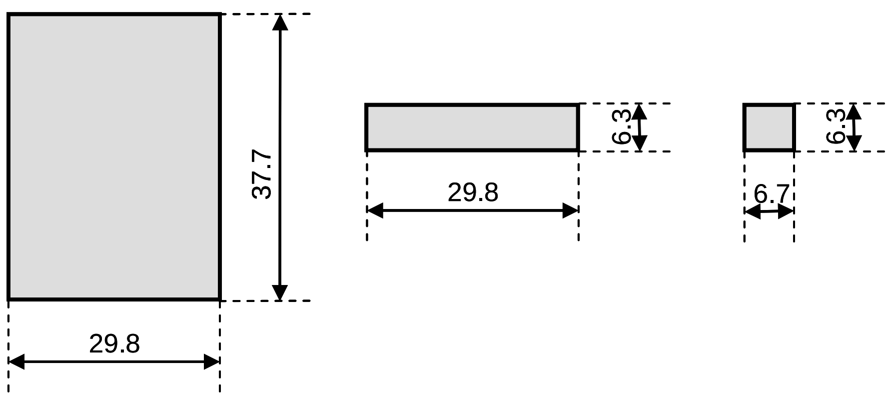


Figure 22. Actual dimensions of a cell for the three sizes of table when mapped on a Wacom Graphire3 A5 (G-630) graphics tablet, as used in this experiment (dimensions in millimetres).

For each table size, the tasks were the same as in the previous study (described in Section 5.2.1.), with the same types of questions: (i) select a row or column with the highest/lowest overall pitch, and (ii) select a quadrant in which there were the highest/lowest data values. For each table size, 12 data sets were prepared (six data sets for each type of question, 36 data sets in total). As just mentioned, exactly the same data sets from the previous study were re-used for the medium-sized tables. For the other two sizes, data were generated with similar characteristics regarding variability in the data and salience of target data subsets, but adapted to the bigger or smaller table size.

6.2.3 Participants

The experiment was run with 18 sighted-blindfolded participants, of whom 11 were male and 7 were female, aged between 16-35 years old. They were all University of Glasgow undergraduate and postgraduate students, from several faculties. Additionally, three partially-sighted participants were also recruited to complete this study. They were students attending Uddingston Grammar School, in Glasgow, aged 16-25. Their degrees of visual impairment required using extra-large fonts and special colour schemes, or screen magnifying software, to be able to access information in computers. The results from these three visually impaired users were used to provide a reference for comparison between populations in explorations across different table sizes.

6.2.4 Results

The results obtained in the experiment are summarised in this section, together with the outcome of the statistical analysis performed on them (one-factor, repeated-measures ANOVA tests).

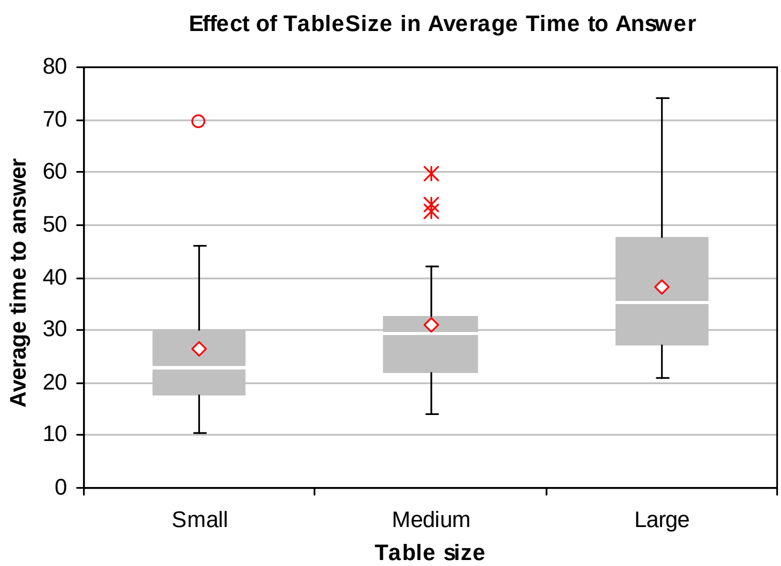


Figure 23. Average time for exploration of tables, in the three table size.

Figure 23 shows the average time to complete task for each table size. Statistical analysis showed that table size had a significant effect on this measure of the efficiency of the explorations (F_{17}=8.17, p=0.001<0.01). At a level of significance of 99%, a Tukey HSD *post hoc* test showed that the average time for exploration was significantly longer for large tables than for small tables, with confidence interval $CI_{99\%}$=(2.67, 21). The same *post hoc* test showed no statistically significant differences between small and medium-sized tables, with $CI_{95\%}$=(-4.42, 13.92), nor between medium and large tables, with $CI_{95\%}$=(-2.08, 16.25), both intervals at 95% of significance level.

Regarding the effectiveness of the explorations (measured as the accuracy of the answers), the results gathered in the experiment (average percentage of correct answers per condition) are presented in Figure 24. An ANOVA showed that table size also significantly influenced the accuracy of the results obtained in the explorations (F_{17}=6.88, p=0.003<0.01). Performing a Tukey HSD *post hoc* test, the decrease in the percentage of correct answers between medium and large tables was found to be significant at 99% significance, $CI_{99\%}$=(1.70, 20.52). However, no significant differences were found between small and medium-sized tables, $CI_{95\%}$=(-3.26, 11.59), nor between small and large tables, $CI_{95\%}$=(-0.47, 14.37), even at significance levels of 95%. Analysis of traces recorded from the explorations showed that 2% of the answers to the questions that required finding a specific row or column were wrongly given due to the pen slipping off from the correct position at the time of requesting details in speech.

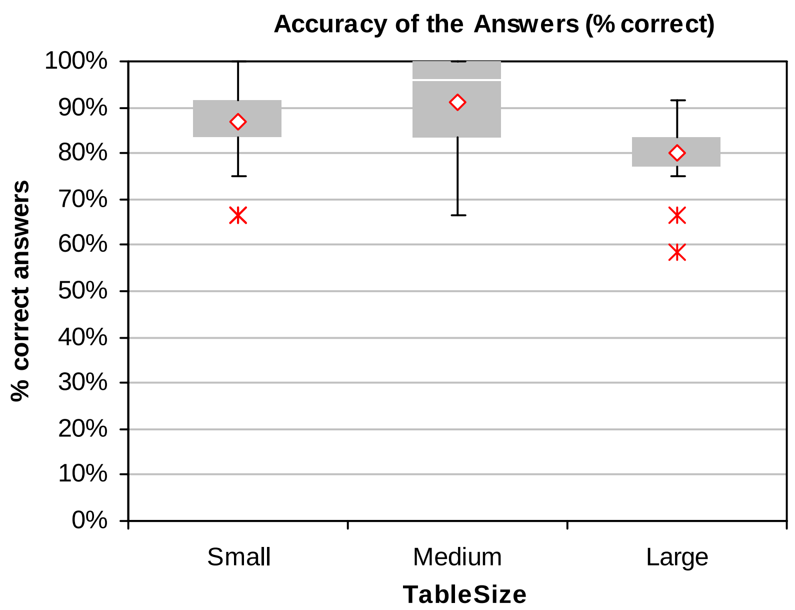


Figure 24. Task solving accuracy (percentage of correct answers), in the three table size conditions.

Finally, averages of the overall subjective workload scores obtained from the modified NASA-TLX questionnaires are shown in Figure 25. An analysis of variance on these results confirmed that the increase in subjective workload due to the increase in table size was significant (F_{17}=6.24, p=0.005<0.01). Tukey HSD *post hoc* tests revealed that the increase in overall mental workload was significant between small and large table sizes, $CI_{99\%}$=(0.23,3.79), with level of significance of 99%. Variations were not significant between small and medium-sized tables, $CI_{95\%}$=(-0.64,2.18), nor between medium-sized and large tables, $CI_{95\%}$=(-0.17,2.65), at levels of significance of 95%.

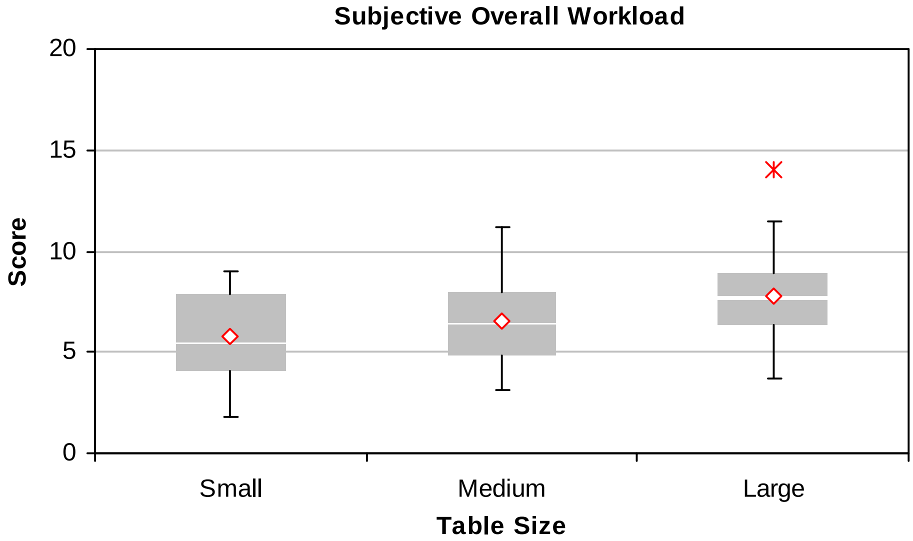


Figure 25. Overall subjective workload in the modified NASA-TLX scales, in the three table size conditions

To find the source of the significant difference in the overall subjective workload, the constituent categories from modified NASA-TLX were inspected. Figure 26 shows all the modified NASA-TLX categories that participants in this study rated. From the analysis of variance performed on these results, only *Mental Demand* showed a significant variation due to change in table size (F=16.25, p<<0.01). According to Tukey HSD *post hoc* tests, the significant increase in mental demand occurred between small and large tables, at confidence level of 99%, resulting in a confidence interval $CI_{99\%}$=(0.135, 7.754). The confidence intervals of the other two non-significant variations, at 95% level of significance, were as follows: $CI_{95\%}$=(-2.31, 5.31) between small and medium sizes, and $CI_{95\%}$=(-1.365, 6.254) between medium and large sizes.

The results of effectiveness and efficiency obtained with the medium-size tables in this experiment (with a population of 18 sighted-blindfolded participants) were compared against the results obtained in the non-speech sound condition in the previous study (with a population of 12 blind and visually impaired participants). Both sets of results were obtained under identical conditions. A summary of this comparison of results is shown in Figure 27 (results about efficiency are displayed on the left and effectiveness on the right).

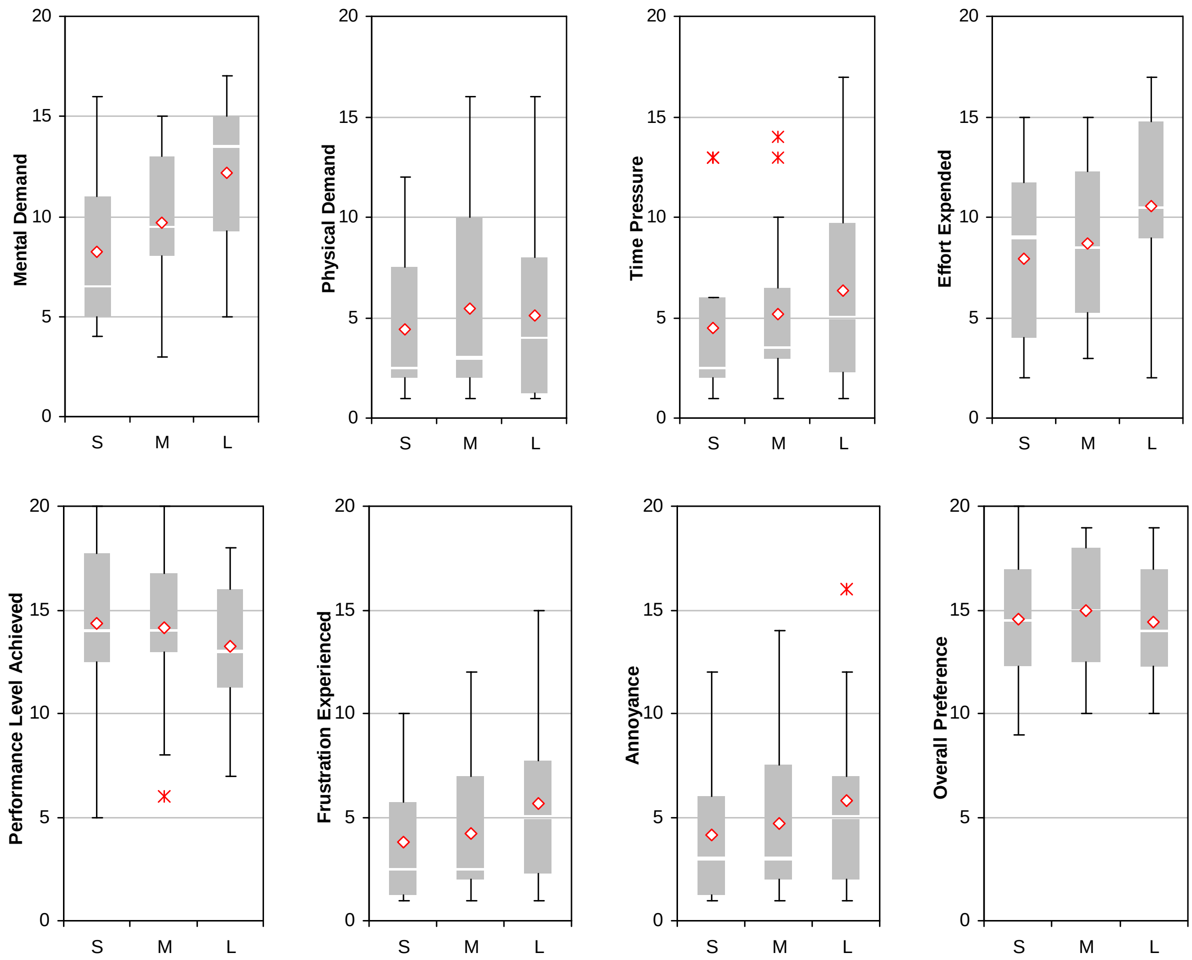


Figure 26. Mean values for modified NASA-TLX workload data in explorations of three different table sizes

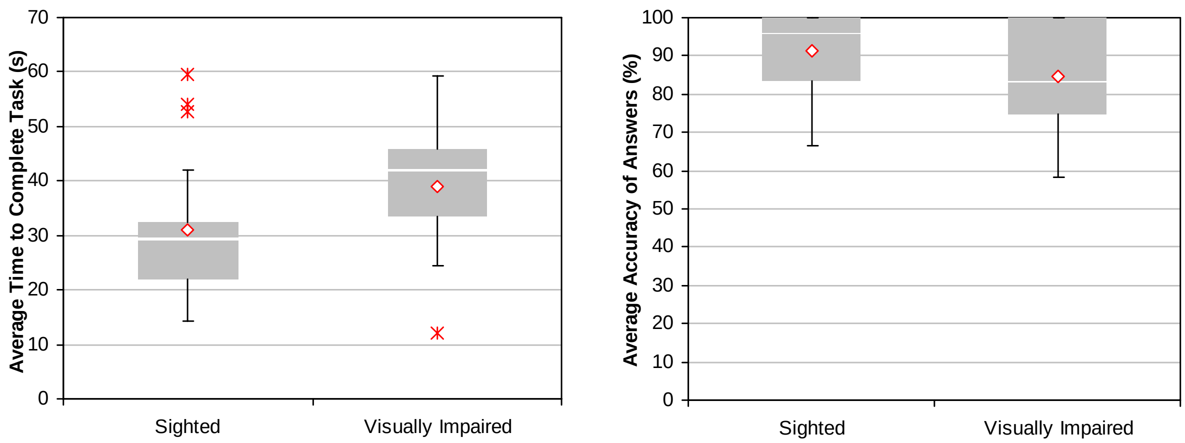


Figure 27. Comparison of performance by a population of sighted-blindfolded and of visually impaired participants, exploring for overview information with TableVis. *Left:* average time for exploration of tables, in seconds. *Right:* Task solving accuracy (percentage of correct answers).

Both sets of results were tested using two-sample T-tests (between-groups, assuming equal variances). At significance levels of 95%, the T-tests failed to find any significance in the differences of results between populations: (T_{28}=1.635, p=0.113>0.05) when comparing results of efficiency, and (T_{28}=1.353, p=0.187>0.05) for the comparison of results of effectiveness.

An additional 3 partially-sighted participants also completed the study with the three table sizes, and a summary of their results is shown in Figure 28. Only data about efficiency and effectiveness were recorded from these participants: time to complete task and percentage of accurate answers. Due to time restrictions at the time of conducting these experimental sessions, these participants did not fill in the modified NASA-TLX questionnaires, and no measures of subjective workload were obtained from them. In this case, because of the small size of this population (3 participants), the differences between conditions were not analysed statistically.

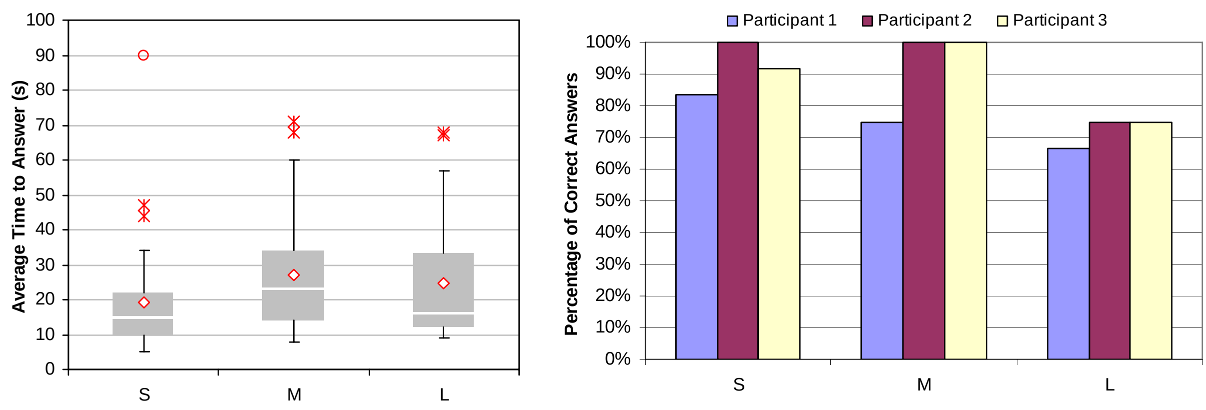


Figure 28. Results from the three partially-sighted users that took part in this experiment. *Left:* time for exploration of tables (seconds), in the three table sizes. *Right:* for each participant, task solving accuracy (percentage of correct answers), also for the three table sizes.

6.3 Discussion

6.3.1 Effect of Table Size

The results presented above confirm that the size of the table explored does influence the performance achieved (both in terms of effectiveness and efficiency) and the workload experienced, when explorations for overview information are performed using TableVis. In order to quantify the sensitivity of these metrics to changes in table size, the results presented above are discussed in this section. First, the results obtained for efficiency and subjective workload, which are similar, are discussed together. Then, the results for effectiveness, presenting a different behaviour, are discussed separately.

The variation in efficiency (time to complete task, Figure 23) was only significant for the largest of differences in table size considered in the experiment, *i.e.* between small and large sizes. Similar conclusions were reached with the analysis of the results about subjective workload obtained from the experiment (Figure 25). Again, a statistically significant variation in subjective overall workload (which was primarily due to differences in mental demand) was only found for the largest of differences in table size. In both cases, the arithmetic means of the scores obtained in the three conditions followed a trend of proportionality with table sizes (as table size increased, effectiveness was lower and subjective workload higher). Also in both cases, only when the table size was increased sufficiently (from the smallest to the largest size) did the scores obtained for these metrics vary significantly.

Considering the ratios between table sizes shown in Figure 21, these results can be put in perspective regarding their robustness to variations in table size. Increasing the number of cells from 28 to 168 (factor of 6 larger), and from 168 to 744 (factor of 4.43 larger) did not produce any significant effect in efficiency and subjective workload. Only when increasing the number of cells from 28 to 744 (factor of 26.57 larger) did the scores for these metrics change significantly. Still in that case, the increase was an average factor of 1.45 times larger for the time required to complete task (from 26.35 seconds to 38.19 seconds in average), and by a factor of 1.35 in the case of the subjective workload (from 5.76 to 7.77 in the 20-points modified NASA-TLX scale). These figures correspond to average slopes[4] of 5.46% and 5.08 % respectively. Bearing in mind that a slope of 0% would mean that variations in table size have absolutely no effect on the metrics, and 100% would indicate that the table size and the metrics varied by the same factor, the figures of these slopes show a very small effect of table size on the metrics. The same comparison of multiplicity factors can be done for the total number of rows plus the total number of columns (R+C) in the tables, also shown in Figure 21. These are the units of information accessed with HDS, and R+C depends not only on the amount of information in the table (number of cells), but also on the shape of the table (how the cells are arranged in the tabular structure). Increasing R+C from 11 to 31 (multiplicity factor of 2.82) and from 31 to 55 (multiplicity factor of 1.77) produced no significant variations of time to complete the task and overall mental workload. Significant increases in these metrics were obtained by increasing R+C from 11 to 55 (a factor of 5). The average slopes for time to explore and subjective workload were only 9.2% and 7% respectively. The small slopes calculated in every case offer a quantification of the sensitivity of efficiency in the exploration and subjective workload to changes in table size. This quantification shows that, even in the cases in which the effect is significant, the efficiency of the explorations and the resulting subjective workload are not very sensitive to changes in size of the tables explored.

The measures obtained for effectiveness (percentage of correct answers, Figure 24) present a different behaviour. With the maximum shift in size (between small and large tables), the variation found in the accuracy of the results was not significant. In addition, the maximum average accuracy was not obtained with the smallest tables, but with the medium-sized ones, although the difference between both was, again, not significant. The only significant difference that the tests identified was the reduction in accuracy when shifting from medium to large tables. The maximum average slopes measured in this relationship were of 25.7% considering the numbers of cells in the tables, and 64.3% in the case of R+C. These slopes suggest a higher sensitivity of the

---

[4] Slope (m) is used here in its calculus meaning, as the ratio between increments in two axes, x and y. The slope as a percentage is defined as: $m\ (\%) = \Delta y / \Delta x \cdot 100$. In the context of this study, slope is calculated as the ratio between multiplicity factors of each of the three metrics and the variations in table size.

effectiveness in the explorations to the size of the tables explored, in comparison to the other two metrics. These results are, however, more difficult to interpret, because they do not show a relationship of proportionality between table size and effectiveness in the explorations. For example, the fact that when considering the whole range of table sizes (from small to large) the variations in effectiveness are not significant, suggest just the opposite conclusion, *i.e.* that effectiveness is less sensitive a metric to variations in table size than the other two metrics.

A possible explanation for this lack of proportionality between effectiveness and table size might be found in some of the qualitative feedback information offered by various participants. Some participants expressed that the medium-sized tables were the most comfortable ones to explore. Large tables were perceived to contain a lot of information per unit of tablet surface, which required fine control over the movement of the pen. Conversely, small tables were perceived to be very empty. This perhaps reflects the finding of Challis & Edwards [32], that blind and visually impaired people find “dead space” in tactile diagrams confusing. Cells in small tables were several centimetres in size, both in width and in height (see again Figure 22 reproducing cells from the three tables with their actual dimensions). During the exploration of a small-size table, after a row or column was entered by crossing one of its borders and the sonified data were heard, the pen had to travel for (what was perceived to be) a long distance on the tablet, before the boundary with the next row or column was crossed and a new sonification was heard. During that interval time, there was no feedback from the interface in response to the movement of the hand. Some participants tried to break this disconcerting silence by moving the pen in very small jumps, repeatedly tapping on the tablet as they advanced across it, and retrieving the same sonification again and again, until a new row or column was entered and the sonification changed. A conclusion that could be derived from these observations is that very large graphics tablets would be more difficult to use in the exploration of small tables, and that providing some kind of feedback during the displacement of the pen within a large cell could be beneficial.

6.3.2 Other observations

The observation of the explorations and the qualitative feedback collected revealed another aspect of exploring the smallest tables which was a source of confusion to some participants. When the number of rows or columns in a table was small, traversing the table side to side repeatedly, without lifting the pen, could produce the sensation that one row or column was missing, and create confusion if the interactive sonification metaphor was not well understood. In fact, this happened for all the table sizes, but was more noticeable in the small ones. A user who traversed the 4 rows of a small table from top to bottom could easily keep count of all the rows that were visited in the first traverse: the first row when the pen was put down on it and the next 3 as the boundaries between rows were crossed. If the pen was not lifted at the bottom of the table when a second traverse was started upwards, only 3 sonifications were retrieved by the time the top edge was reached again, giving some users the impression that one row was missing. Trying to clarify this apparent contradiction, if the user started a third traverse not lifting the pen when changing the direction of movement, once more only 3 chords would be retrieved. This last sequence of 3 chords would not be the sequence heard in the second traverse but in inverted order. Instead, it would be a different sequence, adding to further confusion. Figure 29 illustrates this *border* *effect* graphically, where the dots along the trajectory line represent positions in which a row is sonified. In this example, the three traverses just described would deliver the following sequences of chords:

- *First traverse:* I, II, III, IV;
- *Second traverse:* III, II, I;
- *Third traverse:* II, III, IV;

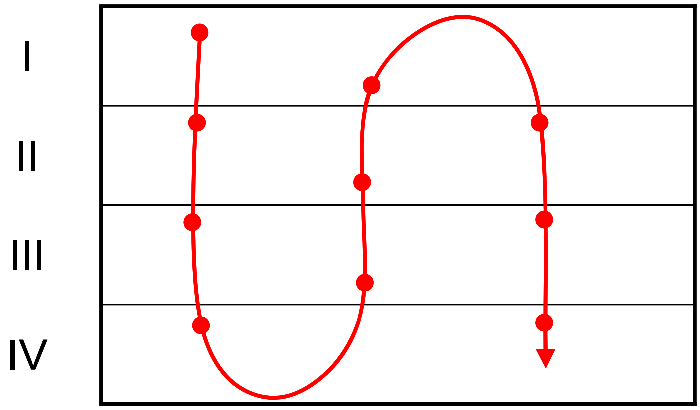


Figure 29. Trajectory of the pen traversing three times a table with four rows, without lifting the pen from the surface of the tablet. The dots signify the positions at which the user hears the sonification of a row.

The fact that smaller numbers of chords could be easily counted and remembered seemed to make this a more common source of confusion when exploring small tables, a problem that disappeared when the number of rows or columns was larger. The way around this problem came from realising that each new full traverse should start by tapping on the initial position. While this aspect of the interaction was pointed out during training, it is possible that some users did not fully understand it until several explorations were completed. This could be the source of some errors in their answers when exploring small tables, particularly because with only 4 rows the answer was often given based on the context ("the third row, *i.e.* afternoon"), rather than by getting the label of the row in speech.

Another observation from the study was that some participants experienced difficulty in completing exploratory tasks when they were comparing several non-adjacent rows or columns. In these cases, some participants reported that they missed being able to landmark positions in order to perform comparisons systematically. This is the same problem that had been reported by some participants in the study described in Chapter 5. This problem is analysed in detail, with the proposal of solutions, in Chapter 7.

6.3.3 Comparison of populations

Another aim of the study described in this chapter was to compare the two different populations (blind and visually impaired participants on one side and sighted-blindfolded participants on the other) in the use of TableVis to obtain overview information, to find out if there are any significant differences in their performance. This comparison was done using two sets of results: the results obtained from a population of 12 blind and visually impaired participants in the study presented in Chapter 5, and the results of 3 visually impaired participants who completed this study in the three table-size conditions. The first set of results can be statistically compared against the results obtained in this study, for the middle-sized table condition. The second set of results cannot be statistically tested due to the small size of the population (3 participants), but it offers a reference of comparison with the main results from this study. The combination of both sets of results provides information about the possible differences between both populations of users, as discussed here.

The comparison of the results by blind and visually impaired users from the previous study, and the results by sighted users in the middle-size condition in this study are summarised in Figure 27. As presented above, the statistical tests that compared both sets of results failed to find any significant differences between the performances of both populations. Regarding the small group of 3 visually impaired participants that completed the study about the effect of table size, their results are shown in Figure 28. On visual comparison of the results of efficiency with those obtained from the main group of participants (comparing Figure 28, left, for visually impaired users with Figure 23 for sighted-blindfolded users), it is noticed that the visually impaired participants completed the tasks in the three conditions faster, on average, than the sighted-blindfolded participants. As in the case of the sighted participants, the three visually impaired participants performed most efficiently with small tables. The efficiency in the exploration of medium-sized and large tables was similar (if slightly lower with large tables). If results about effectiveness are also compared visually (Figure 28, right, for visually impaired users and Figure 24 for sighted-blindfolded users), it is observed that the average levels of accuracy are similar for both populations. The only slight difference can be noticed in the lower level of accuracy obtained with large tables. This lower level of accuracy is perhaps the price paid for the considerably faster explorations users performed with large tables (sighted users required around 50% more time, in average, to explore large tables, and they obtained only around 12% higher accuracy). In summary, while these results have to be taken with caution due to the small size of the population of visually impaired users, they show three instances of at least as good performance by visually impaired users as that obtained by sighted-blindfolded users, reinforcing the statistically analysed results for medium-sized tables, which forecast no significant differences between populations.

To conclude the discussion of the results, it is interesting to compare the number of cases in which the explorations produced the wrong answer due to the pen slipping off at the time of requesting details in speech. While the incidence of this issue was 12% in the case of blind and visually impaired users, as reported in 5.3.3, in the current experiment with sighted-blindfolded participants the incidence went down to 2% of the explorations. This difference supports the hypothesis that users who are visually impaired and who do not regularly hold pens to draw or write find it more difficult to maintain full control of the pen on the graphics tablet, demonstrated at the time of pressing the built-in button on the pen while trying to keep the pen stationary on the tablet. As discussed in the previous chapter, this difference also suggests that training by regular use of pens could reduce the incidence of this problem to the level observed for sighted participants. A different issue that seemed to be common to both populations of users (although no exhaustive data about its incidence was recorded) was the occasional difficulty that participants reportedly found when comparing non-adjacent rows or columns. This problem was already described for visually impaired users in 5.3.3. In the case of the current study, the difficulty was reported when exploring tables with medium and large sizes. An in-depth discussion of this issue is presented in Chapter 7.

6.4 Limitations of this study

As discussed in the introduction to this chapter, this study takes the first steps in understanding how user performance and subjective experience change when exploring tabular data sets with different sizes, shapes and features in the data. While an ultimate goal of this line of research is to build a model that takes all those variables into account, this initial study considered only a limited set of scenarios. In doing so, the results are limited to those cases only. A list of the main limitations is presented here:

- The range of table sizes considered, while comprehensive, is limited to between 4x7=28 cells at the small end of the scale, and 24x31=744 cells at the large end. Without further research, the results from this study cannot be extrapolated to tables with sizes outside this range;
- This range of table sizes is divided into only two intervals, with one intermediate size containing 24x7=168 cells. This small number of intervals within the range considered limits the accuracy with which threshold table-size differences (those leading to significant effects) can be pin-pointed;
- The shape of the tables (the ratio of number of rows to number of columns) was not a separate parameter in this experiment, as each table size had certain specific shape defined by the common metadata used in all the sizes. It is expected that different shapes within the same size would lead to different scores of performance and subjective experience, but this aspect of the research is not covered in this study,
- Yet another aspect that was not parameterised in this study was the size of the working area of the graphics tablet. The results presented here are valid for the particular combination of tablet size and the sizes of tables used in the experiment. Together, they determine the physical sizes of the cells which, and as discussed above, can also have an effect in the scores obtained;
- While a statistically-analysed comparison of performance and subjective workload between populations (visually impaired and sighted) was performed for the middle-sized tables (which did not find any significant differences between populations), only three visually impaired participants completed the study in the three conditions, what can only be taken as anecdotal evidence of a similar behaviour;
- The effect of table size in the scores measured might change with different data sets (data with other characteristics) and with different exploration tasks.

6.5 Conclusions

This chapter described a study which contributed to answering the main research questions in this thesis, defined in Chapter 1:

RQ-1: *How can overview information from tabular numerical data sets be obtained non-visually?*

RQ-2: *What level of performance can be achieved with the techniques proposed in this thesis?*

This was achieved by conducting an experiment in which, using TableVis, sighted-blindfolded participants explored data tables with three different sizes: small (4x7=28 cells), medium (24x7=168 cells) and large (24x31=744 cells). In addition, data gathered in this experiment, combined with results from the previous chapter, were used to perform a comparison of the performance achieved by two different populations of users, visually impaired and sighted-blindfolded, in explorations conducted with TableVis. Keeping in mind the limitations of this study (listed in Section 6.4 above), the knowledge acquired in it is summarised below.

6.5.1 Effect of Table Size

The experiment investigated the effect of table size on the following three metrics: efficiency of the explorations (time to complete task), effectiveness of the explorations (accuracy of the information retrieved) and subjective workload. The scores in these metrics obtained with explorations of three different table sizes: *small* (4 rows x 7 columns = 28 cells), *medium* (24 rows x 7 columns = 168 cells) and *large* (24 rows x 31 columns = 744 cells). The main findings are summarised as follows.

- *Efficiency of the explorations*. The time to complete an exploratory task increases by a factor that is an average 5.46% of the factor by which the number of cells in the table is increased, and an average 9.2% of the factor by which the total number of rows plus the number of columns of the table is increased. This quantification assumes a linear relationship between table size and efficiency, an assumption that is supported by the results from the experiment. A significant increase in the time to complete an exploration was measured between small tables (M=26.35 sec., SD=14.66) and large tables (M=38.19 sec., SD=14.48);
- *Effectiveness of the explorations*. The results from the study do not suggest a linear relationship between table size and effectiveness, as this metric decreased faster in the interval from medium to large tables. In fact, in that interval a significant decrease in effectiveness was measured (as percentage of correct answers), *i.e.* between medium-sized tables (M=91.2%, SD=10.49) and large tables (M=80.09%, SD=7.64). Assuming the relationship is linear within that range, when table size increases, the percentage of correct answers decreases by a factor that is an average of 25.7% of the factor by which the number of cells in the table is increased, and an average of 64.3% of the factor by which R+C of the table is increased. The results also show that, considering the whole range of table-sizes studied, the effect of table size on effectiveness is smaller, as no significant variation in the percentage of correct answers was found between the smallest and the largest size;
- *Subjective workload*. When the size of the table is increased, the overall subjective workload, as measured with the modified NASA-TLX (and particularly due to the increase in mental demand), increases by a factor that is an average 5.08% of the factor by which the number of cells in the table is increased, and an average 7% of the factor by which R+C of the table is increased. A significant increase in the overall subjective workload (scale 1 to 20) was measured between small tables (M=5.76, SD=2.26) and large tables (M=7.77, SD=2.65);

From these findings, it can be concluded that, at least within the range of table sizes and the physical size of the graphics tablet selected for this study (see the discussion about the limitations of this study in Section 6.4 above), there is no need for zooming or computer-aided information filtering in order to complete explorations for overview information. Thus, TableVis provides some of the synoptic quality that is lost when vision is missing, in the sense that the time and effort required to take a glance does not depend strongly on the amount of information explored for overview information.

The results from this experiment also provide further evidence that HDS leads to good estimations of the relative arithmetic means of all the numerical values in each row or column, regardless of the size of the table, at least within the range of sizes considered in this study (between 4 and 31 numerical values that are presented in a single chord). Through using this method to estimate arithmetic means, the average percentage of correct answers in the different conditions ranged between 80% and 91%, similar levels as in the previous studies. Chapter 8 describes a detailed study conducted to understand better how these estimations with HDS are performed and when they can be expected to be most reliable.

6.5.2 Comparison of Populations

A statistically tested comparison of the results from both populations of users (blind and visually impaired users on one hand and sighted-blindfolded users on the other) exploring medium-sized tables found no differences in any of the three metrics used in this study (time to complete task, percentage of correct answers and overall subjective workload). This suggests that the presence of a visual impairment has no significant effect in the performance achieved and in the subjective experience, when exploring data for overview information using TableVis. Additionally, for the complete range of table sizes considered in this study, anecdotal evidence was obtained (from 3 visually impaired users completing the full study) that the size of the table explored has similar effect on the performance achieved by both populations of users.

6.5.3 Other observations

During the experiment, several factors were observed from the participants' interactions and qualitative feedback, which had an effect on the accuracy of the answers obtained from the explorations using TableVis. These observations are summarised here:

- *Border effect.* Novice users of the interface could get confused by the so-called *border effect*, which is particularly noticeable in tables with a small number of rows or columns (*e.g.* 4 rows, like in the case of the small tables in this study) Performing repeated side-to-side full traverses in such a way that the pen is not lifted from the surface of the tablet when the direction of the movement is inverted, can appear as rendering a different set of chords and even a different number of chords in each traverse, as illustrated in Figure 29 . This problem disappeared once the metaphors of the interactive sonification in TableVis are well understood.
- *Dead space effect*. Another negative effect arose with small tables was that of *dead space*. When a table mapped to a graphics tablet resulted in the physical dimensions of the cells on the tablet being in

the region of centimetres rather than millimetres (this effect was reported in the small size condition, in which a cell was 3 cm wide and 3.8 cm high), crossing the area of a single cell with the pen felt like a long distance, during which the movement of the hand produced no apparent response from the interface, resulting in a negative sensation of emptiness. While this effect was not considered to be a serious problem by the participants who reported it, it should be taken into account for future improved design of the interaction.

- *Pen slip off.* A lower incidence of the problem of the pen slipping-off its position when getting details in speech was observed with sighted-blindfolded participants. While in the previous study with visually impaired participants (see Section 5.3.3) 12% of the explorations suffered from this problem, in the current study with sighted-blindfolded participants this problem only affected 2% of the explorations. This difference could be due to sighted participants being more used to handling and using a pen. If this was the case, with enough practice the incidence of this problem among visually impaired users could descend to comparably low levels.
- *Comparisons between remote locations.* As observed in Section 5.3.3, some participants experienced difficulty in completing exploratory tasks when they were comparing several non-adjacent rows or columns. In those cases, some participants reported that they missed being able to landmark positions in order to perform comparisons systematically. This problem is analysed and addressed in Chapter 7.

Chapter 7 EMA-Tactons

7.1 Introduction

During experimental evaluations of TableVis presented in previous chapters, it was observed that the working memory of some participants occasionally reached saturation, and that they found it difficult to complete the task with the functionality included in TableVis. These difficulties were encountered both by participants who were blind and visually impaired and by sighted participants. The difficulties were observed in a small number of cases; indeed, the results from the previous experiments showed that overall the explorations were performed both effectively and efficiently, and with acceptably low levels of subjective workload. However, in spite of the relatively low incidence of these problems, it is interesting to take note of the features that are common to all these cases (both in the tasks and data configurations), in order to understand why they occur and to investigate ways of preventing them.

It is hypothesised that alleviating the problem of working memory saturation encountered during data explorations using TableVis will improve efficiency and effectiveness, and reduce workload. This is addressed in this chapter by investigating how non-visual external memory aids can be implemented with TableVis and quantifying their impact on user performance and subjective workload. This research contributes towards answering the main research questions defined in this thesis: "*How can overview information from tabular numerical data sets be obtained non-visually?*" (RQ-1) and "*What level of performance can be achieved with the techniques proposed in this thesis?*" (RQ-2) This chapter begins with a description of the problem and justification of the need for some form of external memory aid (EMA). EMA-Tactons are designed and justified as a possible solution. Quantitative and qualitative evaluations of EMA Tactons are then reported, including iterations in the design and evaluations, following which conclusions are drawn from this process.

7.2 Description of the problem

During the evaluations of TableVis described in Chapter 5 and Chapter 6, a problem was sometimes observed during the experiments and reported by various participants. In order to find the single highest or lowest row or column, these participants began by scanning the data to narrow down the search to a few most likely candidate rows or columns, then compared the shortlist to find the answer. However, participants sometimes found it difficult to perform comparisons when the candidates were not contiguously located on the tablet. When sliding the pen between the physical locations the user wished to compare, all the rows or columns in between were also sonified, making it very difficult to concentrate only on the sounds of the two target positions.

Under such circumstances, the search algorithm required a lot of information to be retained in the working memory, before the final answer could be inferred:

- Spatial locations of all the short-listed rows or columns (physical positions on the tablet);

- Perceived Overall Pitches (POPs) corresponding to all those locations;
- Ranking position of the data in each particular location, according to the search criterion.

Looking more closely at the third bullet-point, if for example the absolute highest POP was being sought, the user had to discriminate whether the POP obtained at each position was the highest so far, simultaneously keeping track of the comparisons that had already been performed, the locations that had been rejected and the ones that had not yet been tested. Since all this intermediate information was only needed temporarily for the computation of the search algorithm and until the actual result was found, it would be kept in working memory and forgotten once the solution was reached.

It is well known that the capacity to store information in working memory is very limited, both in the number of pieces of information that can be stored (between 5 and 9) [104], and in the lifetime of the information before it fades (15-20 seconds, with little or no rehearsal by repetition) [120]. As described above, the number of pieces of information that were needed only temporarily to reach the answer was often higher than the storage capacity of working memory.

Some of the users reported to have used (and were observed using) simple techniques to overcome these difficulties. One of those techniques was to lift the pen in the air between the two remote positions that were being compared, in order to avoid triggering sounds from data located between them. The other technique was to use the fingers in the non-dominant hand to landmark potentially interesting positions on the tablet, thus to be able to find those positions again more easily and perform comparisons between them without having to look for them each time. Both techniques, while very resourceful, had problems and limitations, which will be discussed more fully in Section 7.3.

7.2.1 Factors influencing memory overload in TableVis explorations

The characteristics of the scenarios in which this problem occurred were determined from the experiences reported by the users and the observation of their practises and strategies in those cases. From the observations in the previous studies, it was concluded that these scenarios had the following elements in common:

i. Data tables containing a minimum of 24 rows and/or columns;
ii. Unsorted data sets with frequent, sudden variations, and where simple trends or patterns of variation in the data are not obvious;
iii. Tasks that require obtaining information with an intermediate level of detail. Participants had no problem in providing general trend descriptions. However analysing the data in finer detail (*e.g.* identifying a particular column in a set of 24) could be challenging enough for working memory limitations to become apparent.

Thus, in order to study this problem, data exploration scenarios with these three characteristics were reproduced in the experimental studies reported in this chapter. As mentioned in the introduction (Section 7.1), the approach taken to addressing this problem was the design and implementation of external memory aids that could alleviate working memory overload using TableVis.

7.3 Requirements for the Design of External Memory Aids

In the context of this study, external memory aids (EMAs) are (based on a definition by Norman) "*external aids that enhance cognitive abilities*" ([114], p.43). In the case of this thesis they are not intended to be aids to facilitate the retrieval of memories, such as shopping lists, as Intons-Peterson describes [72]. Instead, the problem tackled in this chapter concerns the support of working memory, understood as *temporary*, volatile storage for manipulation of information. For example, Card *et al.* [30] report that the use of paper and pencil to perform arithmetic calculations was 5 times faster than mental arithmetic alone. They conclude that the difficulty in mental arithmetic lies not in the mental multiplication itself, but in keeping the partial results in mind while carrying out the calculation. This difficulty of retaining partial results in short term memory is similar to the problem encountered by TableVis users addressed in this chapter. The aim of the research reported in this chapter was to find an external memory aid for TableVis which could be as useful as pen and paper is for sighted users performing similar tasks. In other words, to provide means of making annotations on the data that can help during the process of working out the answer to a data exploration task.

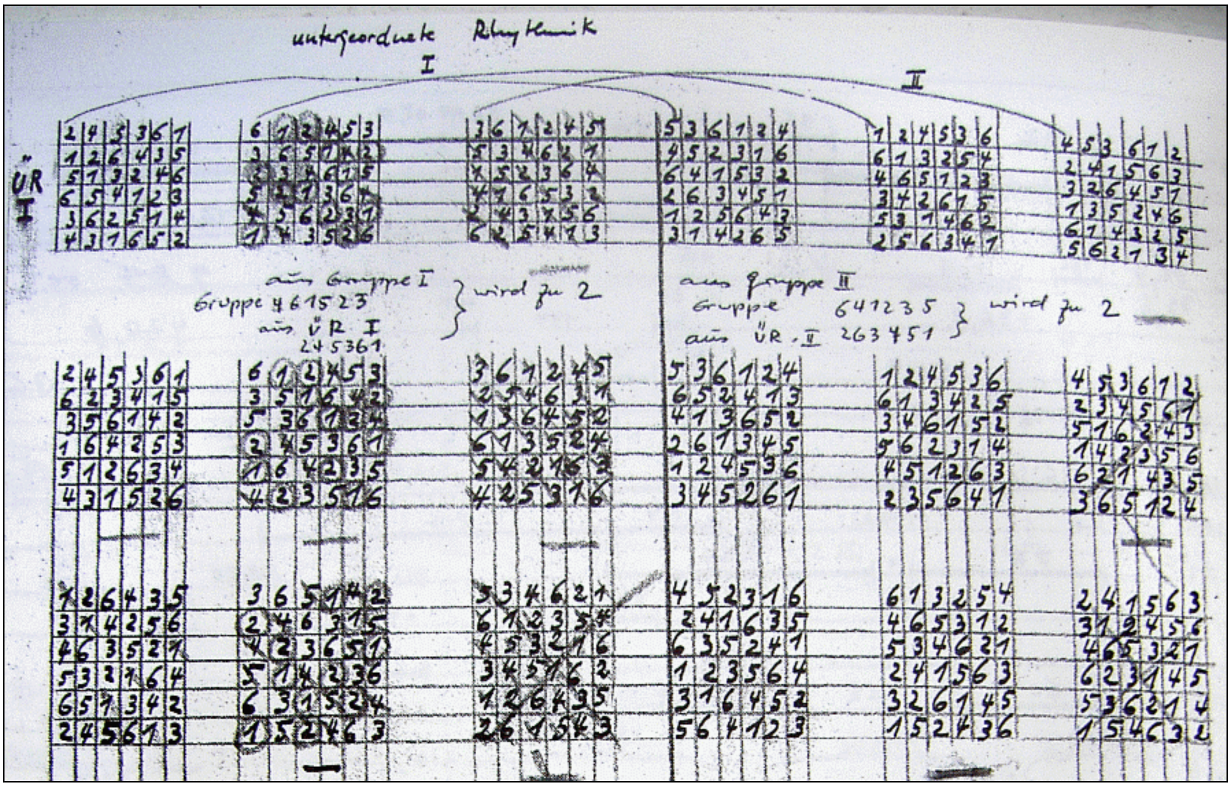
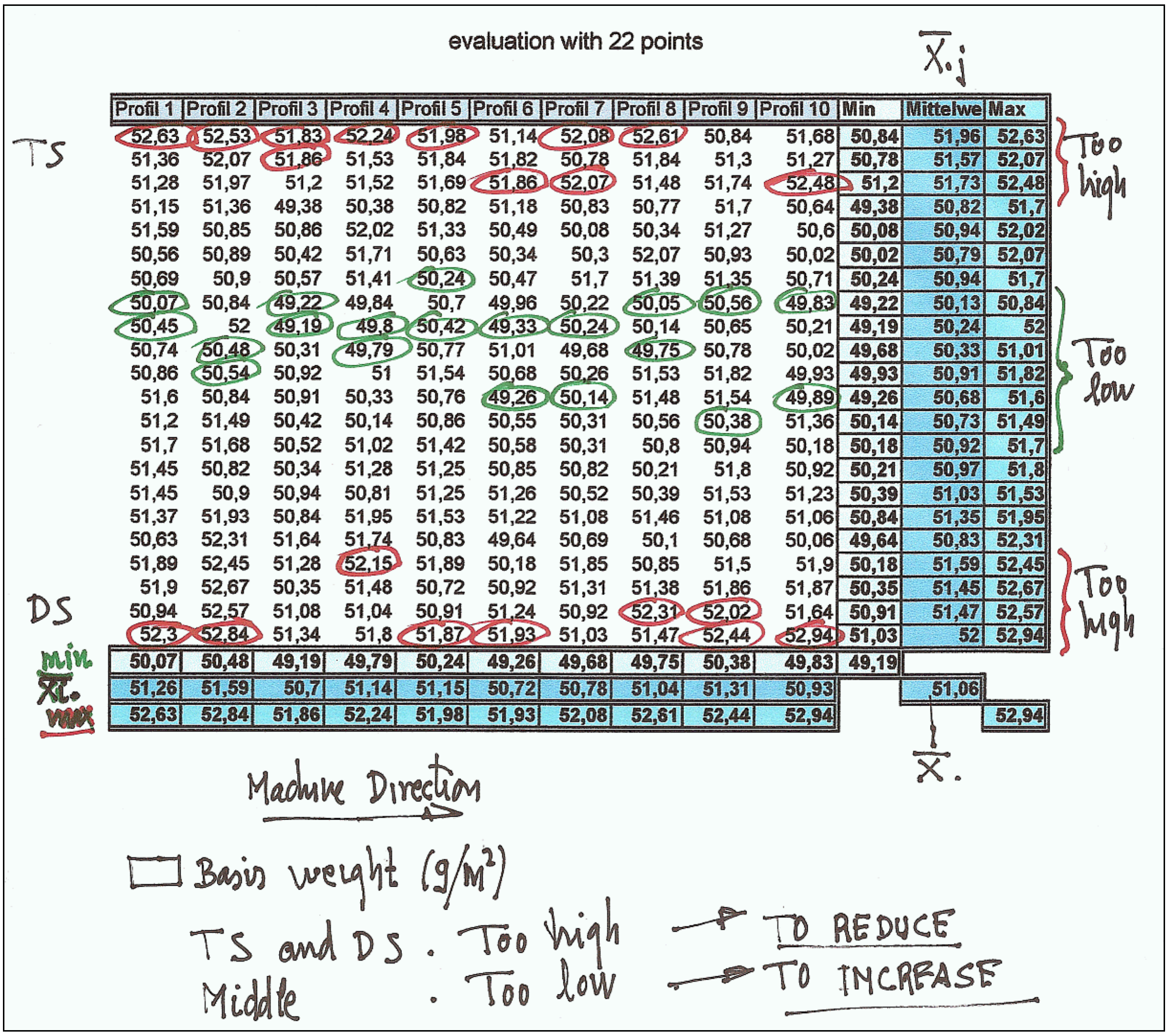


Figure 30. Excerpt from the collection of hand-annotated numerical tables used by the German Composer Karlheinz Stockhausen (1928-2007) to construct his electronic composition "Etude" (1952) [146].

When analysing numerical data visually, it is common to make annotations on a printout of the raw data set or on a graphic representation of it. Figure 30 and Figure 31 show examples of hand-annotated tabular numerical data sets, with very different purposes. Figure 30 is an excerpt written by the *avant garde* German composer Karlheinz Stockhausen to work out the assembling of the samples that would constitute his electronic composition "Etude" (1952) [146].

The tabulated numbers in Figure 31 represent the basis weight (g/m^2) of paper produced in a paper machine, from samples taken at 10 different times of the production and on 22 different positions of the width of the machine. All values oscillate around 50g/m^2, which is the target basis weight of the paper produced in that machine. The person analysing the data roughly highlighted the highest values in red and the lowest ones in green, and concluded that the paper tended to be thinner in the centre of the machine and thicker on both sides.



Profil 1	Profil 2	Profil 3	Profil 4	Profil 5	Profil 6	Profil 7	Profil 8	Profil 9	Profil 10	Min	Mittelwe	Max
52,63	52,53	51,83	52,24	51,98	51,14	52,08	52,61	50,84	51,68	50,84	51,96	52,63
51,36	52,07	51,86	51,53	51,84	51,82	50,78	51,84	51,3	51,27	50,78	51,57	52,07
51,28	51,97	51,2	51,52	51,69	51,86	52,07	51,48	51,74	52,48	51,2	51,73	52,48
51,15	51,36	49,38	50,38	50,82	51,18	50,83	50,77	51,7	50,64	49,38	50,82	51,7
51,59	50,85	50,86	52,02	51,33	50,49	50,08	50,34	51,27	50,6	50,08	50,94	52,02
50,56	50,89	50,42	51,71	50,63	50,34	50,3	52,07	50,93	50,02	50,02	50,79	52,07
50,69	50,9	50,57	51,41	50,24	50,47	51,7	51,39	51,35	50,71	50,24	50,94	51,7
50,07	50,84	49,22	49,84	50,7	49,96	50,22	50,05	50,56	49,83	49,22	50,13	50,84
50,45	52	49,19	49,8	50,42	49,33	50,24	50,14	50,65	50,21	49,19	50,24	52
50,74	50,48	50,31	49,79	50,77	51,01	49,68	49,75	50,78	50,02	49,68	50,33	51,01
50,86	50,54	50,92	51	51,54	50,68	50,26	51,53	51,82	49,93	49,93	50,91	51,82
51,6	50,84	50,91	50,33	50,76	49,26	50,14	51,48	51,54	49,89	49,26	50,68	51,6
51,2	51,49	50,42	50,14	50,86	50,55	50,31	50,56	50,38	51,36	50,14	50,73	51,49
51,7	51,68	50,52	51,02	51,42	50,58	50,31	50,8	50,94	50,18	50,18	50,92	51,7
51,45	50,82	50,34	51,28	51,25	50,85	50,82	50,21	51,8	50,92	50,21	50,97	51,8
51,45	50,9	50,94	50,81	51,25	51,26	50,52	50,39	51,53	51,23	50,39	51,03	51,53
51,37	51,93	50,84	51,95	51,53	51,22	51,08	51,46	51,08	51,06	50,84	51,35	51,95
50,63	52,31	51,64	51,74	50,83	49,64	50,69	50,1	50,68	50,06	49,64	50,83	52,31
51,89	52,45	51,28	52,15	51,89	50,18	51,85	50,85	51,5	51,9	50,18	51,59	52,45
51,9	52,67	50,35	51,48	50,72	50,92	51,31	51,38	51,86	51,87	50,35	51,45	52,67
50,94	52,57	51,08	51,04	50,91	51,24	50,92	52,31	52,02	51,64	50,91	51,47	52,57
52,3	52,84	51,34	51,8	51,87	51,93	51,03	51,47	52,44	52,94	51,03	52	52,94
50,07	50,48	49,19	49,79	50,24	49,26	49,68	49,75	50,38	49,83	49,19		
51,26	51,59	50,7	51,14	51,15	50,72	50,78	51,04	51,31	50,93		51,06	
52,63	52,84	51,86	52,24	51,98	51,93	52,08	52,61	52,44	52,94			52,94

Figure 31. Hand-annotated basis-weight analysis report of paper production samples from a paper machine (2007). The tabulated data represents 10 sample strips from the reel produced in a paper machine, analysed in 22 different points covering the whole machine width [73].

Despite the very different subject domains from which these two examples have been selected, it is interesting to notice that there are strong similarities in the annotations, such as circled data points, augmentation of the information in text form in blank areas, and brackets covering ranges of values.

These examples show that hand-drawn annotations are used at the discretion of the user's needs, where the same annotation techniques (underlining, colour coding, numbering, circling, thickness of the lines or pressure with which the lines are drawn, bracketing and so on) can mean different things. As discussed by Marshall [97], each of these forms of classifying or rating information are only clearly meaningful to the users who have made them (unless they were intended for collaborative work, in which case some coding practices might have been agreed upon). Being able to encode information *ad hoc* is, therefore, also a requirement for the design of non-visual annotation tools.

Thus, from examples of annotated data sets like those shown in the figures above and the specific needs observed during previous evaluations with TableVis, the requirements listed in Table 5 were defined for the design of EMAs for TableVis. During the process of design and evaluation that follows, this list of requirements will be revisited to reflect on the extent to which these goals are attained.

i.	Annotations should be easily added and removed in real time, in an integrated way with the data exploration process;
ii.	Annotation should not be limited in number;
iii.	Each annotation must remain in the same position, unless explicitly moved by the user (persistence of the annotations);
iv.	Annotations must be easy to find, and distinguishable from the original information;
v.	Adding an annotation must leave the information in the data set unaltered;
vi.	An annotation must not obstruct the access to the original information in that position;
vii.	Annotating should combine with other techniques and tools for data exploration and analysis available in the interface, to support the process of information seeking;
viii.	A wide variety of types of annotations should be possible, to be employed and coded at the discretion of the user.

Table 5. Requirements for annotation tools to be used in TableVis.

When participants experienced difficulties in complex explorations that were mentally demanding, as in the cases analysed here, some of them developed two improvised simple techniques to compensate for the saturation of working memory, which were observed by the experimenter and later discussed with them. One technique was to lift the pen in the air between two remote positions on the tablet that were being compared, in order to avoid triggering sounds from data located between them, which made it difficult to mentally isolate and compare the sound events corresponding to the positions of interest alone. This technique was difficult to execute successfully, without triggering sonifications from rows or columns around the positions of interest.

The other technique observed was to use the fingers in the non-dominant hand to landmark interesting positions on the tablet, thus enabling them to find those positions again more easily and perform comparisons between them. In tasks that required identifying a single row or column with certain characteristics, the strategy attempted by participants using this technique was to mark all the candidate positions in the data set and then perform comparisons within that shortlist until the final answer was obtained. Marking positions on the tablet with a finger was quite inaccurate since fingers accidentally moved from their intended positions very easily and the reference often got lost. Additionally, it was difficult to mark multiple positions spread across the tablet, as the hand postures required to do so were uncomfortable. Moving the fingers to annotate a new position often resulted in the loss of some or all of the previous references, and the user had to begin the exploration again.

In view of the relative success that these simple techniques provided, some users suggested that annotations could be made by using physical objects that could be placed on the positions selected on the tablet. The most plausible way suggested to implement this was to employ reusable putty-like adhesive material (commercially available under several names, like Blu-tack, Pritt-tack or others). This technique offered various advantages over the use of fingers, like being able to mark a larger number of positions in a reasonably persistent way, and potentially to create a variety of shapes that could encode information about the type of annotation. However, this technique did also still have some drawbacks, like the fact that the material stuck on the working surface could become an obstacle for the smooth exploration of the surface with the pen, and that the movement of the pen could accidentally displace or remove some of the marks.

Contrasting these solutions with the list of requirements for annotation tools defined in Table 5, while the use of adhesive physical marker complied with more requirements than the use of the fingers in one hand did, there was still one key limitation in this technique, which referred to the requirement number *vii* (combining the annotations with other functionality for information seeking in the interface, due to the annotations not being tracked by the interface). The implication was that these marks could not integrate with the rest of the functionality in TableVis, for not being recognised by the computer. In practical terms, the technology utilised to design EMAs for TableVis had therefore to be computer supported.

One possibility was to provide the means for annotating or simply marking parts of the data that were of interest, in such a way that rendering them auditorily with HDS was highlighted by the mapping to some additional parameter of sound, like loudness or timbre. The auditory rendering generated with HDS was, however, already very rich in information. Research on parametric sonification suggests that the number of simultaneous sound parameters used in auditory displays should be kept to a minimum [50], which discouraged this approach.

Alternatively, the somatic senses were not very intensively used in TableVis except to control navigation and maintain context information through proprioception (see Section 4.5.2). This suggested that haptic interaction techniques could be used in combination with the auditory display to access annotated information. The combination of auditory and haptic modalities for multimodal interaction had proved to be successful in previous research studies [38, 75, 109, 172]. Among the haptic technologies reviewed in Chapter 2, Wall and

Brewster [167] had already researched on combining audio with force-feedback to propose a form of EMA that they called "*beacons*", which helped in the haptic exploration of 3D-graphs. Incorporating force-feedback interaction to TableVis (maybe by replacing the electronic pen on the tablet with the stylus of a PHANTOM force-feedback device [151] to conduct explorations using HDS in a virtual space) would present interesting research questions. However, this approach required performing a completely new implementation of the interface, a task that could not be undertaken within the scope of this thesis. Other options were to use technologies that actuated on the mechanoreceptors of the skin. Two such technologies were reviewed in Chapter 2: pin array displays and vibrotactile displays. Pin array displays had previously been mounted on other input devices, like a mouse in the case of the VirTouch, with two 4 by 4 arrays [159] and on a pen, as demonstrated by Kyung and Park's Ubi-Pen prototype [91]. In these cases, a finger on the hand that holds the pen must remain in contact with the display, to feel the pins. This would affect the ergonomics of using the interface, which requires that the non-dominant hand can be freely placed in areas of the tablet that can provide spatial references for context. Pen ergonomics appeared to be particularly critical for blind users who, as found during initial evaluations of TableVis reported in Chapter 4.

Vibrotactile transducers could also be integrated on the pen of the tablet, and they did not interfere with the ergonomics of holding the pen because vibratory energy could easily be transmitted through the pen and felt on any point of its surface, independently of the way in which the pen was held. Research by Brewster and Brown [22] has shown that with *Tactons* that are generated with simple vibrotactile transducers, rich messages can be transmitted to the user cutaneously (see also the discussion about Tactons in 2.2.4). Brewster and Brown listed the parameters that can be manipulated in the design of Tactons as vibrotactile messages: frequency, amplitude, waveform, duration, rhythm, body location and spatiotemporal patterns. These parameters can be combined in different ways to construct messages: compound Tactons, hierarchical Tactons and transformational Tactons [22]. In the context of this chapter, these properties of Tactons presented them as an appropriate technique to retrieve annotations from the data added as EMAs.

The next section describes the design and implementation of EMA-Tactons. These are vibrotactile messages that users could add as EMA annotations on the data in combination with TableVis. Their compliance of EMA-Tactons with the requirements listed in Table 5, for such annotations is also discussed in that section.

7.4 Design and Implementation of EMA-Tactons

Deriving from the general definition of Tactons given by Brewster and Brown [22], EMA-Tactons are defined as follows:

> *"EMA-Tactons are structured, abstract messages that can be used to communicate external memory aid information non-visually".*

Using EMA-Tactons, the potential for the user to add EMA annotations and to retrieve them via cutaneous stimulation can surpass the mere binary marking of data points or data sub-sets, and it can add more richness to those annotations. In the case of tabular numerical data, such annotations could, for example, inform the user which element of the data table (a single cell or a complete row or column) had been annotated, in addition to ranking position of the annotated piece of data according to several criteria in a particular data exploration. Ideally, the richness of these annotations would approach that of simple annotations that sighted users can make while exploring a data set presented on a visual medium (such as the use of two colours to encode different information in the example in Figure 31).

As outlined above, in the prototype implementation of EMA-Tactons for TableVis, the annotations were retrieved as vibrotactile messages. A Tactaid VBW32 transducer (Figure 32, left) was selected to generate the vibrotactile stimuli [34]. The nominal vibration frequency of a Tactaid (*i.e.* the frequency at which amplitude is highest) is 250Hz. A Tactaid transducer was mounted laterally on the rear end of the tablet's electronic pen (Figure 32, right), using adhesive tape that ensured hard contact between the transducer and the pen. First tests showed that it was important that this contact was not loose, otherwise the external shell of the transducer rattled against the surface of the pen, which could be heard by the user. In a study of augmented stylus-based interfaces, Lee *et al.* [93] mounted a solenoid axially on their Haptic Pen, to accurately simulate a physical event (pressing a GUI button on a touch screen by pushing it with the pen, and feeling the mechanical feedback of that action, as if a physical button had been pressed). EMA-Tactons are abstract information with no physical tangible equivalent to be reproduced, so the attachment was freely chosen with the only aim to make it secure and consistent for all users, as well as functional (*i.e.* the vibration could be easily felt while holding the pen). Thus, the lateral mounting was used for ease and reliability. Since the transducer was mounted at the rear end of the pen, by holding it normally the user's hand would not touch the transducer directly, and the way in which the pen was held would not affect the effectiveness of the prototype because the vibration was transmitted along the pen and felt on any point of its surface that was touched (Figure 32, right). Using a test-bench programmed using Pure Data [121], the pen+transducer assembly was informally tested to observe that at 270Hz the vibration was most noticeable on the pen (the difference in frequency with the nominal frequency of the pen alone was probably due to the change in mass and shape of the pen+transducer assembly as opposed to the transducer alone).

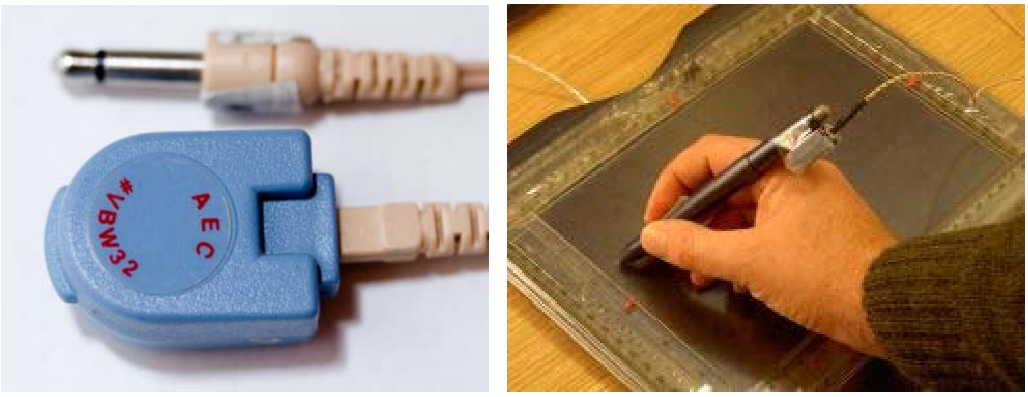


Figure 32. Tactaid VBW32 transducer (left). Tactaid mounted on the pen of the tablet (right).

While Tactons do possess the capacity to convey complex messages, as explained above, the first prototype of EMA-Tactons for TableVis was implemented, as a proof-of-concept, to create the simplest possible messages, which conveyed binary information only. By doing this, the intention was to test the complete setup for usability and to find out if being able to add several one-bit (1 or 0) annotations was enough to relieve the working memory restrictions that had been observed. This would be equivalent to adding any one single type of graphic annotation in a visual print version of the data set, similar to highlighting with a single highlighter pen, for instance.

In the EMA-Tactons prototype this was achieved as follows. During the exploration of a data table, a user could, by pressing a button, mark (highlight or annotate with one bit of information) any position that the pen was pointing at. Depending on the selected navigation mode, only the cell or the complete row or column being pointed at would be marked. As soon as a mark was placed, the vibration would begin and it would continue for as long as the pen remained on the marked cell, row or column. As soon as the pen left the marked area (either because it was lifted from the tablet or because it was displaced to an unmarked area) the vibration would immediately cease. As the exploration progressed, every time the pen re-entered a marked area the vibration on the pen would resume (this behaviour is represented graphically in Figure 39, under "Design-1"). The vibration (a very simple EMA-Tacton) was designed to be of constant-intensity, in the form of a 270Hz sine wave. A user could adjust the actual intensity of the vibration, so that it could be easily noticed yet did not produce any discomfort. The auditory behaviour of the interface would not change in any way due to the presence of EMA-Tactons, and data would be sonified exactly as described in the previous chapters (this adheres to the EMA design requirements defined in Table 5, as discussed below).

In the same way as an EMA-Tacton could be added, it could also be removed, pressing the same button: if when pressing the button the data being pointed at was not marked, an EMA-Tacton would be added; if the data already contained an EMA-Tacton, this would be removed, its effect (the vibration) ceasing immediately. Adding or removing an EMA-Tacton was confirmed by a different, easily distinguishable, percussion sound (a higher-pitch metal sound for an addition and a deeper drum-like sound for a deletion). The actual button to be pressed to add or remove an EMA-Tacton could be customised. A large external button placed on the side of the non-dominant hand was used for this purpose.

In the prototype, data could be marked with EMA-Tactons independently in each navigation mode. Switching between modes did not delete any marks, but made them selectively accessible (*e.g.* in columns mode, only marks affecting complete columns would be accessible). This model could easily be extended to selecting a cell by intersection of a selected row and a selected column, thus complying better with the requirement of integration with other exploration tools already available in the interface (point *vii* in Table 5). Also complying with that requirement, once several data points or sub-sets had been marked, those marks could be used as a criterion to filter out less interesting data. In the prototype implementation, pressing a key on the keyboard or one of the controls on the tablet frame would switch to a filtered condition, in which only the selected data could be sonified. As the user moved the pen on the tablet, he or she would only hear the sonifications of the data that were selected (as well as feeling the corresponding EMA-Tacton). This functionality was not included in the version with which the evaluation was conducted. Those sonifications would still occur on their corresponding physical locations on the tablet, providing consistent context information about their positions at all times. This behaviour of the interface offered explicit support to the problem of comparing two remotely located rows or columns, avoiding the interference of the sounds produced by the data in between, as the pilot study reported in the next section shows. This filtering functionality partly tackles Shneiderman's "zoom and filter" second stage of information seeking. Although this thesis concentrates mainly on obtaining overview information, it does so by also opening paths towards completing the whole exploration process. This filtering functionality is yet another one of those paths. It can be envisioned that with a more complex vocabulary of Tactons than the binary EMA-Tactons implemented here, filtering could take place following different criteria. Investigating how those richer functionalities could be implemented is left for future work. The experimental work reported in the next sections investigates whether the foundational, basic functionality and concepts are correct and successfully implemented. Thus, the version of the application used in the evaluation did not include the filtering functionality.

The prototype complied with the requirements in Table 5. In addition to points *i* and *vii* that have just been discussed, requirements *ii* (annotations should not be limited in number), *iii* (each annotation must remain in the same position, unless explicitly moved by the user), *iv* (annotations must be easy to find, and distinguishable from the original information), *v* (adding an annotation must leave the information in the data set unaltered) and *vi* (an annotation must not obstruct the access to the original information in that position) were also satisfied at design level. Point *viii* (large variety and richness of annotations), is possible in theory and according to other implementations of Tactons reported in literature [27], however the study in the scope of this thesis will not go beyond testing only a very simple binary version of EMA-Tactons.

7.5 Pilot study

The prototype of EMA-Tactons implemented and integrated in TableVis as described above was piloted with a double aim: to validate the design and implementation, and to define a quantitative and qualitative study. Three visually impaired participants took part in this pilot study (two congenitally blind participants and one partially sighted participant with very little residual vision). All three participants were regular screen reader users.

7.5.1 Design of the study

For each participant, the study started with an introduction to the interface and to the concepts involved in exploring data through interactive sonification (which all three participants had used before, when they had taken part in previous evaluations of TableVis). Then, several exploratory tasks were set for the participants, whose explorations were observed, while encouraging them to "think aloud" in order to understand the strategies employed and issues encountered, with the aim to confirm that working memory saturation problems were arising (as intended with the design of a challenging exploratory task). After the first exploration with TableVis, the new EMA-Tactons functionality was introduced and further explorations were conducted, still encouraging "think aloud". Finally, an opportunity for informal feedback was given to gain further insight into the scenarios experienced and the tools used.

7.5.2 Data sets and exploration tasks

The data sets used for the evaluation were tables with 7 rows and 24 columns, the same size as some of the tables used in earlier evaluations of TableVis and with which working memory saturation problems had been observed (as discussed in Section 7.2.1). To isolate the problem under investigation and to preserve simplicity of the task, the only exploration mode available was by columns, and participants were not required to interpret the meaning of the data, thus the functionality of obtaining speech details was not provided. In this way, an exploratory task was reduced to scanning the tablet by moving the pen left to right and comparing the 24 different 7-note-chords that were obtained by doing so. The data presented in the tables were such that each set of 7 values in each column had a different arithmetic mean (all means were approximately equidistant from the nearest higher and lower ones) and all the 24 columns had approximately the same standard deviation. To reproduce the scenarios that led to working memory saturation, columns in each data table were placed in random order, so that the means did not follow any discernible progression (an example of this is graphically represented on the left half of Figure 34). The task for the participants in the pilot study was to *find the column with the highest perceived overall pitch (POP)*, and the accuracy of the answer was judged by selecting the column resulting in the highest arithmetic mean.

For the first two participants, the task of finding the highest single column was simple enough to allow them to complete it without the need for EMA-Tactons. While both participants agreed that EMA-Tactons offered help to complete the task, they completed it making little or no use of them. One of the participants explained that a first scan of all the columns showed where the highest few columns were. This participant claimed to be able to remember one POP and compare it against every other POP until a higher one was found. In this way, only one POP and position had to be remembered at a time, thus working memory was never saturated. This participant was consistently correct in finding the column with the highest arithmetic mean. The performance of these first two participants suggested that, while the task was probably challenging enough for some users (in fact, this scenario had caused difficulties with some participants' working memory in previous studies), it did not always result in the saturation of the users' working memory. It is interesting to observe that Wall and Brewster [167] reached a similar conclusion in their study about beacons (force-feedback EMAs) mentioned earlier, where participants could (and even preferred to) manage the difficulties rather than make use of the tools provided to relieve them.

To further challenge working memory and make sure that it approached its limits, the third participant in the pilot study was required to select the 5 columns with the highest POPs in the set, instead of only the single absolute highest one. The number of comparisons required was much larger, as was the number of intermediate results to be temporarily remembered (positions, pitches associated with those positions and number of positions selected). The procedure to complete the task was that when each one of the five target columns was identified, they were selected by pressing a button while the pen was still on the position of the corresponding column. Columns could be selected and deselected by pressing the same button, and in the end the user had to leave five columns selected, where those five columns produced the five highest POPs in the whole set of twenty four columns. The new task was observed to overload the third participant's working memory very quickly, and she reported that it was definitely easier to complete the task when the EMA-Tactons were available, particularly with the use of the filtering function, when all the non-selected columns could be filtered out so that they would not produce any sounds, facilitating targeted comparisons. In fact, the filtering function was so obviously helpful that it seemed unnecessary to run a complete study comparing performance with or without this function. A more interesting question was to know if marking columns and being able to find them again using EMA-Tactons was providing support for working memory, and how users dealt with the multimodal display of information in the segregated form implemented for this design (audio stimuli for data and vibrotactile stimuli for annotations).

No negative feedback was provided or problems were observed regarding the usability of the prototype. A large external button placed next to the tablet on the side of the non-dominant hand was used to add and remove EMA-Tactons, in order to avoid the difficulties with handling the stylus that had been observed in earlier evaluations. This decision had good acceptance, as the button was easy to find and press (the full final setup can be seen in Figure 33 and is described in the next section). The data selection and de-selection confirmation sounds were considered clearly differentiable and helpful. The vibration itself was easily felt and non-obtrusive, and the option of adjusting both the intensity of the vibration and the loudness of the audio (which was heard through a set of headphones) was considered by the participants to be important.

Having defined a sufficiently challenging task to evaluate the tool and having also pilot-tested its implementation in TableVis, the next sections describe the iterative process of qualitative and quantitative evaluations that was conducted.

7.6 Experimental evaluation of the First Prototype

7.6.1 Experimental Conditions

Two experimental conditions were defined for the evaluation of EMA-Tactons in TableVis:

i. Selecting a column added an EMA-Tacton to that location, which could be felt vibrating when going over it with the pen (this is the normal behaviour of EMA-Tactons as they were implemented in TableVis);
ii. Selecting a column did not add an EMA-Tacton, although the selection/de-selection confirmation sounds were still heard (which alerted the user about trying to select the same column twice, as he/she knew from the confirmation sound that a mark had been removed from an already selected column).

In both conditions, users were asked to carry out exactly the same tasks, derived from the pilot study. The task was to explore the data set in the columns mode (selected by default) and by the end of the exploration to leave selected 5 of the 24 columns in each data set, whose POPs were the highest in the whole set.

7.6.2 Participants

Two blind participants and eight sighted, blindfolded participants took part in this experiment. The blind participants were both students attending RNC in Hereford, both male aged between 16-25. One was congenitally blind and the other was adventitiously blind. The remaining eight participants were undergraduate and postgraduate students from the University of Glasgow, 4 males and 4 females, aged between 16 and 45 years old.

7.6.3 Procedure

The experiment was divided into the following four stages:

- Description and training
- Experimental condition I
- Experimental condition II
- Qualitative feedback

Each of the experimental conditions involved exploring and completing the task for 12 different data sets. The same 12 data sets were explored in each condition, but they were presented in a new randomised order for each condition and for each participant, to account for any order effects. The order in which participants completed both conditions was counterbalanced. At the end of each experimental condition, the subjective workload experienced was assessed using modified NASA TLX questionnaires [65]. To obtain further insight into the answers to the modified NASA-TLX questionnaires, participants were asked if they had any other feedback at the end of both experimental conditions.

The setup and equipment used in this experiment can be seen in Figure 33. The vibrotactile transducer was attached to the tablet's stylus, and plugged into an integrated amplifier. The users could use the main volume control of this amplifier to adjust the strength of the vibration during the training session, from a standard initial setting that was suggested. They were allowed to readjust it at any time, in case it became uncomfortable for any reason (although in practice no participant changed the setting that they chose during the training period). A second integrated amplifier was used to power the headphones used throughout the experiment. Users could also adjust the loudness of the sounds using the main volume control.

Three large push-buttons were located on the sides of the tablet, two on the left and one on the right (in the setup for right-handed participants, as shown in Figure 33; for left-handed participants, a "mirror image" layout was used). While the data were being explored, the non-dominant hand used button #2 to add and remove EMA-Tactons. When the participant considered that the task had been completed (*i.e.* there were five columns selected and they produced the chords with the highest POPs) he/she terminated it by pressing button #1, located further to the left. A textured surface was created on button #1 to avoid confusions with button #2 next to it. This kind of confusion was observed taking place a couple of times during the pilot study, and blind participants reported that adding a texture to one of the buttons made the difference obvious and helped avoid confusion. On the other side of the tablet, button #3 was used to start a new task. A few moments after pressing this button, a brief whistle sound indicated that the system was ready to start and that timing had begun. A maximum time of 120 seconds was allowed for each task, (a duration which the pilot study showed was appropriate), but this time was later extended to 180 seconds, as explained in the next paragraph. If the participant had not completed the task and pressed button #1 before this time was over, the system halted and the participant was notified with five short trumpet notes. Then, the participant could proceed with the next task. The status of the task at the point of timeout was recorded. In between tasks, participants were allowed to rest for a few moments if they needed, and they decided when they were ready to continue with the next task. Participants were asked to prioritise the accuracy of their answers over their speed in answering, while trying not to go over time.

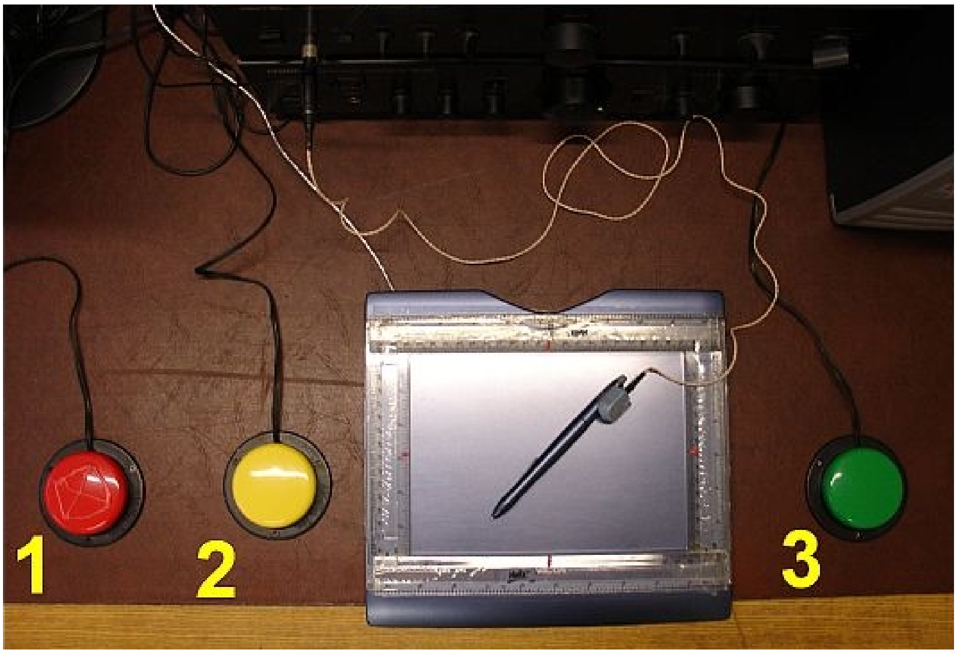
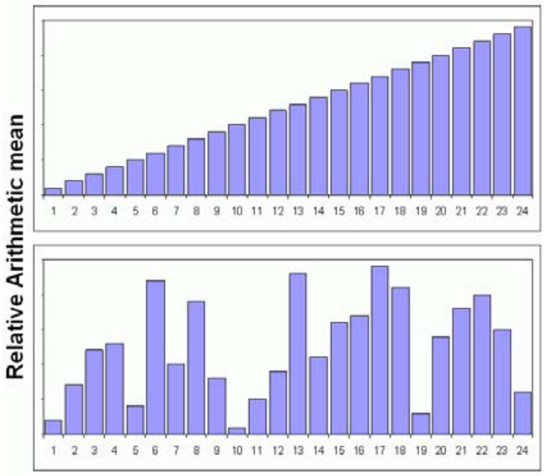


Figure 33. Setup for the EMA-Tactons experiment.

The only difference between the procedures employed with the two blind participants and the sighted participants was that the maximum exploration time allocated to the blind participants was 120 seconds whereas the sighted participants were given 180 seconds. The reason was that the blind participants in this study reported that the limit of 120 seconds imposed some time pressure on them. This had not been found in the pilot study, but it was considered preferable to modify the time allocated for the remaining participants in the study rather than keep the time consistent across both groups, since it was the strategies employed and effectiveness of the performance, and not the efficiency, which the participants were asked to prioritise. The impact of this difference in the results was minimised by rejecting in every case the data from explorations that had timed out, which were only occasional. However, this difference was taken into account at the time of comparing the results from both groups.

7.6.4 Data Sets

Twelve data sets were constructed, in which the 24 columns contained sets of 7 numbers that resulted in different arithmetic means that could be sorted, using HDS, in increasing order (Figure 34, top-left). In each of the data sets, the order of the columns was randomised, so that the five columns with the highest means could appear in any position. A typical example of this is shown in the bottom-left of Figure 34. The correct solution for that example is represented on the right side of Figure 34, where the shaded columns are those selected, over which the pen would display an EMA-Tacton.



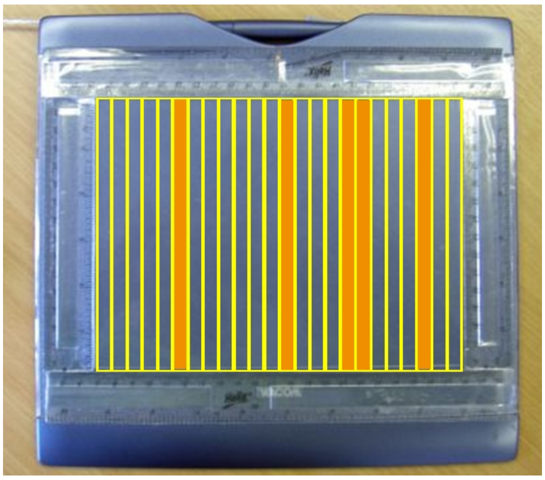


Figure 34. *Left:* graphic representation of the relative arithmetic means of the values in the 24 columns of the data sets used, sorted in ascending order (top) and in random order, typical in a data sets in the experiment (bottom). *Right:* shaded columns indicate the areas on which the pen would vibrate after the task had been completed correctly.

7.6.5 Experiment hypotheses and metrics

In this experiment, it was expected that the task and data sets would create scenarios that led in every case to saturation of working memory. Under such conditions, it was hypothesised that using EMA-Tactons would be beneficial. In particular, the following hypotheses were formulated:

i. The effectiveness (accuracy) in the completion of the tasks will be higher in the condition that uses EMA-Tactons;

ii. The subjective workload will be lower in the condition that uses EMA-Tactons.

No hypothesis was formulated regarding the efficiency of task completion, because the aim of using EMA-Tactons is not to provide faster explorations but to support the user in the completion of long and complex explorations preventing working memory saturation. It was expected that this effect would be reflected in the subjective workload being alleviated and in users being able to implement the search algorithms in full, until the correct answer was obtained, thus increasing effectiveness.

To assess quantitatively the performance of the users in terms of effectiveness, the percentage accuracy of solving the task was divided into two clearly differentiated sub-tasks, each of which provided a metric of effectiveness. A global measure of accuracy was calculated as the multiplication of those two partial metrics, which assessed the accuracy in solving the task as a whole.

- *Sub-task 1 (number of selections).* This part of the task assessed the accuracy in selecting exactly 5 positions on the tablet. 100 % accuracy was only obtained when exactly 5 positions were selected. The further (above or below) the amount of selections was from this target number, the lower would the level of accuracy be. This metric is calculated with the following formula:

$$AccuracySubTask1(\%) = 100 \cdot \left(1 - \frac{|S_s - 5|}{5}\right)$$

Equation 2

- *Sub-task 2 (selecting the highest pitches).* The second subtask considers the accuracy in having selected the positions with the highest POPs. 100% accuracy was obtained only if all the positions selected corresponded to the group of the same number of sounds with the highest pitch. For example, if 7 positions were selected and they were the 7 sounds with the highest pitch in the whole set of 24 sounds then sub-task 2 was 100% correct.

$$AccuracySubTask2(\%) = 100 \cdot \left(\frac{S_c}{S_s}\right)$$

Equation 3

- *Overall task (Combination of sub-tasks 1 and 2).* Metric to assess the accuracy of the overall task, as the product of both sub-tasks. 100% accuracy was only obtained if exactly 5 positions were selected <u>and</u> they corresponded to the 5 highest pitch sounds in the set. This metric is calculated with the following formula:

$$OverallAccuracy(\%) = 100 \cdot \left(1 - \frac{|S_s - 5|}{5}\right) \cdot \frac{S_c}{S_S}$$

Equation 4

In all these formulae, *Ss* is the number of sounds selected and *Sc* is the number of sounds from the selection that are in the group of the Ss sounds with the highest pitch. For example, if 7 sounds are selected, then Ss = 7, and Sc will be the number of those 7 sounds selected that belong to the group of the 7 sounds with the highest pitch.

7.6.6 Results

Since blind users and sighted-blindfolded users are two distinct populations, the results from the two blind users are presented separately from those belonging to the group of sighted participants. Then, the results from the bigger group are compared with those of the two blind participants.

The results from the evaluation with sighted-blindfolded participants are summarised in Figure 35. They show that the effect of using EMA-Tactons was small, as differences were not statistically significant for any of the metrics, according to two-tailed T-tests (paired two sample for means): sub-task 1 (T_7=1.609, p=0.152); sub-task 2 (T_7=-0.378, p=0.717); overall task (T_7=1.27, p=0.245).

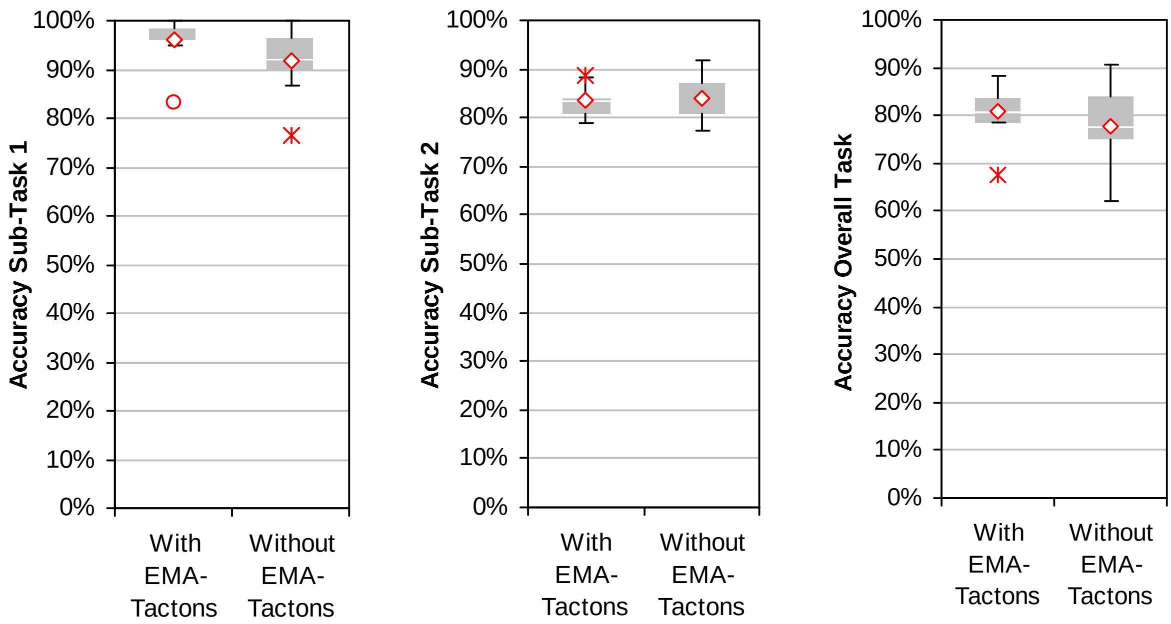


Figure 35. Evaluation of the first EMA-Tactons prototype, with 8 sighted-blindfolded participants. Summary of results, in percentage of task completion accuracy.

The results from the two blind participants for the same experiment (shown in Figure 36) are similar to the results from the group of sighted participants in the trends that they suggest. In the case of both groups, sub-task 1 (number of selections) is slightly more accurate with EMA-Tactons. Also in both cases sub-task 2 (selecting the highest pitches) is slightly less accurate with EMA-Tactons (the differences suggested here not being significant in any of the these two cases).

The hypothesis that effectiveness would be higher with EMA-Tactons could not be proved, according to these results. Among the qualitative feedback provided by the participants, many of them agreed in saying that the vibrotactile information could sometimes help and sometimes interfere in the process of solving the task. Participants generally believed that EMA-Tactons were helpful to keep count of how many columns had already been selected. Several participants, however, reported that sometimes the vibration on the pen could be distracting, stating that it could even get in the way of the data exploration when the user was trying to listen to the sounds. Others said that they found EMA-Tactons helpful in general but that it was very difficult to get information simultaneously from sound and from vibration and that they concentrated on the source of information they needed at each time, ignoring the other source or modality. One participant also reported that vibration and sound sometimes seemed to be two completely unrelated events.

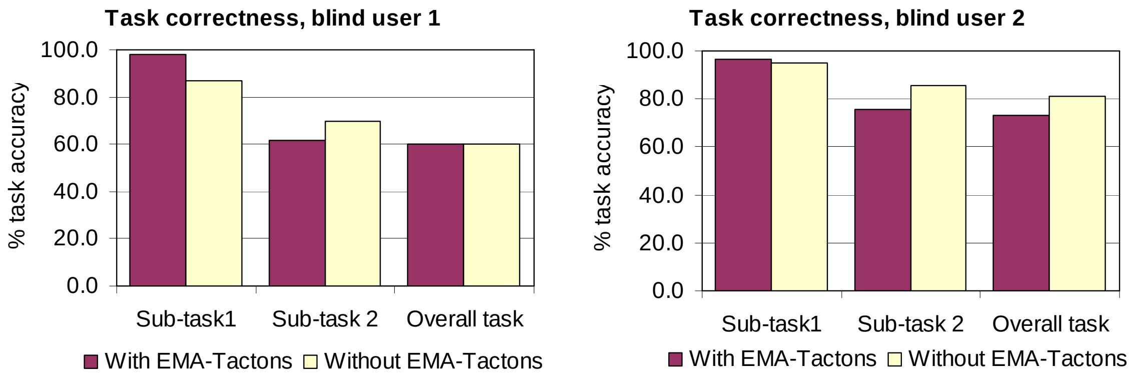


Figure 36. Results (percentage of task completion accuracy) from the two blind users in the prototype evaluation.

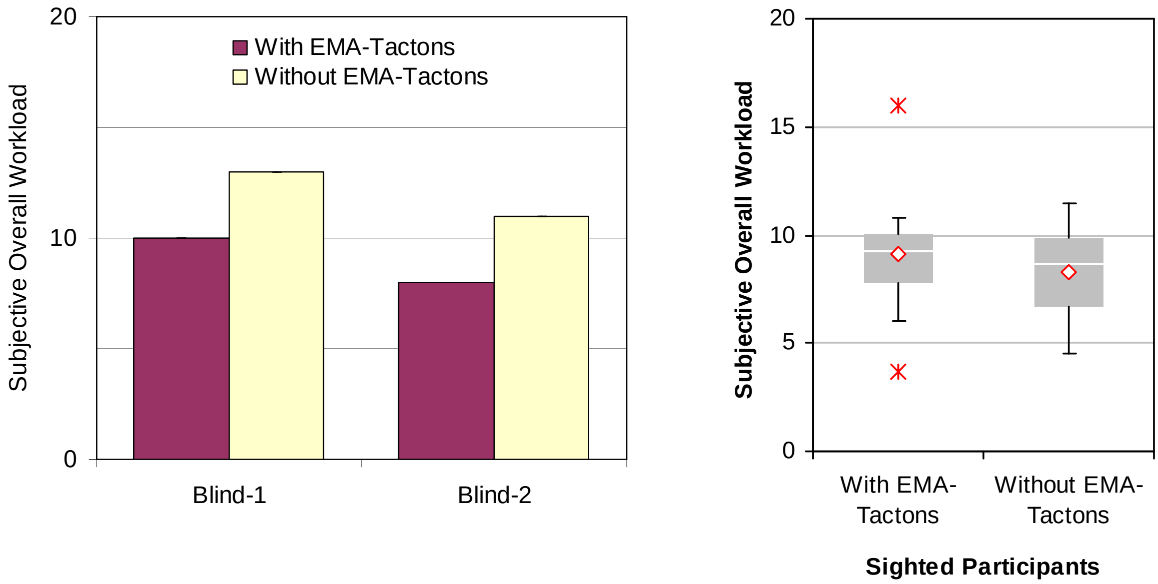


Figure 37. Subjective overall workload, from the modified NASA-TLX questionnaires. The bar chart (left) shows results from two blind participants, while the box plot (right) presents results from 8 sighted blindfolded participants.

Figure 37 shows the overall subjective workload calculated from the modified NASA-TLX measures recorded from all the participants in this experiment. The first two cases ("Blind-1" and "Blind-2") belong to the two blind participants in the experiment, and the third case is the average of the measures collected from the population of sighted-blindfolded participants. Two-tailed T-tests (paired two sample for means) performed on the results of this last group did not show any significant differences (T_7=0.558, p=0.594). The results of one of the participants in this group could be considered an outlier, since the difference between conditions was more than 2σ larger (in absolute value) than the average difference from the rest of participants. Removing this result, the average workload without EMA-Tactons (M=8.76; SD=2.41) turned out to be very slightly larger than with EMA-Tactons (M=8.17;SD=2.5), but still without any significance in the difference (T_6=-0.947, p=0.38). The results from the two blind participants show a clearer reduction in the overall subjective workload when EMA-Tactons were used.

7.6.7 Discussion

A possible explanation for the results described above could be found in the differences between covert[5] endogenous and exogenous spatial attention shifts[6], and in aspects of crossmodal spatial attention. When a participant wanted to count how many sounds were already selected, attention was endogenously (voluntarily) diverted in a covert way to the hand holding the pen, to monitor for vibrotactile cues. This situation produced no conflict, as users were expecting to feel vibrations in the hand while successfully ignoring sounds from data sonification. Attention was focused on the hand, and therefore counting the number of EMA-Tactons was generally easy to do. If, on the contrary, a participant was trying to listen to the sounds for further data analysis while several EMA-Tactons had already been added to the data (*i.e.* attention was endogenously directed towards the auditory events), and suddenly the pen displayed an EMA-Tacton, the participant's attention was exogenously (involuntarily) grabbed and diverted from focusing on the auditory events to feeling the vibration that had unexpectedly appeared on the hand. We could still ask why the vibrotactile stimuli diverted attention from sounds but auditory stimuli did not grab and divert attention so strongly from feeling vibrations in the hand. An explanation for this can be found in the recent literature that has studied crossmodal aspects of human attention. For example, Driver & Spence [43] report significant crossmodal constraints on dual-task performance, where stimuli in two different modalities come from different locations. An unexpected stimulus will generally divert attention away from the stimulus already being attended to. In the case analysed here, sounds are produced every time the hand is moved, while Tactons are only displayed on the pen when it enters an annotated column. In this context, vibrotactile stimuli can be considered more unexpected than auditory stimuli, and for that reason more prone to surprise the user and attract his/her attention. In addition, multiple sensory inputs are processed selectively, and some stimuli get processed more thoroughly than others which can be ignored more easily. In this context, it is possible that under the circumstances of this experiment, vibrotactile stimuli have higher precedence over auditory stimuli, and that the human sensory brain will attend to the haptic modality with higher priority.

---

[5] *Covert* spatial attention shifts involve the mental focussing of attention on a stimulus, without physical movement towards the stimulus. Shifting spatial attention by physically reorienting the senses (*e.g.* turning the head towards the direction of a sound) is *overt* spatial attention shift. [47]

[6] *Endogenous* spatial attention shift is when attention is voluntarily directed to the location of relevant stimuli. *Exogenous* spatial attention shift is when attention is shifted involuntarily to the location of an unexpected stimulus [47].

It can be difficult to attend to simultaneous stimuli from different sensory channels when those stimuli are perceived as unrelated. Spence [142] shows evidence for this in audio visual dual tasks. In [141], he writes: "*(…) humans use the same limited pool of attentional resources to process the inputs arriving from each of their senses (e.g., hearing, vision, touch, smell, etc)*", and adds that "*(…) interface designers should realize that the decision to stimulate more senses actually reflects a trade-off between the benefits of utilizing additional senses and the costs associated with dividing attention between different sensory modalities*" (p.1). Since mental resources to attend to information arriving from different sensory channels seem to be limited, increasing the amount of information transmitted to a user by using other sensory channels contributes to saturating those resources and to conflicts over their allocation. Our interaction with the world around us is, however, richly multimodal, and we suffer no conflicts of this kind in most cases. The problem seems to appear when the different information streams are not perceived to be related, or as "valid co-occurrences" [13]. This could be the explanation for the problems participants in the first evaluation encountered: sound and vibration were perceived to be two unrelated events, not a single multimodal event, and this posed new demands on the cognitive system, while it was trying to relieve others.

The non-conclusive results about perceived overall workload reflect this conflict (Figure 37), where a tool designed with the purpose of reducing working memory demands does not manage to reduce the subjective perception of workload in the slightest. The results were different in the case of the two blind participants, who reported that they perceived overall workload to be higher when no aids were used. However, the results from only two individuals make them inconclusive.

From a usability point of view, some participants reported that when two adjacent columns annotated with EMA-Tactons were traversed with the pen, there was little or no discernible gap in the vibration when the boundary between the two columns was crossed, making it difficult to count the number of columns selected (see Figure 39, Design-1). This fact would strengthen the argument that sensory integration did not take place, and that only one sensory channel was attended to at a time. Otherwise, both separate audio onsets would indicate the presence of two separate vibrating areas. On the basis of the results discussed above, it was concluded that the design of the EMA-Tactons should be reviewed to address the main issues identified.

7.7 Prototype Redesign

The main aim in the redesign of binary EMA-Tactons was to encourage multisensory integration between the vibrotactile and auditory stimuli so that when presented simultaneously they were perceived as a single, multimodal event. Designing multimodal interaction techniques requires an understanding of how humans interpret the world and the stimuli that we receive from it through our various senses. The brain interprets as originating in the same event of the physical world those stimuli that, by their properties, appear most likely to be related. In reality, these processes can be very complex and are not yet well understood. Recent research in neurology that has concentrated on the mechanisms which govern multisensory integration processes in the human brain can help with this understanding. The following discussion concentrates on the aspects that have the most relevance for the problem under study here, and that can shed some light on how to carry out the redesign.

Among the many hypotheses that have been proposed to explain and predict the behaviour of multisensory integration, three in particular have been extensively reported in the literature, and are considered as principles of sensory integration: the *principle of inverse effectiveness*, the *spatial rule of multisensory integration* and the *temporal rule of multisensory integration* [68, 69, 143]. It appears that these mechanisms of multisensory integration tend to enhance the signals derived from the same events and to suppress the stimuli derived from different events [143].

The three principles just mentioned can provide valuable guidelines for the design of multimodal events in human-computer interfaces, when multisensory integration is sought after, as is the case here. For these reasons, the three principles are briefly described here (mainly based on the definitions given in [68]):

- *Principle of inverse effectiveness*. The principle that brain signals from different sensory modalities presented simultaneously will be integrated in inverse proportion to their effectiveness when presented in isolation (thus weak stimuli integrate better than strong ones). [69]
- *Spatial rule of multisensory integration*. Multisensory stimuli are integrated depending on their relative spatial locations. Typically, stimuli presented from the same spatial location will result in enhanced multisensory integration, while stimuli presented from different spatial locations result in decreased multisensory integration, or multisensory inhibition.
- *Temporal rule of multisensory integration*. Multisensory stimuli are integrated depending on their relative temporal parameters. Typically, stimuli presented at the same time will result in enhanced multisensory integration, while stimuli presented at different times result in decreased multisensory integration. A "temporal window of multisensory integration" exists such that stimuli from different sensory modalities will be integrated only if both are presented within this window.

The first prototype implementation of EMA-Tactons was evaluated with regard to these three principles to see how the design could be modified to obtain better multisensory integration. It is difficult to assess the degree to which both auditory and vibrotactile stimuli could tend to integrate into a single perceived event according to the principle of inverse effectiveness, but since either stimulus in isolation was perfectly and unambiguously detectable, it looks unlikely that conditions were optimum for integration. Regarding the spatial rule of multisensory integration, both stimuli originated from different spatial locations: the vibrotactile stimulus was felt on one hand while the auditory stimulus seemed to originate in the participant's head when presented monaurally through a set of headphones. This spatial gap could be bridged if the sound was presented differently, for example using loudspeakers placed near the graphics tablet instead of using headphones. While it would be extremely interesting to compare the differences if this change was made (by using spatial audio, for example), this approach was not taken in the first place because the most common and likely use scenario for TableVis as an accessibility tool is through the use of headphones, preserving privacy and avoiding annoyance in the user's environment (two common problems of auditory displays when loudspeakers are used). Instead, it was decided that in the first redesign of the prototype this spatial dislocation would be maintained as long as there were other aspects that could be modified with the likelihood of improving performance.

The temporal rule of multisensory integration in the first prototype was not implemented to favour such integration either. While both sound and vibration started approximately simultaneously, the vibration remained for much longer than the sound, presenting a strong temporal asynchrony. Meredith *et al.* [102] report the observation of some relatively simple patterns, which apply to pairs of audio and somatosensory stimuli, in reference to temporal factors of stimuli processing. They found that pairs of stimuli that started in temporal synchrony but then moved towards strong asynchrony were initially interpreted as a single multimodal event, but as the asynchrony became apparent the whole experience was interpreted as two wholly unrelated events.

Such findings could be used to hypothesise how audio and vibrotactile stimuli are perceived in TableVis. During exploration of an annotated column, both types of stimuli started simultaneously. After that they became highly temporally disparate, as the sound faded away. Following the temporal rule of multisensory integration described above, this would have explicitly indicated that the two stimuli were not generated by the same event in the physical world.

This reasoning was used to redesign the way in which audio and vibrotactile stimuli were temporally presented. Preserving the synchrony at onset and ensuring it was more or less maintained throughout the life of both stimuli. Thus, as the sound faded, the vibration faded in a similar way (Figure 38). A sharp attack was followed by a 120ms sustain period (so that the presence of the vibration was clearly perceived), and the amplitude of the vibration decayed during the last 80ms. The intention behind fading the vibration instead of simply truncating it was to create a more ecological, natural-feeling combination of sound and vibration, based on the fact (experienced all the time in the physical world) that events that produce sounds also vibrate (and *vice versa*) and they do so in a similar way, so that there is a correlation between loudness of sound and amplitude of vibration. In this redesign, such correlation was not reproduced in detail, but it was roughly suggested in the envelope of the vibration, to approximately fit every chord generated using HDS.

From a usability point of view, this change in the design had a few consequences. Remaining on a marked column was not permanently recognised by the feel of the vibration on the pen, as was the case in the first design. Once sound and vibration disappeared, the user had to tap on that position to be able to receive the EMA-Tacton again. An additional implication of this redesign was that it resolved the ambiguity of adjacent seleted columns. With the new design, even when the table was traversed quickly, a new onset of the vibration would be felt much more easily, signifying a new selected column had been entered. Additionally, if sensory integration was obtained more successfully after the redesign, this division of columns would be emphasised by the simultaneous onset of the chords.

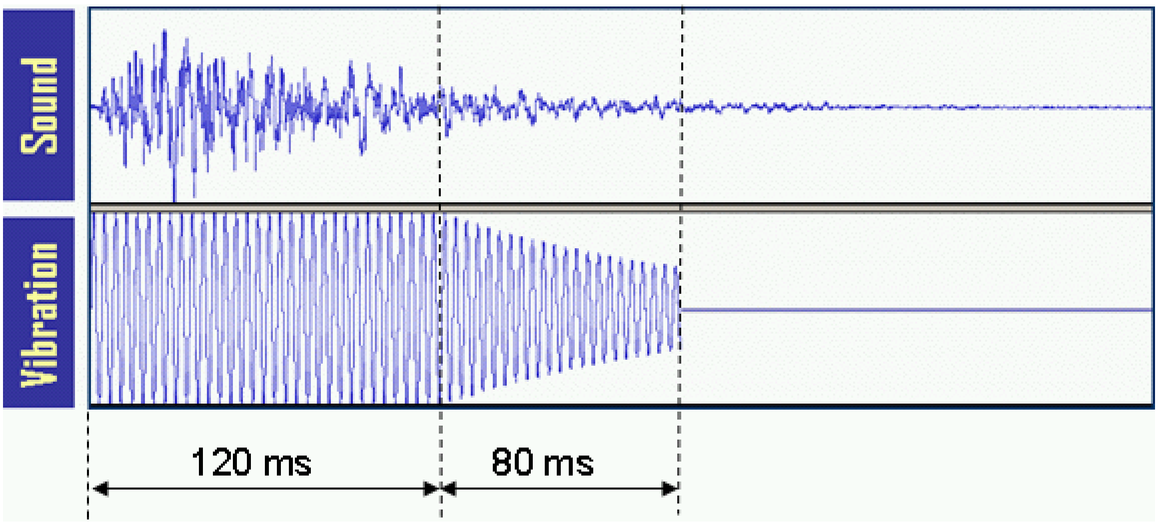


Figure 38. Graphic representation of temporally-synchronised audio and vibrotactile stimuli in the redesigned version of binary EMA-Tactons, as implemented in TableVis. The input wave forms of sound (a typical chord with 7 sounds) and vibration are represented on a timeline (horizontal axis) with origin on the left.

Figure 39 shows a graphic representation of both versions of EMA-Tactons, where these usability differences can be observed side by side. The graph represents a timeline of approximately 1.27 seconds in the middle of a hypothetical data exploration. Along this timeline, the pen first enters a selected column, and while in the first design the vibration is felt during the whole time the pen remains in that column, in the second design the vibration disappears with the sound. The pen then goes on to a non-selected column, where there is only auditory feedback and no differences between designs. The third and fourth stages depict entering another selected column and after that yet another selected one that is located adjacent to the previous one. Notice that while in the first design the gap between selected columns where there is no vibration is extremely narrow, the same border crossing is felt much more easily in the second design.

7.8 Evaluation of Second Prototype

Following the modifications described above, the updated EMA-Tactons design was experimentally evaluated. This evaluation is described and discussed in this section.

7.8.1 Description of the experiment

The experiment in this evaluation involved 12 new sighted-blindfolded participants, 8 males and 4 females, aged between 16-35 years old, who were all undergraduate or postgraduates students from the University of Glasgow. The equipment setup, datasets and procedure followed were identical to those employed in the evaluation of the first prototype with the group of sighted participants reported in Section 7.6. The same two experimental conditions were also used in this case.

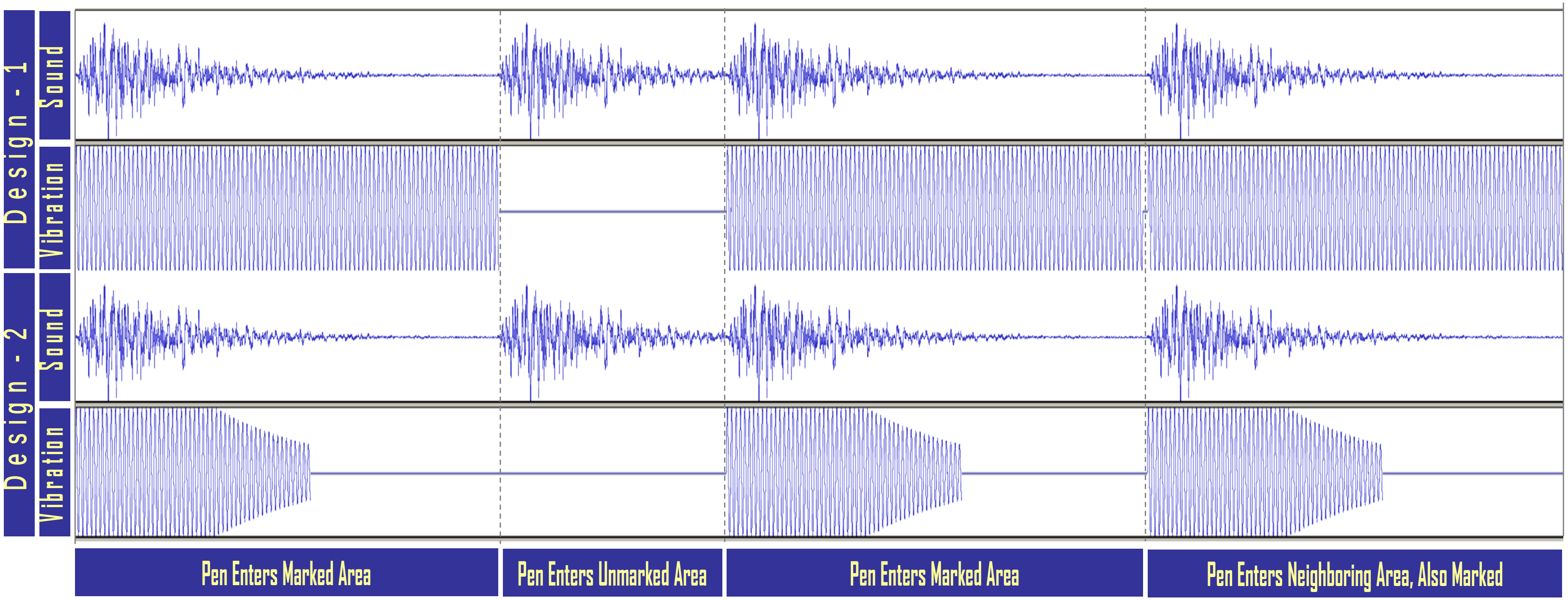


Figure 39. Graphic representation of both versions of EMA-Tactons as implemented in TableVis. For each design, the input wave forms of sound and vibration are represented. The horizontal axis (with origin on the left) represents the timeline (approx. 1.27 seconds in total) in which some usual actions are depicted as taking places: the pen enters a marked column, then it moves to and unmarked one, enters another marked column and then it follows onto yet another marked area that happens to be adjacent to it.

The hypotheses from the evaluation of the first prototype (Section 7.6.5) were still valid for the second prototype:

i. The effectiveness (accuracy) in the completion of the tasks will be higher in the condition that uses EMA-Tactons;

ii. The subjective workload will be lower in the condition that uses EMA-Tactons;

In addition, compared to the results obtained with the first prototype, it was expected that with the use of the second prototype, both stimuli would integrate better, with users tending to perceive them as a single multimodal event rather than as two independent stimuli. Overall mental workload was expected to be reduced further with the second interface; firstly because better sensory integration would reduce conflict between both stimuli competing for the user's attention; secondly, because adjacent selected columns would be easier to differentiate, reducing ambiguity.

7.8.2 Results

The results of the evaluation are shown in Figure 40. Performance in sub-task 1 (accuracy in the number of selections) was statistically significantly better (CI=95%) with EMA-Tactons than without (T_{11}=3.008, p=0.012), participants consistently reached high performance with EMA-Tactons (M=97.16%; SD=4.37).

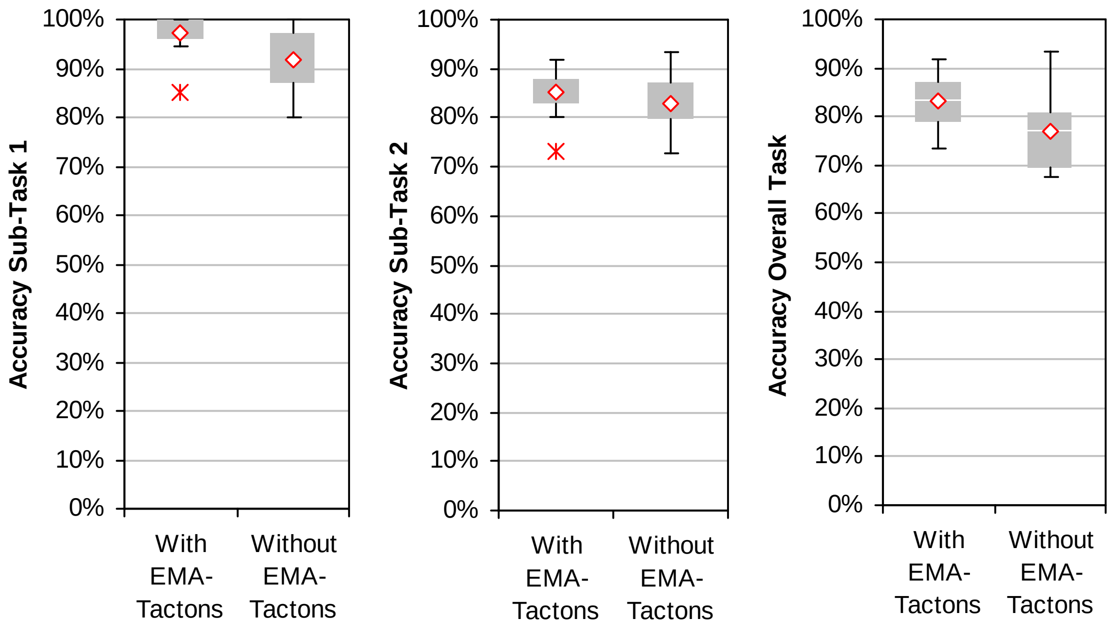


Figure 40. Evaluation of the second EMA-Tactons prototype with 12 sighted-blindfolded participants. Summary of results, in percentage of accuracy in task completion.

There was still no significant difference between the two conditions for performance in sub-task 2, for selecting the highest pitches (T_{11}=1.379, p=0.195). Considering the complete task as a whole, however, the combined results of both sub-tasks lead to the performance with EMA-Tactons being statistically significantly better with a CI of 95% (T_{11}=2.89, p=0.015).

Finally, the subjective overall workload calculated from the modified NASA-TLX questionnaires was significantly lower (with a CI of 95%) when the redesigned version of EMA-Tactons was used than when no aids were available (T_{11}=2.97, p=0.012).

7.8.3 Discussion

The performance results measured with the second prototype are better than those measured with the first. The performance in sub-task 1 (accuracy in keeping count of selected columns) is difficult to improve any further, as all participants scored very highly, close to 100% accurate (M=97.16%; SD=4.37), very slightly higher than with the first prototype (M=96.25; SD=6.61). As expected, in the no-aids condition of the second evaluation, results (M=91.79; SD=11.11) were almost identical to those recorded in the first evaluation (M=91.95; SD=9.89), this coherence confirming the validity of the comparison of results between evaluations.

The difference between conditions for sub-task 2 (accuracy selecting the highest pitches) was not significant. All the results for this sub-task seemed to change very little with any version of EMAs provided. On the other hand the results obtained without any aid and with this deliberately challenging task were higher than expected. One possible reason for this could be that the task was not as challenging as intended. To address this, the experiment was slightly modified to ensure that the level of difficulty was as high as originally intended, and a new evaluation was conducted, a described in the next section.

Despite the fact that performance results were similar for sub-task 2 (accuracy selecting highest pitches) with both prototypes, the significant improvement in accuracy of performance for the tasks as a whole in the second prototype suggests that the redesign decisions were correct. This is also strongly supported by the qualitative feedback received from the participants and the workload measures recorded. Unlike the feedback on the first prototype, comments about confusion between modalities and about the vibrotactile stimuli interfering with the audio signals (diagnosed as spatial attention conflict problems) disappeared from the group of participants that evaluated the second prototype. This suggests that the efforts made to implement the rule of temporal multisensory integration were successful in inducing integration and creating what was perceived to be closer to a single multimodal event. The significant reduction in subjective overall workload provides further evidence that the success in integrating the different sensory stimuli reduced the extra mental processing and attention-splitting burden that the non-integrated stimuli from the first prototype appeared to have generated (see Figure 41).

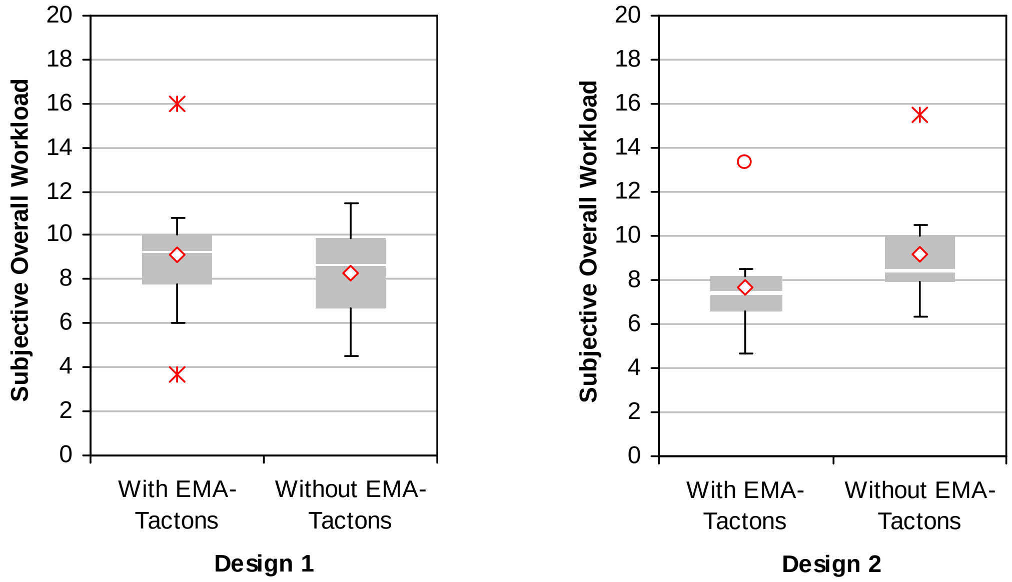


Figure 41. Comparison of average subjective overall workloads (modified NASA-TLX), between the groups of sighted-blindfolded participants in the evaluations of the first and the second prototype.

7.9 Re-evaluation of the second prototype

As reported above, the evaluation of the second EMA-Tactons prototype elicited some unexpected results regarding sub-task 2 (selecting the columns with the highest pitches), which can be summarised as follows:

- Without the help of any EMAs, participants performed unexpectedly well in sub-task 2 with both prototypes, with over 80% average accuracy). This was the case despite the fact that this part of the task had been deliberately designed to be difficult when no EMA was available;
- With EMA-Tactons, performance in sub-task 2 did not improve significantly, despite the redesign for the second prototype, contrary to what had been hypothesised.

The default explanation for these results was that EMA-Tactons were not the right tool to support users in sub-task 2. However, considering other possible reasons to obtain these results led to the design of a modified version of the evaluation of the second prototype, in which the task was modified slightly. This analysis and the subsequent re-evaluation of the prototype are reported in this section.

7.9.1 Analysis of new factors to consider in this study

One reason not to observe significant improvements in the performance of sub-task 2 when EMA-Tactons were added in any of the two prototype versions was that performance without EMAs was already high (around 83% in average). In other words, a ceiling effect provided little room for improvement. The high performance observed without EMAs would mean that the task created for the evaluation was not sufficiently challenging for the participants and that it had not managed to saturate their working memory. However, previous studies presented in this thesis involving simpler exploration scenarios were reported as too challenging by many participants (a fact that motivated the whole of this study). Furthermore, the difficulty of the exploration task had already been increased from the first version used in the pilot study (Section 7.5). If this was the case, the question still remained as to how participants were obtaining such good results without the help of any additional tool.

These two aspects regarding these results for sub-task 2 required closer scrutiny:

- Possible reasons why participants performed well without any EMA:

  a. *Use of pitch memory*. Every data set which each participant explored produced approximately the same chords, arranged in different order (as seen earlier in Figure 34, left). The 24 chords in each data set covered a particular range of POPs, and the five highest-pitched columns that were being targeted covered approximately the same narrow top band of that range in all the data sets. It was possible that after the first few explorations participants learnt what the target columns should sound like, and used that pitch memory from the beginning of each new exploration to filter out every column that sounded clearly outside this top range. If this was the case, the task would not require cross-comparisons between 24 items, as users could learn what the highest pitches sounded like from the few initial explorations. Then, the actual list of sounds from which they selected the highest ones would much smaller, and even simply selecting the whole of that sub-list would provide a high level of accuracy in the answer, without suffering external memory saturation.
  b. *Compromising on accuracy*. In such a complex task it is possible that participants deviated from the expected method for solving it. Indeed, making sure that the answer to the task was correct to the users' best belief requires the systematic, often tedious approach to solving the exploration task repeatedly. However, participants could potentially decide to compromise between the instructions received (which asked them to complete the task to the maximum degree of accuracy) and the amount of thoroughness and effort put into completing it. It is understandable that participants with different levels of motivation decided to use their resourcefulness to optimise their efforts invested to reach a "good enough" answer, rather than what they believed was "the answer". The average 85% accuracy in sub-task 2 (selection of the highest pitches) might well reflect this level of "good enough", and the higher standard deviations associated with these scores could possibly be an indication of the different levels of commitment participants exercised in order to solve the task.

- Possible reasons why using the EMAs did not make any significant difference to performance:

    c. *Ambiguity of pitch discrimination.* HDS has certain limitations as to which extent arithmetic means and perceived overall pitches can be linearly correlated (as discussed in Chapter 8). Thus (and having not conducted the study reported in Chapter 8 that quantifies these limitations), it was thought possible that this factor limited the capacity of participants to achieve higher levels of accuracy even when working memory limitations had been alleviated with EMA-Tactons. In other words, it was possible that with the more easily discriminable pitches from single piano tones participants could achieve higher levels of accuracy in sub-task 2 when working memory saturation did not take place, revealing the benefits of EMA-Tactons.
    d. *Compromising on accuracy.* The same discussion as in b above applies here too.
    e. *The wrong design.* Indeed, the default explanation for not obtaining any improvements in this sub-task could simply be that the tool does not offer the help it is intended to.

Having formulated these possible factors that might explain the results obtained for sub-task 2, the approach adopted was to address the factors that could be controlled and observe whether the results still showed not benefit in the performance achieved for sub-task 2 using EMA-Tactons. The easiest options were to control for *a* (use of pitch memory) and *c* (ambiguity in POP discrimination). The next section describes the modifications made to the experiment and reports the results from the re-evaluation of the prototype.

7.9.2 Description of the experiment

The experiment setup and procedure were exactly the same as for the previous evaluation (Section 7.8), except for the changes implemented to control for the two factors identified above. Twelve new sighted-blindfolded participants were recruited for the experiment. Again, these were University of Glasgow undergraduate and postgraduate students, 10 males and 2 females, aged between 16 and 35.

The differences in the setup with respect to the previous evaluation of the same prototype were the following:

- In order to avoid ambiguity in pitch discriminations, each column produced a single piano note rather than a chord. Thus, participants had to judge the pitches of the single piano notes rather than the POPs from complex chords.
- In order to prevent participants from being able to use pitch memory between explorations, each data set with 24 single piano notes covered a different subset of 23 semitones from the whole range of 65 semitones that TableVis uses (Figure 42). Each of these data sets was, presented in random order to the participants. The single notes in the new data sets were arranged in the same unsorted order as the chords in the data sets employed in the version of

the evaluation with chords (Section 7.8). In this way, every time a new data set was presented the range of sounds encountered was different, and the users could not use pitch memory from previous explorations in any way, having to perform a full exploration each time to identify the highest pitches in the range of frequencies that the data set covered.

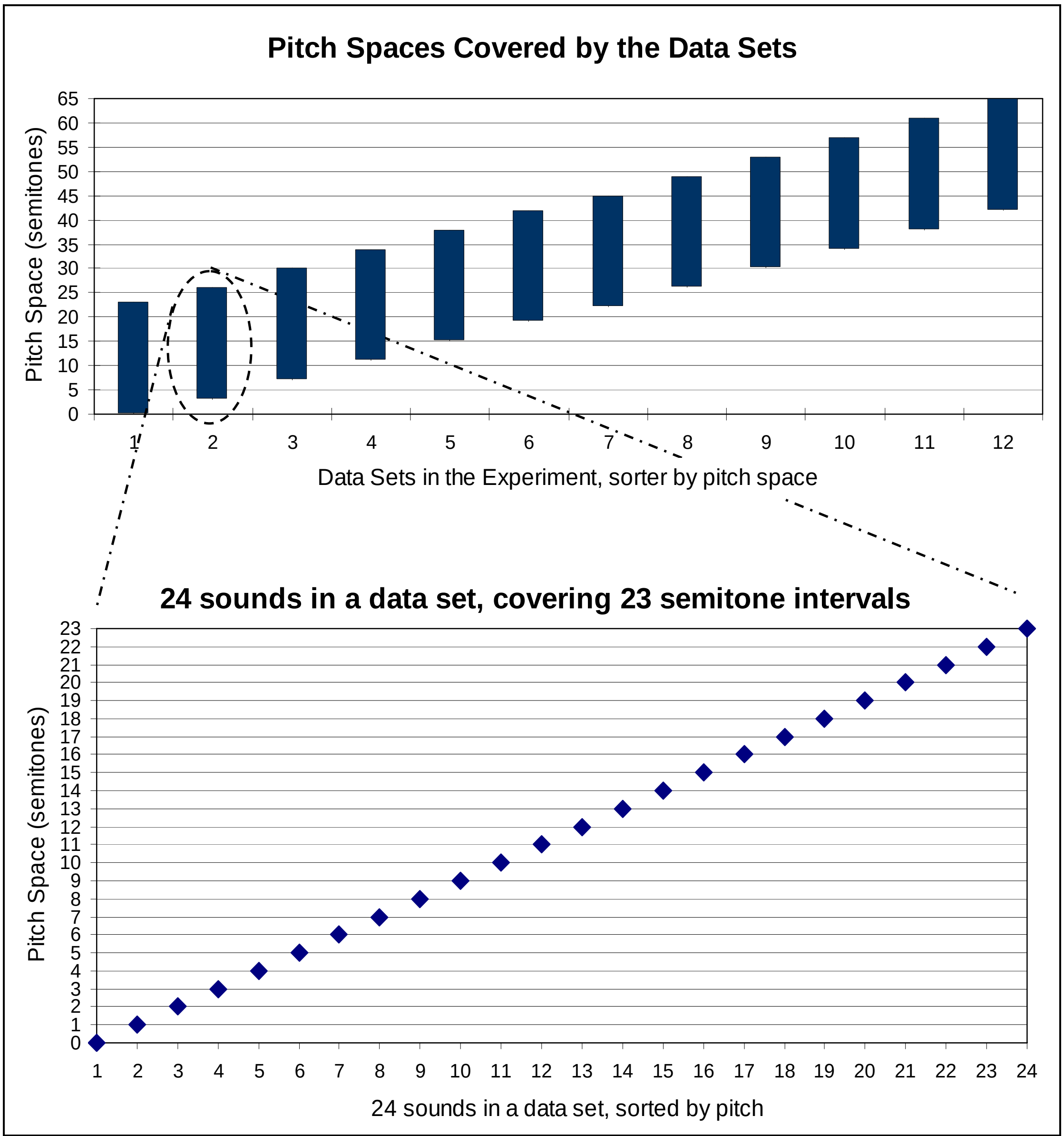


Figure 42. Comparison of pitch spaces between both experimental evaluations of the second prototype. *Top:* the Y axis shows the full extent of the pitch space used by every data set in the previous evaluations with HDS (65 semitones, 61.7Hz to 2.64kHz). Each bar on the graph represents the part of that pitch space (23 semitones) covered by each one of the data sets in the modified experiment. These data sets were presented to each participant in random order. *Bottom:* each data set in the revised experiment produces 24 single notes that if sorted by pitch are in a chromatic scale covering the 23 semitones of one of the intervals in the graph above.

The hypotheses were identical to the previous evaluation.

i. The effectiveness (accuracy) in the completion of the tasks will be higher in the condition that uses EMA-Tactons;

ii. The subjective workload will be lower in the condition that uses EMA-Tactons;

7.9.3 Results

The results of this new evaluation are summarised in Figure 43 below. The improvement in performance for sub-task 1 (accuracy of the number of selections) was not significant (T_{11}=1.892, p=0.085), and average accuracy reached with this sub-task was very high in both conditions: (M=97.4; SD=4.41) with EMA-Tactons and (M=92.5; SD=1.03) without. In contrast, the performance in sub-task 2 (accuracy in selecting the highest pitches) improved significantly, with the use of EMA-Tactons (T_{11}=2.216, p=0.049). Finally, the performance for the overall task also significantly improved when EMA-Tactons were used (T_{11}=2.490, p=0.030).

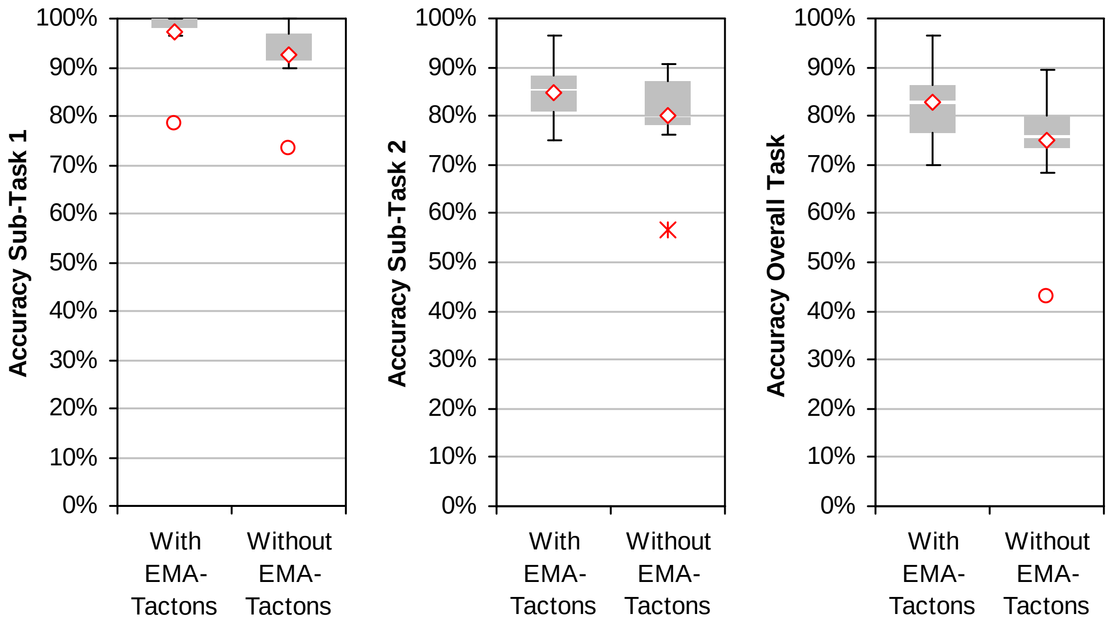


Figure 43. Modified evaluation of the second EMA-Tactons prototype, with 12 sighted-blindfolded participants. Summary of results, in percentage of task. Sounds were single notes instead of chords, and each data set used a different pitch space.

Regarding the subjective overall workload (Figure 44), the level of significance of its reduction in the EMA-Tactons condition was statistically significant, with a confidence interval of 99% (T_{11}=-3.546, p=0.005).

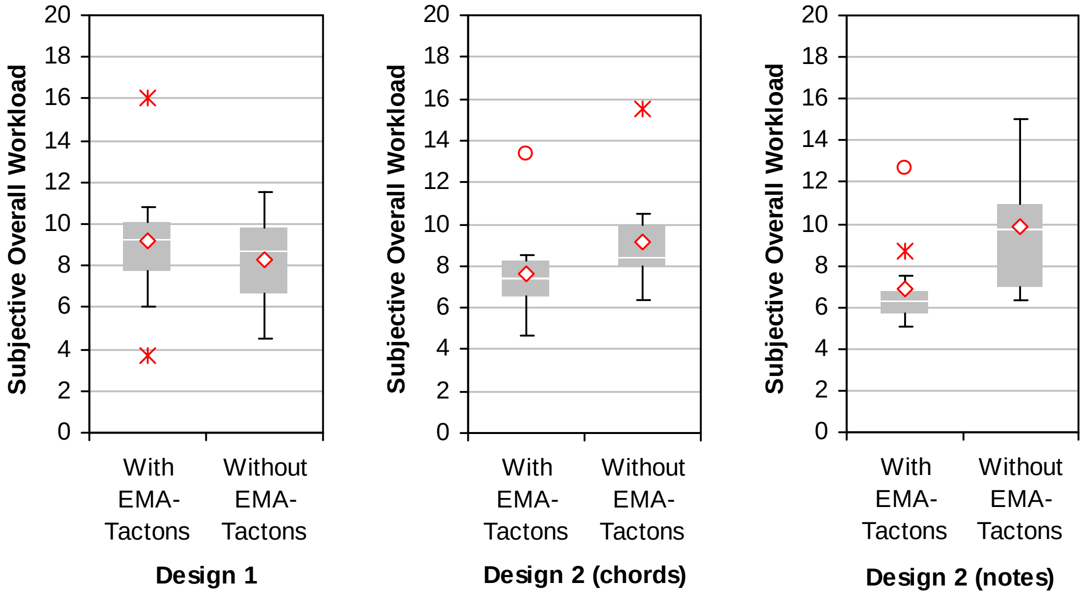


Figure 44. Comparison of average subjective overall workloads (modified NASA-TLX), between the groups of sighted-blindfolded participants in the evaluations of the first and the second prototype.

7.9.4 Discussion

The levels of accuracy achieved for sub-task 1 (keeping count of the number of selections) were very similar to the levels obtained in the previous evaluation of the same prototype, although in this case statistical analysis did not find the difference to be significant. The average level of accuracy reached was, in any case, very high in both conditions, close to 100%. Thus, ceiling effects were likely be present. These results are in line with those obtained in the two previous evaluations which confirms that, despite the efforts to make the task very challenging (and despite the perception of the users that EMA-Tactons helped a lot keeping count of the number of selections), keeping count could be done remarkably well mentally.

As expected, the increase in accuracy for sub-task 2 (selecting the highest-pitched sounds) reached statistical significance. It is difficult to attribute this difference to only one of both of the aspects of the experiment that were modified, but as a result of these the study showed that the second design of EMA-Tactons did assist participants in solving this sub-task more accurately too.

Looking at other results, in all three experiments the time to complete the task was on average longer when EMA-Tactons were used than when they were not (it must be remembered that participants were not trying to finish the tasks fast, but to finish them within 180 seconds, which, as seen in the averages of Figure 45, was normally enough time). This difference was also true for the two visually-impaired participants who took part in the first evaluation, who required on average 90.9 and 75.9 seconds respectively to complete the task with EMA-Tactons, while it took them only 67.3 and 57.8 seconds respectively to complete it without any aids. These results reinforce the hypothesis that EMA-Tactons alleviated working memory restrictions: based on qualitative feedback, the consistently longer explorations of participants when they used EMA-Tactons can be attributed to the fact that, with EMA-Tactons, participants did not reach saturation of working memory, and could be more thorough in their searches, resulting in longer data explorations. This difference between conditions was only significant in the first evaluation of the second design, with a confidence interval of more than 95% (T_{11}=2.789, p=0.018).

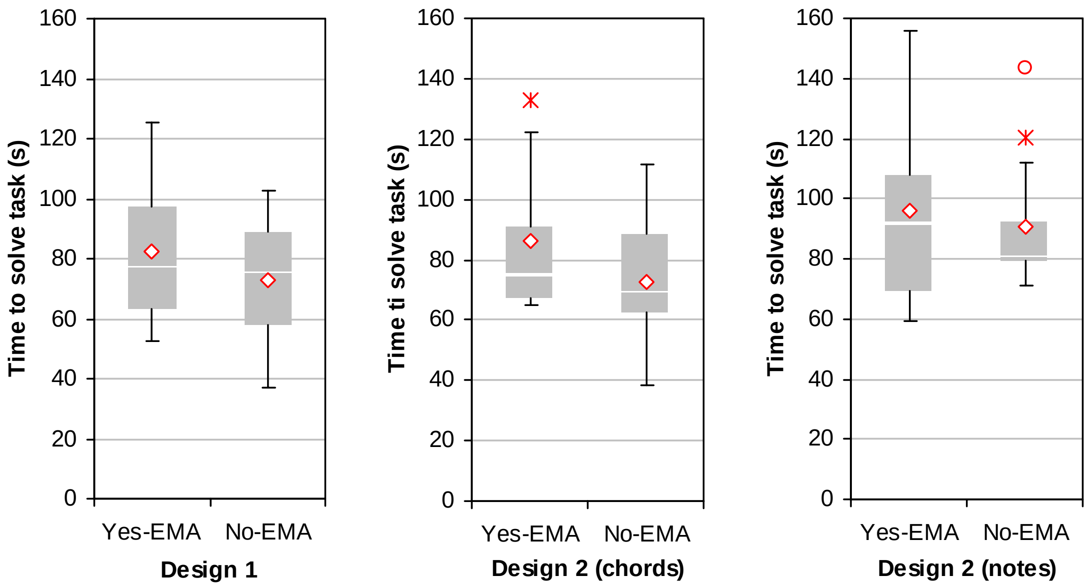


Figure 45. Average time to complete the task in the three experiments.

7.10 Conclusions

This chapter analysed data exploration scenarios observed in previous studies with TableVis where participants encountered difficulties to complete exploration tasks due to working memory saturation. Once the characteristics of such exploration tasks were determined, a set of requirements was defined for external memory aids (EMAs) that could alleviate working memory limitations, based on annotation practices used on print media. The use of EMA-Tactons was proposed as a technique for addressing this problem, which was designed and experimentally evaluated through an iterative design process. The design process was described and justified, resulting in the implementation of EMA-Tactons, a technique that made use of vibrotactile messages to communicate external memory aid information non-visually. After a pilot study conducted with blind and visually impaired participants, an experimental evaluation of the first prototype was conducted, the outcome of which led to the refinement of the design and implementation of a second prototype of EMA-Tactons, which was in turn evaluated in another two experimental studies.

The research questions posed by this thesis are: "*How can overview information from tabular numerical data sets be obtained non-visually?*" and "*What level of performance can be achieved with the techniques proposed in this thesis?*". The work reported in this chapter contributes to answering both of these questions. In answer to Research Question 1, this chapter has confirmed the necessity of external memory aids (EMAs) in supporting the process to obtain overviews through exploration of tabular numerical data sets. In addition, it has defined the key requirements for the design of EMAs to support such exploration tasks. The experimental evaluations of the EMA-Tactons prototypes have shown that it is possible to utilise EMA-Tactons to support working memory in data exploration tasks, thus further supporting users in obtaining overviews of numerical data sets.

In answer to Research Question 2, it was found that in explorations of complex data sets in which trends and patterns are not obvious, and which in unaided explorations lead to working memory saturation, the use of EMA-Tactons permitted users to attain significant improvement in effectiveness of the explorations. In particular, the exploration tasks analysed involved identifying multiple specific features embedded in the data. In addition, by using EMA-Tactons, participants' mental workload was significantly reduced, confirming that EMA-Tactons contributed towards alleviating the constraints on short term memory which they otherwise suffered.

The design process reported in this chapter illustrated that designing multimodal interfaces for good integration of sensory channels is difficult and complex. Subtle changes can make a big difference in perceiving a single multimodal event instead of unrelated events in different sensory channels. Careful redesign of the vibrotactile stimuli following principles of ecological perception produced a better integration of multisensory information, which led to significant improvements in performance.

These experiments have shown that it is possible to utilise EMA-Tactons to support working memory in TableVis. However, only binary EMA-Tactons were implemented in this study, which used one bit of information per annotation. Previous research on Tactons has shown that they are able to encode more complex messages [27]. The combination of rich annotations in different exploratory modalities (rows, columns and cells) has the potential to offer support for complex exploratory tasks that today can only be done visually. It would therefore be interesting to investigate whether semantically richer Tactons could be used to support data explorations with TableVis or other interfaces utilising HDS. This, however, is outside the scope of the current thesis and is therefore left for future work.

The next chapter presents a detailed research study on the perceptual and cognitive aspects related to data mining using High-Density Sonification.

Chapter 8 Understanding Perceptual and Cognitive Aspects of HDS

8.1 Introduction

This chapter reports an in-depth study devoted to understanding perceptual and cognitive aspects related to data mining using HDS. This study contributes to answering Research Question 2, "*What level of performance can be achieved with the techniques proposed in this thesis?*" by investigating the reliability with which relative arithmetic means can be estimated by accessing a collection of numbers in a single auditory event (a musical chord) generated with HDS.

In all the experiments described in the previous chapters, the measurement of the participants' accuracy in carrying out data explorations was based on checking whether they could find the row or column with the highest or the lowest arithmetic mean. As justified in Chapter 5, this way of measuring the accuracy of the answers was chosen because the way in which the arithmetic mean varies is a common and meaningful way of summarising the information in a tabular data set, and because it describes the main features present in the data. In the experiments, users estimated the behaviour of the arithmetic mean (the way in which it changed across a table) by accessing chunks of raw data (all the data in a row or a column), grouped in chords generated using HDS. The relative value of the Perceived Overall Pitch (POP) of such a chord was an estimate of the relative arithmetic mean of the numbers that generated that chord. The precision required in these judgements of pitch was higher with tasks of type I than type II (see Section 5.2.1), in which users had to identify a single row or column from the whole set, often involving the comparison of very similar data sets (and hence very similar chords). Even with such precise demands, the data collected from the experiments show that the accuracy attained was consistently high, as summarised in Table 6. These results show that, with HDS, users could perform basic statistical analysis of tabular data without having to make any numerical computations, mentally or otherwise. In such statistical analysis, users obtained knowledge about relative variations of arithmetic means in both axes of the tables, including an overview of the evolution of the mean in each axis and the precise positions in which extreme values occurred. Without HDS, the conventional way of doing this analysis would involve calculating explicitly the arithmetic means for all the rows and columns in a table and constructing visual graphs in which the evolution of the means in both axes could be seen. Additionally, a user would have to identify the exact positions for the maximum or minimum points of the mean in each axis. With HDS, this analysis was completed by directly accessing the raw data in the cells of the table, in the form of sound, by exploiting the way in which the human auditory system perceives these sounds. Based on the data summarised in Table 6, the accuracy that could be obtained with this technique ranged from 80% to 90% in most of the cases.

In general, the actual level of accuracy achieved with HDS should depend on the configuration of the chords, which in turn would depend on the data configurations in the table. For this reason, estimating the accuracy with which basic statistical analysis can be performed using HDS is equivalent to understanding how overall pitch is perceived in dense, complex chords, and in particular to understand how overall pitch movement is perceived in a series of such chords.

Experiment	**Average accuracy of the worst-performing participant (%)**	**Average accuracy of the best-performing participant (%)**	**Average accuracy in the experiment (%)**
Study-I (Chapter 5) --- Select 1 chord from collections of 7 or 24	83.3%	100%	88.9%
Study-II (Chapter 6) --- Select 1 chord from collections of 4, 7, 24 or 31	59 % (accounting for slip-offs[7]: 68%)	100%	86%
Study-III (Chapter 7) --- Select 5 chords from collections of 24, with EMA	70%	93%	82.7%

Table 6. Reliability of using HDS as a way of estimating relative arithmetic means of sets of numbers, according to results from the three previous studies with TableVis.

The nature of the problem is therefore psychoacoustic: if, on presenting two dense musical chords, a group of users was asked to decide which chord they perceived to have a higher overall pitch (POP, a single pitch that represented the whole sound of the chord), would each user make this choice consistently? If so, under which conditions would such consistency in selecting the highest POP be lost? It would seem reasonable to expect that if most frequencies in one chord were higher than the frequencies in the other, the corresponding POP values would rank in the same way. If that was true, there would be a correlation between the POP from each chord and some measure of average of the frequencies constituting those chords, which could be numerically calculated (for example, the arithmetic mean).

Thus, if a listener can rank the POP of multiple chords, then from Table 6 we can observe that such sorting has a high probability to also be the correct sorting by arithmetic mean of the underlying data sets. The stages in the process from the tabular raw data set to the performance of the statistical analysis using HDS are represented in Figure 46. As shown there, estimating relative values of arithmetic means by listening to chords generated using HDS involves as a prerequisite that relative overall pitch movement is perceived consistently (*i.e.*, non-ambiguously taking place in a particular direction). It is only then that it can be observed if the POP movement is a correct or incorrect

[7] This participant registered many cases of "slip-off" in which, having found the correct answer, the details in speech were collected from a neighbouring row or column, providing an incorrect answer in the task. The second value in this particular cell assumes that those cases were correct.

prediction of relative arithmetic mean values. It should be highlighted than in none of the three stages shown in Figure 46 does the user have to perform any numerical calculations.

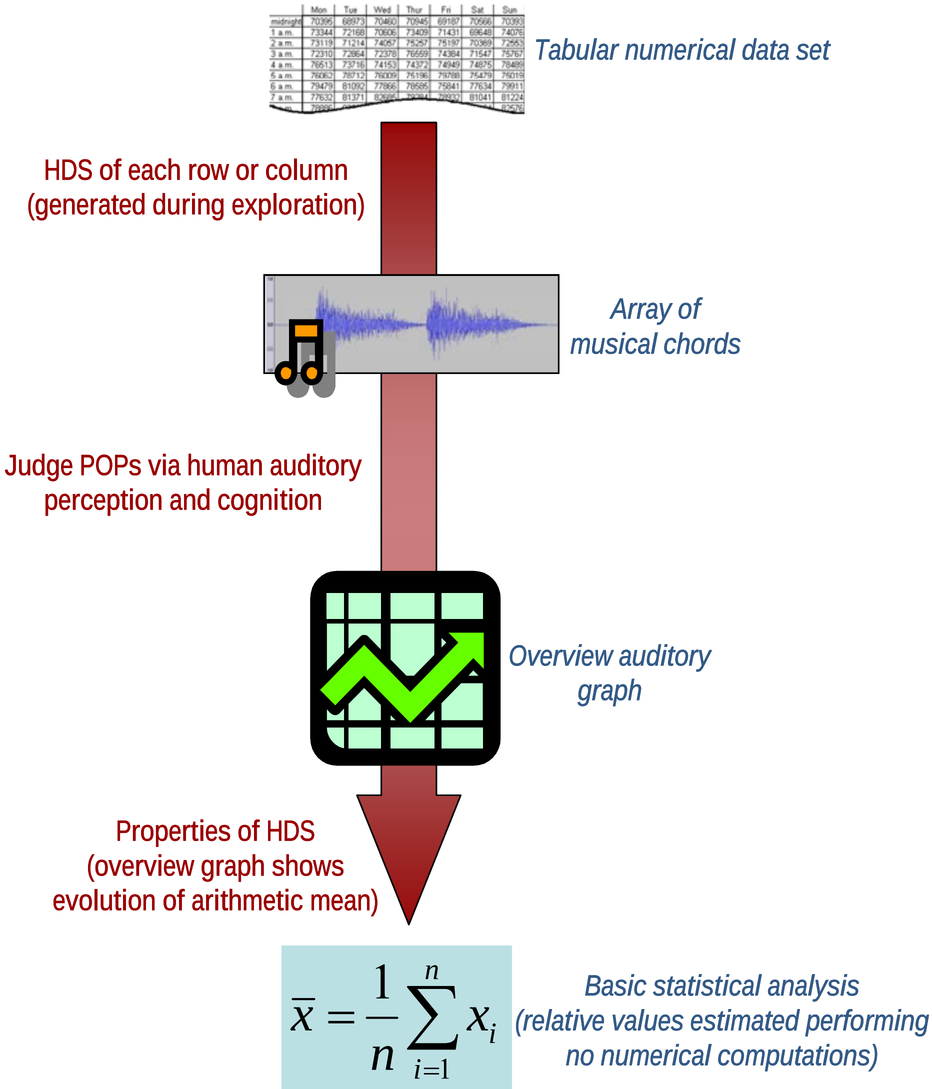


Figure 46. Stages in the process of performing basic statistical analysis from tabular numerical data, using HDS. Perceptual information is mentally processed only in the second step, and at no point do numerical calculations need to be performed.

The research work described in this chapter studied perceptual and cognitive aspects of this data analysis technique, aiming at understanding its limitations and at predicting its reliability based on the configurations of the data sets being explored. In this way, by establishing the level of performance that can be attained with HDS, this work contributes to answering the second research question in this thesis, “*What level of performance can be achieved with the techniques proposed in this thesis?*”.Three consecutive experimental studies, designed in an iterative process, are described and discussed. Conclusions and guidelines about the use of HDS for statistical data analysis are drawn at the end of the chapter.

8.2 First experimental Study

The aim of the work reported in this chapter was to quantify the accuracy with which relative arithmetic means can be estimated by judging the POP of chords generated using HDS, and to understand which factors influence the accuracy that can be attained with this technique. The underlying problem studied to answer this is psychoacoustic: the perception of very dense and complex musical chords, and cognitive aspects that lead to selecting a single pitch as a representation of a whole chord in relation to other similar chords.

8.2.1 Design of the study

The study was designed and conducted in such a way that the actual problem of comparing and extracting information from chords generated using HDS (the process described in Figure 46) was isolated from the rest of the stages associated with exploring tabular data with TableVis (all the steps described in Figure 14). This was achieved by devising a simple, straight-forward task that participants would perform repeatedly: listening and comparing two chords, and judging the relative differences between their POPs. Pairs were presented sequentially, in a random order. Participants were instructed to listen to each pair of chords and identify the one with the highest POP. The POP of a chord was described in the instructions as "*(···) a single pitch that represents the whole sound.*" It was further explained that "y*ou should (···) trust your first impression. (···) If you really find it difficult to make a decision, you can try to sing those two sounds to yourself and see which one you are singing with a higher pitch*". In other words, the task required that the polyphonic sequence of sounds was converted into monophonic, and that the sound event with the highest resulting single pitch was identified.

The equipment employed consisted of a set of high-quality headphones and three large push-button switches arranged in a line and broadly spaced out on a desk. The middle button was yellow and textured, and its function was to request that a pair of chords was played. The buttons on the left and on the right were green and with no surface texture, and they were used to provide the answer for the task. To select the first sound of the pair as the sound containing the highest POP, the participants had to press the button on the left. Similarly, the button on the right was pressed to select the second chord in the sequence as having the highest POP. The buttons were colour-coded to provide visual cues to visually impaired participants, and the texture on the yellow button was intended mainly for blind users, to differentiate it more easily from the other two. By pressing the centre button, participants could request to listen to each pair of stimuli up to four times, before selecting one of the chords. Synthesised speech was used to confirm the answer given by the participant ("first/second sound selected"), and when the participant tried to listen to the same pair more than four times ("please, give your answer now").

At the beginning of each session, demographic data about the participant was collected, including information about their musical training and about any hearing impairments that they were aware of suffering. On completion of the study, participants were given the opportunity to comment on it, on the difficulty of the task and on specific strategies that they had developed.

8.2.2 Stimuli

As a starting point, chord configurations from typical data tables used to evaluate TableVis in the previous studies were chosen as stimuli in this experiment, so as to be able to compare the results from the experiment with the data compiled in Table 6. The average smallest difference between any two chords generated exploring one such typical data table was taken as a reference for the difference between the chords in the pairs constructed for the experiment. The differences between the chords of any pair in the experiment were designed to be the same or smaller than this reference. This would illustrate the most challenging conditions under which participants had estimated differences using HDS (which had led to the performance levels summarised in Table 6). To construct the collection of stimuli for the study, the characteristics of a typical chord (a *standard* chord) were defined and paired with a variety of other chords that were carefully constructed, in order to learn about which features in the chords' configurations led to correct estimations and which did not (details about the configurations of chords used in the study can be found in Appendix G).

The standard chord was defined to have the characteristics listed below. With these characteristics, a series of standard chords was constructed, all of them complying with this description but consisting of slightly different frequency tones. The reason not to always use the identical same standard chord was to avoid any possible cognitive effect that might arise from presenting the same chord many times over the experiment. A piano MIDI instrument was used to produce the sounds.

The standard chord consisted of:

- *Number of tones in the chord:* 24;
- *Precision of the mapping:* maximum. This means that each tone was played with the exact fundamental frequency resulting from the value-to-pitch mapping algorithm (see Equation 1) without rounding to fundamental frequencies corresponding to any musical scale. This was different from the approach taken in the previous studies described, in which the 12-TET scale had been used. The decision to use maximum mapping precision in this case was taken so as to be able to have maximum control over the sounds produced with the data sets that were constructed, without introducing any discrepancies between the arithmetic means in the data sets and the arithmetic means that would correspond to the sounds once they were rounded;
- *Range of frequencies:* 257-633Hz. This range, covering 15.4 semitones, is 24% of the total frequency space used in TableVis (65 semitones). It was an average typical range size for a sonified row or column in previous evaluation studies, centred within the whole frequency space used by TableVis;
- *Distribution:* normal distribution. Mean=402Hz, Standard deviation=1/3 of the total frequency range of the chord;

All the instances of the standard chord were paired with chords such that each pair would belong to one of 8 families of pairs, each family having certain defined characteristics regarding the movement of voices[8] between the first and the second chord in the pair. The names of the families (TU, TD, BU, BD, TUBd, TDBu, TuBD, TdBU) are descriptive of the configuration variations between both chords, and they can be interpreted as follows: T and B refer to the topmost or bottommost voices respectively. The letter that follows (U or D) indicates the direction in which that extreme voice moved (up or down, respectively). Uppercase (U, D) indicates larger shifts in the fundamental frequency, and lowercase (u, d) correspond to smaller shifts (the actual magnitude of these shifts is defined below). For instance, TU ("Top-UP") means that the top voice moved upwards (while the bottom voice, which is not mentioned, remained unchanged). Similarly, TuBD ("Top-up, Bottom DOWN") means that the top voice moved upwards and the bottom voice downwards, in such a way that the size of the top-voice frequency shift was smaller (lowercase "u") than the shift of the bottom voice (uppercase "D"). In all the cases, all the intermediate voices moved proportionally to their position in the chord, creating a gradual change in voice shifts between the extreme voices.

These 8 families are represented graphically in Figure 47, where fundamental frequencies are mapped to the Y axis (the scaling in the Y axis shows the 12-TET, or 12-tone equally-tempered scale). Each oval shape in this figure represents the range in the frequency space covered by a chord. Below each family, the shift in the arithmetic mean of the underlying data is indicated as the percentage of the total range of values that could be presented using the full frequency space in TableVis (the range between the topmost and bottommost dotted lines shown in the figure).

For illustration and better understanding of the shapes shown in Figure 47, an example of one pair of chords picked from the TdBU family is shown in detail to the right. Each chord is represented side by side with the 24 fundamental frequencies that make up each one mapped to the 12-TET scale on the vertical axis. Notice that the fundamental frequencies are not forced to lie on any exact value corresponding to the 12-TET scale. It can also be observed that the second chord is a shrunk version of the first one, where all voices move proportionally. The movement of the bottom voice is dominant in this case, for having a larger shift, and more voices move upwards than downwards. In the underlying raw data this results in the arithmetic mean shifting upwards by 1.4% of the total numerical range.

For each of the families in Figure 47, 5 pairs of chords were constructed. All 5 pairs were different but with the same movement characteristics (40 pairs in total). The symmetric cases (TU and TD, BD and BU, etc.) were constructed by simply inverting the positions of the chords in each pair (*i.e.*, the pairs in symmetric families were identical but in inverse order). The order in which pairs were presented was randomised for each participant.

---

[8] The term "voice" is used in the study of harmony and voice leading to mean a single melodic line. [4]

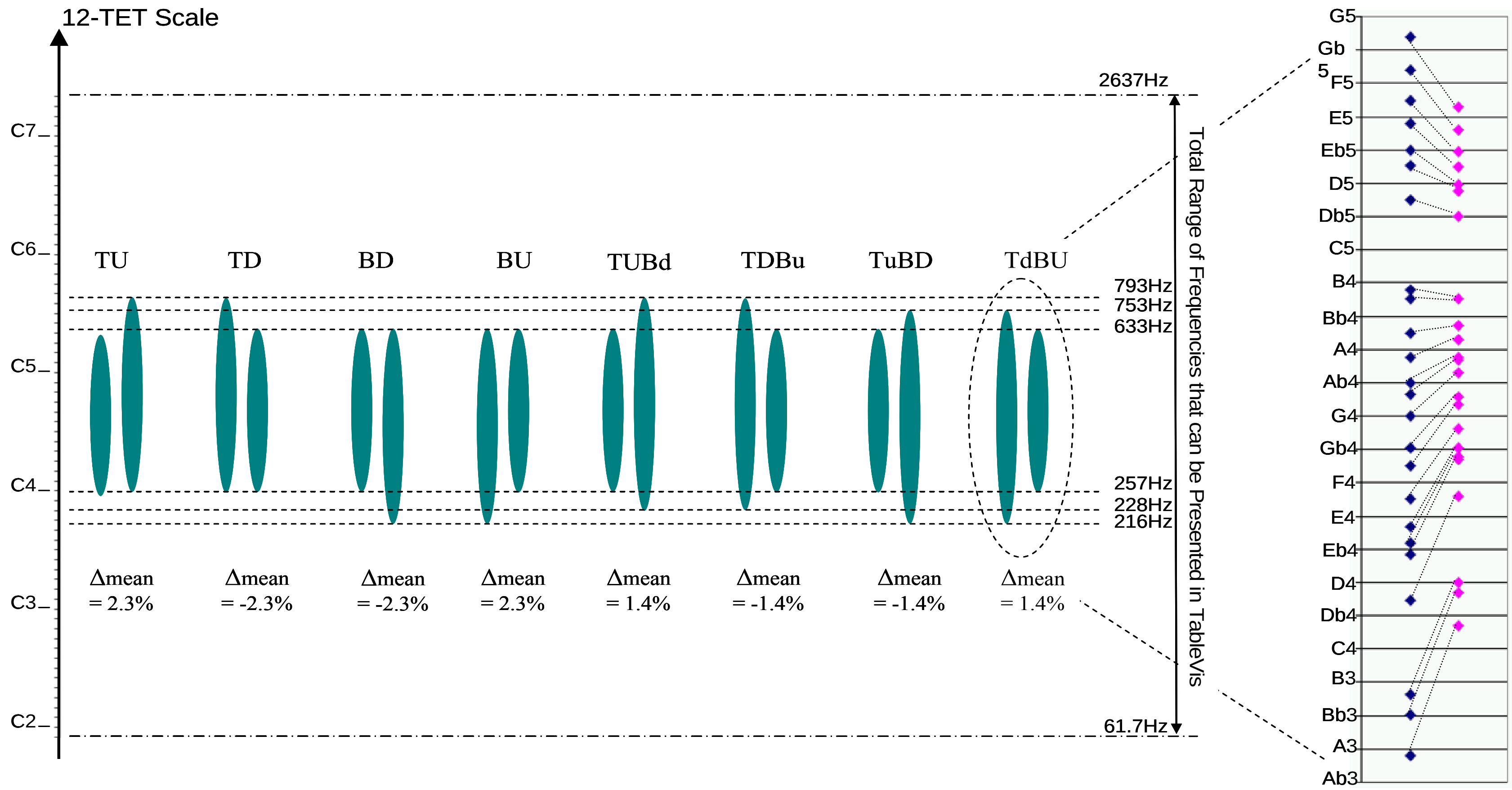


Figure 47. Graphic representation of the families of pairs of chords used as stimuli in the first experiment. Each oval shape, mapped on the Y axis to a fundamental frequency in the 12-TET (12-tone, equally-tempered) scale, represents the relative range in the pitch space covered by a chord. The labels over each pair are the coded name of the family. Below each pair, the shift in the arithmetic mean of the underlying data is indicated, as the percentage of the total range of values that could be presented using the full frequency space in TableVis (range between the topmost and bottommost dotted line in the figure). On the right, a detailed view of the actual structure of the chord in one of the pairs used in the experiment is shown. This example belongs to the TdUB family of pairs.

The complete set of stimuli was presented twice to each participant, in two separate blocks with a pause in between. In each block, all the pairs were presented in a different random order, and participants were not made aware that it was the same set of pairs twice. Presenting each pair of stimuli twice was used to provide a measure of the consistency of the answers.

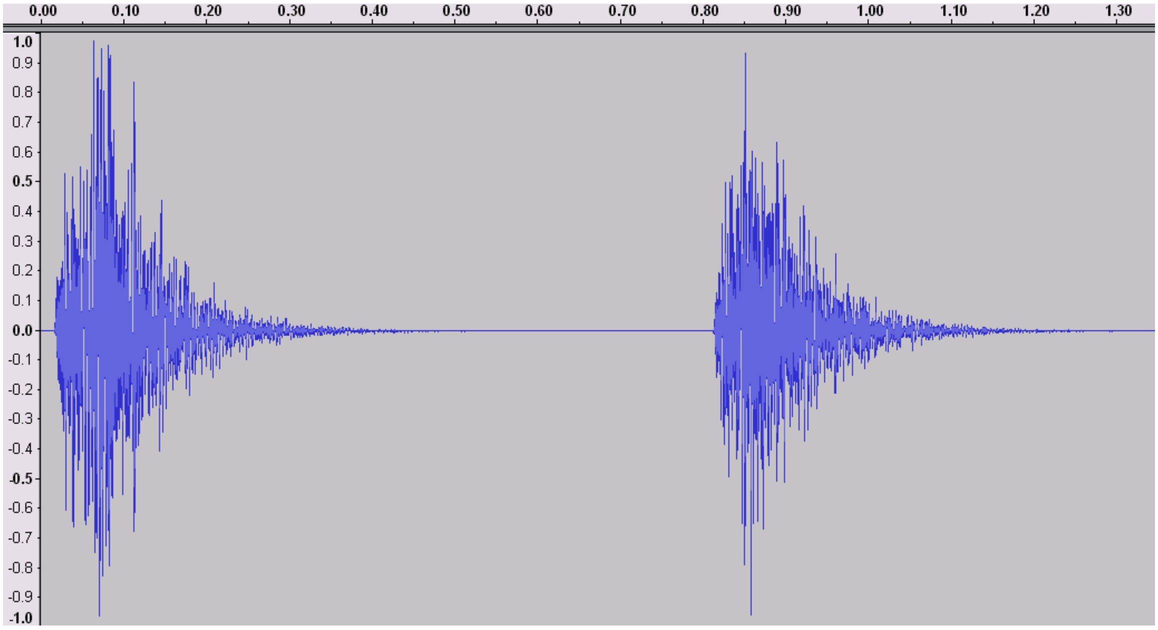


Figure 48. Waveform of a pair of chords in the first experiment test (horizontal axis: time in seconds; vertical axis: normalised amplitude). The time gap between the onset of both chords was 800ms.

Each chord was played at the maximum arpeggio speed, using the default method in which chords were generated with TableVis in the previous studies. The piano MIDI instrument was again used to play the chord. 800ms elapsed between the onsets of the two chords in a pair. Most of this interval was silent (from the extinguishing of the first chord to the onset of the second). The waveform of one such pair of chords is depicted in Figure 48.

8.2.3 Participants

A group of 7 participants was recruited to take part in this first experimental study, of which 5 were blind (3 congenitally blind) and the other 2 were partially sighted. 5 participants were aged between 16 and 25, one was in the 36 to 45 years old age range, and the remaining participant was older than 45. All but one participant had at least some musical training, and three of them were active amateur musicians. None of the participants reported any hearing impairments.

Two of the participants did not complete the second round of pairs of stimuli, where all the pairs were again presented in a different order. This was due to time restrictions, after the extended training sessions required to reach an appropriate level of understanding of the task.

8.2.4 Results

The correctness of the task was measured by observing if the chord that was selected from each pair as having the highest POP was also the chord that had originated from the set of numbers with the highest arithmetic mean. Figure 49 shows the average results for the seven participants in the study for each family of stimuli pairs.

Among the qualitative information collected from informal feedback, it was noted that all participants found some of the pairs harder to rank by POP than others, but it could not be determined from this which those difficult pairs were or whether they belonged to certain specific families. In the more difficult cases, the common strategy followed by the participants was to listen to the stimuli again, expecting the difference in pitch to become more obvious, which reportedly did normally happen. Figure 50 shows, for each family of stimuli, the average number of times users listened to the same pairs of chords before selecting an answer.

8.2.5 Discussion

The results summarised in Figure 49 show that, in the case of most families, the chords identified as having the highest POP were coincident with the chords originated from the data sets with the highest arithmetic means. The average percentage of such correct identifications was above 80% for half of the families, and above 70% for all but one of the families.

A few observations can be made regarding these preliminary results. Firstly, the two families with the lowest percentage of correct arithmetic mean estimations (TdBU and TuBD) have equivalent symmetrical configurations, in which both extreme voices move and the movement of the bottom voice is dominant (the difference between their correctness scores is, however, considerable). In the other six cases, either only one extreme voice moved or it was the movement of the top voice that was dominant. It appears that when only one of the extreme voices moved, the user's attention was drawn to that area of the frequency spectrum, the direction of the movement being identified quite consistently. It might also have helped with the identification of the movement in these cases with unidirectional movement that the shift in arithmetic mean was of the larger type, as seen in Figure 47. It is, however, interesting to observe that in the case of TUBd and TDBu, in which voices moved in two opposite directions and where the shift in the arithmetic mean was of a smaller size, the accuracy in users' estimations was in the same range as for the pairs with unidirectional movement. This difference between both types of bidirectional pairs (TdBU and TuBD with lower correctness scores than TUBd and TDBu) might be due to the movements in the top voice standing out more than the movements in the bottom voice. This may be caused by the way in which the task question was formulated ("which chord has the highest pitch"), which directed the participants' attention to that part of the frequency spectrum.

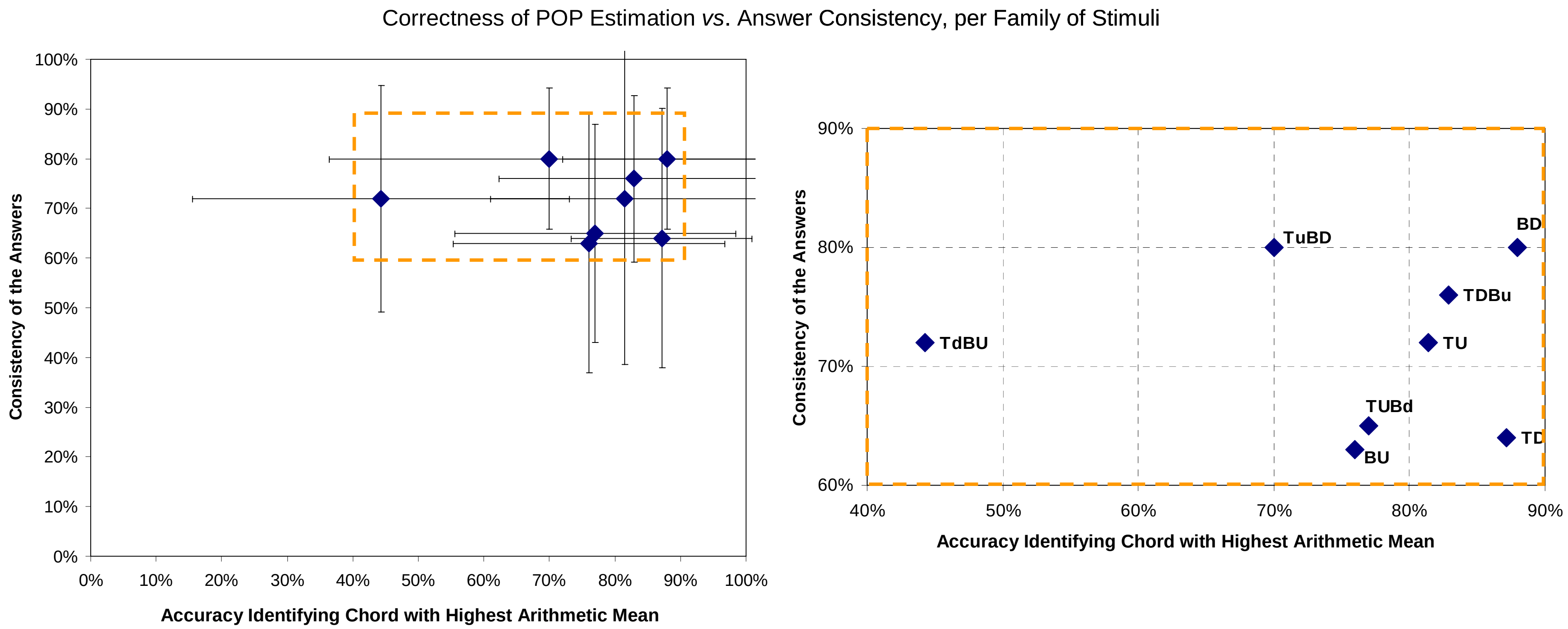


Figure 49. Two views of the same scatter plot at different zooming levels, summarising the results from the first experimental study for each family of pairs of chords. *X axis:* average performance achieved identifying the data set with the highest arithmetic mean by identifying the chord with the highest POP. *Y axis:* for the five participants that completed both blocks in the study, average percentage of the stimuli that were rated consistently both times they were presented to each participant.

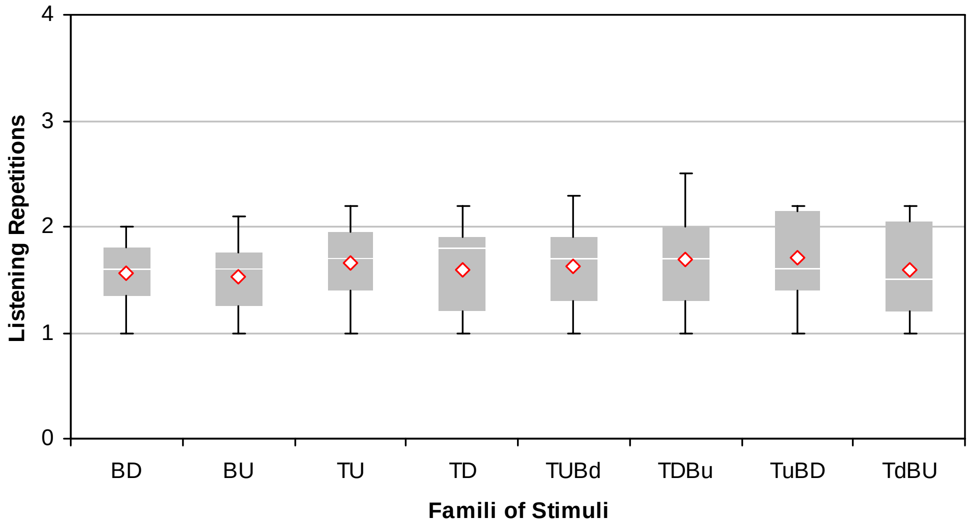


Figure 50. Average number of times stimuli from each family were accessed before an answer was selected**.**

The information related to the consistency of the answers (gathered from the results of the 5 participants that completed both blocks of stimuli) shows that for all the families the average values of consistency were between 64 and 80%. These values are an indication of the confidence with which the answers were given, and therefore an initial measure of the ambiguity perceived by the participants to discriminate POPs in a pair. Interestingly, the pairs from the bidirectional families (with movements in both directions and smaller mean shift) were not perceived to be the most ambiguous ones, and in the case of one family (TuBD) they were the most consistent. Furthermore, the family BUTd that obtained the lowest correctness score was in the middle of the range of consistency scores. These results point towards a lack of correlation between correctness in the estimations, the size of the arithmetic shift mean and the ambiguity with which differences in overall pitch were perceived. Another measure of the general level of ambiguity could be found from the results in Figure 50. While participants claimed that some pairs of stimuli were much easier than others to differentiate by POP, the graph in this figure shows very small differences in the number of times stimuli were listened to between families. This result suggests that there were easy and difficult tasks within every family of stimuli, denoting that the perceptual and cognitive mechanisms that govern the perception and cognition of POP in these types of chords are subtle and complex, with small differences between pairs in the same family having relevance with respect to the subjective level of ambiguity.

Having generated a set of initial results in this first experimental study, highlighting a number of possible perceptual and cognitive patterns, it was necessary to conduct a full experimental study that could provide more accurate data. This experiment is reported in the next section.

8.3 Second experimental study

In order to confirm the patterns that began to emerge from the results of the first experimental study, a larger second experiment was conducted. Limited access to visually impaired participants for a larger study meant that sighted participants were to be recruited for the second experiment. This gave the opportunity to make a comparison between populations in the common task assigned in both experiments. The second experimental study was an expanded version of the first one, were the set of stimuli used in the first case was amended with an additional set, so that the results from both experiments could be compared on the cases that were common to both, an also information about participant's performance with new chord configurations was obtained. This experiment is described in the following subsections.

8.3.1 The experiment

As opposed to the design of the first experimental study, in which there was a single experimental condition, the full experiment consisted of two conditions which differed in the way the experiment task was formulated: in one condition the participants were asked to find the chord in the pair with the *highest* POP (the same as in the first experimental study). In the other condition, participants were asked to find the chord in the pair with the *lowest* POP. The use of these two conditions was motivated by the results in the preliminary study, which suggested that the formulation of the question could be influencing the participants' attention to focus on the higher frequencies, resulting in the movement of the top voices to be perceived as the movement of the whole. Consequently, while the task was identical in both cases, formulating the question differently should inform whether such attention focusing within the frequency spectrum was taking place and about its significance.

8.3.2 Stimuli

A total of sixteen families of pairs of chords were used as the stimuli for the experiment. The set of 8 families from the first experimental study was used again, with the addition of another 8 families of pairs, creating a richer set of cases with which to investigate the patterns identified in the first experimental study. The basic design principles for the stimuli remained the same as described in Section 8.2.2: a standard 24-tone chord fulfilling the characteristics of the same predefined normal distribution as in the first experiment, and a chord that paired with the standard chord, following a particular rule of fundamental frequency shifts. Voice movements involved either shifting only one extreme voice (unidirectional stimuli) or shifting both (bidirectional stimuli), in which case the shift of one of the voices was dominant (larger, determining the direction in which the arithmetic mean would shift). In all the cases, intermediate voices moved proportionally to their position between the extreme voices (as illustrated in the example on the right side of Figure 47). The new additional 8 families were introduced so that in each case (unidirectional and bidirectional) there were pairs with shifts of two different sizes in the arithmetic means of the underlying data sets: 1.4% and 2.3% of the total range of numerical values that could be represented using the full pitch space in TableVis (in the first experimental study, the unidirectional and bidirectional stimuli corresponded to mean shifts of only 2.3% and 1.4% respectively). The same family naming convention as for the first experiment was used in this case, using the following convention for 3 different sizes of peripheral voice movement. From smallest to largest: u, U and Uu for movements upwards, and d, D and Dd for movements downwards. This naming convention can be better understood observing the resulting 16 patterns or families represented graphically in Figure 51. Every two adjacent families in the figure exhibit a symmetrical pattern. The pairs of chords in such symmetrical families were generated by inverting the order of presentation of the chords as was done in the first experimental study. With this design of the families, the following four configuration criteria were combined, each one with two different possible states, giving origin to the 16 different families. These criteria and their possible states were:

I. *Number of directions in which voices moved:*

- *a.* *Unidirectional.* All the voices stayed at the same frequency or moved in the same direction;
- *b.* *Bidirectional.* Some voices moved upwards and others downwards;

II. *Extreme voice with dominant movement:*

- *a.* *Top voice dominant.* The movement in the top voice determined the direction of the shift in the arithmetic mean of the underlying data;
- *b.* *Bottom voice dominant.* The movement in the bottom voice determined the direction of the shift in the arithmetic mean of the underlying data;

III. *Direction of shift in arithmetic mean (the same as the direction of the voice with dominant movement):*

- *a.* *Upwards.* The arithmetic mean in the underlying data shifted upwards;
- *b.* *Downwards.* arithmetic mean in the underlying data shifted downwards;

IV. *Size of arithmetic mean shift:*

- *a.* *Small.* The arithmetic mean shifted by 1.4% of the total range of numerical values that could be displayed (approximately 0.9 semitones in the 12-TET scale);
- *b.* *Large.* The arithmetic mean shifts by 2.3% of the total range of numerical values that could be displayed (just under 1.5 semitones in the 12-TET scale);

Each family consisted of 5 pairs of chords (80 pairs in total), and they were presented in a different random order to each participant. Unlike in the first experimental study, the complete set of stimuli was presented only once to each participant, to keep the experiment time bounded. All the rest of the characteristics of the stimuli were as described for the first experimental study.

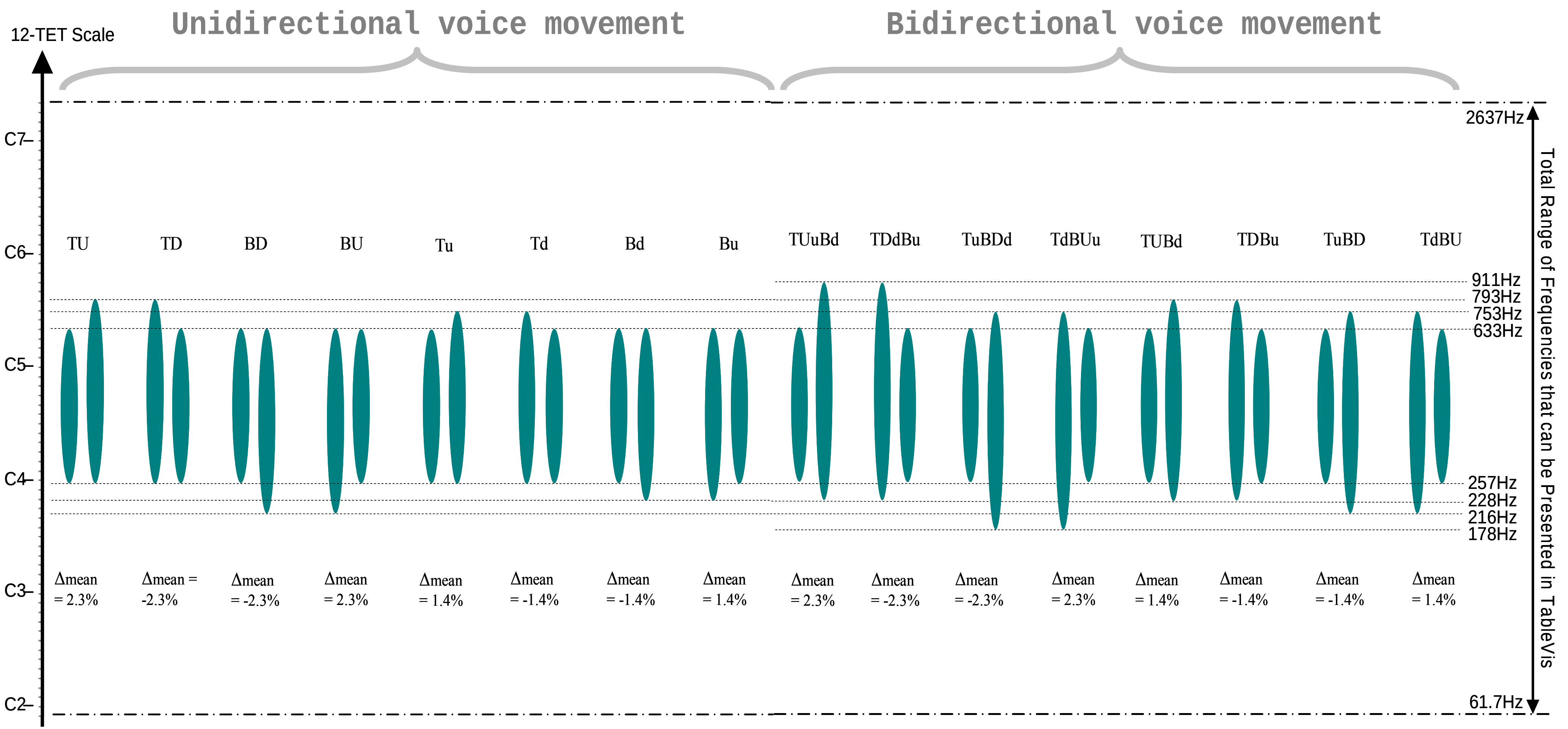


Figure 51. Graphic representation of the families of pairs of chords used as stimuli in the second experiment. Each oval shape, mapped on the Y axis to a fundamental frequency in the 12-TET (12-tone, equally-tempered) scale, represents the relative range in the pitch space covered by a chord. The labels over each pair are the coded name of the family. Below each pair, the shift in the arithmetic mean of the underlying data is indicated, as the percentage of the total range of values that could be presented using the full frequency space in TableVis (range between the topmost and bottommost dotted line in the figure).

8.3.3 Participants

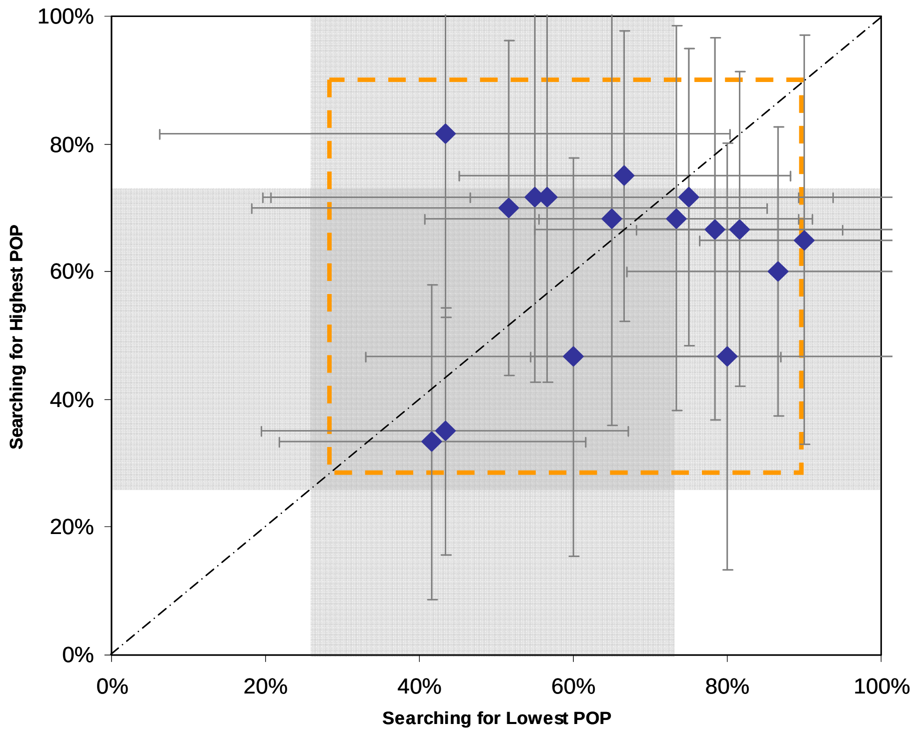
A group of 24 sighted participants was recruited to take part in this first experimental study. 12 of them completed the experiment trying to identify the chord with the highest POP in every pair, and the other 12 looking for the chords with the lowest POP. Unlike in the studies reported in previous chapters, the sighted participants in this case did not wear a blindfold, since there was no spatial navigation involved in the experimental task, unlike during the evaluations of TableVis.

Participant's ages covered the following ranges range: 15-25 years old (12 participants), 26-35 years old (10 participants) and 36-45 (2 participants). 7 were female and 17 male. 7 of the participants were university undergraduate students, 14 were postgraduate students, and the remaining 3 participants were professionals with university degrees. Regarding their musical training, 7 of the participants had received a considerable amount of musical training (several years of formal music theory and could play well at least one instrument), another 7 had some musical training (limited amount of training in music theory and some musical instrument playing), and the remaining 10 participants did not have any formal training or played any instrument, but they all claimed to be able to carry a tune. None of the participants reported any hearing problems.

8.3.4 Results

The results from both conditions in this experiment are summarised in Figure 52. This figure shows a scatter plot of the average scores obtained by the participants in both conditions for each family of stimuli (participants searching for the chord with the lowest and the highest POP on the X and Y axes respectively). Two views of the same graph are shown in the figure, at different zoom levels. Notably, most plots are located in the top-right quadrant of the graph, which corresponds to high scores of correctness in the answers (where judgements on the relative values of the POPs in the pair are correct estimations of the relative arithmetic means in the underlying data). Each condition had a binomial distribution B(12,0.5) and, with 12 participants, significant results were obtained for correctness scores of approximately 75% and above (for a significance level of 95%). This means that for the deviation from the random occurrence of either possible answer to be predictable in 95% of the cases, the mean had to be bigger than 75% or smaller than 25%. There was therefore a band covering 50% of the scale around the centre, in which scores were not significant at the significance level selected. The areas of non-significance for these distributions are shown as shaded rectangles on the left view of Figure 52. It is observed that the plots of half of the families showed significant scores for one of the conditions. Differences between participants in both conditions are revealed in the graph by the distance of each plot to the diagonal line that crosses the graph. The plots above the diagonal line are families in which participants searching for the highest POP obtained a higher average percentage of correct answers than participants searching for the lowest POP. The opposite was true for the plots that are below the diagonal line. The greater the perpendicular distance to the diagonal line, the bigger this difference was.

Average Correctness of Arithmetic Mean Estimation



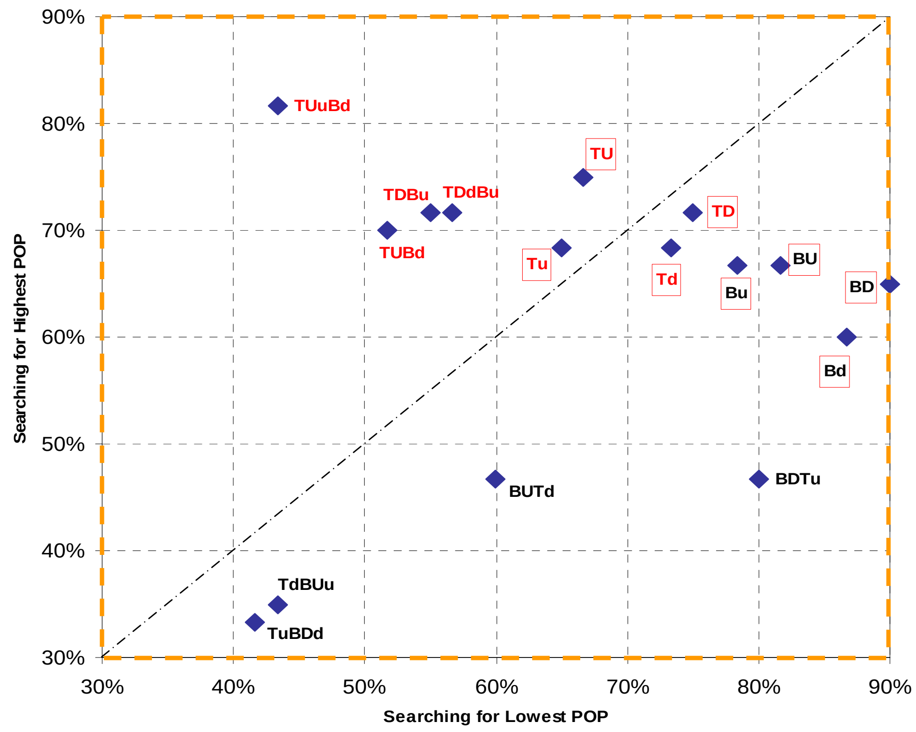


Figure 52. Two views of the same scatter plot at different zooming levels, showing the average correctness of the answers to the experimental task, grouped by family of pairs of chords. The X coordinate of each dot is the percentage of correct answers in pairs of that family, scored by participants that were searching for the chord with the lowest POP. The Y axis shows equivalent information from the participants that were searching for the chord with the highest POP. The view on the left shows error bars representing standard deviations. The two shaded rectangles are the areas of non-significant results in each condition (for each condition, binomial distributions B(12,0.5); significance level = 95%). The view on the right (a magnification of the dotted square on the left view) labels each plot with the name of the family it represents. (Label coding --- *Red font*: top-voice movement dominant. *Black font:* bottom-voice movement dominant. *Framed label:* pairs with unidirectional movement. *Unframed label:* pairs with bidirectional movement).

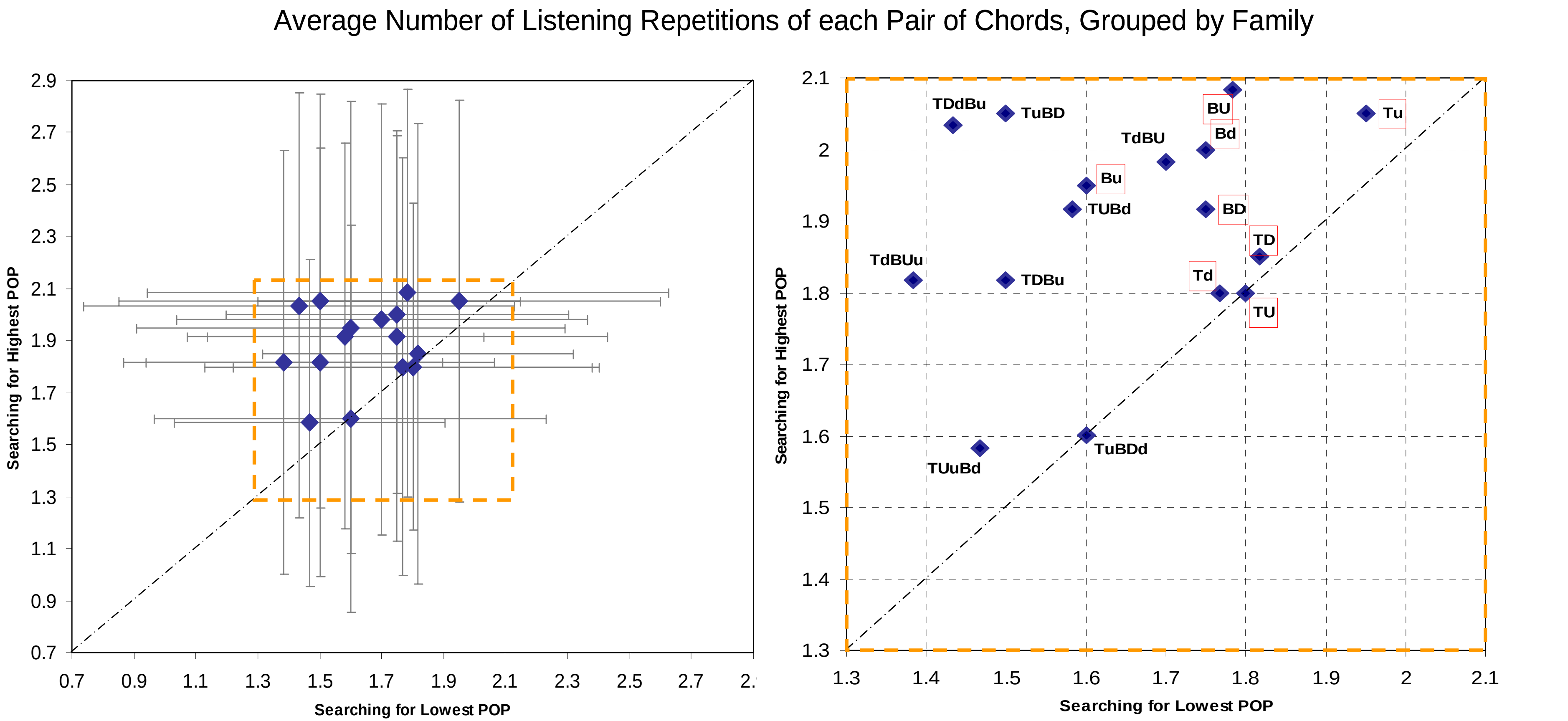


Figure 53. Two views of the same scatter plot at different zooming levels, showing the average number of times participants listened to each pair of chords before selecting an answer. Each plot represents a different family of stimuli. The X coordinate of each dot is the average number of times that pairs of chords was listened by the participants that were searching for the lowest POP. The Y axis shows equivalent information from the participants that were searching for the highest POP. The view on the left shows error bars representing standard deviations. The view on the right (a magnification of the dotted square on the left view) labels each plot with the name of the family it represents. (Label coding --- *Framed label:* pairs with unidirectional movement. *Unframed label:* pairs with bidirectional movement).

The possible effect of the different characteristics of the chord configurations (the four different categories listed in (8.3.2, I to IV) was checked by performing two-factor mixed-design ANOVAs, pairing in each case on of the chord configuration factors (within groups) with the main independent variable, *i.e.* the task of searching for the chords with the highest or the lowest POP (between groups). As expected from the result of the t-test reported above, the differences between groups of users performing the different tasks were again not significant in every ANOVA. Some of the chord configuration criteria were, however, found to affect the scores of correctness:

- *Number of directions in which the voices moved (unidirectional vs. bidirectional).* The scores of correctness obtained by the participants in both conditions for unidirectional pairs of chords were significantly higher than the scores for bidirectional pairs, $F(1,14)=172.983$, $p<0.01$. This strong effect can be observed in Figure 52 (right), where the unidirectional families are shown with their family-name labels framed.
- *Extreme voice with dominant movement (top vs. bottom voice).* The scores of correctness obtained for families in which the movement of the top voice was dominant were significantly higher for the users that were searching for the chords with the highest POP, $F(1,14)=27.807$, $p=0.004$, although it did not affect the scores of the combined population of participants from both conditions, $F(1,14)=3.547$, $p=0.277$. This effect over the participants of one of the conditions can be observed in Figure 52 (right), where the labels of the families with the top voice dominant are printed with red font.
- *Direction of shift in arithmetic mean, or direction of the voice with dominant movement (upwards vs. downwards).* This stimuli configuration criterion did not have any significant effect on the correctness scores obtained by the participants in any of the conditions.
- *Size of arithmetic mean shift (smaller or larger).* Like above, this criterion was not found to have any significant effect on the correctness scores of the participants in any of the conditions.

The comments collected from the participants at the end of the experiment were similar to the comments provided by the participants in the first experimental study: great variability in the ease with which the answer could be chosen was reported. While the answer in some cases was immediately obvious, other pairs were perceived to be much more ambiguous. A frequent strategy used to disambiguate was to listen repeatedly to the stimuli. While some participants reportedly tried to “clear their ears” before every new repeated listening, others tried to rehearse and reflect on the chords just heard before each repetition. The correctness results from the first experimental study (X axis in Figure 49) can be directly compared with the results from this experiment that relate to the 8 families of stimuli common in both studies. This comparison can only be performed considering just the participants from the experiment that searched for the highest POP (which was the task in the first experimental study). Both sets of results (plotted in Figure

54) were not found to be significantly different according to a t-test (two-sample, between-groups, repeated-measures, assuming equal variances): $t_{14}=1.877$; p=0.082.

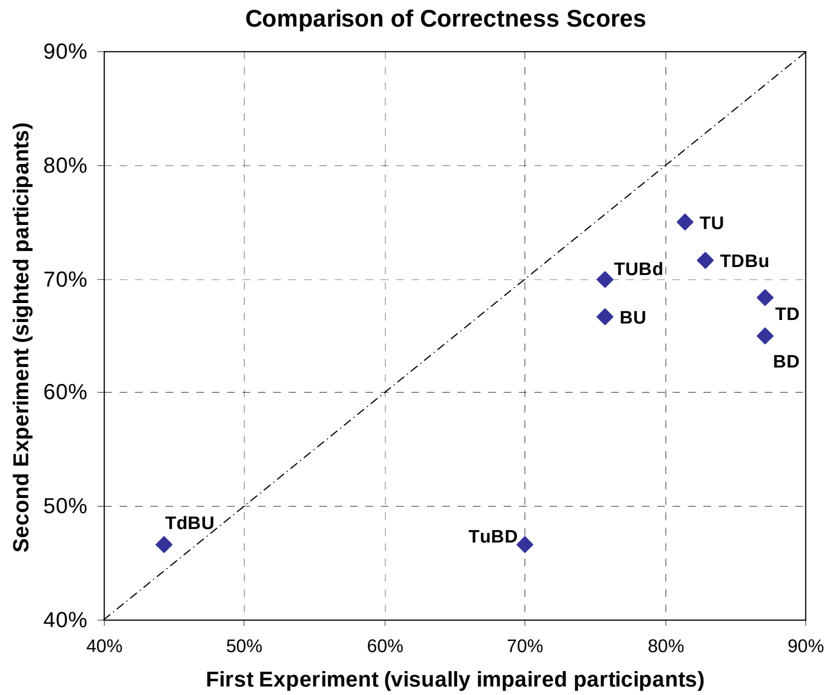


Figure 54. Comparison between the average correctness scores obtained in the first experimental study (visually impaired participants) and in the experiment (sighted participants), with the families of stimuli common to both cases and by the participants performing the same task (find the chord with the highest POP).

8.3.5 Discussion

The results from this experiment provided further confirmation that under a variety of conditions HDS can be used to estimate relative values of arithmetic means by judging relative POPs of complex chords. For most families of data configurations used in the experiment, participants obtained average correctness scores of over 70% in at least one of the conditions (nearing the results observed in previous evaluations of TableVis, summarised in Table 6). However, since the size of the population in each condition was still rather small (12 participants), the required average correctness scores to reach levels of confidence of 95% in each binomial distribution were of 75% or over, which only half of the families obtained.

Significant differences and interactions between the two experimental conditions and certain configuration characteristics in the stimuli were revealed in the results. The following points summarise the most notable of these findings:

- The way in which the question was formulated directed the attention of the participants towards a specific region of the frequency spectrum, influencing their perception of the overall pitch that led to significant differences in correctness scores in the case of some of the families.
- Participants searching for the highest POP performed significantly better with stimuli from the families in which the movement of the top voice was dominant. An equivalent effect, however, was not found for the participants that searched for the lowest POP.
- Participants searching for the highest POP listened to each pair of chords a significantly larger number of times. It is known from the feedback obtained that participants performed repeated listening of the stimuli to disambiguate the least clear cases. It therefore seems that when attending to the highest frequency voices, the level of ambiguity was higher.
- The conscious process of disambiguation by attentive repeated listening and reflection seemed to reinforce the bias introduced by the task towards one of the extreme voices. For that reason, since participants searching for the highest POP listened to the stimuli a significantly larger number of times, their performance with pairs in which the movement of the top voice was dominant was significantly higher.

The phenomenon of selective attention in listening to complex auditory events can help understand some of this behaviour. In the literature, Moore [105] refers to this phenomenon in the following terms: "*It seems that we are not generally capable of attending to every aspect of the auditory input (···); rather, certain parts are selected for conscious analysis.*" He goes on to add: "*In principle, we might think it possible to attend to and compare any arbitrary small group of elements in a complex acoustic signal. However, this does not appear to be the case. Rather, it appears that the complex sound is analyzed into streams, and we attend primarily to one stream at a time. This attended stream stands out perceptually, while the rest of the sound is less prominent*" (pp. 294-5). As hypothesised in the previous section, it is likely that the phrasing of the question in this experimental study, pointing towards the highest or the lowest pitch, tended to direct the participants' attention towards that range of frequencies, contributing to the formation of different streams in both conditions and hence the significant differences observed. Furthermore, the more thorough this conscious analysis was (indicated here by participants listening attentively a greater number of times), the stronger the stream segregation was observed to be.

Other results obtained from the experiment may seem contradictory at first. Participants from both conditions obtained significantly higher scores of correctness with the unidirectional pairs of chords. This was the expected behaviour, since every voice moved in one direction. The apparent contradiction comes from the significantly higher number of times participants from both conditions listened to these stimuli, indicating that they were perceived to be more ambiguous than bidirectional pairs. In addition, there were cases in which the behaviour seemed to respond to different rules. An example is the case of the symmetrical families TdBUu and TuBDd. Participants from both conditions obtained considerably lower correctness scores for these two families than for any other. In fact, these two cases could be seen as illusions in which participants perceived the movements to be predominantly in the direction opposite to the actual POP shift, and these illusions appeared to be quite strong based on the small number of times participants listened repeatedly to stimuli from these two families (see Figure 53).

A study by Allik *et al.* [5] investigating the perception of pitch motion in sequences of simultaneous tones reached some similar conclusions. Their aim was to study the perceptual audio analogy of movement perceived in dot-matrix display that scrolls information (*e.g.* the displays commonly found in trains announcing the name of next station, and other announcements in many public spaces). They did so by investigating the perception of tones from a matrix with a few predefined frequencies that were turned on and off randomly. The resulting sequences of chords in that study were different to the pairs considered in the study reported here: they contained a number of sounds that varied between successive chords (from 1 to 8 sounds in any chord), which were picked from evenly and broadly spaced out patterns, and which were presented in series of varying lengths. For the subset of results they obtained with sequences of 2 chords, the performance they reported was similar to the performance measured in the two experiments described so far in this chapter. They also reported that the peripheral motion (and mainly at the top frequencies) had a stronger influence on the perception of pitch motion. Finally, that study concluded that small variations in frequency contributed more to pitch motion than larger movements. In other words, grouping of voices between two chords into segregated streams took place more strongly between near frequencies. This finding is in accordance with the *gradualness of change* property from the theory of grouping mechanisms in auditory scene analysis [17], which states that a sequence of sounds from the same source tends to change its properties slowly. Thus, sounds from both chords that are most similar in their properties tend to be grouped together, creating segregation of streams. In the study reported in this chapter, given the high density of fundamental frequencies in the chords, it is likely that perceptual one-to-one correspondence of voices was not always preserved and that cross-voice proximity groupings took place, often leading to perception of POP movement in direction opposite to the dominant movement of the actual movement of the voices. Figure 55 shows three representative examples of proximity grouping that could commonly be found in the intermediate voices of the pairs of chords used. Each figure represents two internal voices in a pair of chords drawn on a pitch scale (other voices above and below are not represented in this figure). The arrows show the directions in which the voices move, and the oval shapes represent the groupings by proximity that take place. In (a), voices are grouped in the direction in which they move (grouped by proximity). In (b), the upper voice moves to an ambiguous position, equidistant to the initial position of both voices represented in the example. The bottom voice has the same tendency to group with either note in the second chord, and the actual grouping that takes place will depend on the movements and configuration of the rest of the voices in both chords. In (c), the bottom voice will have a greater tendency to group with the top voice than with the bottom voice, because it moves to a very near position in the pitch space, tending to segregate in an ascending stream.

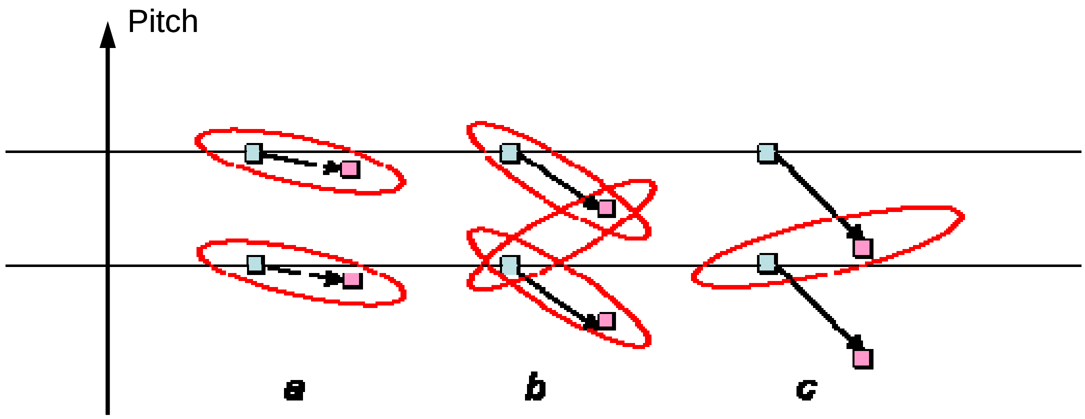


Figure 55. Examples of grouping by proximity in internal voices of chords. Arrows indicate voice movements, and oval shapes represent pairs of sounds that have a tendency to get grouped together.

Also in connection with proximity grouping in series of chords, Bregman and Tougas [18] described that pitch proximity grouping can be formed retrospectively: considering a sequence of three chords, the groupings that take place between the first two are resolved retrospectively after the third chord is heard, *i.e.* new groupings between the second and third chords help confirm the actual cases of groupings between the first and the second, in such a way that the complete series is perceived with the highest continuity. In another related study, Demany and Ramos [37] collected results that reinforced the speculation that the human auditory system contains automatic "frequency-shift detectors" which are sensitive to the direction of such shifts and that can be activated by a pair of successive sounds even when these sounds are separated by a silent delay. Gaps below one second are believed to preserve the functioning of such frequency-shift detectors, although in this reference the existence of shift detectors that work after several seconds of gap is defended. The findings in these studies illustrate the complexity and difficulty to interpret, let alone predict, the groupings that will take place in a sequence of only two chords. Yet, they also confirm the existence and importance of such mechanisms in determining how overall pitch movement is perceived. The influence of such mechanisms in HDS as a technique to perform statistical analysis of tabular data was investigated in a new experiment, described in the next section.

8.4 Third Experimental study

The results from the studies reported above revealed that several factors in the configurations of the chords influence the perception of relative overall pitches; namely the selective attention to specific areas of the frequency space and the tendency of the movement of the extreme voices (highest and lowest frequencies) to stand out from the rest. Additionally, the study showed that these factors alone could not explain all the results, and the mechanism of proximity grouping by frequency was identified as an additional factor that could explain the data as they had been observed.

To advance the understanding of the perceptual and cognitive aspects of data mining using HDS, and having verified the influence of the first two factors in POP movement (attention-driven selective listening and segregation of extreme voices into separate streams), this experiment was designed to focus on isolating the third factor (stream segregation by proximity grouping in internal voices), with the aim of testing whether this aspect alone could convey a sense of POP movement and finding out the degree of concordance of such POP movement with the actual arithmetic mean shifts.

8.4.1 Experiment design

To isolate the factor of stream segregation by proximity grouping in internal voices, the following changes were introduced in the design of this experiment, with respect to the previous one:

- The pairs of chords were constructed so that there was no frequency shift in the extreme voices. The top and the bottom voice remained at the same frequency in both chords. Thus, all the pitch movements happened in internal voices only;
- There was only one task (only one condition), which was phrased in such a way that it did not drive the attention of the user to any specific region of the pitch space, as happened in the previous experiment. The new "open" question asked participants which direction the POP moved (upwards or downwards). The question was phrased as follows:

  In which direction does the "Overall Pitch" move?

- To encourage the participants to give their answers based on their first impression, preventing them from performing conscious analysis of the auditory events, each pair of chords could only be listened to once.

During the experiment, participants interacted with the interface shown in Figure 56. Participants clicked on "**LISTEN**" to hear a pair of chords once, with no option to listen to it again. The pairs were presented in a random order. Then, if the POP was judged to be moving upwards, participants clicked on "**Pitch goes UP**". On the contrary, if overall pitch was perceived to go down, participants selected "**Pitch goes DOWN**". Having made the selection, they could proceed to listening to the next pair of chords by clicking on "**LISTEN**" again.

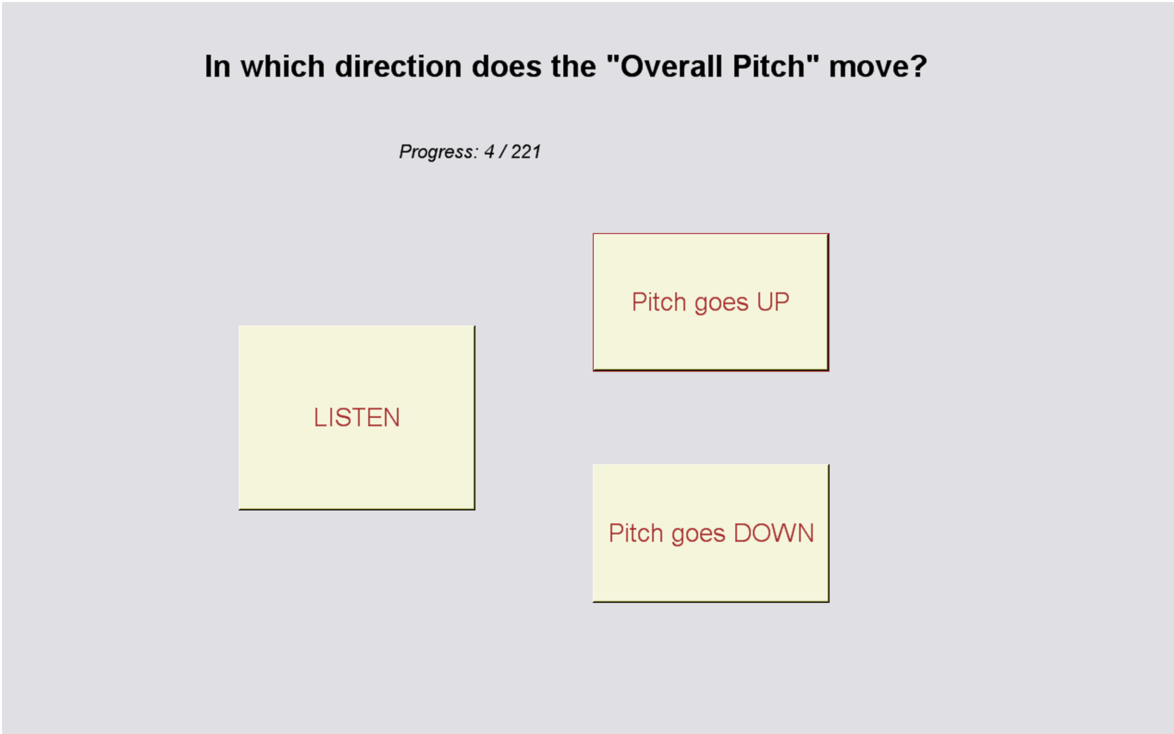


Figure 56. Screenshot of the interface used to run the third experiment

The experiment was divided into two parts. In each part, half of the pairs of sounds were presented, with a pause of several minutes in between both halves of the experiment to avoid mental exhaustion and attention drop. Pairs of stimuli were presented in random order, which was different for each participant. Before the actual experiment, participants had time to read the instructions, which included an image of the graphic user interface, and they could ask any questions until the task was clear.

8.4.2 Stimuli

Each chord in a pair was sustained for 500ms, after which all MIDI piano notes were muted and they decayed rapidly. A 200ms decay time was allowed between sending the mute message to the notes of the first chord and onset of notes from the second chord. Figure 57 shows the wave-form of a typical pair of chords in the experiment. Each chord was sustained (unlike the previous experiment and in the evaluations conducted with TableVis, where the MIDI notes were muted as soon as they were set on). The reason for doing this was that the nature of this experiment was primarily psychoacoustic, unlike the previous one that tried to reproduce scenarios encountered by the users of HDS in data explorations. Sustaining the sound events that are to be compared is a common practice in related psychoacoustic experiments [5, 37, 118]. The chords presented consisted of tones that were not constrained to any musical scale, and could contain any fundamental frequency from the continuum of pitch space.

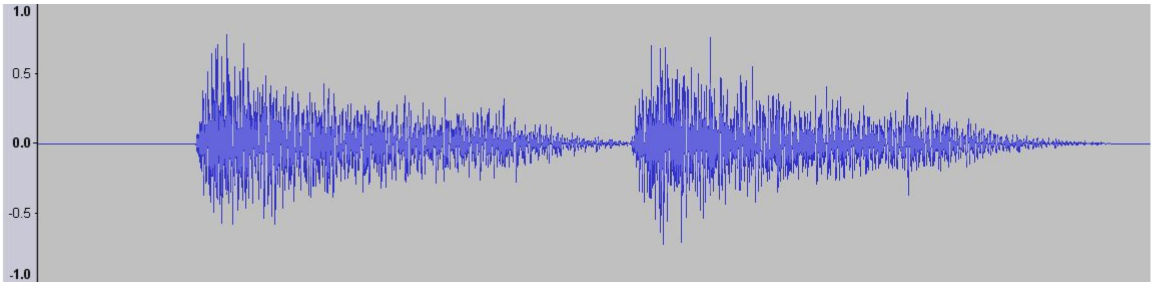


Figure 57. Wave form of a typical pair of chords in the experiment. The time interval between the beginning of each onset was 700ms. Each chord was sustained for 500ms before muting.

A collection of 442 pairs of stimuli was created for this experiment. All the chords covered the same range in the frequency space, from 81.4 Hz (E2 in the western 12-tone scale) to 783.99 Hz (G5, also in the western 12-tone scale). In other words, the lowest and the highest note were common to every chord. This range of frequencies (narrower than the range used in all the previous studies) provides the highest salience for complex sounds (like the piano sound used in this study), *i.e.* pitches are easiest to distinguish in that range [71]. Thus, the differences in pitch discriminability are minimised in the range of frequencies considered for the study.

Adjusting to the narrower pitc h space, all chords consisted of 15 sounds. Since the frequencies of both extreme voices remained unchanged, only the 13 internal voices could undergo any pitch movements. These movements were designed carefully in order to have a broad range of cases to study. Differences between chords were small in every case, and contradictory effects were attempted in order to provoke ambiguous effects (for example, implementing cases as illustrated in Figure 55. The movements of voices between two chords normally happened in both directions (some voices moved upwards while others moved downwards or remained unchanged). In the case of 114 of the 442 pairs, while both chords were different, the arithmetic means of the underlying collections of numbers remained constant. Among the remaining 328 pairs in which the value of the arithmetic mean did change, half of the pairs shifted upwards and half downwards, to have the same likeliness of correct answers in any direction. Furthermore, the variety of mean-shift sizes was the same in both directions.

8.4.3 Participants

A larger participant population was used in order to obtain a narrower band of non-significance for the binomial distribution to which the results of the experiment would fit. Thus, 41 participants completed the experiment successfully (the results from another two were considered invalid due to technical problems during the experiment). 23 of the participants were female and 18 male. Their ages were 25 years old or below in 35 of the cases, between 26 and 35 in 4 cases and above 35 in the remaining two cases. 15 of the participants did not have formal musical training, 13 were amateur musicians (they had some musical training), and the remaining 13 participants had extensive musical training. All participants had normal hearing, and reported to be able to carry a tune even when they had received no musical training.

8.4.4 Results

All the participants' answers in this experiment are summarised in Figure 58. This graph plots, for each of the 442 pairs of chords presented in the experiment, the proportions of participants that judged POP moving upwards. Since there were only two possible answers in each case (binomial distribution), the remainder of the score shown in the graph indicates the opposite, *i.e.* the percentages of participants reporting that POP moved downwards. Thus, higher values in the Y axis indicate high proportions of participants reporting upward POP movements, whereas low scores correspond to pairs in which a high percentage of the participants reported the POP to be moving downwards. The shaded area around the scores in the middle range is the band of non-significant results (binomial distribution B(41,0.5); significance level = 95%), where the difference in the proportion of participants perceiving each direction of the movement was not significant. Plots at either side outside the shaded band, indicate that the difference in that proportion of participants was significant, and they comprise 61.8% of all the pairs of stimuli in the study.

The direction of POP movement reported for each pair of chords could either be in accordance or opposite to the direction in which the arithmetic mean in the underlying data sets shifted. In the first case (when the reported direction of POP movement and the direction of mean shift were coincident), HDS would lead to a correct estimation of the statistic. In the other case (when both directions were not coincident), the estimation obtained with HDS would be misleading, and would result in an incorrect estimation. This comparative analysis between direction judgement of POP movement and the actual direction in which the arithmetic mean changed only made sense for pairs of chords in which the arithmetic mean did actually change. This was only the case in 328 of the pairs, and the results of this comparison are shown in Figure 59. Because both types of information are not independent, all the plots in this scatter plot lie on two diagonal lines. This means that if for a pair of chords, 75% of the participants said that POP moved upwards, if that was the correct estimation for the direction of movement of the mean then 75% of the participants would have obtained the correct prediction, or 25% of the participants in the opposite case. However, no other scores would be possible, plots always lying on one of both diagonal lines across the graph. The size of the bubbles (the area of the circles) represents the relative number of pairs of chords that share the same position on this graph. The vertical axis shows the same scores as in Figure 58 (proportion of participants reporting that the POP moved upwards), and the horizontal axis provides information about the correctness of such judgement as an estimation of the relative arithmetic mean values. The shaded rectangle in the middle is the area of non-significance common to both binomial distributions B(41,0.5) significance level = 95%, where the differences between proportions of participants were not significant.

59% of the pairs of chords in this graph are located outside the gray area, *i.e.* they were rated with a particular direction of movement by a significantly larger proportion of the participants. Of those 193 pairs with significant directionality ratings, 85.5% were correct estimations of the relative values of arithmetic means.

The results corresponding to the 114 pairs in which the arithmetic mean value did not change (and where consequently there were no correct of incorrect estimations of mean shift direction) are shown in Figure 60. These plots are a subset of the information shown in Figure 58, and as in that case, the shaded band contains non-significant differences in proportions of participants, B(41,0.5); significance level = 95%. The proportion of these stimuli outside the band of non significance is above 70.2%. In this case, the pairs of chords are sorted by percentage of participants that perceived the POP moving upwards.

8.4.5 Discussion

The first conclusion extracted from the inspection of the results in Figure 58 is that by simply varying the internal voices in subtle patterns (without movement in the peripheral voices and without explicitly directing the participants' attention towards any particular region in the frequency spectrum), a strong sense of directionality in POP movement was conveyed in the majority of the cases (61.8% of the pairs were rated to be moving in a particular direction by a significant proportion of the participants, at a significance level of 95%). Interestingly, in the subgroup of stimuli in which the arithmetic mean remained constant, the proportion of pairs with significant directional ratings was higher (70.2%) than in the group of pairs in which the mean did actually change (58.8%). These results reinforce the hypothesis that local groupings of internal voices have a determinant effect for creating a consistent sense of movement in a particular direction, rather than a more holistic view of the movements.

The next conclusion is extracted from inspecting Figure 59, which shows only results about the pairs in which there was an actual shift in arithmetic mean. From the pairs that in this graph are represented outside the shaded area (193 pairs that were perceived to be moving in a particular direction according to a significant majority of the participants), 85.5% are in regions B and D, *i.e.* they were correct estimations of the relative values of the arithmetic means in the underlying data. In other words, if the direction of the POP movement was judged consistently, that judgement was also a correct estimation of the behaviour of the arithmetic means. The remaining 14.5% of the cases represented in regions A and C, although much smaller a group, are just as interesting from a psychoacoustic point of view: 28 cases in which the judgements about the direction of POP movement were highly consistent, but that predicted the wrong direction of change for the arithmetic means.

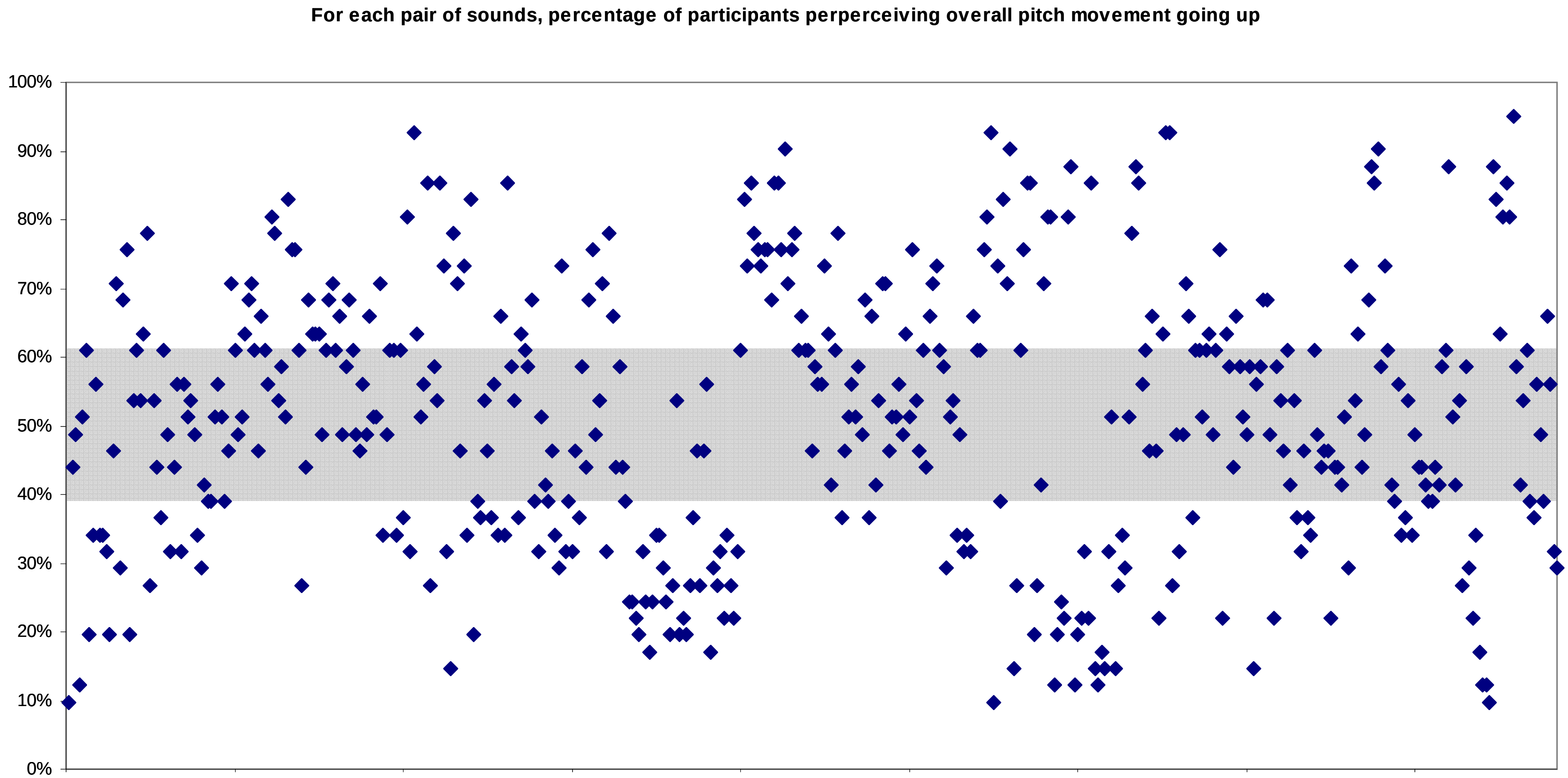


Figure 58. Each dot represents the percentage of participants that considered that POP moved upwards in particular pair of chords. The shaded area is the band of non-significant results (binomial distribution B(41,0.5); Significance Level = 95%). The dots above the shaded band correspond to pairs of chords for which the proportion of participants perceiving overall pitch as moving upwards was significantly higher. Significantly higher percentages of judgements about POP moving downwards are plotter below the shaded band.

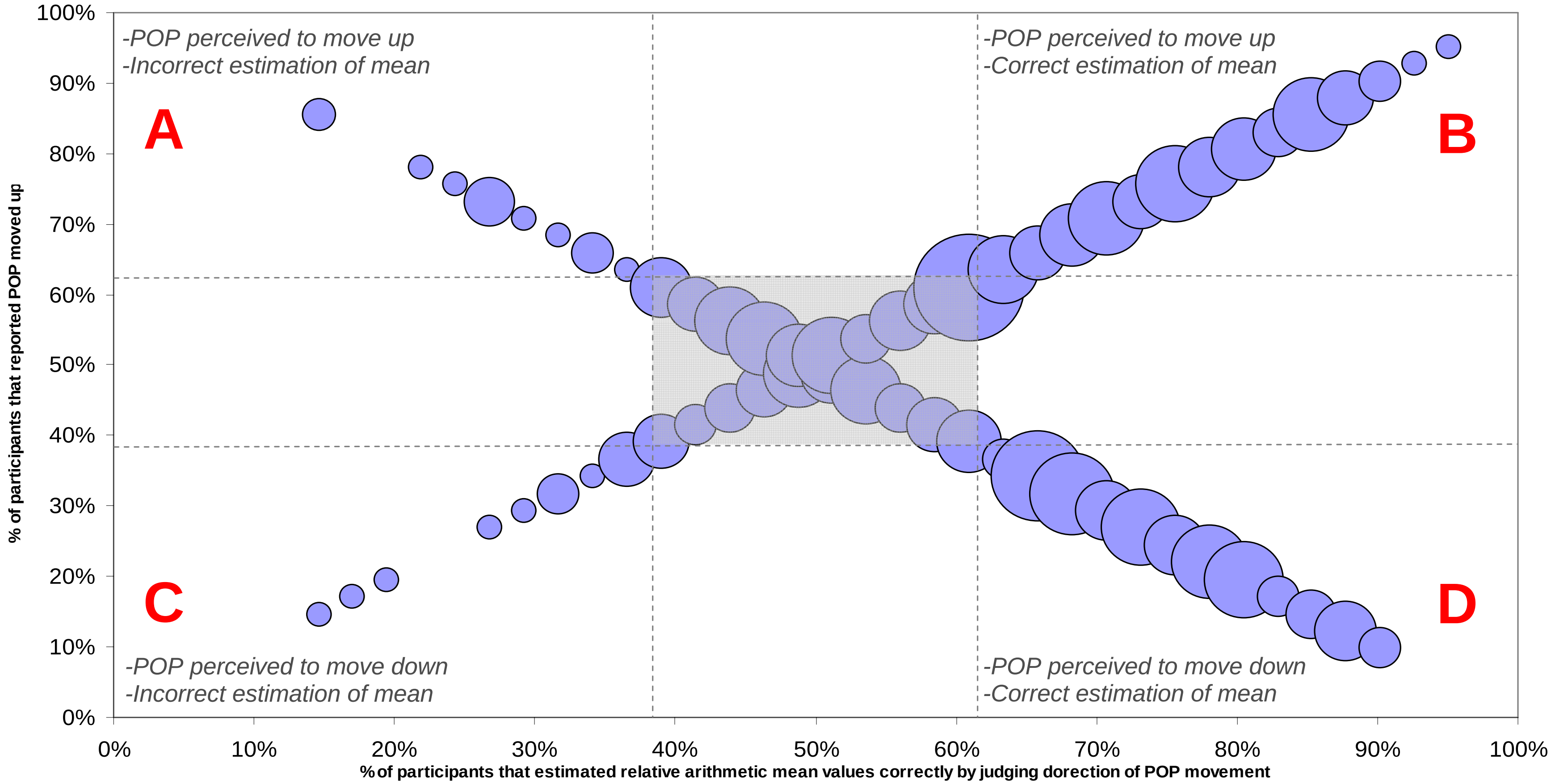


Figure 59. Scatter plot showing, for each pair of chords with different arithmetic means, the proportion of participants that reported the POP to be moving upwards (or downwards, inverting the vertical axis scale), and whether that judgement was a correct estimation or not, in the horizontal axis. The size of each bubble represents the number of pairs that share the same position on the graph. The shaded rectangle contains the non-significant differences for either variable. Regions A, B, C and D contain significant scores in both binomial distributions, B(41,0.5) with significance level = 95%.

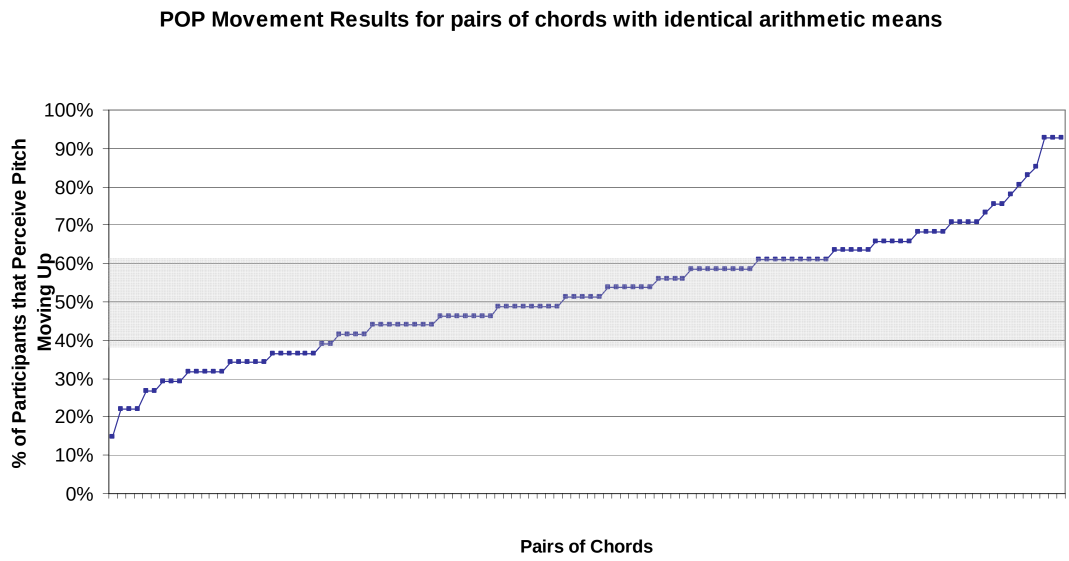


Figure 60. For each pair of chords in which the arithmetic mean remained constant, this graph shows the percentage of participants that reported POP to be moving upwards. Pairs are sorted by score, in ascending order. The shaded area is the band of non-significant results (binomial distribution B(41,0.5); significance level = 95%),

In summary, the results from this experiment confirm that proximity grouping mechanisms play a decisive role in POP movement, along with the other factors identified earlier (selective attention in the frequency space and movement in the peripheral voices). The experiment also offered a rich set of results, showing that: *(a)* even with very subtle variations in internal voices, in most cases (above 60% in this experiment) there is a defined direction for POP movement that is consistently reported; *(b)* in the cases in which the consistency reporting the direction of POP movement was significant, a majority of the direction judgements (more than 85% in this experiment) are correct estimations of the relative values of the arithmetic means in the underlying data sets. As discussed in the introduction to this section, identifying and fully understanding the actual mechanisms that lead to results so different as in regions A and C *vs.* B and D is beyond the scope of the research questions in this thesis, and it is left for future work.

8.5 Limitations

The number of possible combinations of chord configurations was so high that the study could only be designed by selecting a comprehensive subset of configurations, representative of a broad range of possible cases. As for the extent of the analysis conducted on the results, this was limited to extracting conclusions that could contribute to answering the thesis research questions, without attempting to explain the results fully from a psychoacoustic and auditory scene analysis point of view.

A broader approach could have also been taken for the study of the kinds of statistical analysis that HDS allows. For instance, while it was assumed (supported by the results from other evaluations, summarised in Table 6) that the measure of average best estimated with HDS was the arithmetic mean, other statistics of average (like modes, medians, etc.) could have been considered too. Likewise, there was some anecdotal evidence that information about data dispersion could be extracted from judging sound events generated with HDS, for instance estimates of relative values of arithmetic means or other dispersion statistics. Such a study, similar to the one reported here, could be conducted as future research work, that would provide more complete knowledge about the data mining capabilities that HDS offers.

8.6 Conclusions

This chapter was devoted to studying the perceptual and cognitive aspects related to extracting knowledge from tabular data sets using HDS, with the aim of contributing answers to the second research question, defined in Chapter 1: "*What level of performance can be achieved with the techniques proposed in this thesis?*" (RQ-2). This was achieved by determining factors that affect the reliability of using HDS as a technique to perform basic statistical analysis (the estimation of relative values of arithmetic means) by solely using properties of the human auditory system and without performing any numerical computations.

As shown in Figure 46, reliably predicting relative values of arithmetic means using HDS requires that relative overall pitch movement is perceived consistently (*i.e.*, non-ambiguously taking place in a particular direction). Only then, the movement of Perceived Overall Pitch (POP) can be tested to see if it predicts relative arithmetic means or not. In this sense, the analysis of the data gathered from the experiments resulted in the identification of the following three factors as influencing the perception of POP movement:

- *i.* Movement of the peripheral voices, mainly the top one, tend to stand out as a segregated stream;
- *ii.* Selective attention to specific regions of the frequency spectrum, so that the features in the attended region dominate the overall perception;
- *iii.* Proximity groupings in intermediate voices, which form segregated streams that tend to stand out.

If in cases of ambiguous POP movement a user was given the opportunity to analyse the auditory events, results suggested that they were rationalised by magnifying the influences of these three factors. In relation to *ii*, it was seen that the frequency region on which attention focused could be task dependent. Thus, if a participant was looking for the highest values, attention would be likely to focus on the higher frequencies and generalise the behaviour of those frequencies as the behaviour of the whole of the sound events.

With regard to the reliability of HDS as a technique to estimate relative arithmetic mean values, the results from the experiments suggested that in general reliability was above 70%. In more controlled situations where there was an absence of conscious rationalisation of the perceived sound events (single listening), with no attention bias by the data mining task (open search question) and with no movement cues in the peripheral voices, it was found that once the perception of POP movement was consistent, there would be a correct estimation of relative arithmetic mean values in 85% of the cases.

From these results, it could be expected that users could train themselves to use this technique optimally by attempting to grasp the sound events as globally as possible, by approaching the listening with no specific focus on any frequency region in particular, and by refraining from rationalising the sound events (*i.e.* by repeated listening or mental rehearsal). This hypothesis could form the basis for further research in this direction.

Bearing in mind once more that the pairs of chords in the experiments were designed to be very similar and even to contain voice movements that would group to produce misleading perceived pitch movements, these reliability estimations could be expected to be conservative (as it was observed in some of the previous studies in this thesis, which provided even higher scores of reliability). The results could also be conservative compared to longer series of chords, where it may be easier to predict trends. However, these point should also be confirmed by further research.

In summary, the research conducted in this chapter has confirmed that HDS can be used to perform estimations of relative arithmetic means of the numerical values in a data table (an important aspect of overview information), by simply judging the relative perceived overall pitched of the chords that it renders. Thus, it permits the users to perform such analysis using auditory perception in a reliable way, freeing them from performing any computations.

Chapter 9 Conclusions

9.1 Introduction

This thesis has investigated the problem of obtaining overview information from complex numerical data sets non-visually. In an attempt to reconcile the spatial metaphors employed in visual data representations with intrinsically non-visual forms of data representation, the work in this thesis focused on the particular case of tabular numerical data structures, a middle ground between the raw data and the data perceptualisations constructed from them. The work in this thesis answers this problem with the introduction of High Density Sonification (HDS), as its main contribution. This new interactive sonification technique facilitates access to overview information of the data by making use of properties of the human auditory perception and cognition, which present information grouped in single auditory events in such a way that users can perform accurate estimations of relative arithmetic means of data sets without performing numerical computations, as studied in detail in Chapter 8.

HDS was implemented in an interface for data exploration, called TableVis. This implementation is utilised in the thesis as a platform to investigate aspects that are relevant to the exploration of non-visual tabular numerical data, such as providing a non-visual implementation of the focus+context metaphor in data explorations, analysing the effect that the size of the data set has in the performance of the user that explores it with HDS, the comparison of HDS with speech synthesis technologies to obtain overview information, and an analysis of the limitations of the user's working memory in explorations conducted with HDS in TableVis: The study mentioned last derived in the development and evaluation of EMA-Tactons, a technique to annotate information using vibrotactile cues, as means of supporting working memory during the explorations.

This work is motivated by the need to enable blind and visually impaired users to easily access and analyse numerical data. In spite of the implementation of accessibility policies and development of accessibility tools such as screen readers, there still exist significant barriers for blind and partially-sighted people to access overviews of numerical data, putting them at a significant disadvantage in relation to their sighted peers by limiting their freedom to pursue studies or professions in disciplines which involve the analysis of numerical data. The need to obtain an overview of data represented in auditory displays has been stressed by various authors [51] [149] [173] [113], however most work on access to numerical data has focused on single auditory graphs, in which data points from one-dimensional arrays are retrieved one by one [95] [23] [63]. This research is novel in that it tackles the problem of obtaining overview information from complex numerical data sets quickly and easily, by taking a non-visual approach in the solution proposed, while preserving spatial metaphors that are necessary for collaboration with users that utilize visual representations of the same data.

The aim of this research was to investigate how to enable blind and visually impaired users to obtain overview information from tabular numerical data sets, and to understand the level of performance which can be achieved by users of the proposed solution, based on the use of HDS. In Chapter 1, the thesis statement was as follows:

> "*Obtaining an overview is a necessary first step in data analysis, for which current non-visual data accessibility methods offer little support. High-Density Sonification, in combination with proprioception and vibrotactile cues, forms a non-visual multimodal interaction technique that permits a user to extract overview information from tabular data sets efficiently and effectively*."

This statement has been defended by answering the following two research questions:

RQ-1: *How can overview information from tabular numerical data sets be obtained non-visually?*

RQ-2: *What level of performance can be achieved with the techniques proposed in this thesis?*

These two questions have been addressed through a review of the literature on non-visual accessibility tools and techniques, a process of requirements capture conducted with blind and visually impaired computer users, the iterative design of a novel interface to enable obtaining overview of tabular numerical data sets, and a series of empirical experiments evaluating the performance and subjective workload which could be achieved by users of that interface in typical data analysis tasks. This chapter summarises the work reported in this thesis and discusses how the findings answer the two research questions above. Following this, a set of guidelines for the design of interfaces to support non-visual access to tabular numerical data is presented. Limitations of this research and potential areas for future work are then discussed. Finally, some general conclusions are drawn from this research, and the main contributions of this work are stated.

9.2 Summary of the Thesis

This thesis describes the research conducted in response to the need for non-visual ways of obtaining overview information from tabular numerical data sets. High Density Sonification (HDS) was introduced as a solution, and implemented in TableVis, an interface for non-visual navigation, supported by EMA-Tactons, a vibrotactile data annotation technique.

This thesis began by introducing the motivation for the work, namely that blind and visually impaired users need to be able to access overview information of numerical data sets to help ensure equality of educational and professional opportunities. The importance of obtaining overviews of numerical data was justified through the literature review presented in Chapter 2. This chapter also provided a review of accessibility technologies for blind and visually impaired users, highlighting the strengths and weaknesses of these current tools. This was followed by a discussion of previous research on non-visual techniques for the exploration of numerical data. The chapter identified that properties of the auditory system can be successfully used to provide overview information from collections of data rendered in sound. Based on the motivation outlined in Chapter 1, this thesis took a user-centred approach in the design of an interface to enable blind and visually impaired users to access numerical data non-visually. Chapter 3 thus reports a requirements capture process which involved focus groups, interviews and classroom observations conducted with blind and visually impaired computer users. Feedback obtained highlighted that it was difficult to obtain quick overviews of more complex numerical data sets with current accessibility technologies such as screen readers and raised paper. Positive and negative aspects of current technologies were also established in reference to user needs. This process led to the development of a list of requirements for consideration in the design of solutions for non-visual overviews of tabular numerical data.

The information from these initial chapters led to a consideration of the approaches, existing exploratory techniques and sensory modalities that could be best employed towards answering RQ-1: how to obtain overview information from tabular numerical data non-visually. Chapter 4 reported the outcome of this. A novel data sonification technique called High-Density Sonification (HDS) was defined, which helped highlight general overview features in the data, while filtering out detail perceptually. HDS was implemented and used in an interface called TableVis in combination with other input and output techniques, such as speech synthesis and a tangible user interface. A pilot evaluation of the interface showed that HDS, as implemented in TableVis, could be used to obtain overview information from tabular numerical data sets. It also validated the usability of the interface, while highlighting some issues which were then addressed by a further design iteration.

Chapter 5 reported an experimental study which compared performance in data exploration tasks carried out using HDS and speech synthesis, which showed that exploring tabular numerical data with HDS is more effective, more efficient and produces lower subjective workload than exploring with speech. Qualitative data collected from this study were also presented. An analysis of participants' exploratory strategies and procedures was conducted, which provided further insight into the process of obtaining an overview of tabular numerical data sets using TableVis.

While HDS was shown to be effective in the study reported in Chapter 5, this was only tested with tables of one size (24 columns and 7 rows). In order to ensure that HDS and TableVis could potentially be used in real-world situations, it was necessary to establish their utility for data sets of varying sizes. Thus in Chapter 6, an experimental study was reported in which the main aim was to measure the effect of table size on user performance and subjective workload, when conducting explorations with TableVis. The results of this experiment provided evidence of the low impact of table size on these two metrics, and suggested that, within the range of table sizes considered in the study, there is no need for zooming or computer-aided information filtering in order to complete explorations for overview information using TableVis. Thus it was concluded that the time and effort required to take a glance does not depend strongly on the amount of information explored. In addition, the performance of blind and visually impaired and of sighted-blindfolded participants was compared, and no significant differences were found.

One problem which was highlighted by qualitative feedback and observations of user interactions in the first two experimental studies reported in Chapters 5 and 6 was that when attempting to compare rows or columns which were not contiguously located on the tablet, users' working memory sometimes became saturated. In an attempt to resolve this problem, Chapter 7 analysed it and described the design and implementation of EMA Tactons as vibrotactile external memory aids that the user can add as annotations to the data during an exploration, to relieve cognitive demands. Experimental evaluations were conducted, which showed that the use of EMA-Tactons permitted users to attain significant improvement in the effectiveness of the explorations. In additions, participants' mental workload was significantly reduced.

The experimental studies reported in Chapters 5-7 indicated that HDS leads to good estimations of the relative arithmetic means of all the numerical values in each row or column, regardless of the size of the table, achieving 80%-90% accuracy in most cases. Thus, in order to more fully understand the potential for HDS and TableVis to support effective data exploration, Chapter 8 reported an in-depth experimental study which investigated the reliability with which relative arithmetic means can be estimated by accessing a collection of numbers in a single musical chord with HDS. A series of experiments was conducted, which showed that HDS could be reliably used to estimate arithmetic means in 70% of cases, and under controlled conditions accuracy reached 85%, even with very complex chords. As a consequence, HDS was shown to be a technique with which users could perform basic statistical analysis from raw data without having to perform mental computations, by utilising properties of the human auditory system.

In the following section, the novel contributions of this thesis are outlined in further detail, and the ways in which this work can answer the two research questions posed by this thesis are considered.

9.2.1 Research Question 1

Research Question 1 asked:

> *How can overview information from tabular numerical data sets be obtained non-visually?*

RQ-1 was addressed firstly through a review of the literature on current accessibility tools and technologies available for visually impaired users, as well as new accessibility techniques for numerical information that are under development. Secondly a user-centred requirements capture process (conducted in the form of focus groups, interviews and classroom observations with blind and visually impaired computer users) provided a list of requirements. These requirements were later applied in the design of the techniques that form the main contributions of this thesis. Experiments reported in Chapters 5 to 8, in addition to providing quantification of the performance attainable with these techniques (thus responding to RQ-2 of the thesis, as summarised in Section 9.2.2 below), also provided information that contributed to further answering RQ-1. These findings are summarised below.

Literature Review

The need to preserve context information during the exploration process was a concern which appeared repeatedly. Some of the technologies and techniques had *a priori* more limitations in enabling users to access overview information from electronic sources in a quick and easy way, due to accessing small units of information at a time, forcing sequential access or not supporting the presentation of information dynamically. One conclusion that emerged was that the properties of the auditory system can be utilised to provide overview information from collections of data rendered in sound. The high temporal resolution and fine pitch discrimination capabilities that the auditory system offers have been used successfully for this purpose and could thus help to provide at least part of the synoptic properties of human vision, enabling the user to take a summarising "glance" of a collection of information. This is further supported by the fact that many researchers working on haptic techniques for data representation have also found it useful to implement some form of data sonification, to assist users in obtaining an overview. Steps taken in the field of auditory graphs provided some of the key factors to consider in presenting numerical information. The need to obtain an overview of data represented in auditory displays has been stressed by various authors [51, 113, 149, 173], however most work on access to numerical data has focused on single auditory graphs or one-dimensional arrays of data [63, 95, 162]. Initial attempts to convey overview information of 2D tables as collections of one-dimensional graphs rendered at higher speed [80] led to the identification of rendering speed as a parameter to control the level of detail conveyed in a graph.

Requirements capture

The requirements capture process conducted with blind and visually impaired computer users and professionals involved in their education confirmed the need for tools and techniques to explore tabular numerical data in search for overview information and identified a list of further requirements for their design:

- The user needs to be able to obtain a quick overview of the data;
- The user needs to have control over the data exploration and retrieve details on demand;
- The interface should enable the user to retain a sense of context to avoid "getting lost" in the data;
- The interface should not overload the user's working memory;
- Two-handed exploration may be preferable for tactile interaction;
- Data need to fit in one "page" or workspace, and the spatial layout of the data should be preserved;

- Users need to be able to access real world data.
- The solution should be as simple to use and as affordable as possible;
- The entire data exploration should be possible without sighted help;

Qualitative evaluations of TableVis

The requirements collected from the literature and from the target user population led to the definition of High Density Sonification (HDS) to extract overview information from the data and to the design of TableVis to facilitate interactive explorations with the help of non-visual focus+context and vibrotactile data annotations (EMA-Tactons) to manage demands on mental workload. The evaluations of each of these techniques provided quantification of the suitability of these techniques and a broad set of qualitative results that also contributed to answering RQ-1. Thus, in Chapter 5, the first full evaluation of TableVis conducted with blind and visually impaired participants compared effectiveness, efficiency and subjective workload in explorations of tabular data to extract overview information using HDS and using speech. Scores were significantly better in the three metrics when HDS was used, confirming the suitability of the technique proposed to obtain overview information from numerical data tables. In the same chapter, exploratory strategies and procedures employed by the participants were analysed. As a result, a general structure of every exploration performed with TableVis was observed to be divided in three stages:

I. Interpret the question (exploration task) in terms of tabular data structure, and devise a strategy for exploration
II. Interactive data exploration;
III. Combine all the information retrieved through interaction with the system, and obtain the answer to the question.

Exploration traces and other interaction features (like the use of exploration modes and of the speech functionality) that had been recorded during stage II in the explorations were also analysed, breaking down the complete interactive process and enabling an understanding of the approaches taken by the users. This analysis provided information about practices (both predicted at the design stage and others adopted by the participants) that should be supported in future iterations of the design process, in order to more effectively facilitate exploration for overview. These can be summarised as follows:

- *Use of speed of scan* (speed of the hand while scanning data); the use of a constant speed to scan the data and compare rows or columns offers better results than frequently varying speeds. The actual speed preferences vary between participants, without affecting performance.
- *Number of traverses*. A single traverse of the data in rows or columns mode was normally enough to obtain a general overview that permitted identify the position of the areas of interest.

- *Direction of scan*. New scans of the data normally started at specific corner of the table, and participants returned to that point by direct pointing rather than by navigating sequentially, what showed the value of maintaining context of the navigation through proprioception.
- *Use of tangible, physical borders*. While these were provided because they are essential to maintain context through proprioception, it was observed that many participants used them as guides for the pen, to conduct horizontal and vertical scans. The possibility of doing this with any tangible border should therefore be preserved.
- *Shape of lines*. Even without using the tangible borders as guides, and although this was in no way required to explore data sets successfully, participants followed very straight lines in their exploration paths, which were remarkably well aligned with the horizontal and vertical axes.
- *Use of speech details-on-demand*. This functionality was generally only used when strictly required, which was taken as a further indication of the validity of HDS as a technique to obtain overview information.

The experimental study comparing the performance achieved exploring tables with different sizes (reported in Chapter 6) also identified common issues encountered during the explorations, the awareness of which contributed to answering RQ-1:

- *Border effect.* Most noticeable with small sizes of tables. Novice users of the interface could get confused when performing repeated side-to-side full traverses in such a way that the pen was not lifted from the surface of the tablet when the direction of the movement is inverted. After the first traverse, it could appear as if one row or column was missing. This problem disappeared once the metaphors of the interactive sonification in TableVis were well understood.
- *Dead space effect*. This effect appeared when exploring small tables with comparatively large graphics tablets. If the physical dimensions of the cells mapped on the tablet were several centimetres tall and wide, crossing the area of a single cell could produce a sensation of emptiness, where the movement of the hand seemed to produce no feedback. A way of compensating for this effect is to increase the speed of scanning the table.
- *Pen "slip-off"*. This sometimes occurred when participants tried to click on the button of the pen while keeping the pen on a fixed location. Slip-offs hardly ever happened with sighted users, which suggests that it is related to being familiar with holding and manipulating pens, and therefore it should disappear with practice.
- *Comparisons between remote locations*. Some participants experienced difficulty in completing exploratory tasks when they were comparing several non-adjacent rows or columns. This problem was analysed and addressed in Chapter 7.

The research conducted in Chapter 7 contributed to answering RQ-1, while studying the problem of memory saturation in some data explorations. Three main characteristics of exploration scenarios were observed to lead to working memory saturation:

- Data tables containing a minimum of 24 rows and/or columns;
- Unsorted data sets with frequent, sudden variations, and where simple trends or patterns of variation in the data are not obvious;
- Tasks that require obtaining information with an intermediate level of detail (*e.g.* identifying a particular column in a set of 24.

Based on common practices for adding visual annotations on print media and applied to the specific case of TableVis, a list of requirements for the design of an annotation tool was defined. The design process led to the definition of EMA-Tactons, as a tool to add vibrotactile annotations to be used as external memory aids during the explorations. The process for designing EMA-Tactons was iterative, resulting in two prototypes and three full experimental evaluations. During this process, the importance of applying ecological principles of multisensory integration to the design of multimodal events was observed, and three principles of multisensory integration collected from literature on multisensory integration research were incorporated as design guidelines: *principle of inverse effectiveness, spatial rule of multisensory integration* and *temporal rule of multisensory integration*. Redesigning the first prototype according to the third principle resulted in a tool that alleviated the problem of working memory saturation significantly.

Finally, the study reported in Chapter 8 provided detailed insight into perceptual and cognitive aspects related to chords generated using HDS, in the case of complex data sets. In studying the relation between perceived overall pitch (POP) from chords generated with HDS and estimations of relative values of arithmetic mean in the data, it was found that HDS facilitates performing relative arithmetic mean value estimations. As such, this finding contributes to answering RQ-1, by informing that HDS can not only provide overview information about the data sets, but at the same time it helps with the performance of basic statistical analysis of the data, without imposing additional cognitive demands on the user.

9.2.2 Research Question 2

Research Question 2 asked:

> *What level of performance can be achieved with the techniques proposed in this thesis?*

RQ-2 has been addressed through the empirical studies reported in Chapters 5, 6, 7 and 8. These chapters reported a series of experiments which investigated the performance which could be attained in the exploration of tabular data sets for overview information using HDS as implemented in TableVis. This was done by comparing HDS with the main current non-visual accessibility technique (speech synthesis), and by quantifying the effect of different sizes of data sets on user performance. An in-depth study was carried out which investigated the capabilities of the human auditory system to extract meaning from auditory representations generated by HDS, providing further information about the suitability of this technique to obtain information from tabular data sets, and particularly as a method to perform basic statistical analysis (estimation of relative arithmetic mean values) from the data explored, using properties of human audition. In addition, levels of subjective workload during exploration tasks using TableVis were investigated, resulting in the proposal of EMA-Tactons, vibrotactile annotations that the user can add to the data in order to prevent working memory saturation in the most demanding data exploration scenarios. The findings of each experimental study, and how these relate to RQ-2, are summarised below.

HDS versus Speech Synthesis

In a general evaluation of TableVis, user performance (effectiveness and efficiency) and subjective workload were compared using the same interface with two different information rendering techniques: HDS and speech synthesis.

It was found that exploring tabular data sets to obtain overview information is more successfully done using non-speech sound generated with HDS than when information is rendered in speech: efficiency and subjective workload were significantly higher and overall subjective workload was significantly lower when HDS was used. The average accuracy of the answers in the exploration tasks reached 80% with HDS, whereas it stayed below 60% where information was retrieved in speech. Regarding the efficiency of the explorations, the exploration tasks were completed in an average time of less than 40 seconds, while using speech the average time was measured to be around 70 seconds. Thus, shorter exploration time produced more accurate results with HDS. Regarding the average subjective overall workload, average scores (in the scale from 1 to 20) were 6.3 for explorations using HDS and 10.7 when speech was used. More specifically, among the modified NASA-TLX categories, time pressure, annoyance and mental demand were significantly higher in the speech condition, and performance level achieved was significantly lower.

These results contribute to answering RQ-2 by quantifying effectiveness, efficiency and subjective workload obtained with explorations of tabular numerical data using HDS, and by showing that the scores measured for the three metrics are significantly better than the scores achieved when information was retrieved in speech.

Effect of Table Size

A range of table sizes was considered (represented by tables with sizes 4x7, 24x7 and 24x31), to measure the differences in effectiveness, efficiency and subjective workload that were achieved obtaining overview information from them. The results from this study, obtained experimentally with sighted-blindfolded participants, were also used to compare against the results previously obtained with blind and visually impaired participants who explored the same tables with size 24x7.

The results from the experiment showed that, within this range of sizes, the three metrics considered did not change very much. Significant changes in the scores of the three metrics were observed, i.e. bigger table sizes did affect the three metrics, by making effectiveness and efficiency lower, and overall mental workload higher. However, these changes became significant only after a large increase in the amount of information explored. This was quantified by measuring the slopes of the ratios between the changes in the scores measured and the changes in table size. This was done by considering two measures of table size: *R* x *C* and *R* + *C* (where *R* is the number of rows and *C* is the number of columns in the table). Keeping in mind that a slope of 0% means that changes in size do not affect the scores of the metrics and that 100% means that doubling the size results in also doubling the scores of the metrics, the results were as follows (notice that for each metric two different slopes were considered, one for each measure of table size):

	Efficiency (time to complete task)	**Effeciveness (% correct answers)**	**Subjective overall workload (1-20 scale)**
R x C	5.46%	25.7%	5.08%
R + C	9.2%	64.3%	7%

Table 7. Effect of the size of the table explored in effectiveness, efficiency and subjective workload, using HDS in TableVis. Figures in the table represent slopes (ratios of change), as explained in the main text. (R = number of rows; C = number of columns).

These results contribute to answering RQ-2 by providing quantifications of the three metrics considered (effectiveness and efficiency of the explorations, and overall subjective workload experience by the users). They show that, within the range of sizes considered, sensitivity of the scores in these metrics is low, particularly for efficiency and subjective workload, demonstrating that HDS as implemented in TableVis provides non-visually some of the synoptic properties that vision offers when inspecting data visualisations.

Using EMA-Tactons to improve performance

EMA-Tactons were designed and implemented in Chapter 7, with the aim of relieving cognitive demands in explorations of complex numerical data sets. A series of experiments was conducted to quantify the effect of EMA-Tactons on effectiveness (accuracy) and mental workload in data exploration tasks. Efficiency (speed) was not measured as the aim of implementing EMA-Tactons was to improve accuracy in completing data exploration tasks. Effectiveness was measured by two metrics: the ability to keep count of the number of columns selected and the accuracy of selecting the highest pitch in the table. Sub-task 1 was thus to select exactly 5 columns, and sub-task 2 was to identify the column containing the highest value in the table. The study found that both tasks were performed significantly better using EMA-Tactons, showing that the implementation of EMA-Tactons supported improved performance in terms of effectiveness (accuracy) of explorations. In addition, participants' mental workload was significantly reduced, confirming that EMA-Tactons contributed towards alleviating the constraints on short term memory which they otherwise suffered.

Perceptual and Cognitive Aspects of HDS

This study analysed in detail perceptual and cognitive aspects related to the audio representations of the data generated using HDS. In particular, it quantified the reliability with which HDS can be used to estimate relative values of the arithmetic means resulting from the numbers in the rows and columns of the tables explored.

A series of experiments provided insight into the complex interactions between the perception of pairs of very dense and similar chords, to investigate under which chord configurations and with which minimum differences between chords participants could judge about the direction in which perceived overall pitch (POP) moved (upwards or downwards) and in which cases their judgements were correct estimations of the relative arithmetic mean values of the data sets that originated the chords. As seen in the figures reporting the results in Chapter 8, their complexity makes it difficult to summarise them in detail. In practical terms, the results from this study contribute to answering RQ-2 by asserting that, even with data sets so complex and with so small and intricate differences in the values between rows or columns as used for the experiments, in 61.2% of the cases participants agreed on the direction in which POP moved. From those reliable judgements of POP movement, 85% were correct estimations of the relative values of the arithmetic means of the underlying data sets. In conclusion, HDS is a reliable technique for estimating relative arithmetic mean values from data sets by judging relative POP from chords. This is true in the majority of the cases (including the cases of the data sets used to evaluate HDS throughout this thesis), where data collections in the rows and columns of tables will produce bigger POP movements, and therefore higher reliability of the technique.

9.3 Guidelines

In addition to answering the research questions as described above, this thesis has also developed novel techniques to facilitate and support non-visual exploration of numerical data sets. The development and evaluation of these techniques has led to the formulation of a list of guidelines which designers can refer to in the design of other tools to support non-visual data exploration. These are summarised as follows, indicating the chapter from which each is derived.

9.3.1 Guidelines for the design of interfaces to facilitate non-visual exploration of tabular numerical data

- Provide control over choosing a variety of different speeds to scan the data. (Chapter 4)
- Provide means for pointing directly to meaningful positions in the data, by supporting ways of maintaining context during navigation. Using proprioception is a good way of doing this. (Chapter 4)

- To obtain a setup in a non-visual interactive exploration interface, which permits the use of proprioception to maintain context while focusing on specific areas of the data, the following set of three invariants must be observed:
    - *i.* The complete data set must be always on display;
    - *ii.* Information must remain stationary in its spatial layout;
    - *iii.* The size of the workspace on which information is laid out must be fixed and its borders must be tangible. (Chapter 4)
- If pen-based interaction is used, it can be awkward for users who do not normally use pens to click the buttons on the electronic pen, which can interfere with the exploration. As a guideline, provide the option for accessing the same function outside the pen, e.g. on a computer keyboard. (Chapter 5)
- When mapping a data set onto a spatial interface such as a graphics tablet, be aware of the "dead-space" effect, where small data sets may appear relatively "empty". Enable the user to adjust the speed of navigation to compensate for this. (Chapter 6)
- With explorations of data sets where data do not follow simple trends and patterns, some exploration tasks can lead to saturation of working memory, because the user has to remember too much intermediate information before the answer is found. For these cases, offer some form of annotation on the data, to serve as external memory aids. The work in this thesis proposes the use of EMA-Tactons. (Chapter 7)
- When designing multimodal interfaces that should generate multimodal events, such as combining sound and tactile feedback, the temporal rule of multisensory integration should be observed, i.e. ensure onset and cessation of the event is replicated across the different modalities. (Chapter 7)

9.3.2 Guidelines for the design of EMAs for data exploration interfaces

As discussed in Chapter 7, EMA-Tactons were designed and implemented as a way of relieving cognitive demands on users making comparisons between sonifications during data exploration tasks using TableVis. Based on experimental observation as well as guidelines from literature on annotation in other contexts, the following guidelines were assembled for the design of external memory aids for data exploration interfaces (these are reproduced from Table 5):

- Annotations should be easily added and removed in real time, in an integrated way with the data exploration process;
- Annotations should not be limited in number;
- Each annotation must remain in the same position, unless explicitly moved by the user (persistence of the annotations);
- Annotations must be easy to find, and distinguishable from the original information;
- Adding an annotation must leave the information in the data set unaltered;
- An annotation should not obstruct the access to the original information in that position;

- Annotating should combine with other techniques and tools for data exploration and analysis available in the interface, to support the process of information seeking;
- A wide variety of types of annotations should be possible, to be employed and coded at the discretion of the user.

9.4 Limitations and Future Work

The limitations introduced by the design decisions that led to the proposed solutions have been discussed throughout this thesis. Similarly, the scope of the thesis in terms of data sets considered (*i.e.* types and variety of data in a table, and sizes and shapes of data sets, as well as number of dimensions in the tabular structures) has been discussed as a limitation of the scope to which the findings in this thesis apply.

During the course of the research, a number of areas were identified which seemed worthy of further investigation, but which it was not possible to encompass within the scope of the current research. These are summarised and discussed in the following subsections.

HDS for other data structures

In the evaluations conducted in this thesis only a small subset of all the possible configurations of data tables have been considered: 2-dimensional tables with no empty cells, within a limited range of sizes. It should be possible to extend the techniques used in TableVis to more general cases, like tabular structures with more than 2 dimensions, tables with empty cells, tables with ranges of merged cells etc.

Non-canonical HDS

As noted in Chapter 4, the canonical HDS of a tabular data set is by no means the only possible form of HDS. Other examples of HDS could include grouping data points that share values in two or more axes, data spanning over certain ranges on one or more axes, or data subsets that are not arranged in parallel to any of the axes. Another form which could be particularly useful is grouping data points in concentric round shapes (*e.g.* concentric spheres in 3D-tables, or concentric circles in 2D-tables) that expand around a certain point of interest. Future work could examine how these other forms could be implemented, and their utility in terms of performance and subjective workload.

HDS and TableVis for real-world data exploration tasks

In order to maximise the possibility for HDS and TableVis to be developed to support real world data exploration tasks, real or realistic data sets have been utilised for experiments evaluating the proposed interface as a whole (Chapters 5 and 6). However, by definition an experimental setup is not the same as real-world experience of interacting with tabular numerical data. In order to establish whether HDS and TableVis can be effective in real world data exploration it would be necessary to conduct real world evaluations, which might include using these techniques to analyse data from numerical-intensive disciplines.

Utilising other sound parameters

Research on sonification in general and auditory graphs in particular has demonstrated the utility of mapping values to pitch. As discussed in Chapter 4, these findings were the basis for the decision to use pitch as the parameter exploited with HDS and TableVis. The techniques developed by this thesis are intended to cater for data with any meaning, which fit into a 2D-table. However, as Walker and others have observed, different sound parameters may be appropriate for different kinds of data [160] [111]. A potential area for future investigation with reference to HDS and TableVis is the use of different auditory parameters to encode different data attributes and to customise design decisions on polarity and scale factors to the particular meaning of the data. Additionally, it is also an interesting research question to investigate how to best signify other features related to numerical data in the tables, like changes in the sign of the numbers.

Exploratory Strategies and Procedures

As reported in the analysis of participants' exploratory strategies and procedures using TableVis in Chapter 5, it was noted that most people started scanning the data from the top-left of the tablet. Whether this reflected the nature of the data set itself or typical practice in the normal reading of a document was not established. Further studies could compare exploratory procedures utilising different types of data set in order to see if this affected the browsing procedure. Understanding the factors underlying such exploratory procedures could provide an insight into the mental representation of the information during the exploration (*i.e.* whether the participant maintains a sense of the meaning of the data during the exploration or if this is not considered once the query has been mentally formulated in terms of data structure). Answers to these questions could help in the improvement of interaction designs.

Quantifying efficiency achieved with EMA-Tactons

As discussed in Chapter 7, EMA-Tactons were implemented in TableVis to alleviate mental workload and improve accuracy in data exploration tasks. As a result, efficiency (speed) of explorations was not statistically analysed, although it appeared that tasks took longer to complete when EMA-Tactons were used than when no annotation support was available, presumably because participants could continue the exploration for a longer time without reaching memory saturation. Future work could be conducted to establish performance in terms of efficiency with EMA-Tactons, with implications for future design iterations.

Investigation of more complex EMA-Tacton designs

Tactons are capable of encoding more complex messages than the EMA-Tactons implemented in Chapter 7, which used only one bit of information. It could be envisioned that using Tactons constructed with more complex structures, a more complex vocabulary of Tactons than the binary

EMA-Tactons implemented in this thesis, filtering of data could take place following different criteria, and users could encode richer information in their annotations. This warrants further investigation into how richer functionalities could be implemented with EMA-Tactons to aid data exploration and analysis using TableVis.

Further work on multisensory integration

Chapter 7 discussed the role of multisensory integration, focusing on temporal integration whereby the onset and fade of audio and vibrotactile stimuli were synchronised in order to be perceived as part of the same multimodal event. A further aspect which could have been investigated was the spatial rule of multisensory integration, which affects the likelihood of two stimuli being perceived as a single multimodal event when these originate from different spatial locations.

Future work on the perception and cognition of HDS

The study described in Chapter 8 initiated a line of research that aims at understanding perceptual and cognitive aspects of HDS. The focus of the research reported in that chapter was on determining the conditions under which HDS can provide accurate estimations of relative arithmetic mean values in a data set. This study provides a quantification of the relative number of instances in which such estimations are reliable, instances in which reliable estimations are difficult to perform and instances in which estimations reliably lead to wrong estimations. What the study does not attempt to answer is the reasons why each instance belongs to either of these groups. Future work should attempt to understand the perceptual and cognitive reasons for which estimations of arithmetic mean movements based on listening to chords generated with HDS work differently for different data configurations.

A related line of research for future work is to investigate perceptual estimation of other data descriptors and to understand when such estimations are reliable and why, in order to improve the sonification technique itself towards the support of such descriptor estimations.

9.5 Final Remarks

This thesis was motivated by the need to enable blind and visually impaired people to obtain quick and easy overviews of numerical data. Following a user-centred design process, this thesis has proposed a new sonification technique, High-Density Sonification (HDS), and provided a detailed experimental investigation into the implementation of HDS within TableVis, a novel interface developed to enable obtaining overviews of tabular numerical data sets non-visually. HDS enables subsets of information to be represented as single auditory events, which can be compared against each other revealing overview information and even enabling the user to perform basic statistical analysis of the data by exploiting properties of the human auditory system. The experimental studies reported in this thesis have established that High-Density Sonification as implemented in TableVis, supported by speech synthesis, proprioception and vibrotactile cues, supports effective and efficient data explorations and minimises mental workload.

In addition to answering the research questions posed by this thesis, in Section 9.3 a set of guidelines has been provided to aid other researchers or interface designers in the design of non-visual interfaces to facilitate obtaining overviews by exploring data sets, as well as the design of annotations such as EMA-Tactons. In addition, High-Density Sonification, as defined in full in Section 4.3, offers flexibility for implementation in a wide variety of auditory interfaces and for a wide range of data sets and structures. It can therefore be used by designers developing novel auditory interfaces and by researchers wishing to investigate aspects of usability and user performance, as well as perceptual and cognitive properties of different implementations.

APPENDICES

Appendices A to G, due to the large amount of data presented in them, are included in separate electronic folders. In the case of the print copy of this thesis, these folders are included in a CD-ROM.

- *Appendix A. Transcriptions from the Requirements Capture (Chapters3 and 4)*
- *Appendix B. Modified NASA-TLX questionnaire form*
- *Appendix C. Prototype Evaluation of TableVis (Chapter-4)*
- *Appendix D. Experimental Evaluation of TableVis: HDS vs. Synthesised Speech (Chapter-5)*
- *Appendix E. Experiment about the Effect of Table Size in the Use of TableVis (Chapter-6)*
- *Appendix F. Experimental Evaluations of EMA-Tactons (Chapter-7)*
- *Appendix G. Experiments on Perceptual and Cognitive aspects of HDS (Chapter-8)*

REFERENCES

1. *UN Convention on the Rights of Persons with Disabilities*, U. Nations, Editor. 2006.
2. *General Braille Information Center*. [website] 2008 [cited 2008 28 July]; Available from: http://www.nfb.org/nfb/Braille_general.asp?SnID=2.
3. Adapt-IT. *Braille Maker for Windows*. 2008 [cited 2008 15 August]; Available from: http://www.adapt-it.org.uk/products/braillemakerforwindowsstandard.asp.
4. Aldwell, E. & C. Schachter, *Harmony and Voice Leading*. 2nd Edition ed. 1989, New York: Harcourt. 624.
5. Allik, J., E.N. Dzhafarov, A.J.M. Houtsma, J. Ross & N.J. Versfeld, *Pitch motion with random chord sequences.* Perception & Psychophysics, 1989. **46**(6): p. 513–527.
6. Alty, J.L. & D.I. Rigas. *Communicating graphical information to blind users using music: the role of context*. in *Proceedings of Conference on Human Factors in Computing Systems (CHI 98), 18-23 April 1998*. 1998. Los Angeles, CA, USA: ACM.
7. Bachem, A., *Absolute Pitch.* Journal of the Acoustic Society of America, 1955. **27**(6): p. 1180-1185.
8. BBC. *BBC Weather --- Average Conditions, Cardiff, UK*. 2004 [cited 2008 25th September]; Available from: http://www.bbc.co.uk/weather/world/city_guides/results.shtml?tt=TT003770.
9. BBC. *BBC Weather --- Average Conditions, Nairobi, Kenya*. 2004 [cited 2008 25th September]; Available from: http://www.bbc.co.uk/weather/world/city_guides/results.shtml?tt=TT000300.
10. Bederson, B.B. & J.D. Hollan, *Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics*, in *Proceedings of the 7th annual ACM symposium on User interface software and technology*. 1994, ACM Press: Marina del Rey, California, United States. p. 17-26.
11. Berman, L. & K. Gallagher. *Listening to program slices.* in *12th International Conference on Auditory Display*. 2006. , London, UK, June 20-23, 2006.
12. Bernareggi, C., C. Comaschi, A. Marcante, P. Mussio, L.P. Provenza & S. Vanzi. *A Multimodal Interactive System to Create and Explore Graph Structures*. in *CHI*. 2008. Florence, Italy.
13. Bertelson, P. & B. de Gelder, *The Psychology of Multimodal Perception*, in *Crossmodal Space and Crossmodal Attention*, C. Spence and J. Driver, Editors. 2004, Oxford University Press: Oxford. p. 141-178.
14. Bertin, J., *Semiology of graphics : diagrams, networks, maps*. 1967/1983: Madison, Wis. : University of Wisconsin Press.
15. Blattner, M.M., D.A. Sumikawa & R.M. Greenberg, *Earcons and Icons: Their Structure and Common Design Principles.* Human-Computer Interaction, 1989. **4**(1): p. 11-44.
16. Bly, S. *Presenting information in sound*. in *Conference on Human Factors in Computing Systems*. 1982. Gaithersburg, Maryland, United States: ACM.
17. Bregman, A.S., *Auditory Scene Analysis: Hearing in Complex Environments*, in *Thinking in Sound. The cognitive psychology of Human Audition*, S. McAdams and E. Bigand, Editors. 1993, oxford University Press. p. 10-36.
18. Bregman, A.S. & Y. Tougas, *Propagation of constraints in auditory organization.* Perception and Psychophysics, 1989. **46**(4): p. 395-396.
19. Brewster, S., *Overcoming the Lack of Screen Space on Mobile Computers.* Personal and Ubiquitous Computing, 2002. **6**(3): p. 1617-4909.
20. Brewster, S. & A. King, *An Investigation into the Use of Tactons to Present Progress Information*, in *Human-Computer Interaction - INTERACT 2005*. 2005, Springer: Berlin / Heidelberg. p. 6-17.
21. Brewster, S.A., *Providing a structured method for integrating non-speech audio into human-computer interfaces*, in *Department of Computer Science*. 1994, University of York: York, UK. p. 292.
22. Brewster, S.A. & L. Brown. *Tactons: Structured Tactile Messages for Non-Visual Information Display*. in *AUIC 2004*. 2004. Dunedin, New Zealand: Australian Computer Society.
23. Brown, A., R.D. Stevens & S. Pettifer. *Audio Representation of Graphs: A Quick Look.* in *International Conference on Auditory Display*. 2006. London.
24. Brown, L. & S.A. Brewster. *Drawing by Ear: Interpreting Sonified Line Graphs.* in *ICAD*. 2003. Boston, MA.

25. Brown, L., S.A. Brewster, R. Ramloll, W. Yu & B. Riedel. *Browsing Modes for Exploring Sonified Line Graphs*. in *16th British HCI Conference*. 2002. London.
26. Brown, L.M., S.A. Brewster & H.C. Purchase. *A First Investigation into the Effectiveness of Tacton*. in *WorldHaptics 2005*. 2005. Pisa, Italy: IEEE Press.
27. Brown, L.M., S.A. Brewster & H.C. Purchase. *Multidimensional Tactons for Non-Visual Infomation Display in Mobile Devices.* . in *In Proceedings of MobileHCI 2006*. 2006 Espoo, Finland: ACM Press.
28. Brown, L.M., S.A. Brewster, R. Ramloll, M. Burton & B. Riedel. *Design Guidelines for Audio Presentation of Graphs and Tables.* in *Proceedings of ICAD 2003 Workshop on Auditory Displays In Assistive Technologies*. 2003. Boston, MA.
29. Bureau, U.S.C. *International Data Base (IDB)*. 2005 [cited 2005 24-7-2005]; Available from: http://www.census.gov/ipc/www/idb.html.
30. Card, S.K., J.D. Mackinlay & B. Shneiderman, *Information Visualization*, in *Readings in Information Visualization. Using Vision to Think*, S.K. Card, J.D. Mackinlay, and B. Shneiderman, Editors. 1999, Morgan Kaufmann Publishers, Inc. p. 1-34.
31. Castellano, C., *Making It Work: Educating the Blind/Visually Impaired Student in the Regular School*. 2000: IAP.
32. Challis, B.P. & A.D.N. Edwards. *Design Principles for Tactile Interaction*. in *Haptic Human-Computer Interaction: First International Workshop*. 2000. Glasgow, UK: Springer-Verlag.
33. Cohen, R.F., A. Meacham & J. Skaff, *Teaching graphs to visually impaired students using an active auditory interface.* ACM SIGCSE Bulletin, 2006. **38**(1): p. 279-282.
34. Corp, A.E. *Tactaid VBW32 transducer*. 2008 [cited 2008 15 August]; Available from: http://www.tactaid.com/.
35. Corporation, W.T. *Wacom Graphire3 A5, G-630), .* 2008 [cited 20 September]; Available from: http://www.wacom.com.sg/price/graphire3/graphire3_A5.html.
36. Cove, J.F. & B.C. Walsh, *Online Text Retrieval Via Browsing.* Information Processing & Management, 1988. **24**(1): p. 31-37.
37. Demany, L. & C. Ramos, *On the binding of successive sounds: Perceiving shifts in nonperceived pitches.* The Journal of the Acoustical Society of America, 2005. **117**(2).
38. DiFilippo, D. & D.K. Pai. *Contact Interaction with Integrated Audio and Haptics*. in *ICAD*. 2000: International Community for Auditory Display.
39. Dolphin. *Dolphin Easyreader Digital Talking Book Player*. [Web site] 2008 [cited 2008 7 September]; Available from: http://www.yourdolphin.com/productdetail.asp?id=9&z=1.
40. Dolphin. *Dolphin SuperNova - combined screen reader and magnifier with Braille support*. [Web site] 2008 [cited 2008 7 September]; Available from: http://www.yourdolphin.com/productdetail.asp?id=1.
41. Dombois, F. *Using audification in planetary seismology*. in *2001 International Conference on Auditory Display*. 2001. Espoo, Finland.
42. Douglas, G., C. Corcoran & S. Pavey, *The role of the WHO ICF as a framework to interpret barriers and to inclusion: visually impaired people's views and experiences of personal computers.* British Journal of Visual Impairment, 2007. **25**(1): p. 32-50.
43. Driver, J.S., C., *Crossmodal Spatial Attention: Evidence from Human Performance*, in *Crossmodal space and crossmodal attention*, C.D. Spence, J., Editor. 2004, Oxford University Press: Oxford.
44. Duxbury. *Duxbury Braille Translator (DBT)*. 2008 [cited 2008 15 August]; Available from: http://www.duxburysystems.com/dbt.asp.
45. DuxburySystemsInc. *Megadots*. 2008 [cited 2008 15 August]; Available from: http://www.duxburysystems.com/megadots.asp.
46. Edwards, A., R. Stevens & I. Pitt, *Representation non visuelle des mathematiques*, in *Le Defecit Visuel*, A.B. Safran and A. Assimacopoulos, Editors. 1995, Editions Masson: Paris. p. 169-178.
47. Eimer, M., *Electrophysiology of Human Crossmodal Spatial Attention*, in *Crossmodal space and crossmodal attention*, C.D. Spence, J., Editor. 2004, Oxford University Press: Oxford. p. 221-246.
48. Fernström, M., E. Brazil & L. Bannon, *HCI Design and Interactive Sonification for Fingers and Ears.* IEEE Multimedia, 2005. **12**(2): p. 36-44.
49. Finlayson, J.L. & C. Mellish. *The 'audioview': providing a glance at java source code*. in *11th International Conference on Auditory Display*. 2005. , Limerick, Ireland.
50. Flowers, J.H. *Thirteen Years of Reflection on Auditory Graphing: Promises, Pitfalls, and Potential New Directions*. in *First International Symposium on Auditory Graphs (AGS2005),*

in the Eleventh Meeting of the International Conference on Auditory Display (ICAD-05). 2005. Limerick, Ireland.

51. Flowers, J.H., D.C. Buhman & K.D. Turnage, *Data Sonification from the Desktop: Should Sound be Part of Standard Data Analysis Software?* ACM Transactions on Applied Perception, 2005. **2**(4): p. 467-472.
52. Foulke, E., *Reading Braille*, in *Tactual Perception: A Sourcebook*, W. Schiff and E. Foulke, Editors. 1982, Cambridge University Press: Cambridge. p. 168-208.
53. Franklin, K.M. & J.C. Roberts. *A Path Based Model for Sonification*. in *Eighth International Conference on Information Visualization*. 2004: IEEE Computer Society.
54. Frauenberger, C., T. Stockman & M.-L. Bourget. *A Survey on Common Practice in Designing Audio in the User Interface*. in *HCI 2007: People and Computers XXI - HCI... but not as we know it*. 2007: British Computer Society.
55. Freitas, D. & G. Kouroupetroglou, *Speech technologies for blind and low vision persons.* Technology and Disability, 2008. **20**(2): p. 135-156.
56. Fricke, J. & H. Baehring, *Design of a Tactile Graphic Tablet and its Integration into a Personal Computer System for Blind Users.* Journal of Microcomputer Applications, 1993. **16**(3): p. 259-269.
57. Fritz, J.P., T.P. Way & K.E. Barner. *Haptic Representation of Scientific Data for Visually Impaired or Blind Persons*. in *Eleventh Annual Technology and Persons with Disabilities Conference*. 1996. California State University, Los Angeles, California.
58. Gaver, W.W., *The SonicFinder, an Interface that Uses Auditory Icons.* Human-Computer Interaction, 1989. **4**(1): p. 67-94.
59. Goldstein, E.B., *The Somatic Senses*, in *Sensation and Perception*. 1996, Brooks/Cole Publishing Company: Pacific Grove, CA. p. 460-497.
60. Gray, G. & S.M. Wilkins, *A snapshot of 2003-4: blind and partially sighted students in Higher Education in England and Northern Ireland.* British Journal of Visual Impairment, 2005. **23**(4): p. 4-10.
61. Griffin. *Griffin Technology PowerMate USB Controller*. 2008 [cited 2008 15 August]; Available from: http://www.griffintechnology.com/products/powermate.
62. GW Micro, I. *GW Micro Windows Eyes*. [Web site] 2008 [cited 2008 7 September]; Available from: http://www.gwmicro.com/Window-Eyes/.
63. Harrar, L. & T. Stockman, *Designing Auditory Graph Overviews: An Examination of Discrete vs. Continuous Sound and the Influence of Presentation Speed*, in *Proceedings of the 13th International Conference on Auditory Display (ICAD2007)*. 2007, Schulich School of Music, McGill University: Montreal, Canada.
64. Hart, S. & L. Staveland, *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*, in *Human Mental Workload*, P. Hancock and N. Meshkati, Editors. 1988, Elsevier Science Publishers B.V. p. 139-183.
65. Hart, S. & C. Wickens, *Workload assessment and prediction*, in *MANPRINT, An Approach to Systems Integration*, H.R. Booher, Editor. 1990, Van Nostrand Reinhold. p. 257-296.
66. Hermann, T. & A. Hunt, *An Introduction to Interactive Sonification.* 2005.
67. Hermann, T. & H. Ritter. *Listen to your Data: Model-Based Sonification for Data Analysis*. in *Advances in intelligent computing and multimedia systems*. 1999. Baden-Baden, Germany: International Institute for Advanced Studies in System Research and Cybernetics.
68. Holmes, N.P., G.A. Calvert & C. Spence, *Multimodal Integration*, in *Encyclopedia of Neuroscience (to be published)*, M.D. Binder, et al., Editors. 2008, Springer: Heidelberg.
69. Holmes, N.P. & C. Spence, *Multisensory Integration: Space, Time and Superadditivity.* Current Biology, 2005. **15**(18): p. R762-R764.
70. Hunt, A. & T. Hermann, *The Importance of Interaction in Sonification*, in *ICAD 04 - Tenth Meeting of the International Conference on Auditory Display*. 2004: Sydney, Australia.
71. Huron, D., *Tone and Voice: A Derivation of the Rules of Voice-leading from Perceptual Principles.* Music Perception, 2001. **19**(1): p. 1-64.
72. Intons-Peterson, M.J., *Memory Aids*, in *Basic and Applied Memory Research*, D. Hermann, et al., Editors. 1996, Lawrence Erlbaum Associates: Hillsdale, NJ. p. 317-332.
73. Irizar, J.A., *Basis-weight analysis of paper production samples from a paper machine*. 2007: Germany (for the data production). Tolosa, Spain (for the annotations).
74. Janata, P. & E. Childs. *Marketbuzz: Sonification of Real-Time Financial Data*. in *Tenth Meeting of the International Conference on Auditory Display*. 2004. Sydney, Australia.

75. Jeong, W. & M. Gluck. *Multimodal bivariate thematic maps with auditory and haptic display*. in *8th International Conference on Auditory Display (ICAD2002)*. 2002. Kyoto, Japan: Advanced Telecommunications Research Institute.
76. Jones, R.W. & I.E. Careras, *The Empirical Investigation of Factors Affecting Graphical Visualisation.* Behavior research methods, instruments & computers, 1996. **28**(2): p. 265-269.
77. Keil, S., D. Parris, R. Cobb, A. Edwards & R. McAllister, *Too little, too late: Provision of school textbooks for blind and partially sighted pupils*. 2006, RNIB.
78. Kennel, A., *Audiograf: A Diagram-Reader for the Blind*, in *Assets '96: Proceedings of the second annual ACM conference on Assistive Technologies*. 1996, ACM Press: Vancouver, Canada. p. 51-56.
79. Kildal, J., *Browsing 2D tabular numerical information using non-speech sounds for visually-impaired users*, in *Department of Computing Science. MSc Dissertation*. 2001, University of Glasgow: Glasgow, Scotland.
80. Kildal, J. & S.A. Brewster. *Explore the matrix: Browsing numerical data tables using sound.* in *ICAD2006*. 2005. London, Ireland.
81. Kildal, J. & S.A. Brewster, *Exploratory Strategies and Procedures to Obtain Non-Visual Overviews Using TableVis.* International Journal on Disability and Human Development, 2006. **5**(3): p. 285-294.
82. Kildal, J. & S.A. Brewster. *Non-visual Overviews of Complex Data Sets*. in *Extended Abstracts of CHI 2006*. 2006. Montreal, Quèbec, Canada: ACM Press.
83. Kildal, J. & S.A. Brewster. *Providing a Size-Independent Overview of Non-Visual Tables*. in *12th International Conference on Auditory Display (ICAD2006).* 2006. Queen Mary, University of London.
84. Kildal, J. & S.A. Brewster. *EMA-Tactons: Vibrotactile External Memory Aids in an Auditory Display*. in *IFIP Interact 2007*. 2007. Rio de Janeiro, Brasil.
85. Kildal, J. & S.A. Brewster, *Interactive Generation of Overview Information Using Speech*, in *CHI '07 extended abstracts on Human factors in computing systems*. 2007, ACM Press: San Jose, CA, USA. p. 2489-2494.
86. Kramer, G., *An Introduction to Auditory Display*, in *Auditory Display: Sonification, Audification, and Auditory Interfaces*, G. Kramer, Editor. 1994, Addison-Wesley. p. 1-77.
87. Kramer, G., B.N. Walker, T.L. Bonebright, P.R. Cook, J. Flowers, N. Miner, J.G. Neuhoff, R. Bargar, S. Barrass, J. Berger, G. Evreinov, W.T. Fitch, M. Gröhn, S. Handel, H. Kaper, H. Levkowitz, S. Lodha, B. Shinn-Cunningham, M. Simoni, andS. Tipei, *Sonification Report: Status of the Field and Research Agenda. Prepared for the National Science Foundation by members of the International Community for Auditory Display*. 1999, International Community for Auditory Display (ICAD): Santa Fe, NM.
88. Krueger, R.A. & M.A. Casey, *Focus Groups : A Practical Guide for Applied Research*. 3rd Spiral edition ed. 2000: SAGE Publications. 320.
89. Kuniavsky, M., *Focus Groups (Chapter 9)*, in *Observing the User Experience: A Practitioner's Guide for User Research*. 2003, Morgan Kaufmann. p. 201-257.
90. Kwasnik, B.H., *A Descriptive Study of the Functional Components of Browsing*, in *Proceedings of the IFIP TC2/WG2.7 Working Conference on Engineering for Human-Computer Interaction*. 1992, North-Holland. p. 191-203.
91. Kyung, K.-U. & J.-S. Park, *Ubi-Pen: Development of a Compact Tactile Display Module and Its Application to a Haptic Stylus*, in *Proceedings of the Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. 2007, IEEE Computer Society.
92. Landau, S. & L. Wells, *Merging Tactile Sensory Input and Audio Data by Means of The Talking Tactile Tablet*, in *EuroHaptics Conference*. 2004: Trinity College Dublin.
93. Lee, J.C., P.H. Dietz, D. Leigh, W.S. Yerazunis & S.E. Hudson, *Haptic pen: a tactile feedback stylus for touch screens*, in *Proceedings of the 17th annual ACM symposium on User interface software and technology*. 2004, ACM Press: Santa Fe, NM, USA. p. 291-294.
94. Lopresti, D.P. & G. Nagy, *A Tabular Survey of Automated Table Processing*, in *Selected Papers from the Third International Workshop on Graphics Recognition, Recent Advances*. 2000, Springer-Verlag. p. 93--120.
95. Mansur, D.L., *Graphs in Sound: a Numerical Data Analysis Method for the Blind*. 1985, University of California.
96. Marchionini, G., *Information seeking in electronic environments*. First Hardback Edition ed. Cambridge Series On Human-Computer Interaction. Vol. Volume 9. 1995, New York, NY, USA: Cambridge University Press. 224.

97. Marshall, C.C. *Annotation: from paper books to the digital library*. in *2nd ACM international conference on Digital libraries*. 1997. Philadelphia, PA USA.
98. McGookin, D. & S. Brewster. *Graph Builder: Constructing Non-Visual Visualizations*. in *HCI 2006 - People and Computers XX*. 2006. London, UK: Springer.
99. McGookin, D. & S.A. Brewster. *An Initial Investigation into Non-Visual Computer Supported Collaboration*. in *ACM CHI*. 2007. San Jose, CA, USA: ACM Press.
100. McGookin, D.K. & S.A. Brewster. *Contextual Audio on Haptic Graph Browsing*. in *ICAD 2006*. 2006. London.
101. McGookin, D.K. & S.A. Brewster. *SoundBar: Exploiting Multiple VIews in Multimodal Graph Browsing*. in *NordiChi*. 2006. Oslo,Norway: ACM Press.
102. Meredith, M.A., J.W. Nemitz & B.E. Stein, *Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors.* The Journal of Neuroscience, 1987. **7**(10): p. 3215-3229.
103. Miller, G.A., *The Magical Number Seven Plus or Minus Two: Some Limits on Our Capacity for Processing Information.* The Psychological Review, 1956. **63**: p. 81-97.
104. Miller, G.A., *The Magical Number Seven Plus or Minus Two: Some Limits on Our Capacity for Processing Information.* The Psychological Review, 1956. **63**: p. 81-97.
105. Moore, B.C.J., *Auditory Pattern and Object Perception*, in *An Introduction to the Psychology of Hearing*. 2004, Academic Press. p. 269-298.
106. Moore, B.C.J., *Frequency Selectivity, Masking, and the Critical Band*, in *An Introduction to the Psychology of Hearing*. 2004, Academic Press. p. 65-126.
107. Moore, B.C.J., *An Introduction to the Psychology of Hearing*. 5th ed. 2004: Academic Press. 413.
108. Morgan, D.L., *The Focus Group Guidebook*. The Focus Group Kit, ed. D.L. Morgan and R.A. Krueger. Vol. 1. 1998: SAGE Publications, Inc. 102.
109. Murphy, E., G. McAllister, P. Strain, R. Kuber & W. Yu. *Audio for a Multimodal Assistive Interface*. in *Eleventh Meeting of the International Conference on Auditory Display*. 2005. Limerick, Ireland.
110. Neuhoff, J.G., *Ecological Psychoacoustics: Introduction and History*, in *Ecological Psychoacoustics*, J.G. Neuhoff, Editor. 2004, Elsevier Inc. p. 1-13.
111. Neuhoff, J.G. & L.M. Heller. *One Small Step: Sound Sources and Events as the Basis for Auditory Graphs*. in *First International Symposium on Auditory Graphs (AGS2005), at the 11th Meeting of the International Conference on Auditory Display (ICAD05)*. 2005. Limerick, Ireland.
112. Newell, A.F., *Extra-ordinary human-computer interaction*, in *Extra-ordinary human-computer interaction: interfaces for users with disabilities*, A.D.N. Edwards, Editor. 1995, Cambridge University Press: Cambridge & New York. p. 3-43.
113. Nickerson, L.V. & T. Stockman. *Sonically Exposing the Desktop Space to Non-Visual Users: an Experiment in Overview Information Presentation*. in *Eleventh Meeting of the International Conference on Auditory Display*. 2005. Limerick, Ireland.
114. Norman, D.A., *Things that Make Us Smart: Defending Human Attributes in the Age of the Machine*. 1993: Addison-Wesley Pub. Co. 290.
115. Parkes, D. *Nomad: an Audio-Tactile Tool for the Acquisition, Use and Management of Spatially Distributed Information by Partially Sighted and Blind Persons*. in *Second International Symposium on Maps and Graphics for Visually Handicapped People*. 1988. King's College, University of London.
116. Parkes, D., *Tactile Audio Tools for Graphicacy and Mobility: "A circle is wither a circle or it is not a circle".* British Journal of Visual Impairment, 1998. **16**(3): p. 99-103.
117. Pavey, S., G. Douglas & C. Corcoran, *Transition into adulthood and work - findings from Network 1000.* British Journal of Visual Impairment, 2008. **26**(2): p. 202-216.
118. Peretz, I., J. Ayotte, R.J. Zatorre, J. Mehler, P. Ahad, V.B. Penhune & B. Jutras, *Congenital Amusia: A Disorder of Fine-Grained Pitch Discrimination.* Neuron, 2002. **33**(2): p. 185-191.
119. Pérez-Quiñones, M.A., R.G. Capra & Z. Shao. *The Ears Have It: A Task by Information Structure Taxonomy for Voice Access to Web Pages*. in *Human-Computer Interaction -- INTERACT'03*. 2003. Zürich, Switzerland: Published by IOS Press, (c) IFIP.
120. Peterson, L.R. & M.J. Peterson, *Short-Term Retention of Individual Verbal Items.* Journal of Experimental Psychology, 1959. **58**: p. 193-198.
121. PureData. *Pure Data (PD) - http://puredata.info/.* 2008 [cited 2008; Available from: http://puredata.info/.

122. Ramloll, R. & S. Brewster. *An environment for studying the impact of spatialising sonified graphs on data comprehension*. in *Proceedings Sixth International Conference on Information Visualisation, 10-12 July 2002*. 2002. London, UK: IEEE Comput. Soc.

123. Ramloll, R., S. Brewster, W. Yu & B. Riedel. *Using non-speech sounds to improve access to 2D tabular numerical information for visually impaired users*. in *People and Computers XV - Interaction without Frontiers. Joint Proceedings of HCI 2001 and IHM 2001, 10-14 Sept. 2001*. 2001. Lille, France: Springer-Verlag London.

124. Resnikoff, S., Pascolini, D., Etay'ale, D., Kocur, I., Pararajasegaram, R., Pokharel, G.P. & Mariotti, S.P., *Global data on visual impairment in the year 2002.* Bulletin of the World Health Organization, 2004. **82**: p. 844-851.

125. Rigas, D. & J.L. Alty. *The Use of Music in a Graphical Interface for the Visually Impaired*. in *INTERACT 97*. 1997. Sydney, Australia.

126. RNIB, *Work Matters: Enabling blind and partially sighted people to gain and retain employment*. 2002, Royal National Institute of Blind People: London.

127. RNIB. *Data tables - Web Access Centre*. 2008 [cited 2008 23 August]; Available from: http://www.rnib.org.uk/xpedio/groups/public/documents/PublicWebsite/public_datatables.hcsp.

128. RNIB. *How do I produce Braille?* 2008 [cited 2008 23 August]; Available from: http://www.rnib.org.uk/xpedio/groups/public/documents/publicwebsite/public_prodbraille.hcsp.

129. Roberts, J.C., *Visualization Equivalence for Multisensory Perception: Learning from the Visual.* Computing in Science and Engineering, 2004. **6**(3): p. 61-65.

130. Roffler, S.K. & R.A. Butler, *Localization of Tonal Stimuli in the Vertical Plane.* Journal of the Acoustical Society of America, 1967. **43**(6): p. 1260-1266.

131. Rusconi, E., B. Kwan, B. Giordano, C. Umilta & B. Butterworth, *The Mental Space of Pitch Heighta.* Annals of the New York Academy of Sciences, 2005. **1060**(The Neurosciences and Music II: From Perception to Performance): p. 195-197.

132. Saue, S. *A model for interaction in exploratory sonification displays*. in *International Conference on Auditory Display (ICAD)*. 2000. Atlanta, Georgia, USA.

133. Schroeder, F.K., *Perceptions of Braille Usage by Legally Blind Adults.* Journal of Visual Impairment & Blindness, 1996.

134. Scientific, F. *Focus Braille Displays*. 2008 [cited 2008 2nd of September]; Available from: http://www.freedomscientific.com/fs_products/displays_focus40-80.asp.

135. Scientific, F. *Freedom Scientific JAWS for Windows*. [Web site] 2008 [cited 2008 7 September]; Available from: http://www.freedomscientific.com/fs_products/software_jaws.asp.

136. Sellen, A.J. & R.H.R. Harper, *The Myth of the Paperless Office*. 2002, Cambridge, Massachussetts: The MIT Press.

137. SensAble Technologies, I. *PHANTOM Desktop Haptic Device*. 2008 [cited 15th August]; Available from: http://www.sensable.com/haptic-phantom-desktop.htm.

138. Shneiderman, B. *The eyes have it: a task by data type taxonomy for information visualizations*. in *IEEE Symposium on Visual Languages*. 1996. Boulder, CO, USA: IEEE Comput. Soc. Press.

139. Sloboda, J.A., K.J. Wise & I. Peretz, *Quantifying Tone Deafness in the General Population.* Annals of the New York Academy of Sciences, 2006. **1060**(The Neurosciences and Music II: From Perception to Performance): p. 225-261.

140. Smith, S., R.D. Bergeron & G.G. Grinstein. *Stereophonic and Surface Sound Generation for Exploratory Data Analysis*. in *CHI 1990*. 1990.

141. Spence, C. *Crossmodal attention and multisensory integration: implications for multimodal interface design (Invited talk)*. in *5th international conference on Multimodal interfaces* 2003. Vancouver, British Columbia, Canada.

142. Spence, C. & L. Read, *Speech shadowing while driving: on the difficulty of splitting attention between eye and ear.* Psychological Science, 2003. **14**(3): p. 251-256.

143. Stein, B.E., P.J. Laurienti, T.R. Stanford & M.T. Wallace. *Neural mechanisms for integrating information from multiple senses*. in *2000 IEEE International Conference onMultimedia and Expo (ICME 2000)*. 2000. New York, NY, USA: IEEE.

144. Stevens, R.D., *Principles for the Design of Auditory Interfaces to Present Complex Information to Blind People*, in *The Department of Computer Science.* 1996, University of York.: Heslington, York, UK. YO1 5DD. p. 316.

145. Stevens, S.S., J. Volkmann & E.B. Newman, *A Scale for the Measurement of the Psychological Magnitude Pitch.* Journal of the Acoustic Society of America, 1937. **8**(3): p. 6.
146. Stockhausen, K., *Electronische Musik 1952-1960*. Stockhausen Complete Edition. Vol. 3. 2001. 184.
147. Stockman, T. *The Design and Evaluation of Auditory Access to Spreadsheets*. in *Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display*. 2004. Sydney, Australia.
148. Stockman, T., G. Hind & C. Frauenberger. *Interactive Sonification of Spreadsheets*. in *ICAD 05-Eleventh Meeting of the International Conference on Auditory Display*. 2005. Limerick, Ireland.
149. Stockman, T., L.V. Nickerson & G. Hind. *Auditory Graphs: a Summary of Current Experience and Towards a Research Agenda*. 2005. Limerick, Ireland.
150. Taylor, S.E., H. Frackenpohl & J.L. Pettee, *Grade level norms for the components of the fundamental reading skill.* EDL Research and Information Bulletin, 1960. **3**.
151. Technologies, S. *SensAble Technologies - Haptic Devices*. 2008 [cited 2008 20-1-2008]; Available from: http://www.sensable.com/products-haptic-devices.htm.
152. Tillsley, C., *Gaining Access to Employment Opportunities.* British Journal of Visual Impairment, 1997. **15**(2): p. 67-71.
153. Tukey, J.W., *Exploratory Data Analysis*. 1977, Don Mills: Addison-Wesley Publishing Company.
154. University, O. *Oxford English Dictionary*. 2008 [cited 2008; Available from: http://www.oed.com.
155. van Scoy, F.L., T. Kawai, M. Darrah & C. Rash, *Haptic display of mathematical functions for teaching mathematics to students with vision disabilities: design and proof of concept*, in *Haptic Human-Computer Interaction* 2001, Springer: Berlin / Heidelberg. p. 31-40.
156. van Scoy, F.L., D. McLaughlin & A. Fullmer. *Auditory Augmentation of Haptic Graphs: Developing a Graphic Tool for Teaching Precalculus Skills to Blind Students.* in *ICAD 05 - Eleventh Meeting of the International Conference on Auditory Display*. 2005. Limerick, Ireland.
157. Vanderheiden, G., *Nonvisual Alternative Display Techniques for Output from Graphics-Based Computers.* Journal of Visual Impairment and Blindness, 1989.
158. Vickers, P. & J.L. Alty. *Musical Program Auralisation: Empirical Studies*. in *Sixth International Conference on Auditory Display (ICAD 2000)*. 2000. Atlanta, GA.
159. VirTouch. *VTPlayer Mouse*. 2008 [cited 2008 2 September]; Available from: http://www.VirTouch2.com.
160. Walker, B., N. & G. Kramer, *Mappings and metaphors in auditory displays: An experimental assessment.* ACM Trans. Appl. Percept., 2005. **2**(4): p. 407-412.
161. Walker, B.N., *Magnitude Estimation of Conceptual Data Dimensions for Use in Sonification.* Journal of Experimental Psychology: Applied, 2002. **8**(4): p. 211-221
162. Walker, B.N. & J.T. Cothran. *Sonification Sandbox: a Graphical Toolkit for Auditory Graphs*. in *International Conference on Auditory Display (ICAD)*. 2003. Boston, MA, USA.
163. Walker, B.N., M.T. Godfrey, J.E. Orlosky, C. Bruce & J. Sanford. *Aquarium Sonification: Soundscapes for Accessible Dynamic Informal Learning Environments*. in *12th International Conference on Auditory Display*. 2006. London, UK.
164. Walker, B.N. & G. Kramer, *Ecological Psychoacoustics and Auditory Displays. Hearing, Grouping, and Meaning Making*, in *Ecological Psychoacoustics*, J.G. Neuhoff, Editor. 2004, Elsevier Inc. p. 149-174.
165. Walker, B.N. & D.M. Lane. *Psychophysical Scaling of Sonification Mappings: A Comparision of Visually Impaired and Sighted Listeners*. in *Proceedings of the 2001 International Conference on Auditory Display,*. 2001. Espoo, Finland.
166. Walker, B.N. & M.A. Nees. *An Agenda for Research and Development of Multimodal Graphs*. in *First International Symposium on Auditory Graphs (AGS2005), in the Eleventh Meeting of the International Conference on Auditory Display (ICAD-05)*. 2005. Limerick, Ireland.
167. Wall, S. & S. Brewster. *Providing External Memory Aids in Haptic Visualisations for Blind Computer Users*. in *Proceedings of International Conference Series on Disability, Virtual Reality and Associated Technologies (ICDVRAT)*. 2004. New College, Oxford, UK.
168. Wall, S. & S. Brewster. *Feeling What You Hear: Tactile Feedback for Navigation of Audio Graphs*. in *CHI 2006*. 2006. Montreal, Quebec, Canada: ACM.

169. Wies, E.F., J.A. Gardner, M.S. O'Modhrain, C.J. Hasser & V.L. Bulatov. *Web-based touch display for accessible science education*. in *Haptic Human-Computer Interaction. First International Workshop, 31 Aug.-1 Sept. 2000*. 2001. Glasgow, UK: Springer-Verlag.

170. Yesilada, Y., R. Stevens, C. Goble & S. Hussein, *Rendering tables in audio: the interaction of structure and reading styles*, in *Proceedings of the ACM SIGACCESS conference on Computers and accessibility*. 2004, ACM Press: Atlanta, GA, USA. p. 16-23.

171. Yu, W. & S. Brewster, *Evaluation of multimodal graphs for blind people.* Journal of Universal Access in the Information Society, 2003. **2**(2): p. 105-124.

172. Yu, W., R. Ramloll & S.A. Brewster, *Haptic Graphs for Blind Computer Users*, in *Proceedings of the First International Workshop on Haptic Human-Computer Interaction*. 2001, Springer-Verlag. p. 41--51.

173. Zhao, H., C. Plaisant, B. Shneiderman & R. Duraiswami. *Sonification of Geo-Referenced Data for Auditory Information Seeking: Design Principle and Pilot Study*. in *International Conference on Auditory Display (ICAD)*. 2004. Sydney, Australia.