An Investigation of Interactions with Multimodal Graphs on Mobile Devices

Zico Pratama Putra

Submitted in partial fulfilment of the requirements of the

Degree of Doctor of Philosophy

School of Electronic Engineering and Computer Science

Queen Mary, University of London

August 12, 2020

Statement of Originality

I, Zico Pratama Putra, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below. I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material. I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis. I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university. The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date of Submission: 12th August 2020

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

GUI - Graphical User Interface

HCI - Human-Computer Interaction

MAG – Mobile Auditory Graph

RMSE - Root Mean Squared Errors

SCE - Semantic Congruence Effect

SNARC - Spatial-Numerical Association of Response Codes

UI - User Interface

VI – Visually Impaired

$Y_{Estimate}$ – The value of the user estimation

Y_{Max} - The maximum Y value

Abstract

The increasingly broad spectrum of interaction contexts has pushed back the limitations of the Graphical User Interface. However, while the benefits of multimodal computing are increasingly to be found, visually impaired users remain faced with many challenges that prevent them from fully exploiting the benefits of graphical information. This thesis aims to contribute to the research area of accessible graphical information, and to propose a methodological framework for improving multimodal graph interaction.

The experiments described in this thesis employ mobile "tablet" devices, as these are an already well established tool within education, and their form factor appears to be well suited to undertaking tasks represented on a surface which is both accessible and provides sufficient space to afford a fair degree of graphical resolution. Three central questions examined in this thesis are as follows:

1) How accurately can visually impaired users estimate the values of data points rendered in auditory graphs presented on a mobile device?

2) Are there modes of interaction which can improve the ability of visually impaired people to perform point estimation tasks presented on a mobile device?

3) What format should the auditory display take to enable accurate understanding and efficient processing of auditory graphs?

An analysis of point estimation errors and the correlation between the predicted and actual data points was used to examine the first question. The way in which RMSEs and correlation values vary, generally worsening, as the numbers of data points in the presented auditory graphs are increased is described in detail.

Multi touch gestures are then investigated as an alternative approach to passive listening as a means of making point estimation tasks more active and engaging, which in turn might lead to improved performance (question 2). The investigation showed that the additional touch modality enabled visually impaired users to perform point estimation tasks with higher correlations with actual values and lower point estimation errors. The analysis reveals that combining audio playback with user interaction offers an advantage over auditory graph presentation requiring only passive listening. In the final two studies of the thesis, we examine different approaches to the presentation of Y coordinates in auditory graphs (question 3), including the representation of negative numbers. These studies involved both normally sighted and visually impaired users, as there are applications where normally sighted users might employ auditory graphs, such as the unseen monitoring of stocks, or fuel consumption in a car.

A mixed methods approach was employed combining quantitative statistics with qualitative data from interviews and informal feedback to form a rounded picture of the results of the studies. The experiments employed tablet-based prototypes and data was captured primarily using audio recordings, notes on a laptop and digital timing data. Participants were recruited appropriately from the visually impaired and normally sighted populations, and were mostly resident either in London or Jakarta.

Multi-reference sonification schemes are investigated as a means of improving the performance of mobile non-visual point estimation tasks. The results showed that both populations are able to carry out point estimation tasks with a good level of performance when presented with auditory graphs using multiple reference tones. Additionally, visually impaired participants performed better on graphs represented in this format than normally sighted participants. This work contributes to the introduction of a new multimodal approach, based on the combination of audio and multi-touch gesture interaction, contributing to more accurate point estimation and graph reproduction tasks, improving the accessibility of tablet and smartphone user interfaces.

Chapter 1. Introduction

Researchers have performed various investigations to comprehend human interpretation of visual graphs in reading, comprehending, and interpreting displayed data. They have found several teaching methods to enhance the skill of understanding and delivering visual graph (Bowen & Roth, 1998; Fischer, 2000; Shah & Freedman, 2011). Full-scale graph comprehension theories to make predictions for future studies and to create applications with a visual graph have been proposed (Bowen & Roth, 2005; Pinker, 1990). However, little has been done to study the individual perception and interpretation of auditory graphs as contrasted with the broad literature and information resources available for visual graphs (Bonebright, 2005).

The utilisation of auditory graphs has received some attention in recent years in a fair range of application scenarios. AudioGraf was one of the early systems to make graphics readable by using a tap panel and an auditory display (Kennel, 1996). King et al. (2004) developed the TeDUB project to present the Unified Modelling Language (UML) in graphs accessible for the visually impaired (VI) listener. Cohen et al. (2005) designed PLUMB support people with visual impairments (VI) to understand graphs and data structures using auditory cues. A recent study has developed further using a Graph Sketching tool to include VI users in computing and other science, technology, engineering, and mathematics (STEM) disciplines in which graphs are essential (Balik et al., 2014).

Audio in the interface is, in general, becoming more important as technologies get smaller and portable. With shrinking screen sizes, there is less space in which to display information, thus increasing the importance of using audio to convey information. Despite these clear reasons to support research on mobile auditory graphs, little has been done to explore multimodal mobile graphing systems on mobile devices. Investigating how to expand the functionality of auditory graphs then implement it into a user-friendly Mobile Auditory Graphs application (MAG app) could help VI users to have a better interpretation of the shape of graphs presented on a portable device. This MAG app could perhaps serve as a good alternative for VI users instead of the portable traditional embossed graphs.

1.1. Motivation

The use of many kinds of mathematical and statistical graphs is prevalent in STEM areas. Their visual nature, however, is still a major challenge for VI users, presenting one of the key barriers to their having an equitable learning environment with their sighted peers. This PhD project is aimed at improving the accessibility of graphs for VI users in a computing environment which is already well accepted and widely used in mainstream education.

Many studies have been proposed to improve STEM education for VI students using sonification sounds in learning (e.g., Walker (2010), King, et al., (2004), Cohen et al., (2005), Balik, (2014), Goncu and Marriot, (2015)). For example, Walker (2010) proposed a systematic study of creating useful and usable auditory graphs. Moreover, Goncu and Marriot (2015) have developed sonified graphs presented on a web-based service. In an open discussion forum at the ICAD 2018 conference, the possibility of developing a system that could automatically generate auditory graphs for any graphs in Wikipedia was discussed. This was discussed as one possible way forward in helping to give the sonification of data a higher profile and bring it more into the mainstream.

Despite these efforts, there is still a lack of systematic means of creating useful audio graphs for mobile interaction, even though audio in the mobile interface is becoming more important as technologies get smaller, portable, and ubiquitous.

The form factor of tablets appears to provide a good compromise in providing portability while presenting sufficient surface area for VI users to comprehend the layout of a graph or chart. Therefore, the overall objective of this PhD project is to provide auditory graphs in a form that is both portable and highly usable to people with VI studying or working in STEM disciplines. For this purpose, we developed the Mobile Auditory Graphs (MAG) app which will give VI users the possibility to use multimodal and cross-modal interaction mechanisms employing audio and gestures which are now becoming a common standard in mobile interaction, e.g., voice recognition, tapping, and swiping as multimodal inputs with the help of a screen reader. To the author's knowledge, this study is the first attempt to incorporate the auditory graph embedded in a portable device system using a multimodal approach.

For the studies in this project, we deliberately take a task-based approach. This is for several reasons:

1) Using specific tasks enables us to collect quantitative data for comparative purposes, for example, between sighted and VI users, or between the same users under different experimental conditions.
2) Employing the same, or very similar tasks as previous researchers such as Walker (Walker 2010) and Metatla (Metatla 2016) facilitates comparisons with previous work.
3) The tasks chosen, primarily point estimation and graph reproduction, are highly relevant to many scenarios in STEM subjects in education and employment, helping to frame the research in a context more likely to be relevant to real use scenarios.

This focus on tasks is not, however, to detract from the value of qualitative data in canvassing reactions, opinions, suggestions, and other comments. All studies reported in this work also rely on qualitative data to capture these elements which form a crucial part of studies examining usefulness and usability.

1.2. Research Question

This thesis aims to improve the accessibility of graphs for VI users in a Computing environment that is viable and accepted in many educational and employment settings. It aims to develop a methodological framework for improving multimodal graph interaction. To meet these objectives, this study will address three main research questions:

1) **How accurately can visually impaired users estimate the values of data points rendered in auditory graphs presented on a mobile device?**

   We start with an exploratory study, reported in chapter 4, to test the system's general feasibility with sighted users. This exploratory study is then used to provide a foundation for further implementation of the MAG app for use by VI users. We then explore this question for point estimation and graph reproduction tasks using sonification in chapter 5.

2) **Are there modes of interaction that can improve visually impaired people's ability to perform point estimation tasks presented on a mobile device?**

   We introduce a new multimodal approach based on multi-touch gesture interaction. We then address this second research question in the study described in chapter 5 by examining whether VI users can employ multi-touch gestures to obtain an improved interpretation of sonified graphs compared with interpretations using passive listening.

3) **What format should the auditory display take to enable accurate understanding and efficient processing of auditory graphs?**

This question is explored in some depth in the 3rd and 4th studies, respectively reported in chapters 6 and 7.

Chapter 6 introduces the idea of multi-reference sonification, describing an approach used by Metatla (Metatla 2016). We propose a new approach that addresses two weaknesses of Metatla's approach: (a) that it becomes lengthy in duration and (b) that it becomes increasingly cognitively demanding when numbers far from 0 are involved. Chapter 6 goes on to describe an experimental evaluation of this new approach.

Chapter 7 describes a comparative evaluation of four different sonification schemes. A possible criticism of the multi-reference sonification approach is its complexity, a valid question being "is there not a danger of throwing simplicity out of the window in the search for sonifications that provide additional information to improve contextual understanding of the data?".

For this reason, the sonification schemes evaluated in chapter 7 include simple schemes such as pitch-only and single-reference sonifications. For comparison with these, and with each other, the study also includes two variants of the approach we proposed and evaluated in chapter 6. Detailed descriptions of all four sonification approaches employed are given in full in chapter 7.

1.3. Contributions of the work

The main contributions of this thesis are:

- Introduction of an accessible and functional prototype demonstrating multimodal interaction with auditory graphs.
- The ability of sighted users to use the MAG 1.0 app to perform point estimation and graph reproduction tasks with a fair level of accuracy. The prototype helped the sighted users obtain more information to identify the auditory graph shape in a mobile device with relatively limited space.
- Introduction of a new multimodal approach based on the combination of audio and multi-touch gesture interaction, leading users to have a more accurate mental model of the graphs and improving the accessibility of tablet and smartphone user interfaces'.
- Comparison of point estimation performance between passive listening and multi-touch gesture interaction. Study 2 shows that the RMSEs of multi-touch gesture interaction were

distributed equally for almost all conditions, while the results obtained for passive listening tended to increase. Unlike passive listening mode, which transmits the auditory graph unidirectionally from the device to the user, a key feature of multi-touch interaction is the bi-directional flow of information to and from the user, allowing the user to perceive and actively engage with the system. Touch sensations combined with audio effectively close a feedback control loop between the system and the user, providing cues to the user, enabling them to actively and intuitively control the interaction. Therefore, this contribution increases the understanding of this aspect of human perception and interaction.

- The provision of new information on user perceptual skills, listening experience and the accuracy of auditory graph perception and comprehension.
- Introduction of design practice to improve the performance of non-visual point estimation tasks by implementing multi-reference sonification mapping in auditory graphs. In general, the study show that the multi-reference mode generated more accurate results compared to the single point modality. The evaluation confirms previous research that adding context to auditory graphs such as multiple reference tones enhances auditory graph perception.
- The accuracy of VI users can surpass that of sighted users. This finding seems to be particularly the case when there is sparser (less contextual) information available in the sonification approach used to render the auditory graphs. Similarly, the ability of VI users appears to be more robust to changes in the sonification approach employed.

It is hoped that this approach can trigger a discussion in this community on how the process of point estimation in the mobile domain - no matter in which application area - can be demystified and embedded in a fundamental framework. The mobile auditory graph was explicitly designed to allow these communities to create a common basis for an interaction model that can be used to leverage previous work. In the long term, this work should impact the exchange of best practices and make effective mobile auditory graphs more widely used in everyday technology.

1.4. Outline of the report

This report is organised as follows:

Chapter 1 gives a brief introduction to the project, the motivation behind it, some initial background, and the research questions to be addressed.

Chapter 2 describes several related research fields that contribute to building the foundation for understanding the work described in later chapters.

Chapter 3 describes the research methods employed in each stage of the project and the rationale behind them. It also describes the development of the successive versions of the MAG app.

Chapter 4 presents an exploratory study to test the usability of the MAG app for sighted users. This chapter describes our investigation of whether the complexity of audio graphs influences sighted users' ability to perform graph reproduction tasks.

Chapter 5 presents an exploratory study to test the usability of the MAG app for VI users. The study in this chapter includes investigating the effect of adding complexity or additional modalities on the auditory graphs to improve point estimation and graph reproduction tasks.

Chapter 6 describes the evaluation of our proposed multi-sonification approach with sighted users undertaking point estimation tasks, focusing on the size of errors and task completion times.

Chapter 7 describes a comparative evaluation of four sonification schemes by sighted and VI users undertaking point estimation tasks. The chapter includes an evaluation of a representation of negative numbers. This representation is evaluated across all four sonification conditions and both sighted and VI populations.

Chapter 8 summarizes the work undertaken, describes the original contributions to research, and develops some ideas for future work.

1.5. Associated Publications and Presentations

Putra. Z, Setiawan. D, Point Estimation with Markers for Effective Mobile Auditory Graphs, iJET (International Journal of Emerging Technologies in Learning), 2020.

Putra. Z, Exploring the Feasibility of Mobile Auditory Graph on Graph Reproduction Task, JESTEC (Journal of Engineering Science and Technology), 2020.

Setiawan. D, Putra. Z, Designing a Multimodal Graph System to Support Non-Visual Interpretation of Graphical Information, Journal of Physics, IOP, 2019.

Putra. Z, Setiawan. D, Multi-Touch Gesture of Mobile Auditory Device for Visually Impaired Users, 2nd International Conference on Broadband Communications, Wireless Sensors and Powering (BCWSP), Yogyakarta, Indonesia, 2020.

Putra. Z, Setiawan. D, Investigation of Multi Reference Point Estimation with Timbre for Effective Mobile Auditory Devices, 2020 International Conference on Computer and Information Sciences (2020 ICCIS), Kingdom of Saudi Arabia, IEEE KSA, 2020.

Chapter 2. Literature review

2.1. Overview of the chapter

This chapter discusses the roots and conceptual problems of the interdisciplinary and progressive field of auditory graphs. It begins by discussing the history of sonification, the types and classifications of auditory display with sound, also their benefits and limitations. It also explains how the characteristics for a better portable design of auditory graphs were proposed. This lead to the implementation of auditory graphs on mobile devices using touchscreen as a medium of interaction. This idea is followed by a discussion on the concept of touchscreen and the theory behind its interaction technique.

The projects also investigate and identify the challenges for VI users to understand and interpret an audio graph presented on mobile devices. While mobile devices offer interactive features with different modalities, which are sometimes not available on computer desktops, the use of sonification using a tactile representation for displaying graphs on mobile devices opens many possibilities to be implemented. Therefore, reviewing some literature closely related to this topic was intended to construct a solid background to support this idea and its technical implementation. For that purpose, this review will be broken down into five different main topics:

> Section 2.2 deals with the concept of acoustic and sound perception. This section discusses some terms on auditory research regarding pitch, loudness and timbre. It also includes discussion on the diatonic scale based on the intervals formed by natural tones that are neither flat nor sharp, and is constructed on seven whole levels of perfect fifths: C - G - D - A - E - B - F.
>
> Section 2.3 reviews the early history of auditory display and its development. This section is followed by a section discussing the advantages and challenges of auditory displays. It also includes different types of auditory displays and techniques for rendering its representations. Parameter Mapping has been the predominant technique for representing sonification, while one implementation of the parameter mapping techniques is the auditory graph.
>
> Section 2.4 discusses the relevant factors in the comprehension of auditory graphs by human listeners. The display design of auditory graphs begins with how to represent some quantitative data. Discussion on the graphical context that describes several

aspects of the auditory graphs to interpret data is especially critical in this stage, followed by decisions regarding mapping, scaling, and polarity.

Section 2.5 reviews the important issue of how to estimate the data values mapping to the sound dimensions. The section discusses further auditory recall techniques that are used to identify memory structure and to retrieve a list of elements in the order of their presentation. Furthermore, we discuss the role of negative number for auditory graphs as part of the fundamental concept for our work in Chapter 7.

Section 2.6 takes a closer look at the study of auditory graphs on mobile device implementation. This section starts with presenting the accessibility features of the mobile device that support auditory graph representation, followed by discussing the shortage of recent approaches in this area.

Section 2.7 provides a review of the current state of the art of touchscreen accessibility for people with visual impairments and identifies new directions for research. This section discusses touchscreen interaction strategy, i.e., pointing strategy, scrolling strategy, navigation, and cross-selection. While these strategies may not gain their optimum functionalities on a screen reader, we extend the discussion further by adding the multimodal interaction techniques to improve the performance and accuracy of interpreting data. As the work of Chapter 5 involves VI users accessing the multi-touch gesture, this section discusses the effort made in haptic interaction for VI users and its challenge related to the perception and manipulation of objects using the senses of touch.

2.2. Acoustics and Psychoacoustics

Acoustics refers to the investigation of the mechanical motion waveforms that travel through gases, liquids, and solids. Acoustics' main and fundamental component is the study of propagation and physical properties of sound waveforms (Rossing, 2014).

Psychoacoustics involves exploring sound perception. The perception and low-level interpretation of sounds lead to comprehension of sound that differs from the actual sound produced. Psychoacoustics also investigates, to some extent, the way people make sense of perceived sounds.

2.2.1. Acoustic Properties

It is common for researchers to generate sounds for studying. A sine wave is one of the basic sounds to control the waveform properties. In a single-audio speaker, a sine wave generated has a specified frequency and amplitude. Frequency is the compression wave sound amount at a particular listening point over the course of a second, also called Hertz (Hz) (Davison, 2013).

Amplitude refers to compressive change of the wave, visually referred to as the height of the sine wave. Amplitude is often expressed in decibels (dB), a logarithmic scale (Davison, 2013).

The third property in acoustics is the complexity of the sound (Schiffman, 1977). Sounds of nature rarely sound like a sine wave. These sounds frequently have several frequencies with different amplitudes and waveforms that change over time (Schiffman, 1977).

Redundant acoustic cues and complex MIDI notes were used to illustrate differentiation in this document's graphic representation.

2.2.2. Perception

Schiffman (1977) defined pitch as the perception of frequency. Pitch, measured in mels[1], varies in a complex way, depending not only on frequency but also amplitude and complexity. With a knob, Stevens et al. (1937) summarizes the relationship to frequency by asked the listeners to modify the frequency and specify when the frequency perception was halved, see Figure 2-1.

[1] *Mel* was chosen as the name for the subjective pitch unit. The name was taken originally from the root of the word *melody* (Stevens et al., 1937).

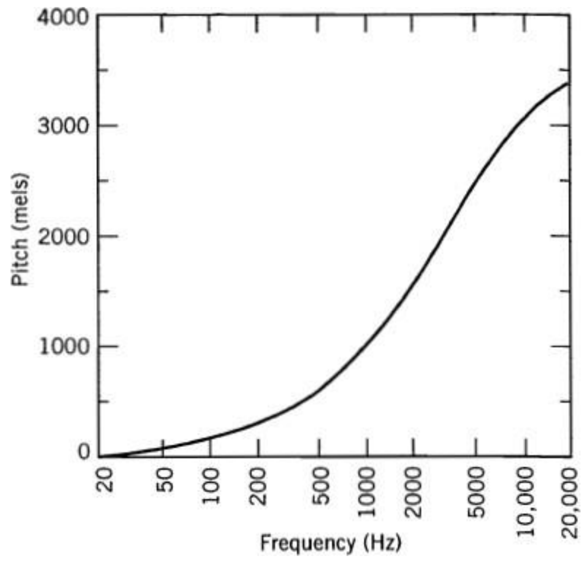


Figure 2-1 The function between frequency and pitch, measured by Stephens et al. (1937). At low frequencies, the pitch changes rapidly, while at high frequencies, it changes more slowly. The frequency is logarithmic.

The pitch can, in practice, be regarded as a musical note doubled every 12 semitones, which is equal to 1 octave. A musical frequency assignment can be written as:

$$f = 2^{\frac{N}{12}}\ x\ 220Hz \qquad (1)$$

N equals the semitone difference of A3 (e.g., B3 is +2 because there are two (A) semitones. A3 is defined as 220 Hz. For each octave, the frequency will be doubled (Schiffman, 1977). For an octave of 12 semitones, each consecutive pitch is calculated by dividing (decreasing) or multiplying (increasing) the pitch of the preceding one by the 12[th] root of two, with f as the resulting frequency.

Loudness or volume is the perceived amplitude. A sone is the volume of a 1000 Hz sine wave at 40 dB SPL. Threefolding the intensity (10dB increase) doubles the volume. The loudness is also strongly influenced by the frequency of the sound (Schiffman, 1977). Noise between 200-5000 Hz is more sensitive to humans and they are less sensitive to sounds with the same amplitude but outside this frequency range.

In music, timbre or tone color is defined as the perceived sound quality of a musical note, sound or tone. The fundamental frequency, as Schiffman (1977) states, "mainly determines the pitch of a complex sound, while its overtones determine its timbre". Musical instruments are distinguished from each other mainly by their timbre.

A diatonic scale in music theory is a heptatonic scale that contains five complete tones (complete steps) and two semitones (half steps) for every octave (Bordman, 1876). The two half tones being divided by two or three complete tone steps depending on their location in the scale. In a diatonic scale comprising more than one octave, this pattern verifies that at least two whole steps separate all semitones. Therefore, F gives us F—C—G—D—A—E—B. Place them in the correct order and they will give F G A B C D E, better known as "F lydian". That is the same thing "2221221" has given in whole and semi steps (with '2' as the symbol for the whole tone steps and '1' for the half tone steps). Its pattern turns out to be conveniently the same as the major scale "2212221", except that it has been transposed a few positions further. A diatonic scale is formed by each sequence of seven consecutive natural tones, such as C-D-E-F-G-A-B, and each transposition of them (Clough, 1979).

2.3. Auditory display and sonification

2.3.1. History of auditory display and sonification

The study of auditory displays, which deals with the use of non-speech sound to display information, has become very active in the last thirty years. The implementation of auditory displays has been engaged in various complex work environments, ranging from computer applications, aircraft cockpits, medical workstations, and control centres of atomic reactors (Brewster, 2002). The main goal in designing these auditory displays is to optimize the level of conformity between the intended information and the information obtained by listeners using their cognitive experience.

A specific type of auditory displays is sonification, a technique used to typically map data sets to acoustic parameters to represent the data audibly. It usually represents information in non-speech audio (Walker & Nees, 2011). It can help users analyse the trend of data and its distribution by hearing the sound as representing the rendered acoustic data. Data presented using sonification has a benefit to be perceived as it is broader and clearer than the speech sound which is precise and demands more focus (Brewster, 2002).

The Geiger counter, invented in 1908, is one of the earliest and highly successful applications of sonification. Geiger counter presented data using sounds by relying on change in pitch or alteration in the rate of audible clicks or parameters range modification, i.e., amplitude or frequency (Mcgee, 2009). The Geiger counter meter had a low-pressure gas tube; each detected particle generates a current pulse, which can create an audio click when the particle ionizes the gas. The original version was only able to recognize alpha particles.

Meanwhile, the quantitative assessment of the perception of sonification was carried out by Pollack (Pollack, 1952), who was mainly concerned with evaluating information transmission properties of hearing stimuli. He evaluated two different mappings of multidimensional data on parameters of sound. His results showed that multi-dimensional displays, i.e., displays that use several parameters of sound, surpass one-dimensional displays. Therefore, dividing the display dimensions into more detailed levels does not so much improve the transmission of information but rather increases the number of display dimensions. A decade later, Speeth (1961) experimented with auditory data to improve ways to distinguish earthquakes from underground bomb explosions based on seismic measurements.

Bly (1982) conducted a more extensive study of auditory representation in which she analysed auditory representation for three data classes: multivariate, logarithmic, and time-varying. When looking at multivariate data, her main focus was to distinguish unordered quantities of multivariate data points and assign either one or the other to an unknown data point. By comparing displays that were either sound only, graphics only, or bimodal, she experimented with different mappings and training methods. In her design, she found that auditory representation was as effective as visual representation and the bimodal display was better than either mode alone.

Two years later, Mezrich et.al. (Mezrich, Frysinger, & Slivjanovski, 1984) designed both a dynamic visual and an auditory representation of multivariate time series displays. Time-series data are typically presented graphically in x-y diagrams, whereby the different dimensions are either overlaid, stacked, or presented on separate axes. Although time-series presentation was initially designed for oil drilling, a kind of multivariate time series commonly used in the petroleum industry, the nature of this data often made it hard to access meaningful examples. Therefore, they used economic indicators statistically similar to well protocols and generally available to the public. The Mezrich et al. (1984) display lets the user hear and the undisturbed indicators without experiencing "sensory overload" and allows interaction with the data display that was not available before. The analyst is presented with a multivariate sample from the time series at any given time in their scheme instead of the entire data set.

According to Frysinger (2005), the individual economic indicators are univariate time series that describe variations over time of things like car sales and housing construction. When taken individually, they are usually not attractive, but they provide insight into an economic situation when used in combination. In practice, they are only visually represented and analysed for "interesting behaviour". However, visual representations are often difficult to make meaningful visual inspections, especially when the range of indicators increases. Therefore, indicators are commonly arranged into an index using weighted linear combinations, which are then used as economic predictors. The problem is that in the absence of a reasoned interaction model between the indicators, the weighted combinations have questionable validity and tend to have the information available in each time series.

However, Frysinger (2005) point out that until the 1980s and early 1990s (e.g., when the International Conference on Auditory Display (ICAD)[2] series began), relatively little progress has been made in this area. He argues that at least one technological reason for this; due to the introduction of sound cards for personal computers in the mid-1980s and the development of the MIDI standard making digital sound generation possible.

The first ICAD was held later in 1992. Founded by Gregory Kramer, the ICAD was working together on the development of this discipline and has evolved into a forum for researchers from various disciplines interested in using sound to deliver information through its events and peer-reviewed proceedings. Researchers and practitioners have equipped the field with taxonomies, tools, techniques, and methods to support the creation, development, and presentation of auditory displays.

[2] www.icad.org

2.3.2. Benefits and difficulties of auditory display

Human hearing is very efficient, and it can distinguish different data sound from the pressure waves entering ears. Thus, sounds give us the perception of the environment and the situation around us. The sounds are transitory and need time to be rendered, but sounds give a better overview than textual descriptions. Although the latter is more precise, they are harder to generate and need a longer time to be listened to than the sounds (Baddeley, 1997).

Non-speech sounds could enhance speech information as visual icons enhance text information (Brewster, 2002). For instance, icons can display data in a narrow area instead of displaying it on texts, therefore non-speech sounds may broadcast data in a shorter time than speech sounds.

However, the non-speech sound can mask or confuse other acoustic signals, such as voice communication, which can create difficulty in using acoustic displays for particular applications. In addition, the acoustic output can be annoying or disturbing. Therefore, more works need to be considered on graphs representing non-speech sounds than on graphs being represented using speech sounds.

Studies have shown that VI people can read line diagrams that are sonified with a musical note by rendering each data point (Mansur, Blattner, & Joy, 1985; Sahyun, 1999). The auditory display allows both blind (using a screen reader) and sighted users who use their eyes for other tasks to use them without the need for their vision. Quick recognition of acoustic signals and the all-around perception of hearing can contribute to an auditory display's effectiveness even when vision is present. While this is the case, to verify that sonification and auditory graphs are practical and effective, the audio display designer must consider the end user's perceptual and cognitive expectations as discussed in the next section.

2.3.3. Types of auditory display

2.3.3.1. Audification

Audification is an auditory display technique for translating a sequence of data values as sound (Hermann, 2002). In audification, the data sequence, typically a time series, is interpreted as an acoustic signal waveform by mapping the incoming data to the pressure level of a sound, measured in decibels (dB). Often different waveform handling procedures are applied to highlight unusual data characteristics.

Audification is particularly suitable for large amounts of data with periodic components. Therefore, many data values are required to perform an audification, so that audification enables the audience to hear periodic components as frequencies (Hermann & Ritter, 2004).

According to a 2005 study by Pauletto, users auditing time series data could identify properties ranging from noise, periodic vibration, discontinuities, repeating patterns, and signal strength to an extent like that found in visual spectra analysis (Pauletto & Hunt, 2005).

Some of the areas of application are the seismic records audification (Dombois, 2001) and the verification of neurophysiological human signals (Olivan, Kemp, & Roessen, 2004).

2.3.3.2. Auditory icons and earcons

Auditory Display can include various sound representations, such as the use of acoustic signals ("earcons") as tracking instruments for presenting data as directly as possible. Mynatt (1994) has investigated a way to design auditory icons defined by Gaver (1986) as everyday sounds mapped into several computer events. Blattner (1989) uses the term ‘earcons' to refer to the auditory message sent by applying general, synthetic tones in structured combinations used to present information about several computer objects, operations or interactions. Earcons are a popular feature of computer platforms and programs, from beeping when an error occurs to the customizable sound schemes of Windows 10 that feature start, shutdown, and many other events. Compared to visual symbols, the earcons are abstract. Therefore their meaning has to be learned (Brewster, 2002).

2.3.3.3. Parameter mapping sonification

The sonification techniques used so far focus both on high volumes of data (audification) or discrete signals (auditory icons and earcons). Still, the method of parameter mapping is a reliable representation. It is the best-used sonification technique to represent large-scale data as sound (Hermann, 2002; Hermann & Ritter, 2004). Sonifications for parameter mapping can be described as acoustic scatter graphs (Flowers, Buhman, & Turnage, 1997) or as parameter mapping of n^{th}-degree. The concept of mapping is based on the data plotting technique: Scatter plots add graphical symbols (elements) to the area of display and the attribute symbols (x, y coordinate, position, dimension, color, or symbol type) can be controlled using adjustable values of the displayed data set. Data dimensions are usually mapped to sound parameters such as cognitive complexes (timbre, rhythm), psychophysical (pitch, volume), or physical (frequency, amplitude) (Worrall, 2009).

In the framework of de Campos (2007) sonification design spaces map, auditory graphs is implemented by using parameter mapping sonification techniques that contained three large groups of sonification approaches, i.e., event-based, continues-based, and model-based. Event-based displays work mainly symbolically; continuous displays mostly analog, when data are time series and signal, can be directly translated into sound; and model-based displays mostly analog, but in a discontinuous way. Model-based approaches focus strongly on the user's active manipulation of sonification and tend to be highly data-dimensional (de Campo, 2007).

2.4. Auditory graph design

In the auditory graph design, the question is centered on how the sound dimensions can be mapped to the displayed data. The main mapping issue includes whether pitches should be increased or decreased in response to changes in the associated data. Auditory graphs can be considered a group of sonified displays that use audio to display numerical information. This mean that when quantitative information is modified, all modifications are mapped to reflect alterations in one or more dimensions of the sound. As part of the auditory display framework, auditory graphs may solve the audio clutter arising from an attempt to listen to many numeric values in speech. Imagine the difficulty of trying to remember ten or more data values spoken out loud. In comparison, non-speech sounds make audio graphs easier to follow by merely listening to the trend of a constant sound whose pitch is changing according to the values in the dataset.

As in the visual graphs, the auditory graph characteristics need to be set up properly so that the listener can understand the meaning of data. While the properties of the visual graphs (i.e., spatial area, colour, trend, and size) are regularly changed, the audio properties in sonification, like tempo, loudness, pitch, pan, and timbre may be changed. These properties describe some mappings to the sound attributes such as loudness (identifies with the sound's amplitude), pitch (a feature that relates to the frequency of sound) and timbre (a characteristic of a sound that identifies it from the various reference of a similar pitch and volume) (Nesbitt & Barrass, 2004).

Walker et al. (2000) investigated these questions by comparing polarities, scale-functionality, and data-to-display mappings to correlate data values with associated audio characteristics for sighted and VI people. They discovered that in some circumstances, VI listeners might prefer opposite polarities to sighted listener perceptions. They found that for a specific mapping, for example, mapping coin size to pitch, VI people tended to have an opposite mapping compared to sighted people.

Brown et al. (2003) explored the question of mapping sound and formulated guidelines for the design of auditory graphs based on line graph sonification experiments and procedures for the sonification of diagrams, including up to three data sequences.

Furthermore, Walker (2010) has demonstrated that sighted people may draw the mental model trend of graphs while listening to them. Therefore, they suggest that evaluation with VI users would be possible because VI people would probably better than the sighted people due to their regular use of hearing in educational and work contexts.

An important characteristic separating visual and auditory displays is that visual displays can be permanent, while audio is transient in nature. In this way, various investigations have been led to compare several new auditory display techniques equally with the visual display techniques in terms of their effectiveness in presenting the information.

Peres and Lane (2003) have explored several ways to present box plots using sound. Their finding has suggested that spatial location was poor for mapping technique of sonifying statistical graphs. Though, the temporal mapping may offer better performance than pitch or panning mapping.

Meanwhile, Nesbitt and Barass (2004) have observed stock market data sonification with a graphical representation using the data and the audio-visual representation. Their study shows that the trading trends of stock-market can be identified not only on visual displays but also in auditory graphs.

Brown & Brewster (2003) attempted to assess the accuracy level of sighted people after listening to the various combinations of instruments by drawing sketch graphs to represent their non-speech sounds. This experiment, however, found no significant differences in the presentation of data between the same instruments or different instruments like a piano and a trumpet, although they had high precision (over 80% on average). Finally, as part of the MultiVis Project, their findings suggested that the sonified charts containing two data series can be perceived and drawn by sighted people (Brown & Brewster, 2003).

Another research in auditory graphs included VI people to access graphs by creating the combinatorial graphs which are often conveyed as node-link diagrams. These graphs enable VI screen reader users to set up and access graphs as hub connect diagrams and distribute them to the individual in real-time (Balik et al., 2014) or to use interactive online stock market charts (Zou & Treviranus, 2015). Harrar (2007) researched the design of auditory graphs by comparing discrete sounds and continuous sounds, and found that continuous sounds were more accurate to gain an overview of data series.

2.4.1. Historical development of auditory graphs

A low vision and VI users have to address issues of graphic literacy. In contrast to graphic designs, that are designed, edited and altered with low-cost media such as paper prototypes, non-speech auditory feedback is more difficult to articulate. This includes, in particular, how a certain form or colour can be rendered audibly or by touch, or even how one can respond to an auditory or tactile object. Traditionally, there are two primary approaches to assisting users living with VI to understand data analysis of graphs: a tactile graphics and a textual description.

AudioGraf was the early research in using a touch panel and auditory display, to make graph diagrams accessible (Kennel, 1996). Conventional graphics editing software with no additional tactile graphics features were available for years, and there were also a variety of tools to assist a graphic author in producing tactile graphics. Embossing printers are one approach made by ViewPlus to produce the tactile graphics. Tools such as the BrlGraphEditor (Batusic & Urban, 2002), Sparsha (Lahiri, Chattopadhyay, & Basu, 2005) and PictureBraille ("PictureBraille," n.d.) are designed to create graphics for text-only Braille system. However, the displays have several drawbacks in providing accessible versions of graphics which are costly to print and take a long time to produce. The printers capable of printing graphics are expensive, typically costing upwards of 3,000 pounds.

Moreover, they are not suitable for a collaborative study as they need more copies to be produced. Synchronization of the tactile graphics is also a potential issue when the team members are separated geographically. They could not easily change the data either.

The sound is considered to have more potential for non-visual graphs such as those implemented in software Triangle (Gardner, Lundquist, & Sahyun, 1996) and the Sonification Sandbox (Davison & Walker, 2007). The use of sound in data analysis has many advantages, as it is much cheaper since sound cards can produce it in standard computers. It is also flexible that if the auditory graph software is programmed correctly, the user can be offered a range of options for which data variables he can graphically display and in which sound parameters he can integrate them. Auditory graphs can be implemented on various mobile platforms, and with the advance of network support, it can support collaboration needs.

2.4.2. Anatomy of the auditory graph

There are some characteristics of the auditory graphs that need to be adjusted as follows:

2.4.2.1. Data notes

The auditory graphs present some quantitative information such as a sine wave. A sine wave may have a specific frequency and amplitude with one audio speaker.

The data type and its properties can be important for the perception of an auditory graph. Data presented in an auditory display can be categorized roughly as qualitative (verbal) or quantitative (numerical). The auditory display design for recording quantitative data can differ significantly from the design of a display showing qualitative information.

Factors that the auditory graph designer must consider include: which parts of the data are relevant to the user's task; points that the user must complete for the task; what type of display the user must use; how the data can be manipulated (such as filtering); and how much information the user needs for the task (Walker & Nees, 2011). Other information, in any case, is challenging as real-time data in a stock market, which may pose problems for an auditory graph designer. The playing back data as sounds may be difficult to interpret since the stock market does not follow physical-acoustic laws (Nesbitt & Barrass, 2004). Moreover, it is also hard to follow because it can change quickly in real-time and may not follow an easily discernible trend.

2.4.2.2. Context

Smith and Walker (2005) refer the context as the intended addition of non-signal data to a display. The context of a graph includes elements to be part of the chart and related elements like axes, labels, and background (Nees & Walker, 2007).

An empirical study has identified the importance of a cognitive model of graph interpretation (Shah, 1997). As initial theory and studies in understanding visual graphs, auditory graph study focused on the actual sound dimensions applied to quantify data (e.g., basic tone sequences). Later, researchers started to think about creating context and constructing an auditory graph to frame the data.

Extra information in the visual displays such as axes and tick marks may improve the understanding and increase the perception by allowing more efficient top-down processing (Tufte, 1990). A visual graph without axes or tick marks presents no alternative to determine the value at any point (Smith & Walker, 2005). The listener would never discover the value that temperature had been changed or the price on a particular date on a stock market. Moreover, a lack of context seriously prevents those with visual impairments from collaborating in any significant way.

The shape of the line presents some related context that may allow an observer to produce a trend analysis, such as the trend of the drop and recovery in a Finance graph, as seen in Figure 2-2.

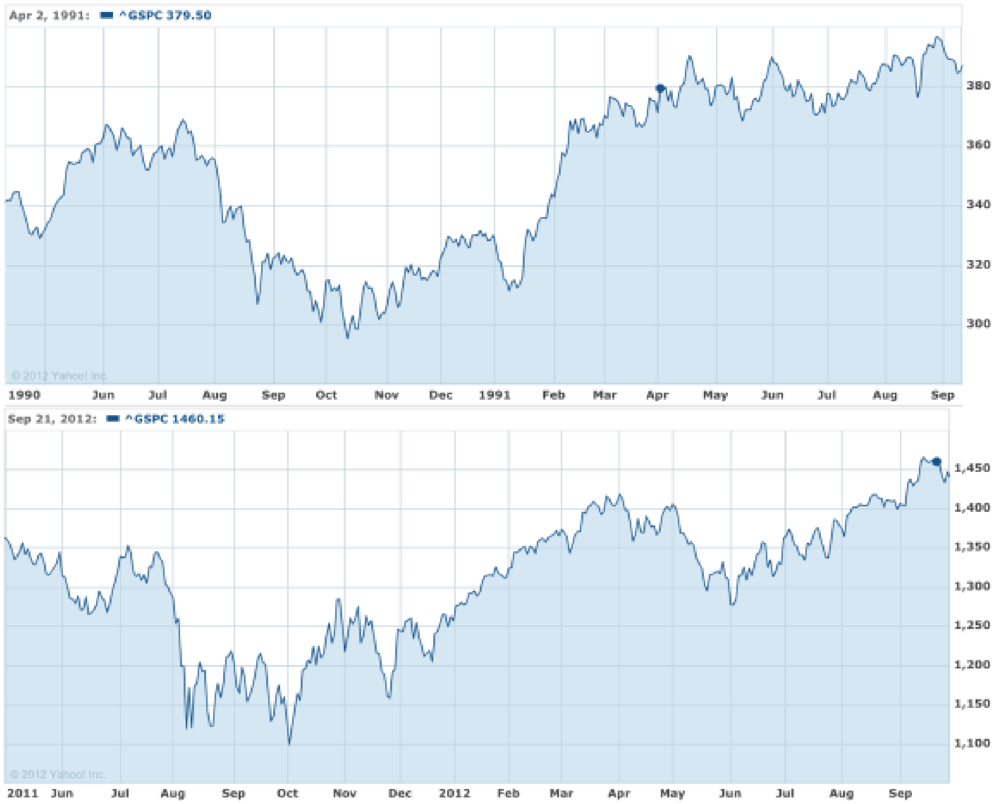


Figure 2-2 Comparing Yahoo Finance graph of the S&P 500 value in 1991 and 2012. Though the decline and recovery trend is similar, however, there are differences in scale (Davison, 2013)

Smith and Walker (2005) argue that the Y-axis as reference notes can improve the point estimation task output related to the auditory graph. Researchers also found that the X-axis context was often useful as rhythm snaps or beats (Bonebright et al., 2001). Metatla (2016) showed that adding reference tones as context information for data sonification increases user accuracy. However, this addition results in slower task performance times than if no reference tones are specified.

2.4.3. Auditory graph fundamentals

The topics of mapping, scaling and polarity are critical when creating a typical parameter-mapped sonification, such as displaying precipitation and average daily temperature over the previous year (Walker, 2002). These topics are based on several questions like which sound parameter is most useful for displaying some data, such as temperature? A further question is whether some mappings are considered outstanding or fairly obvious, others that are considered acceptable, and some that are regarded as poor mappings. When a designer decides which sound dimension to use to display the data, the third question is: How much change in pitch is required to convey a particular change, e.g., in temperature? This psychophysical scaling function is crucial in using sonifications for accurate comparisons and absolute assessments.

2.4.3.1. Mappings

Mapping defines most data sets to the sound that the features are related to the known maximum and minimum data in the display design. Furthermore, sonification depends on a specific sound dimension to represent a given data dimension. This reason is that among the listeners, some consensus seems to exist about how good the sound attributes are when representing certain data dimensions (Walker & Nees, 2011). For instance, Walker (2002) argued that pitch is generally good for representing temperature, while tempo is not effective. Therefore researchers have researched these issues in detail (Davison, 2013; Nees & Walker, 2007; Smith & Walker, 2005; Walker & Mauney, 2010). According to Nees (2007), mapping refers to a sound measurement that determines the change in the auditory graphs' data. In the auditory graphs, data change is often mapped based on the frequency of sounds.

Data to frequency mapping has often been used in auditory graphs, and existing theoretical methodologies might be implemented (Nees & Walker, 2007; Walker & Mauney, 2010) to estimate trends of tones in frequency over time. For the frequency mapping used in the auditory graphs, a display designer must also choose the type of sound (pure tones, MIDI instruments, and so forth) into which the data is to be mapped.

2.4.3.2. Scaling

Once a mapping is chosen, determining the variation in the pitch sound is important to carry a particular change, e. g. in temperature. According to Walker et al. (2010), the scaling is the degree of sound dimension change required to display a unit of data dimension that corresponds to the line slope on a visual chart.

In their verification, size estimation was used to identify the desired scaling values, mapping, and polarity for several data-to-display mappings. In a practical example, frequency changes are most effective to display the average daytime temperature in the range of 0-30° Celsius (Walker & Mauney, 2010). The temperature data could be upscaled to cover the entire hearing range (at best about 20 Hz to 20,000 Hz), but it is much more effective to scale the data to the range where the hearing is most sensitive, for example, between 1000-5000 Hz. The scaling slope for the same quantitative data changes may vary depending on the data type (e.g., temperature, size, and pressure) displayed and the sound attribute that is being varied. Whenever possible, the auditory graph scaling variable should match the preferred scaling for the conceptual dimension represented.

Researchers also argued that the auditory graph's data scaling should not be lower than at least MIDI note 35 B1 (equal to 61.7 Hz) or be higher than MIDI note 100 E7 (equal to 2637 Hz) (Brown et al., 2003). This scale is suggested based on what most humans can easily hear.

2.4.3.3. Polarities

Another essential idea for the auditory graph configuration is the mapping frequency polarity and the data represented. Brown et al. (2003) suggest a linear mapping of the Y-axis of graphs to the pitch of musical notes whereby increasing the y-value increases the pitch. Their research has demonstrated that positive polarity mapping (increasing-increasing) is more natural for most target groups across the majority of data dimensions, see Figure 2-3. Walker et al. (2010) reported that they came across instances where Polarities preferred by VI people were reversed compared to those preferred by sighted people.

Walker (2010) gave the example of coin mapping. The mapping preferred by sighted people of coin value to pitch on the Y axis followed the usual convention of up-up, that is a higher pitch sound would represent a higher coin value. On the other hand, the mapping preferred by some VI users was the reverse of this. The VI users who preferred this reversed mapping explained that they associated larger (higher value) coins with a lower pitch, because a lower pitch is heard when these larger coins are dropped on the floor, compared to the higher pitch when smaller coins are dropped. On this basis, some VI users preferred an up-down polarity mapping of coin value to pitch.

Overall, in auditory graph design, the aim should be to make choices of dimension, scale, and polarity that present natural representations to the target user population (Walker et al., 2000).

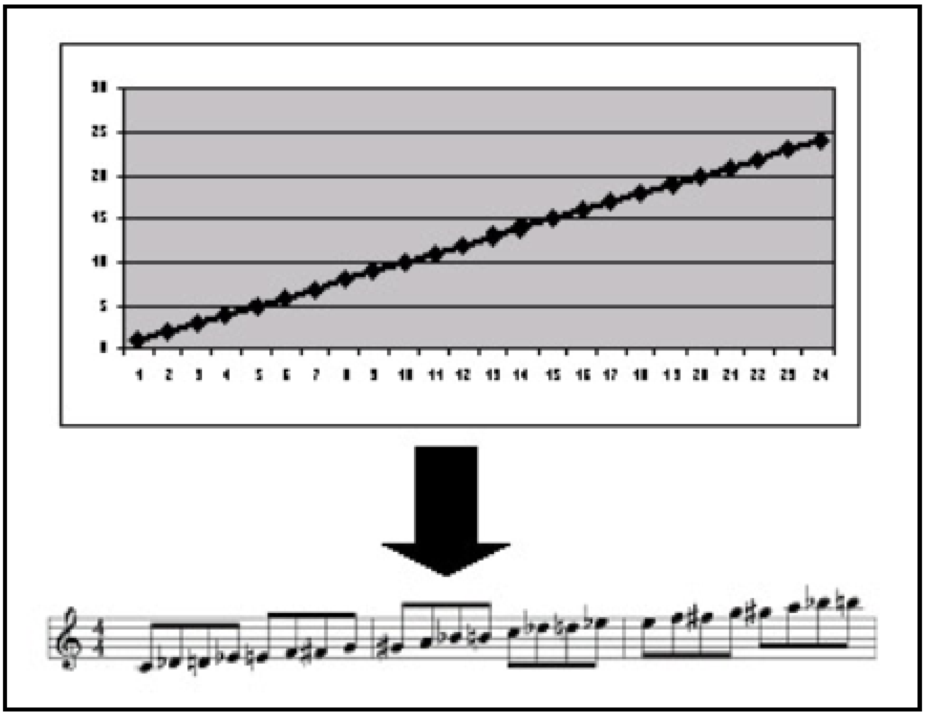


Figure 2-3 If the y-value increases, the musical note pitch is increased (Brown et al., 2003)

2.5. Auditory graph's point estimation

Researchers found that creating valid auditory graphs is more than just solving the problems of audio data mapping. In addition to the representation of numeric data, there is also a wealth of information in visual graphs for improving the readability and understanding of this type of data. While presenting numeric data visually, extra features including axis, annotations, and markings increase reading and understanding of this type of data by enabling more effective top-down processing (Bonebright et al., 2001; Flowers et al., 1997; Smith & Walker, 2002; Tufte, 1990). For example, in visual line graphs, it is the line that provides a certain context, but only the random context associated with the observation that some data points are further to the right or above or below the data points to the left (Brewster, Wright, & Edwards, 1994). This inherent context could allow an observer to perform a trend analysis of the data (e.g., whether the line is rising or falling in general), but the exact extraction of a particular value is impossible. If a visual graph has no proper context (for example, no axes), then there is nothing we can do to estimate the figure at a particular point. This type of feature provides visual graphs a benefit over other data display tools, for example, text in linear form (Larkin & Simon, 1987).

Typically, a way to add the context of the x-axis to a sonification is by using multiple ticks or creating percussion tones. Bonebright et al. (2001) studied rhythmical markers as the sound of clicking and examined the students' ability to match auditory displays against the actual visual graphs. Research by Smith and Walker (2005) explored the potential of adding a variety of contextual information in such cases to improve the tasks of non-visual point estimation. They investigated the use of audible clicks to indicate x-axis context and the insertion of markers to indicate scaling on the vertical axis. They concluded that adding auditory contexts improves the interpretation of auditory graphs.

Tasks for reading graphs, e.g., point estimation, can be severely influenced by the insufficient context and information for the user's reference. For example, studies have examined the relationship between data density and the number of individual data points displayed each second and trend reversal for point estimation and trend identification (Nees & Walker, 2008). They found that in the point estimation tasks, user efficiency decreased as data density and trend reversal grew.

Metatla et al. (2016) exploring further support of non-visual point estimation tasks using another form of sonification by integrating multiple tones as references to represent a note. The study by Metatla et al. showed that using multiple references in auditory graphs could improve the accuracy of point estimation tasks. It showed that displaying data in auditory graphs using contextual information is more effective if it is properly designed. More research is necessary to explore potential approaches for context implementation that will allow users to perform trend analysis tasks and perform point estimation tasks effectively.

2.5.1. Auditory recall technique

Different recall techniques were used to identify memory structure and organization. Among the recall techniques, the immediate serial recall is used to retrieve a list of elements in the order of their presentation. Recency is a feature of serial recall, in which recalling the last or last few items in a list is more effective than recalling items in the middle of the list (Surprenant, Pitt, & Crowder, 1993). However, recency effect size varies according to the type of presentation: At the end position in a list, better performance is observed mainly in the auditory display and is significantly weaker in elements presented visually (Conrad & Hull, 1968). There were performances of differential recency and suffix effects even inside the auditory modality, based on what stimuli were used. Studies on auditory memory for verbal material such as numbers (Stigler, 1978), words (Murdock & Walker, 1969), for example, consistently reported robust recency effects.

The size of the recency effect differed from experiment to experiment even when similar non-speech stimuli were used. Greene and Samuel (1986) achieved a greater effect than Foreit (1976) while each of them used a vocabulary with widely spaced pitches or similar stimuli. Therefore, it is well known that the recency effect was difficult to grasp for non-speech stimuli. According to Surprenant (1993), the recency effect level obtained with ordinary speech sounds is far greater than the recency effect perceived with non-speech sounds.

The opposite perspective on tone pitch separation -between wide and narrow- is offered by viewing piano tone sequences as having potential for melody. It is well known that the perception and processing of melody are supported by sequences with narrow pitch jumps between adjacent tones (Bregman, 1990). Even commonly used melody sequences in octave-coded versions, for example, are difficult to identify directly (Dowling & Hollombe, 1977). Therefore, as far as the quasi-melodic processing by our stimuli takes place, we should expect that "auditory streaming" would favour the narrow state (Surprenant et al., 1993). We use this quasi-melodic approach to design our final auditory graph study (see section 3.11.6).

Greene and Samuel (1986) found out from their recall of pitches experiment that musicians had a better overall recall for pitch sequences than non-musicians, while both groups had the same recency.

2.5.2. Directional cues for stream segregation

It is challenging to display multidimensional data by auditory display. Based on how many auditory streams can be detected, the tasks are classified into divided and selective attention. With divided attention, simultaneous signals must be extracted, and for selective attention, only one important audio stream must be isolated from concurrent auditory messages (Anderson, 2005; Anderson & Sanderson, 2004). In the study by Bregman (1990), for the auditory scene analysis, the principles about the listener's auditory ability to distinguish particular strands of meaning from the mixture of sounds and group the entire content are integrated into the perceptual framework. These multi-stream situations often cause confusion and mixing of auditory streams. Therefore, for mapping strategies, both data information and clarity requirements must be met for the listener. For this reason, the main investigation concerns the way in which sound dimensions in auditory graphs could be formed in order to achieve a clear representation of the information contained in them, and the effectiveness with which these dimensions for further understanding may be used.

In contrast to the directional orientation, where the structure for the auditory representation is apparently interpretable, the semantic expression is provided by the timbre, emphasizing the auditory stream's identity. According to Hawkins (1986), three-dimensional separation offers a better "force and semantic structure" to reduce "problems of peripheral sensory masking” and focus attention on sound sources than pitch. By combining spectral and temporal characteristics of binaural signals, directional cues that act as spotlights improve sound processing and accelerate discriminatory responses by providing an interpretative context that gives the sound a special structure. Based on experience, directions could clarify audio streams' subjective mental representation (Song & Beilharz, 2007).

2.5.3. The role of negative number for auditory graph

Most studies to date that examined numbers have concentrated on positive numbers, whereas those working on negative numbers are rare. The question that researchers debated is whether negative numbers are represented mentally to their components values or to their holistic values (Fischer, 2003; Pinhas & Tzelgov, 2012; Shaki & Petrusic, 2005; Tzelgov, Ganor-Stern, & Maymon-Schreiber, 2009).

The component representation implies that the polarity sign and the digit of the negative numbers are initially processed separately. Then the meanings of negative numbers are assembled at a later point in time. If we process negative numbers, only the digit component is represented on the mental number line, on which the numbers are represented ordered according to their size, i.e., smaller numbers are to the left of the larger numbers (Rugani, Vallortigara, Priftis, & Regolin, 2015). Thus, the negative numbers whose numerical value is large (e.g., -1 and -2) would be shown to the left of the negative numbers whose numerical value is small (e.g., -8 and -9), which is opposite to the representation of positive numbers.

In contrast, the polarity and magnitude information of negative numbers is integrated and displayed on the left extension of the entire mental number line (Fischer, 2003; Mende, Shaki, & Fischer, 2018; Varma & Schwartz, 2011). During the processing of negative numbers, the meanings of negative numbers are retrieved and not put together. In this way, if the negative numbers have a large numerical value (e.g., -1 and -2), they would be displayed directly to the right of the negative numbers have a small numerical value (e.g., -9 and -10), and this would be equivalent to positive numbers. Answering this question, most researchers examined *the distance effect*, *the SNARC effect* (spatial-numerical association of response codes), and also the *semantic congruence effect* (SCE).

The distance effect is concerned with the reliable observation that the response delays for quantitative comparisons correlate negatively with the quantitative differences (Moyer & Landauer, 1967). There should be a distance effect when negative and positive numbers are perceived along a sequence, in pairs consisting of a positive and a negative value (a mixed polarized pair). No such *distance effect* was observed in previous studies (Tzelgov et al., 2009), indicating that participants were not aware of the absolute size of the numbers and used only the character information that was in line with the concept of component representation.

The SNARC effect, which is found for positive numbers, refers to the results that small numbers' reactions are faster with the left side, while the reactions to large numbers are faster with the right side. In the study by Fischer and Rottmann (2005), participants were asked to judge the parity of the number displayed, whether it was positive or negative (experiment 1) and whether the displayed number was lower or higher than zero (experiment 2). The results showed a distance effect and a *SNARC effect* for positive numbers. No distance effect, but an inverse *SNARC effect* was recorded for negative numbers, suggesting the component representation of negative numbers.

The SCE, on the other hand, agrees with the finding that the responses in number matching tasks out of two large numbers are faster when the greater number is selected rather than when the smaller numbers are selected. In opposite, the responses are faster if the smaller number among two small numbers is chosen rather than if the larger number is chosen (Banks, Fujii, & Kayra-Stuart, 1976). Shaki and Petrusic's study (2005) showed that participants were quicker to respond to positive numbers if they were asked to select the larger one than the smaller one. On the contrary, participants were quicker to respond to negative numbers if they were asked to select the smaller one than the larger one. *The SCE*, based on polarity, tended to support a holistic view of negative numbers.

The SNARC effect and *the distance effect* inconsistent results in the size comparison tasks are also observed. Fischer (2003) reported that negative numbers were compared quicker using the left hand compared to the right hand, positive numbers were compared with the right hand faster than with the left hand. For negative numbers, this SNARC effect proposed a number line running to the left, supporting the holistic presentation of negative numbers. However, the similar SNARC effect in the study by Krajcsi and Igács (2010) was not replicated. According to Shaki and Petrusic (2005), it was only applied if positive and negative numbers were mixed.

However, Ganor-Stern and Tzelgov (2008) achieved another result pattern with a larger number of negative and positive numbers from -99 to 99. Participants' reactions to positive numbers were faster in the numerical comparison task when selecting which numbers were larger than when selecting which were smaller. However, their reactions to negative numbers were not influenced because of the instructions. The results suggest that processing negative numbers is not about remembering their meaning, but about integrating the polarity sign with the sizes of the digits. According to Tzelgov et al. (Tzelgov et al., 2009), there is no SCE for negative numbers because the negative polarity sign is considered to be " low" and the number itself is considered to be " high", and both balance one another out. Therefore, these results are the same as the representation of the components.

While those mentioned above proposed to display negative numbers concerning their elements in the visual modality, the study by Kong et al. (2012) investigated how negative numbers are processed in the auditory modality and how it is influenced by context. In one of their investigations, a stimulus recognition task was used in which negative as well as positive numbers were combined as indicators. In this study, an inverse attention SNARC effect was obtained for negative numbers. Their results indicate that in auditory modality, negative numbers are constructed from the set of positive numbers, which supports a representation of components concept.

2.6. Auditory graph implementation on mobile devices

Research related to the auditory display on a mobile device has been evolving in recent years due to mobile communication's extensive use (Lin & Scott, 2012; Shin et al., 2013; Su et al., 2010). Researchers have exploited the use of accessibility features such as screen readers and voice commands for communication purposes such as to create, record, send and receive emails; use of maps for navigation; and modify a document (Strumillo, 2010; Su et al., 2010).

An earlier study has developed a system on a tablet PC called exPLoring graphs at UMB (PLUMB), which was designed to support people with visual impairments to understand graphs using auditory cues (Cohen et al., 2005).

Researchers from Monash University has been developed GraCALC, as an approach for implementing numerical and statistical graphics to VI (Goncu & Marriott, 2015). The system presents a graphic from a mathematical function as a line graph displayed on a web-based service. Graphics are displayed on a visual screen on the iPad. The interface has been intended to utilise the VoiceOver screen reader and the standard iOS framework. As the user explores the display, the system determines a combination of speech, non-speech audio, and tactile (through vibration motors attached to the fingers) feedback to allow them to explore the screen.

Eight of ten VI students had an impression of being faster in understanding the graphs when presented with tactile feedback. However, as the authors say themselves, it needs to be confirmed using a user study with more participants.

Moreover, the participants confirmed that they prefer to use a mobile device rather than a traditional tactile display due to its form factor and portability. Although their web-based applications have a standard code across multiple mobile platforms, it is not designed for collaborative work study. They have a limited boundary to access mobile-specific functions such as Bluetooth, SMS, and GPS. This shortcoming motivates us to develop an Android application that could do the task not also when connected to the network, but also offline.

2.7. Touchscreen technology for accessible interaction design

Johnson (1965) first outlined a concept for developing a touch display as early as 1965. Followed by the Elograph, an opaque touch sensor invented in 1971, was an important milestone in touchscreen technology (Buxton, 2010). A touch-sensitized graphic digitizer with a transparent surface and tailored to a computer screen was then the first real touchscreen developed in 1974. As a kind of display, the touch screen has a touch-sensitive, transparent panel that covers an LCD or CRT panel. Three components make up a touch screen system - a sensor panel, a controller and a software driver. The sensor panel and controller capture the touch event and contact position to process a user's input, and the software driver then passes the touch coordinates to the OS of the computer. Touchscreens can benefit from the precision of a stylus as well as the ease-of-use of fingertips. Typical for a small screen, a stylus is often used to press the small controls.

2.7.1. Interaction technique theory of touchscreen

Touch screen interaction is the most direct form of user experience in which information display and control are bonded into one interface. Touchscreen operation is intuitive for beginners due to the zero offset between input and output, hand movement and eye vision, also control and feedback. Albinsson (2003) argued that touch screen interfaces have other advantages in addition to directness. In the first place, because its user interface is superimposed on the screen, no additional input device or space is required for interaction with the touch screen. Furthermore, touchscreens are more solid compared to moving input devices like the mouse. Therefore, various interaction techniques are used to precisely map the finger action to the desired target.

Pointing Technique

Since the touchscreen is located directly between the control unit and the screen, it also has special restrictions. Before 1988, touchscreens had the poor image of being inaccurate. Most books on user interfaces suggested that touchscreens were only available for subjects greater than the human finger. In those days, the selection was made so that when a finger came across a target, it would select it and immediately perform the appropriate action. Errors were frequent and frustrating for users due to parallax or calibration problems.

In the first place, the fingers of the user can cover part of the screen. Furthermore, as a pointing device, the user's finger has a very low "resolution". As a result, pointing at targets smaller than the width of the finger is difficult. Potter et al. (1988) have already recognized and addressed these limitations. They reviewed three sets of pointing strategies, i.e., *first-contact*, *land-on*, and *take-off*. *Land-on* provides the simplest strategy that only registers the position of the first touch. A decision is made as long as a target exists under the first touch.

In the first-contact strategy, users drag their fingers on the screen to a selectable area, which is then selected, and the appropriate process is initiated. Again, all extra contacts are ignored until the finger is pulled away from the screen. This strategy improves the *land-on* strategy by allowing the user to pull his finger to the target of the *first-contact* (Sears & Shneiderman, 1991).

Take-off strategy uses a cursor on the user's fingertip when touching the screen with a fixed offset. Drag the pointer to the destination and pull the finger to select the desired location. To cope with very small targets, Potter and colleagues (Potter et al., 1988) applied techniques based on knowledge of the system's targets.

Using more recent capacitive touch screen technology, Ljubic et.al. (2015) evaluate Fitts' law's predictive power of finger-based pointing on mobile touch screens with different screen sizes and appropriate interaction styles. Their results showed that the models of Fitts' Law are indeed effective for pointing tasks and that the smaller screen size is suitable for predicting pointing times on mobile touch screen devices.

Scrolling Technique

Scrolling on a touch screen can be performed by moving continuously along a single axis, e.g., by dragging a scrollbar image displayed on a screen. In some cases, repeated touches on the screen are required to scroll an object on the screen.

Previous research in scrolling on a touch screen has been concentrated on the *flick and ring-scrolling techniques* for a more efficient interaction (Aliakseyeu et al., 2008; Moscovich & Hughes, 2004; Wherry, 2003). *The flick and ring-scrolling technique* has been developed on the model used by Tu (2014). As shown in Figure 2-4a, the flick technique has two points, p2(x2, y2) and p1(x1, y1), indicating both the actual and prior point in a motion path. The scroll distance corresponds to the exact value of (y2 - y1). Use the algebraic sign (y2 - y1) to determine the direction of the document: The document is scrolled forward if the sign is negative. Otherwise, the document will scroll in reverse.

In a gesture path, there are at least three points p1, p2 and p3 (p1 is a preceding point of p2, and p2 is a preceding point of p3) as shown in Figure 2-4b. θ refers to the vector's angular rotation (p1, p2) to the vector (p2, p3). The scrolling distance is θ × R/2(R is a constant-value of 220 pixels). The direction of scrolling depends on the sign of the dot product of the vector (p1, p2) and the vector (p2, p3): A positive sign scrolls the document forward. Doing so causes the document to scroll backward.

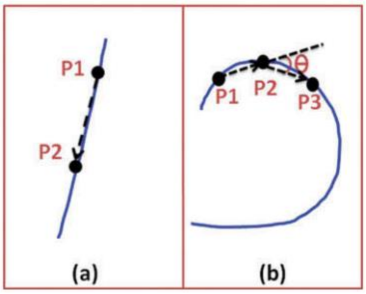


Figure 2-4 (a) Flick Scrolling. P1 and P2 each represent the actual and the previous point of a gesture path. (b) Ring Scrolling. There are at least three points-P1, P2 and P3 (P1 is a preceding point of P2, and P2 is a preceding point of P3) - in a gesture path. θ refers to the angle that moves from the vector (P1, P2) to the vector (P2, P3). (Tu et al., 2014).

Navigation Technique

For many systems, users have a limited view of the larger graphical workspace. These systems often require the users to find and select destinations in the workspace that are not visible in the current view, commonly referred to as off-screen locations (Irani, Gutwin, & Yang, 2006). A wide variety of navigation techniques have been developed by researchers to improve performance when working with large workspaces on small devices. Scrolling, zooming or panning are included in these navigation techniques. These techniques have become very common and are strongly embedded in map browsers. The capabilities of these navigation techniques have, however, only been investigated with a limited range of tasks (Cockburn & Savage, 2003; Gutwin & Fedak, 2004; Igarashi & Hinckley, 2000).

Crossing Selection Technique

In a crossing selection interface, an operation is triggered by the user moving a pointer over a boundary instead of typing within a destination (Forlines & Balakrishnan, 2008). This solution can be used as an alternative to point-and-click techniques not only on fingertip touchscreens, but also on pen-based computing. Apitz and Guimbretiere (2004) proposed using a crossing selection that allows a smooth transition from one action to another, e.g., the selection of a hierarchical menu is supported by the interaction of *target distance* and *contact continuity*. For closely spaced targets, continuous crossing can lead to fewer selection events than for targets further apart.

2.7.2. Touchscreen: benefit and weakness

The most important features of modern touch screens, which affect the user-friendliness and design of the touch screen interface, are its support for pointing directly at objects. There are no inter-mechanical devices and no displacement between input and output, control and feedback, hand movement and vision (Liu, 2012). The user does not need a physical input device before choosing it. Furthermore, only one interface linking information display and selection of menus means the control options of the user are limited. This helps to avoid learning curves, reduces menu selection and increases the efficiency and accuracy of the user. It is also quicker to access than manually moving a mouse wherever the user needs it on the screen. Easy learnability and user-friendliness make touchscreens particularly suitable for inexperienced users. With the combination of display and input surface, compact I/O device design reduces space requirements. Due to its easy-to-clean design, touchscreens are ideal for hygienic applications.

However, the main drawback of the touchscreen is the false activation that can be caused by accidental contact. Touch may not be accepted, or a double response may be given, depending on user pressure and touch sensitivity. Also, the size of the touch controls is limited by the size of the human finger. Because the finger may be large enough to point to small objects, screen controls must be of a minimum size. Soiling can reduce the sensitivity of a touch screen. The screen may also be soiled by fingerprints, making the contents of the screen unclear.

By nature, input on a touch screen is sequential: a one-finger click. Compared to keyboard input, this slows down input by not allowing multiple fingers to be used in parallel. The technique can strain the arm muscles under heavy use, especially if the display is in a vertical position. Typing a lot of numbers or letters with fingers is very tedious and exhausting. For this reason, it is pointless to use touchscreens at workplaces where a lot of text or numbers have to be entered.

On touch screens, there is no analogy to mouse movements. Mouse-over the selected element (e.g., by highlighting), user can drag the cursor over a display element and validate their choice by pressing the mouse button. In the contrary, touchscreen users point directly to a screen element. When they are lucky, they can pull back their finger if the wrong screen element has been touched. For other touch screens, contact will immediately trigger action and there is no way to cancel the action.

Given the advantages and disadvantages of touchscreens, they are most suitable for applications that do not require much or very little training, also no precise location, and minimal textual and numeric data entry. By not using physical buttons, knobs or sliders, touchscreens are particularly useful for simplifying pointing operations in applications exposed to vibration or motion, such as in workshops and cockpits (Liu, 2012). Its easy cleaning and sealing allow the use of touchscreens in locations where cleanliness is important (e.g., medical center) or where soil or fat is present (e.g., restaurants, workshops).

2.7.3. Screen readers on mobile device

These are software programs that recognize the elements presented on a screen and return it to the users using text-to-speech or Braille output devices. They can scan a page visually like sighted people, but VI people need screen readers to locate text, navigation elements, graphics, titles, page sections, links, and so forth.

Using a screen reader requires considerable concentration and increases the cognitive load on the user[3]. It is important to pay attention to how the text is spoken to find out about the content of the page, find out what is of interest to the users, and determine if an element is workable. As opposed to visual web pages, screen readers also display information in strict consecutive order: users have to patiently listen to the page description before they meet the

[3] The World Wide Web Consortium (W3C) has a useful introduction regarding the cognitive accessibility user research. See https://www.w3.org/TR/2020/WD-coga-usable-20200717/

element they are interested in; they cannot select the most interesting one without first looking at the elements that precede it. However, a certain degree of direct access is possible. By expecting the headlines to be in the middle of the page, users can place a finger in this general area of the screen so that the voice reader skips the page elements before that position, saving time for listening to the whole page. If the shopping cart is expected to be in the top right corner, the user can directly tap this part of the screen. Some screen readers also provide the ability to jump directly between on-screen elements of a particular type, for example, headings on a page, frames, forms or items in a list.

In terms of accessibility, Apple has added VoiceOver as a screen reader that uses a gesture-base to make the IOS devices speak for what has been written on display. Google's Eyes-Free Project has developed talkback as the built-in screen reader for Android devices. TalkBack has been introduced in Android 1.6 (Donut) in 2009, and it is available on all Google Android later versions. Spiel by Nolan Darilek is an alternative screen reader to TalkBack which offers roughly the same accessibility level of spoken, auditory and haptic (vibration) feedback. Android 2.2 then introduced virtual keyboards developed by Darilek and Google using voice input (Johnsen, Grønli, & Bygstad, 2012).

TalkBack utilises spoken feedback to describe the results of actions, such as opening an app, navigating a device, describing user gesture and activating an event like notification. However, although many tools rely on the use of speech recognition sensor and text-to-speech (TTS), there has been less work on presenting non-text material such as graphs in audio (Balik et al., 2014; Su et al., 2010).

2.7.4. Using a touch screen with a screen reader

Most of the touch screen tasks are very visual for VI users that make them difficult to complete the tasks. Most interactions, such as menu navigation and text input in particular are easier with hand-eye coordination. However, it makes them difficult for people with VI to interact with them on mobile devices. Solutions are available to make these actions generally possible using a speech-based screen reader, however, these solutions come with a learning curve and some VI people find them hard to adopt.

The way interactions work with the mobile device are quite different when a screen reader is running. For example, on the pointing strategy, the sighted users will apply this strategy to select the location of the icon by first viewing the icon's position. Then they tap the icon once to run such an application. Different strategies apply to users with VI who rely on screen readers. Using a screen reader, this tap will instead cause the device to display the information directly under the fingertip. Touch in different places of the display and let the user listen to the different icons or parts of the descriptive text.

Suppose the display was touched and the screen reader detected the playback icon on the mobile device's music player, then double-tap (tap twice in quick succession) anywhere can be used on the screen to activate the playback control. The screen reader will interpret this gesture as if the user is a sighted user tapping a single finger on the control identified.

In the scrolling strategy, the users need to scroll one direction vertically or horizontally with one finger from the start to the end position. Different strategies apply to users with VI who rely on screen readers. Using a screen reader, this scroll will instead cause the device to move and read the information after the previous icon or section. Scrolling vertically or horizontally may result similarly as the screen reader will only translate both gestures to move either to the previous icons or to the next icons. Double-scroll with two fingers, the screen reader will interpret this gesture as if the user is a sighted user scrolling a single finger on the control identified.

These limitations encourage the implementation of multimodal interaction to improve the performance and accuracy.

2.7.5. Mobile multimodal interfaces for VI user

Multimodal multi-sensor interfaces may connect one or multiple user input modalities and extract information from sensors (e.g., camera, microphone, touch screen, position, acceleration, proximity, tilt) (Oviatt et al., 2017). Sensory cues allow users to analyse the physical state, state of health, psychological state, current context, commitment to activities, as well as various other types of information (Oviatt et al., 2017). Users can intentionally perform actions when using sensor controls, for example, flipping a screen to rearrange its orientation. In addition, sensors can also serve as "backend control" for the interface to adapt automatically and without the user's intentional intervention (e.g., dimming the telephone screen when not in use). The purpose of sensor input is to make the interaction between user and system and the adaptation to the needs of the user transparent.

Non-Visual Mobile Multimodal Interaction

The navigation on a mobile device, which benefits both sighted and VI users, has been investigated in several studies. With the Mobile ADVICE system, Amar (2003) combines a scroll wheel with acoustic and tactile feedback to browse through the menus of mobile phones. The EarPod, incorporating a circle-shaped touchpad combined with acoustic feedback to facilitate non-visual interaction with multi-level menus, was investigated by Zhao et.al. (2007). They claimed that the system outperformed visual menus in practice. Most of these studies require the designer to customize the hardware that is not widely practical for general users. Without modifying the hardware, Sanchez (2007) has designed desktop and mobile applications based on pointing gestures to help VI users travel on the city rail. The BlindSight system by Li et al. (2008) is based on the physical keyboard of the phone and is designed to provide access to a non-visual menu during the phone call.

In Kane et.al. (2008), slide rules were developed that enable multi-touch gestures in mobile device interaction. In addition, Kane et al. (2011) have considered a new set of rules to improve access to mobile devices using gestures on mobile touch screen devices. A preliminary study by Metatla et.al (2014) investigated non-visual menu navigation regarding completion times and mental workload. The study showed that by using an audio-tactile menu display, users were significantly slower in finding a menu item than when using visual or audio-only displays.

Chapter 3. Research Methods and Prototype Development

3.1. Introduction

In this chapter, we present the overall methodology of the thesis and the development of the successive prototypes of the Mobile Auditory Graph (MAG) app. Section 3.2 discusses the overall approach taken. Section 3.3 outlines the studies carried out. Section 3.4 describes the main research methods employed, followed by a discussion of the study design in section 3.5. The instruments employed in the user studies are discussed in section 3.6, including a justification of the task choices made in the point estimation and graph reproduction tasks. A description and justification are provided of the methods used to gather data in section 3.7. In section 3.8, we summarize the qualitative and quantitative measures used to analyse the data. The chapter continues with a discussion of the participant sampling criteria in section 3.9 and the participant recruitment process and ethical considerations are discussed in section 3.10. The final section, 3.11, covers the development of the mobile auditory graph (MAG) application from version 1.0 to 4.0.

3.2. Overview of Approach Taken

The nature of the multidisciplinary research in Human Computer Interaction (HCI) commonly employs quantitative and qualitative research methods to evaluate both technological and human behavioural concerns. In the field of HCI, quantitative methods are considered by some researchers to be a gold standard, generally task-based (Barkhuus & Rode, 2007), which involved users directly. In using them, researchers aim to answer specific questions by numerically measuring the usability of a certain device, technique, or system and its usage by exploring the extent to which a technology is beneficial compared to previous alternatives. It also allows researchers to make comparisons between groups and/or between interactions performed under different conditions. Qualitative methods have also played an important role to evaluate HCI characteristics which are concerned more with how and why questions. (Lopes, 2016). They focus on getting data related to users' motivations, opinions, expectations, and behaviours about the evaluated interface, which are valuable to interaction designers.

To the best of our knowledge, our study is the first to evaluate interactions by visually impaired (VI) users with auditory graphs presented on a mobile device.

Driven by its explorative nature, we employed mixed methods, quantitatively and qualitatively, to understand users' interactions with the MAG application, evaluate the impact of each version, and identify and resolve specific problems.

While the studies described in this thesis employ a novel platform, the focus is on generic graphing tasks such as point estimation and graph reproduction. The aim has been to examine these fundamental tasks and make lasting research contributions by exploring how might they be made easier, quicker and more intuitive in the context of a mobile, multimodal device that has already seen substantial take up in a wide range of educational and work settings.

Qualitative methods such as questionnaires and semi-structured interviews were employed to understand performing auditory graphing tasks on a mobile device and the user’s behaviours while undertaking the assigned tasks.

Quantitative studies were performed to evaluate the tasks through the use of measures including accuracy of the estimated points, trend in the accuracy across different conditions, groups, features (e.g. modalities), note duration and length of pause between the musical notes used to render the graphs..

3.3. Outline of the Work Done

We present an outline of the main stages of the work as follows.

Study 1

We carried out an exploratory, observational study to explore our first MAG prototype's usability with 12 sighted participants, as described in Chapter 4. We experimented to determine how accurately these users could perform graphing tasks. We also observed the users' behavior regarding their choice of pause length between notes while performing the point estimation tasks. The study explores point estimation errors, the relationship between the number of points estimated and errors, length of the pause, and the possibility of learning effects. The quantitative and qualitative measures obtained, including direct user feedback, allowed us to develop our understanding of the interactions involved in these point estimation and graph reproduction tasks. This in turn, led to changes that fed into the development of the next version of the MAG application.

Study 2

After making some changes in the first MAG prototype, described in section 3.11, we conducted an observational study to evaluate our second version with visually-impaired participants, as described in Chapter 5. Our main goal was to examine to what extent VI users can retain in memory the details of auditory graphs using either passive listening or multitouch gestures to audition the graphs. The participant's interpretation of the graphs is determined by the accuracy of their point estimation and graph reproduction tasks across three graph complexity levels (simple, medium, and complex), the same leveling concepts employed in Study 1. We investigated whether multitouch gestures, an additional feature of our second prototype, provided better or worse participants' performance than passive listening. We conducted semi-structured interviews to gain participants' feedback about the issues and challenges encountered while performing these tasks. Their feedback was useful to improve and refine further MAG app development.

Study 3

The study by Metatla et al. (2016) showed that using multiple references could improve the accuracy of point estimation tasks in auditory graphs, but at the expense of more time being required to complete the tasks. Similar too, but with some important differences from the work of Metatla et al. (2016), we developed a multiple reference sonification approach for auditory graphs for use in this third study. Full details of the approach, including how it differs from that of Metatla et al. (2016) are given in chapter 6. However, in brief, the algorithm works for positive Y values from 0 up to a maximum value (Y_{Max}). The approach is to play notes in multiples of 10^{th} of Y_{Max} up to and including the value of the point being estimated ($Y_{Estimate}$). Further, two different timbres are employed to assist users in distinguishing between points in the lower and higher halves of the range of Y values from 0 to Y_{Max}.

The results of point estimation tasks for all participants were calculated by taking the RMSE between the estimated values and the true values. Using this approach, the multiple reference sonification we had developed was compared with the single pitch approach. This comparison between the two approaches was undertaken because we were interested in exploring the relationship between point estimation tasks' performance and the methods used to sonify the points to be estimated.

Study 4

In study 4, we performed an observational study to examine the fourth version of MAG with both sighted and VI participants (see Chapter 7). Four different sonification mappings are compared in this study: 1) single point; 2) single reference; 3) Multiple references with a fixed step size of 20^{ths} of YMax and 4) Multiple references with a fixed step size 10ths of YMax.

Because we wished to have a usable representation of negative numbers for use in these tasks, we also evaluated whether there was any difference across the four conditions on point estimation tasks and a polarity sign task for each group of participants. The choice of the representation employed for negative numbers is fully explained in chapter 7. The results achieved by sighted and VI participants is also compared.

3.4. Methods

User Experimental Studies

In study 1, we explored the basic feasibility of presenting auditory graphs using a mobile application with sighted participants, evaluating our first version of the MAG app to determine how accurately point estimation and graph reproduction tasks could be performed. The first question will be addressed by examining the hypothesis one and two on point estimation and graph reproduction tasks. Hypothesis one predicted that participants would make more point estimation errors as the task's complexity increases (i.e. more data points). We evaluated this by calculating the root mean squared errors (RMSE) between the estimated (predicted) values to the true values. To test hypothesis two, we assessed the correlation between the estimated and the true values for all tasks. Our performance analysis is performed by comparing RMSE values and the correlation means obtained from graphs with more data points with fewer data points.

In study 2, we experimented to test the second version of MAG with VI-participants. We refined the first version by adding multitouch gestures and limiting the number of plays back repetitions to a maximum of 3. The difference in performance between using multitouch gestures and passive listening modes was assessed. We also investigated the correlation coefficient between the estimated and actual values to see how this varied due to the number of points on the auditory graphs.

In study 3, we compared the single reference sonification approach with a new approach to multiple reference sonification following Metatla et al.'s (2016) work. We addressed how employing a multi-reference sonification mapping in auditory graphs could improve the performance of non-visual point estimation tasks, while still being as time-efficient as those using pitch only graphs. In this study, our participants were asked to perform 10-point estimation trials per mode, i.e. either pitch-only or multi-reference.

In study 4, we explored four alternatives means of representing Y coordinate values using single and multiple reference tones as well as negative numbers using the fourth version of MAG.

Semi-Structured Studies

Semi-structured qualitative studies primarily focus on the development of an exploratory understanding of a situation (Blandford, 2013). The approach includes several qualitative and quantitative research methods that involve questionnaires, observational studies and interviews. Qualitative methods such as structured observations and semi-structured interviews allowed us to understand the process of point estimation and graph rendering as well as user behavior during interaction. In all four studies, we conducted semi-structured interviews to discuss the issues encountered while performing the tasks. Interviews are best suited to understand people's perception of a situation and provide an opportunity for the researcher to explore people's experiences in more detail (Denzin & Lincoln, 2000).

Quantitative data such as point estimation errors, completion times, the correlation coefficient between estimated and actual values, etc. are essential measures to understand the performance of the prototype.

Usability Evaluations

The Human Computer Interaction (HCI) discipline deals with many aspects of sonification systems. Variables may have domain-specific dependencies between data and measurement tool in sonification system, and for that reason, knowledge of the domain can help design effective sonifications (Hermann, 2002). Sonification designers have to acknowledge that not all mappings are generated equally. They have to use guidelines and usability evaluations to ensure that the message they intend to convey is received by the listener (Walker & Nees, 2011).

In study 1, the speed option was used to evaluate the users' listening experience with the interface. The aim is to determine how the participants require long pauses between notes in a series of auditory graphs. Further in study 3, a questionnaire was distributed to evaluate the participants' comfort levels when reproducing the graphs. To examine the extent to which participants believed that certain conditions would be better when performing the point estimation and polarity sign tasks, a questionnaire was developed using a five-point Likert scale which was used in study 4.

3.5. Study Design

We recruited different participants for each study. The objective here was to reduce the potential for learning effects that may influence the analyses' results, as some procedures and tasks could have several similarities across different studies.

In all four studies, all participants were trained by illustrating the mappings of notes used to represent the minimum value and then increasing linearly to the maximum value. Following exposing participants to the range of notes to represent the corresponding range of values to be sonified, they are then trained to listen to and estimate a series of random values represented using the specific sonification approach or approaches to be employed in that study.

In the first study, we conducted 14 tests for each participant, with an increasing number of notes for each subsequent test. In studies 2 to 4, participants were paired to use a different mode for each of the graphs in random order, as a counterbalance to the possible learning effect of using two modes with the same graph. In studies 2-4, we also limited the number of repetitions allowed of sonifications to a maximum of three.

The participants were randomly assigned to one of two groups. Each of them was asked to estimate two graphs (Graph A and Graph B) with one of the two modes: passive listening or multitouch gestures for study 2 and single point or multi-reference mode for study 3. They were paired to use a different mode for each of the graphs to eliminate the learning effect of having two modes for the same graph. Participants listened to the sounds via the built-in speakers on the tablet. The volume on the tablet was set to maximum for all tests.

In Study 4, participants were asked to perform all four conditions (single point, single reference, multi-reference for steps of 20 and multi-reference for steps of 10) in random order. Therefore, one could start with condition three, for example, but the next participant could start with condition four.

3.6. Instruments

3.6.1. Point Estimation and Graph Reproduction Tasks

Point estimation is relatively easy to perform with visual data in tabular form (J. Meyer, 2000); however, data is often presented in graphical form (Zacks et al., 2002). Extracting information about a single datum is, therefore, a task that may need to be solved from a display of the data rather than with a table. With the accessibility of STEM subjects and work opportunities for VI people in mind, researchers have started to study the feasibility of point estimation using auditory displays of quantitative data, such as auditory graphs (Walker & Nees, 2011).

Unlike *trend identification*, which concerns with increases and decreases in quantitative data (Walker & Nees, 2011), *graph reproduction* is a more holistic listening task whereby a user attempts to identify the overall set of notes representing quantitative data. This technique is typically tested by asking sighted participants to reproduce the auditory graph graphically or by extracting properties from the graph, including maximum and minimum values, and by estimating other points of interest (Harrar & Stockman, 2007). The reproduction process may be undertaken either during or after hearing the audio representation of the graph and might take the form of a verbal description (particularly for VI participants), a drawing, or some combination of the two.

In studies 1 and 2, we asked the participants to reproduce the auditory graph (verbally for visually impaired participants) by estimating the set of notes representing each graph. In the design of the exploratory tasks used in these studies, we have been influenced by Harrar (2007), who introduced the three complexity levels (i.e., simple, medium, and complex) based on the increase in the number of data points and also the speed and mode of presentation of auditory graphs.

On the other hand, studies 3 and 4 both involved non-visual point estimation tasks employing a range of different sonification methods to represent points on the graph, including multi-reference sonification approaches following Metatla et al. (2016). Full details of each of the sonification approaches employed are given in the corresponding chapter (chapter 6 for study 3 and chapter 7 for study 4). Study 4 also included a usability study of the representation of negative numbers.

3.7. Data Collection Techniques

We used the following methods of data collection during the experimental studies.

3.7.1. Questionnaires

Questionnaires provide an effective means of assessing the attitudes, behaviour, preferences, opinions, and intentions obtained from a relatively large number of people less expensively and more efficiently than interviews (Guest, Namey, & Mitchell, 2017). From study 1 to study 4, questionnaires were administered before the experimental session to gather data about participants' demographics, their experience of using technology such as mobile applications, auditory graphs, and (for VI users only) screen readers. In studies 2, 3, and 4, we also asked about participants' familiarity with musical notes and musical training experience. In studies 3 and 4, questionnaires were used to collect data on accessibility and usability satisfaction following the experimental sessions. The VI participants received digital versions of the questionnaires to complete.

3.7.2. Semi-Structured Interviews

Semi-structured interviews were used in all studies after the experimental session to collect data about the user's perception of the MAG app's corresponding version. The semi-structured interviews were conducted in a one to one format. These interviews' general objective was to examine issues encountered while using the interface and gain participants' perspectives about ways to enhance the MAG app design.

3.7.3. Graphical grids

In study 1, sighted participants were given a sheet of paper containing several empty X-Y axes, which help the participants to mark their estimated points. The participants selected the grid positions corresponding to the Y-axis points and wrote them on the paper grid. In studies 2, 3, and 4, VI participants estimated the value of a point by simply answering verbally so that the researcher could note it down.

3.7.4. Videotaping and Audio Recording

The sessions in study 2,3 and 4 were recorded and the researcher then transcribed the conversations. Recording both audio and video is a powerful tool for studying phenomena that are complex and relatively unknown, like communications (Morse, 1994). Thus, the main limitations of video recording are the lack of contextuality and the fact that the researcher does not have the opportunity to test his interpretation as an active participant on-site (LicNSc, 2001).

Recordings facilitate a repeated review of the data so that different behaviour and relationships between the phenomena of interest can be observed. Repeated replay allows the researcher to focus on various factors to analyse multiple participants and their interactions and multiple factors in the environment and establish relationships between them. Qualitative sequential analysis is also useful for linking conditions and events over time and evaluating the history and consequences of phenomena of interest (Williams, Herman, & Bontempo, 2013). For example, Williams et al. (2013) have utilized video observations of nurse-resident interactions at mealtimes to identify the sequential associations between nurse person-centered actions, nurse task-centered actions, and resident behavioral symptoms, and the time variation in these associations in dementia care facility.

Researchers are generally expected to record identifiable information about participants only when it is required for research purposes. Recording audio or video and taking photographs of participants usually results in collecting identifiable information about participants. Current scientific advances in technology support the recording of observation data on video. There is mixed evidence, however, to demonstrate the added value of video over audio recordings, and a careful assessment of whether the benefits of video recording outweigh the extra costs and data management requirements in relation to the research question is justified (Howe, 1997; Weingarten et al., 2001; Williams et al., 2013).

Recording audio is generally less intrusive than recording video because of the lack of visual images. Audio recording also eliminates the need for an investigator to operate a camera. In general, audio equipment is less expensive, and making audio recordings is generally simpler than making a video recording.

The process of video recording is intricate and demands attention in both audio and visual recording of the data. Since an operator is required and can be easily observed by the participants, video recording can alter naturally occurring communication more than audio recording. Video file sizes can be excessively large, with additional costs for storage, time, and material required to record and archive data.

In this thesis's work, audio recording was used when interviewing participants in all studies; interviews were subsequently transcribed by the researcher. We also video recorded part of the experimental sessions for verification that everything was working correctly. Still, for several of the limitations mentioned above, the quantity of video recording was limited to only what was required for verification purposes. Participants gave their consent for all video and audio recording.

3.8. Data Analysis

3.8.1. Pre-processing data and normality test

Figure 3-1 discusses the statistical analysis testing flowchart for this thesis. Before we decided to use parametric or non-parametric analysis, we assessed each variable's normality (univariate normality) by inspecting the respective Shapiro-Wilk Normality test and also skewness and kurtosis values. If the p-value of the Shapiro-Wilk test is less than 0.05, then the data violate the normality assumption. When required, we transformed the data to improve its normality distribution using inverse normal transformation (INT) generated by log transform and arcsine for negative values (Derrick, White, & Toher, 2017). It was preferred in our studies as it is arguably a more advanced method for transforming the data than the commonly used log and arcine transformation.

However, researchers often use the skewness and kurtosis values, which are less conservative than the Shapiro-Wilk test. The skewness occurs when responses are more frequent at one part of the measurement scale and affect the variance-covariance among variables. Kurtosis reflects the flatness in the data distribution. The further the value of skewness or kurtosis is from zero, the more likely it is that the data are not normally distributed. The standardized values of skewness and kurtosis are obtained by merely subtracting the mean of the distribution (the skewness values or excess kurtosis) and then divide by the standard deviation of the distribution (converted into z-scores).

$$Z_{skewness} = \frac{S-0}{SE_{skewness}} \quad \text{and} \quad Z_{kurtosis} = \frac{K-0}{SE_{kurtosis}} \tag{2}$$

We followed the threshold values of 3.29 as recommended by (Field, Miles, & Field, 2012). The skewness, kurtosis values, and their respective z-scores are discussed further for study 1 and 4 to further establish the principles that underpin all of the statistical analysis that follows Chapters 4 to 7. In particular, to clarify whether or not inferential tests (ANOVA, t-Test, Mann-Whitney, etc.) are: paired (repeated measures) or unpaired; and if they are one or two-tailed tests.

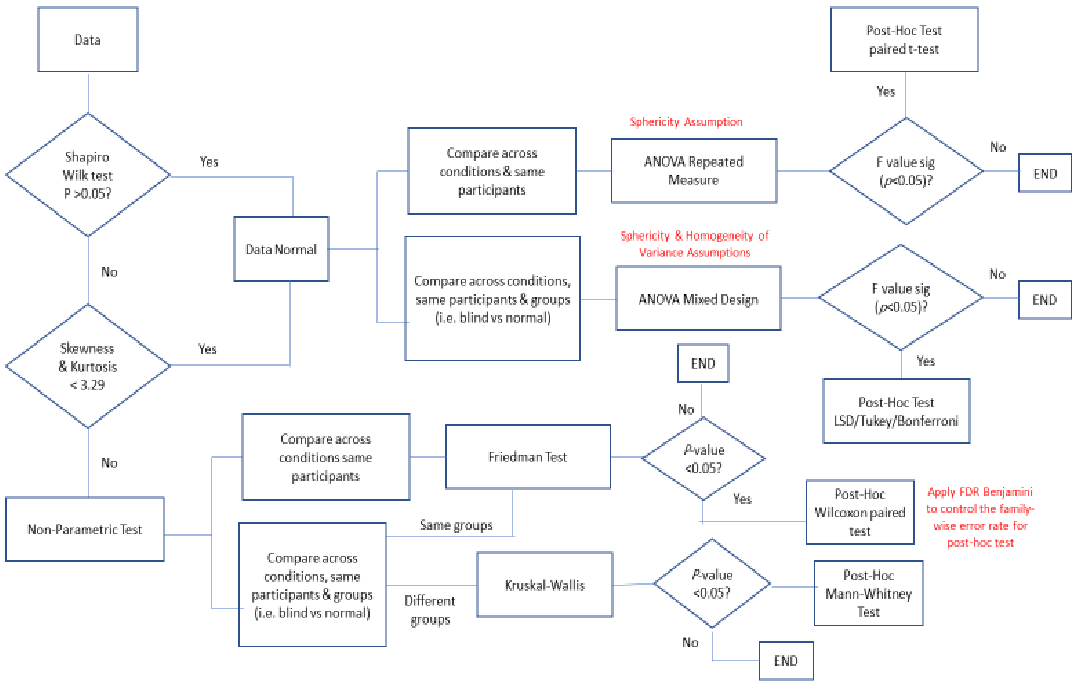


Figure 3-1 Flowchart of the statistical analysis testing. The significance level was set at alpha 0.05 of the two-tailed test

If the data meet the normality assumption, we determine to use the parametric test and non-parametric test otherwise. To evaluate whether any significant differences of the variable of interest between groups (i.e., RMSE on point estimation task, correlation coefficient on graph reproduction task, and time interval on length of pause), we will conduct ANOVA repeated measure or Friedman test if data distribution was not normal. Both analyses compare three or more groups where the participants are the same in each group.

3.8.2. Sphericity test

In addition to the normality distribution assumption, the ANOVA repeated measure requires the sphericity assumption, as shown in Figure 3-1. Sphericity (denoted by ε and sometimes referred to as "circularity") is a more generic compound symmetry condition. Compound symmetry holds when both the variance across conditions and the covariances between pairs of conditions are equal. Sphericity refers to the equality of variances of the differences between treatment levels. Suppose Mauchly's test statistics are significant (p-value< 0.05), we should conclude that there are significant differences between the variances of differences; the sphericity condition assumption is violated. When the result violates the sphericity assumption, the valid F-ratio can be produced using the Greenhouse-Geisser correction if sphericity estimates is less than 0.75. When the estimate is greater than 0.75, the Huynh-Feldt correction should be used. (Field et al., 2012).

3.8.3. Statistical Testing

Before analyzing inferential statistics, we presented descriptive statistics to summarize the central tendency (means, median), variability (standard deviation, inter-quartile range), frequency, and distribution of each variable of interest. Plots of data are also displayed to visualize the tendency, identify outliers, and anomalies in the data (Kelly, 2009). We used the root-mean-square error (RMSE) to calculate the error of a model by taking the differences between values (sample or population values) predicted by a model (standardized estimated or forecasted values = f) and the values observed (true values = o). The formula followed the Barnston (1992) equation as follow:

$$\boldsymbol{RMSE_{fo}} = \left[\sum_{i=1}^{N} \left(z_{fi} - z_{oi}\right)^2 \Big/ N \right]^{\frac{1}{2}} \tag{3}$$

Where:

- Σ = summation ("add up")
- $\left(z_{fi} - z_{oi}\right)^2$ = squared differences
- N = sample size.

To investigate how closely the users' mental models matched the actual plot, we calculated the correlation between the true and observed values for all participants using the Kendall tau non-parametric test. Kendall's tau is a robust technique to outliers and the normality assumption and is an alternative to the Pearson correlation parametric test (Newson, 2002). The mean values of Kendall's tau correlation between the true and observed values for all participants are measured for each condition (i.e. simple, medium and complex) to group them later on the graph and the box plot of the mean values of Kendall's tau correlation coefficient. A boxplot is useful in our analysis to indicate how the values are distributed in the data.

Moreover, we performed inferential statistics to answer specific hypotheses as formulated in detail in each study. To evaluate the differences between more than two conditions in study 1 to study 4, for example, across all graphs (simple, medium, complex), we performed the Kruskall Wallis test, a non-parametric test which is robust to outliers and normality assumptions. If a significant difference was found, a post-hoc multiple comparison analysis was conducted to examine which levels of the independent variable differ from each other. In this case, we applied the Benjamini-Hochberg correction method to control the familywise error rate (FER) or false discovery rate (FDR). Furthermore, a Mann-Whitney test was also carried out to examine the difference between two groups (e.g. sighted and VI participants).

3.9. Sampling

In this thesis, we employed purposive sampling which involves selecting a sample of participants who are likely to use our design based on several inclusion and exclusion criteria (Hanson et al., 2005). The inclusion criteria for both sighted and VI participants recruited in the study were experienced mobile applications users. For VI participants, we included only those who were completely blind and used a speech-based screen reader on their mobile device.

Study 1: we recruited 16 sighted participants with four other participants dropped from the study because they were not able to finish all the tasks.

Study 2: we recruited 15 VI participants and they were randomly assigned to two groups in a within-subject experimental design.

Study 3: we recruited 20 sighted participants and they were randomly assigned to two groups of ten in a within-subject experimental design.

Study 4: we recruited 20 participants for both VI and sighted group and they were randomly assigned to perform all four different conditions (single point, single reference, multi-reference for steps of 20 and multi-reference for steps of 10).

3.10. Recruitment and Ethical Consideration

The sighted participants were recruited through a posted announcement through the university mailing list and by asking our colleagues in Queen Mary University of London. VI participants were recruited through Indonesian blind association and a posted announcement through social media. The participants in study 1 to study 4 were different. In all of the studies, both sighted and visually impaired participants received cash incentives for their participation.

3.11. The Development of the MAG app Prototype

The Mobile Auditory Graph app was developed on the Android operating system. There are two main reasons why we chose this support software environment: primarily, it does not have a lot of accessibility barriers (Grussenmeyer & Folmer, 2017; Mascetti et al., 2016). Android app development can be performed on multiple operating systems, and it is open source. By contrast, development for iOS requires one of Apple's development languages, which are only used for Apple-centric development (iOS and OS X), and they cannot be applied to other operating systems. Secondly, the Android app does not have any dependency issues and it can be easily integrated with other programming languages. It uses the platform-independent language, Java, with the help of a rich set of libraries (Friesen, 2010). Therefore, the MAG app can easily be integrated into other software.

3.11.1. Hardware

The first prototype (i.e. MAG app 1.0) was developed based on the Android Operating System version 6.0.1 on Google Nexus 7 tablet for MAG app 1.0, which comes with a 7.02-inch (180 mm) display. It runs the TalkBack screen reader for VI accessibility option. The TalkBack screen reader is a feature that helps blind and VI users to receive the data by using spoken text when touching, selecting, or activating objects on the screen.

The MAG app 2.0 was developed on a 9.7-inch screen Samsung Galaxy Tab S2 with the Android 7 operating system. It has a larger screen than the previous MAG app 1.0 and runs the Voice Assistant screen reader, which replaces the TalkBack feature on older mobile phone models and operating systems.

3.11.2. Software

The MAG app has been developed on Android Studio which is an integrated development environment (IDE) based on Java for the development of Android platforms. This IDE is a cross-platform tool for creating apps for any Android device. Navigating the MAG app on the device is performed by locating the MAG app icon while the screen reader is activated or in a visual mode. The audio is presented through the device's built-in speakers. The MAG app reads values entered by the users and has additional functionality to read from a file, then automatically generates sound for auditory graphs based on this information.

3.11.3. Overview of MAG app 1.0

In the first design, the position points were mapped to the Y coordinate following positive polarity. Thus, when the Y value increases, the pitch will also increase. The sound parameter, i.e., the frequency, varies during playback of a data set by mapping the data to sounds. F5 notes produced the sounds with a 698 Hz frequency (Sengpiel, n.d.; Suits, 1998).

The value 0-100 are mapped in linearly scaling data array from 69.8 Hz to 698 Hz: the higher the data point, the higher the frequency, as displayed in Table 3.1. When points on the graph do not fall into one of the defined Y values presented in Table 3.1, the frequencies dynamically change using linear distribution. For example, the point on position 35, which is located between 30 (209.4 Hz) and 40 (279.2), would play MIDI half of the sum of these two frequencies, i.e., 244.3 Hz.

Y	Frequency (Hz)
0	No sound
10	69.8
20	139.6
30	209.4
40	279.2
50	349
60	418.8
70	488.6
80	558.4
90	628.2
100	698

Table 3.1. Reference Mapping of Note Frequency Values to Y-Axis Values

The musical note represents the corresponding Y value and represents the sound of the line graph.

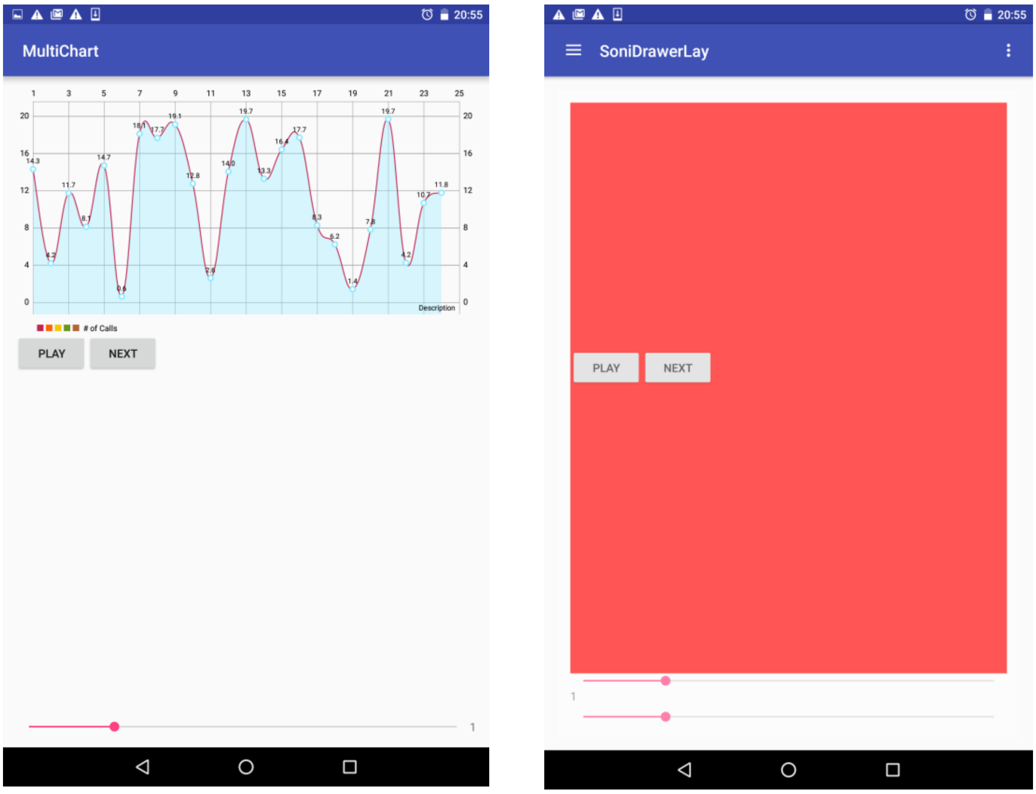


Figure 3-2 MAG on Normal View (Left) and MAG on Blindfolded View (Right)

This first MAG app design has two different interfaces to present the graphs: the uncovered normal view graphs and the blind view covered graphs, as shown in Figure 3-2. The application will show a graph with multiple data on the X-axis and Y-axis. The graphs were created based on the number of notes, and then for the covered graphs, these numbers were covered with a red layer over the graphs so that sighted participants could not see these numbers. This scenario is intended to hide the numbers from participants taking part in experiments.

The "drawer menu" has been set up for navigation to other task menus to enable users to change the category of graphs between simple, medium, and complex graphs. A speed control SeekBar was added to change the time interval between the data points. The speed is divided into five steps, starting with step 1 for the interval 0.5 s; step 2 for the interval 1.0 s; and so on in increments of 0.5s up to step 5 for intervals of 2.5 s.

3.11.3.1. Graph complexity

The graphs were arranged as follow:

- Four simple graphs, ranging from 6 to 10 points.
- Six medium graphs, ranging from 11 to 25 points.
- Six more complex graphs, ranging from 25 to 40 points.

A previous study by Harrar (2007) has also categorized three levels of complexity by having "low" for 12 points, "medium" for up to 30 points with 3-4 peaks, and "high" for 30 points with 7-8 peaks.

Our points for each successive graph are gradually increased to build up the complexity of the graphs. All 14 graphs can be opened in a sequence of tasks by pressing the "next" button before proceeding with the next task. In this mode, users need to press the Playback button to hear the series of notes corresponding to all the points in the currently selected graph.

The MAG app interface has been structured according to the navigation menu, map area, button area, and scroll bar. The navigation menu has a window on the left side, which shows the MAG app's primary navigation options. It is hidden by default, but it will be displayed when the users swipe their fingers from the left side of the screen or touch the program icons on the menu bar. It has three menu lists, i.e., simple, medium and complex to represent the three categories of graphs.

The graph area on the normal view displays X and Y axes with coordinates from 0 to 100 and displays a line graph corresponding to the currently selected graph. When the "blindfolded" mode is selected in the app menu, the visual representation of the graph is obscured so that participants cannot see the values during experiments. The buttons "Play" and "Next" are located after the graphs and are not hidden from the user's view. There is a SeekBar for changing the audio frequency setting and another SeekBar for setting the audio interval between the five different levels.

3.11.3.2. How the literature influenced the design of the MAG app

The design of the MAG app was influenced by several previous studies. First, several works by Walker et.al (Nees & Walker, 2008; Smith & Walker, 2002; Walker & Mauney, 2010) introduced systematic methods for the creation and evaluation of useful and usable auditory graph. These studies also placed the use of parameter mapping in auditory graphs on a more scientific basis. In addition, Walker (2005) proposed that adding modalities into auditory graph results in a better understanding of quantitative information. Therefore, we further develop the multi-touch modality as an alternative to passive listening in study 2.

A previous study by Harrar (2007) influenced the design of studies 1 and 2 in the use of three complexity leveling by having "low" for 12 points, "medium" for up to 30 points with 3-4 peaks, and "high" for 30 points with 7-8 peaks. Harrar (2007) researched the design of auditory graphs by comparing discrete sounds and continuous sounds for graph reproduction. Harrar et al. tested this technique by asking participants to reproduce the auditory graph visually or by extracting properties from the graph and by estimating other points of interest.

Third, the work of (Metatla et al., 2016) showed that adding multi reference tones when sonifying data could increase point estimation accuracy. This work inspired the design of studies 3 and 4. However, (Metatla 2016)'s results have a drawback that user performance become slower as the point to be estimated gets further away from zero. Therefore, in study 3 we investigated an alternative scheme in which we present fewer reference tones and examine the effect of this on user performance.

3.11.4. Overview of MAG app 2.0

The second design used the same pitch mapping as described in MAG app 1.0 in table 3.1. On MAG app 2.0, a swipe interaction was added to help the users locate the points on the X-Y coordinates. The interface is a mobile graph application with multimodal input by allowing gesture interaction and haptic feedback.

The TalkBack screen reader can access the menus and the primary navigation around the MAG app interface. In contrast to preliminary study 1, where accessibility is not required, the screen reader mode for accessibility was activated on MAG app 2.0 to let the VI user navigate the app.

The sound parameter, i.e., the frequency, varies to transmit a data set to the listener by mapping data to sounds. The data values increase according to the scale from 0 to a maximum value of 10.

The purpose of the series of studies is to test the effectiveness of the prototype with the help of sonified graphs employing two types of multimodal interaction: passive listening (playback) and multitouch gestures (swipe) as follows:

Playback Modality

This study's playback modality is designed as a tool to sonify each data point with a musical note within a specific time interval. The user can interact with this modality using a simple gesture, by physically touching the playback icon. When the user double-taps the play button, a series of tones are played representing all of the points on a line graph, the pitch of each tone being proportional to represent the Y coordinate of the respective point on the graph. In terms of the touchscreen interaction technique (see section 2.7.1), this modality is classified as a pointing technique. To trigger the playback action mechanism, VI users should locate the playback button using the TalkBack screen reader and double-tap it. The button is placed under the graph.

Swipe Modality

Swiping is a haptic interaction commonly found on smartphones and tablets that enables users to interact through the sense of touch. Compared to the playback modality, the swipe mechanism for MAG app is designed to let users perceive the data notes by sequencing the Y-data manually, i.e., by touching each note along the x-axis. When the screen reader is activated, two fingers are required to scroll, as described in section 2.7.4. This method can increase errors as the finger contact area's size is at least twice larger than using one finger, while the target coordinates are very small.

In the swipe modality, the pointing and scrolling techniques are combined with a continuous X-axis movement. This interaction on the computer device involves selecting an icon with a mouse and dragging it to the target.

Obtaining user feedback

The third prototype was built to test our approach to multi-reference sonification and learn more about the effect of increasing graph complexity (number of data points) on user tasks.

3.11.5. Overview of MAG app 3.0

Multiple Reference Sonification

The third prototype uses the same pitch assignment as in the second prototype. Instead of hearing only one pitch note, the user listens to several consecutive reference tones with different frequencies, which are all points before the present position and the source reference. This version only works for positive Y values. We think of the Y values going from 0 up to a maximum value (Y_{Max}). The idea then is to play notes leading up to the value of the point the user is trying to estimate ($Y_{Estimate}$). When $Y_{Estimate} > Y_{Max} * 0.5$, we only start playing them from the point $Y_{Max} * 0.5$, so the user will never have more than five notes played before the value of $Y_{Estimate}$ is played. This is because research found that if many reference tones were played, people lost track of them and became less useful (Metatla et al., 2016).

3.11.6. Overview of MAG app 4.0

The fourth prototype was built with the aim to enable the comparison of 4 different sonification approaches, to learn how performance with these varied with graph complexity (number of data points) and also to perform a usability study on the representation of negative numbers.

The prototype, therefore, has four conditions as follows:

1) Single point, where the point to be estimated is represented by one single note representing its position on the Y-axis
2) Single reference, which is similar to condition 1, but each note is preceded by a reminder of the pitch corresponding to Y = 0.

3) An approach where notes are played representing each 5th of the distance from 0 to Y-max, up to the value just below the point to be estimated, and then a final note representing the actual point itself.

   For example, if the notes can range from 0 to 100, to represent a point to be estimated at 65, points would be played representing the following values: 0, 20, 40, 60, 65.

4) Multiple references with a fixed step size of a 10th of YMax. In this condition, if $Y_{Max} = 100$, notes will be played from 0, in fixed steps sizes of 10, up to the value just below the point on the Y-axis to be estimated, and then the note itself. For example, the representation of 65 would include notes for Y = 0, 10, 20, 30, 40, 50, 60 and 65.

   We set Y-max to 100; one instrument will be used for all displays rather than the two used in the previous study. Our exploratory study 3 reveals that using a separate timbre to display audio in this particular study could lead to errors since participants could either ignore it or did/could not hear the timbral difference.

A pause of 500 ms was used -based on the outcome from study 1- between all the notes representing points in conditions 2, 3 and 4, in all sonification conditions.

Non-Linear Human Hearing

In study 4, we used a different pitch mapping scheme -as discussed in section 2.2- that differs from the previous three prototypes that use linear mapping. According to Zwicker (1938), frequency range and mapping were designed to conform to the human auditory range, following the exponential distribution, with the following frequencies changing by a fixed coefficient rather than by a constant term. Therefore, humans can determine the pitch and timing of a sound signal with greater accuracy than conventional linear analysis allows.

Table 3.2 shows the mapping reference ranging from 0 to 100 following the keys of a modern 88-key piano frequency standard, ranging from the 28th key, the C3, tuned to 130.81 Hz for 0 note to the 79th key, the D♯7/E♭7, tuned to 2489.01 Hz for 100.

Y Note	Key Number	Piano Key	Frequency (Hz)
0	28	C3	130.8128
10	34	F♯3/G♭3	184.9972
20	40	C4	261.6256
30	43	D♯4/E♭4	311.127
40	49	A4 A440	440
50	55	D♯5/E♭5	622.254
60	61	A5	880
70	64	C6 Soprano C	1046.502
80	70	F♯6/G♭6	1479.978
90	76	C7 Double high C	2093.005
100	79	D♯7/E♭7	2489.016

Table 3.2 Mapping reference value of Y-note with a respective key number, piano key, and its frequency in Hertz

The range and mapping were selected to match the human auditory range and the extent that quasi-melodic processing is engaged by listener stimuli, which is a wider pitch range compared to our previous prototype that was implementing linear mapping.

Representation of Negative Numbers

Negative numbers may be represented as *componential representation*, as two separate components (one digit and one sign) or *holistically*, as discussed in section 2.2. For the fourth version of the MAG app, we chose the *componential representation* approach to represent the mapping by having the same positive mapping reference for the digit as described in Table 3.2 and adding one sign before the digit with "sonar" sound. The sonar sound is intuitively chosen because it brings to mind a submarine positioned below 0 meters on the land, suggesting the perception of a value < 0.

Chapter 4. Research study one: an exploratory study of point estimation and graph reproduction tasks

4.1. Introduction

Auditory graphs offer the capability to make data more accessible for diverse user populations. Although the main target population for this study is VI people, it is useful to know how the interaction model of the MAG app works among sighted users. This can provide an interesting baseline for comparison with VI users and provides evidence one way or the other concerning the app's general usability.

This chapter's study is an exploratory study to evaluate the usability of our first MAG app prototype using both the uncovered normal view graphs and the blind view covered graphs. We investigate how well sighted participants perform graph reproduction tasks by using a touch screen mobile device. This study's overall objective is to better understand the users' interaction with the MAG app during point estimation and graph reproduction tasks. Through an experimental design, we will determine to what extent the accuracy decreases with the addition of the number of points. We also intend to identify the challenges that occur in the app development's basic stages as a foundation for further refinement study with VI users.

As discussed in chapter 2, it has been suggested that auditory graphs could improve the understanding of information for blind and partially sighted students or students with sight whose learning style leads them to use audio.

4.1.1. Point estimation and graph reproduction tasks

Walker and Nees (2011) have defined five steps so that a listener can perform a point estimation task using pitch-based parameter mapping as follows:

"1) listen to the sonification; 2) determine in time when the datum of interest occurs; 3) upon identifying the datum of interest, estimate the magnitude of the quantity represented by the pitch of tone; 4) compare this magnitude to a baseline or reference tone; and 5) report the value" (Walker & Nees, 2011).

Graph reproduction tasks (in this context) involve the listener auditioning a series of points rendered in audio, or possibly auditioning a continuous sonification used to represent a graph's shape. In some form reproducing the shape of the graph so represented. The reproduction process may be undertaken either during or after hearing the audio representation of the graph. It might take the form of a verbal description, a drawing, or some combination of the two.

4.2. Motivation

This exploratory study is motivated by the need to investigate the interaction between sighted users and the MAG app's first version, studying how closely participants' mental model formed from auditioning the graph matches the actual visual plot.

This study will highlight the challenges faced during the interaction, the pattern of behaviour that emerges during the interaction and the effect of altering the playback speed setting of the data notes.

4.3. Research Questions

Our study attempted to address the following specific research questions:

1. How well can sighted users estimate points on the Y-axis of a graph?

2. How long should the pause be between the presentation of successive notes in the sonification?

We formulated the hypotheses as follows:

H1:	Participants will make significantly more point estimation errors when listening to more data points (i.e. more complex audio graphs)
H2:	The simple graph will have the highest correlation coefficient when being compared to medium and complex graphs
H3:	Participants will require a significantly longer pause between notes when assessing more complex graphs

4.4. Study Design

To address all of the questions, we run an experimental design to investigate the effects of adding pitch to a line curve along the X-axis for each point. The experiment focused on adding sounds to the MAG app by mapping the graphs' y-coordinates to pitch, with the y-coordinates ranging from 0 to 100. The value 0-100 are mapped in linearly scaling data array to the frequency of sampling rate 698 Hz: the higher the data point, the higher the frequency, as displayed in Table 3.1.

The first question will be addressed by examining the hypothesis one and two on point estimation and graph reproduction tasks. Hypothesis one predicted that participants would make more point estimation errors as the task's complexity increases (i.e., more data points). We evaluated this by calculating the root mean squared errors (RMSE) between the estimated (predicted) values to the true values. To test hypothesis two, we will assess the correlation between the estimated and the true values for all tasks. We also plan to analyze whether participants' performance is similar or different between simple, medium, and complex graphs. Previous research in graph reproduction task has implemented these three-complexity levelling and they found that the estimation accuracy dropped when the note is rising to 30 (Harrar & Stockman, 2007). Our performance analysis will be implemented by comparing RMSE values and the correlation means obtained from graphs with more data points with fewer data points.

The second question will be answered by observing the users' behavior when they set the interval using the seek bar in the app when attempting to achieve better point estimations. A seek bar is provided so that the users can set the pause length between the played MIDI notes to display the notes' positions along the Y-axis. The X-axis points are mapped to time, and the position of the seek bar determines the time interval between notes.

4.4.1. Participants

A total of 16 participants were recruited from undergraduate or postgraduate students from various universities in London. Four participants dropped from the study due to task complexity reasons. The final sample consisted of 12 participants (9 males, 3 women) with ages ranging from 18 to 44. All participants were experienced in using the touch screen on mobile devices but did not have any experience in using such an auditory graph (e.g. MAG) on a mobile device.

4.4.2. Apparatus

We conducted the test in quiet rooms, free from disturbing noises at several London locations: the Queen Mary University of London and Nansen Village, Woodside Avenue.

The MAG app was presented on Google Nexus 7 running on the Android 6.0.1 operating system. The graphs were delivered in ascending order of complexity: the graph that has smallest number of points was presented first. The sounds were created from piano notes of audio .wav files with frequencies as described in Table 3.1. The result of the point estimation task from each participant was recorded as points on the XY-axis on-grid papers.

4.4.3. Experimental Procedure

At the beginning of the experiment, a demographic questionnaire was administered to participants. We then conducted a series of initial training sessions to get familiar with the MAG app and the tasks. The training relating to sonification took approximately ten minutes, followed by an explanation of all the controls available to the app user.

We started by explaining the idea of audio graphs and introducing them to the range of sound pitches they would hear. We explained that the range of pitches would represent Y values ranging from 0 to 100, with 0 corresponding to no sound and 100 corresponding to the highest pitch. In general, the lower the sound, the lower the Y value, and vice versa. We showed them the low and high ends of the range of pitches so that they could start to form a mental representation of the pitch range to be employed. We demonstrated how to perform a graph reproduction task using both blind and normal view on the MAG app. The participants were allowed to explore the application to become familiar with the playback modality and the frequency ranges used to represent points on each graph's Y-axis.

To evaluate the extent to which the accuracy of participants' mental models generated from the sonification matched the actual plot, participants were asked to draw their perceived plot for each data set. They were provided a sheet of paper containing several empty X-Y axes that helped mark the estimated points using a pen. After playing and listening to the audio graph, they were asked to draw the perceived plot by selecting the grid positions corresponding to Y values' points along the X-axis. They could listen to the audio graphs as often as possible and change their prediction of point locations on the Y-axis while repeating the audio. If they started to listen to the audio again, they could continue to complete the plot on the same set of X-Y axes on the paper grid.

Participants could move to the next graph test by tapping the "next" button. They could continue to the other tasks by navigating the menu button then selecting the series of graphs presented in order. The participants could use the touch screen by tapping the "Play" button so that the system would playback using the default speed for graphs with intervals of 500 milliseconds for each point. For example, assuming the user run the first task of simple graphs consisted of 6 data points, the entire audio playback graph would be rendered for 3 seconds each time they tapped the button. They could also change the length of the pause between notes with the “seek bar” provided on the application. If they changed the pause duration between notes, the application would be remembered when they moved to their next graph.

The tick position marked on the grid is based on the interpretation of sound from the audio graphs. The X-point represents its point in time in the playback relative to other notes and the Y-point is the participant’s estimate of that point, scaled in the range 0-100. During the analysis, which took place after the participant had left, their estimates were compared with the actual values for further analysis.

In the main experiment, participants worked their way through from simple to complex graphs. They filled the blank sheet by marking the points from the beginning to the end of the line curve one after the other. After completing the first series of simple graphs, they were offered a 5-minute break, then moved on to the six medium tasks, followed by another 5-minute break. Finally, they completed the last four complex graphs. At the end of the experimental tasks, they were asked for their feedback by completing a questionnaire.

The experiment was initially planned to be completed within 30-45 minutes: starting from opening the first graph and working through the graphs sequentially. Because the first participant took 90 minutes to complete all tasks, we removed the last two complex tasks, changing the number of complex tasks from 6 to 4.

4.4.4. Statistical Analysis

We used the root-mean-square error (RMSE) to calculate the error of a model by taking the differences between values (sample or population values) predicted by a model (standardized estimated or forecasted values = *f*) and the values observed (true values = *o*). Moreover, we also conducted a correlation analysis to examine the accuracy of the mental model on graph reproduction task and the difference between all the difficulty categories. Finally, the choices of user's when selecting the pause duration will be analyzed.

4.5. ANOVA statistical testing

4.5.1. Normality test

To further establish the principles that underpin all of the statistical analysis that follows this study, we follow the statistical analysis testing flowchart as discussed in 3.8. In particular, to clarify whether or not inferential tests (ANOVA, t-Test, Mann-Whitney, etc.) are: paired (repeated measures) or unpaired; and if they are one or two-tailed tests. Before we decided to use parametric or non-parametric analysis, we assessed each variable's normality (univariate normality) by inspecting the respective Shapiro-Wilk Normality test and also skewness and kurtosis values. If the p-value of the Shapiro-Wilk test is less than 0.05, then the data violate the normality assumption. However, researchers often use the skewness and kurtosis values, which are less conservative than the Shapiro-Wilk test. The skewness occurs when responses are more frequent at one part of the measurement scale and affect the variance-covariance among variables. Kurtosis reflects the flatness in the data distribution. The further the value of skewness or kurtosis is from zero, the more likely it is that the data are not normally distributed. The skewness, kurtosis values, and their respective z-scores for each condition is provided in Appendix A.

If the data meet the normality assumption, we determine to use the parametric test and non-parametric test otherwise. To evaluate whether any significant differences of the variable of interest between groups (i.e., RMSE on point estimation task, correlation coefficient on graph reproduction task, and time interval on length of pause), we will conduct ANOVA repeated measure or Friedman test if data distribution was not normal. Both analyses compare three or more groups where the participants are the same in each group.

4.5.2. Sphericity Test

In addition to the normality distribution assumption, the ANOVA repeated measure requires the sphericity assumption, as discussed in 3.8. The sphericity of RMSE, Correlation Coefficient, and Length of Pause is provided in Appendix A.

	Mauchly's	Sig (p-value)	Greenhouse-Geisser
RMSE into 3 conditions	.448	.018	.645
RMSE into 14 conditions	.000	< 0.001	.384
Correlation Coefficient into 3 conditions	.448	.018	.645
Correlation Coefficient into 14 conditions	.000	< 0.001	.318
Length of Pause into 3 conditions	.459	.020	.649
Length of Pause into 14 conditions			

Table 4.1. Sphericity of Data Study 1

4.5.3. Point Estimation Task

A. Parametric Test- ANOVA Repeated Measures

We grouped the tasks into three categories based on their complexities and we could not find any significant difference using ANOVA repeated measure (F=1.074, p=0.359), as shown in Table 4.2.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Task	Sphericity Assumed	19.755	2	9.878	1.074	.359
	Greenhouse-Geisser	19.755	1.790	11.034	1.074	.354
	Huynh-Feldt	19.755	2.000	9.878	1.074	.359
	Lower-bound	19.755	1.000	19.755	1.074	.322
Error(task)	Sphericity Assumed	202.262	22	9.194		
	Greenhouse-Geisser	202.262	19.694	10.270		
	Huynh-Feldt	202.262	22.000	9.194		
	Lower-bound	202.262	11.000	18.387		

Table 4.2. ANOVA Repeated Measure on Point Estimation Task (RMSE) for 3 task groups

However, when we examine more carefully, by grouping the tasks according to a specific number of data points (14 groups: simple 1, medium 1, complex 1, etc.), we found that the participants showed lower performance at specific points, as shown in Table 4.3.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Task	Sphericity Assumed	1702.069	13	130.928	4.035	.000
	Greenhouse-Geisser	1702.069	4.990	341.090	4.035	.003
	Huynh-Feldt	1702.069	9.631	176.730	4.035	.000
	Lower-bound	1702.069	1.000	1702.069	4.035	.070
Error(task)	Sphericity Assumed	4640.658	143	32.452		
	Greenhouse-Geisser	4640.658	54.891	84.543		
	Huynh-Feldt	4640.658	105.940	43.805		
	Lower-bound	4640.658	11.000	421.878		

Table 4.3. ANOVA Repeated Measure on Point Estimation Task (RMSE) for 14 task groups

Note: If we want to be consistent with the aforementioned analysis, thus replace Kruskal-Wallis in your preliminary analysis with Friedman-test as a non-parametric alternative test for ANOVA repeated measure.

This analysis is aligned with the study purpose that we aimed to examine at which point the participants' performance started to decline. Furthermore, in the initial analysis, we determined the number of data points arbitrarily, so we were not grouped into three bigger groups.

B. Non-Parametric Friedman Test RMSE

When grouped the tasks into three categories based on their complexities, we could not find any significant difference using Friedman-test (Chi-Square=1.500, df =2, p=0.472), as shown in Table 4.4.

Test Statistics[a]

N	12
Chi-Square	1.500
df	2
Asymp. Sig.	.472

a. Friedman Test

Table 4.4. Friedman Test on Point Estimation Task (RMSE) for three task groups

However, when we examine more carefully, by grouping the tasks according to a specific number of data points (14 groups: simple 1, medium 1, complex 1, etc.), we found that the participants showed lower performance at specific points, as shown in Table 4.5 (Chi-Square=35.622, df=13, p<0.001).

Test Statistics[a]

N	12
Chi-Square	35.622
df	13
Asymp. Sig.	.001

a. Friedman Test

Table 4.5. Friedman Test on Point Estimation Task (RMSE) for 14 task groups

C. Non-Parametric-Post-hoc test-Wilcoxon

To follow up on which condition part is different, we conducted the Wilcoxon-paired test, as discussed in section 4.6.1.

4.5.4. Graph Reproduction Task

For coefficient correlation of graph reproduction task across three conditions, we can use ANOVA because normality assumption was met (see Table 4.6). However, we should not use ANOVA for 14 conditions because the normality assumption was violated (see Appendix A).

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
cortask	Sphericity Assumed	.148	2	.074	7.649	.003
	Greenhouse-Geisser	.148	1.289	.115	7.649	.011
	Huynh-Feldt	.148	1.387	.107	7.649	.009
	Lower-bound	.148	1.000	.148	7.649	.018
Error(cortask)	Sphericity Assumed	.213	22	.010		
	Greenhouse-Geisser	.213	14.180	.015		
	Huynh-Feldt	.213	15.256	.014		
	Lower-bound	.213	11.000	.019		

Table 4.6. ANOVA Repeated Measure on Graph Reproduction Task (Correlation Coefficient) for 3 task Groups

For the graph reproduction task, we found significant differences across three conditions and 14 conditions, as shown in Table 4.7.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
cortask	Sphericity Assumed	1.074	13	.083	4.346	.000
	Greenhouse-Geisser	1.074	4.128	.260	4.346	.004
	Huynh-Feldt	1.074	6.917	.155	4.346	.000
	Lower-bound	1.074	1.000	1.074	4.346	.061
Error(cortask)	Sphericity Assumed	2.719	143	.019		
	Greenhouse-Geisser	2.719	45.406	.060		
	Huynh-Feldt	2.719	76.086	.036		
	Lower-bound	2.719	11.000	.247		

Table 4.7. ANOVA Repeated Measure on Graph Reproduction Task (Correlation Coefficient) for 14 task Groups

Non-Parametric Friedman Test Coefficient Correlation

To be consistent, since coefficient correlation data were not normally distributed, we performed the non-parametric Friedman Test (an alternative for ANOVA repeated measure). In parallel with ANOVA repeated measure results, there is a significant difference in coefficient correlation across three or either 14 conditions, as shown in Table 4.8 and Table 4.9.

Test Statistics[a]

N	12
Chi-Square	10.667
Df	2
Asymp. Sig.	.005

a. Friedman Test

Table 4.8. Friedman Test on Graph Reproduction Task for 3 task groups

Test Statistics[a]

N	12
Chi-Square	45.067
df	13
Asymp. Sig.	.000

a. Friedman Test

Table 4.9. Friedman Test on Graph Reproduction Task for 14 task groups

4.5.5. Length of Pause

For the length of pause analysis of the three conditions, we can use ANOVA because the normality assumption was met. However, we should not use ANOVA for the length of pause analysis of the 14 conditions because the normality assumption was violated (see Appendix A). There is no significant difference in length of pause or time interval across all conditions (3 or 14 conditions), as shown in Table 4.10 and Table 4.11.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
pausetask	Sphericity Assumed	443576.389	2	221788.194	3.496	.048
	Greenhouse-Geisser	443576.389	1.298	341686.345	3.496	.074
	Huynh-Feldt	443576.389	1.400	316937.168	3.496	.069
	Lower-bound	443576.389	1.000	443576.389	3.496	.088
Error(pausetask)	Sphericity Assumed	1395543.981	22	63433.817		
	Greenhouse-Geisser	1395543.981	14.280	97725.982		
	Huynh-Feldt	1395543.981	15.395	90647.451		
	Lower-bound	1395543.981	11.000	126867.635		

Table 4.10. ANOVA Repeated Measure on Length of Pause for 3 Conditions

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
pausetask	Sphericity Assumed	2340773.810	13	180059.524	2.096	.018
	Greenhouse-Geisser	2340773.810	2.862	817980.179	2.096	.123
	Huynh-Feldt	2340773.810	3.974	589058.281	2.096	.098
	Lower-bound	2340773.810	1.000	2340773.810	2.096	.176
Error(pausetask)	Sphericity Assumed	12284226.190	143	85903.680		
	Greenhouse-Geisser	12284226.190	31.478	390245.991		
	Huynh-Feldt	12284226.190	43.711	281030.810		
	Lower-bound	12284226.190	11.000	1116747.835		

Table 4.11. ANOVA Repeated Measure on Length of Pause for 14 Conditions

Non-Parametric Friedman Test Length of Pause

When grouped the tasks into three categories based on their complexities, we could not find any significant difference using Friedman-test (Chi-Square=1.500, df =2, p=0.472) or either 14 conditions as shown in Table 4.12 and Table 4.13.

Test Statistics[a]

N	12
Chi-Square	10.667
df	2
Asymp. Sig.	.005

a. Friedman Test

Table 4.12. Friedman Test on Graph Reproduction Task for 3 task groups

Test Statistics[a]

N	12
Chi-Square	45.067
Df	13
Asymp. Sig.	.000

a. Friedman Test

Table 4.13. Friedman Test on Graph Reproduction Task for 14 task groups

4.6. Results

Before conducting statistical analysis, we determined all data's normality by examining the respective histogram's skewness. Data were not normally distributed when its significant value α less than 0.05. Non-parametric tests will be applied as alternative to parametric tests when data deviates from the normal distribution. The significance level of tests was set at $\alpha < 0.05$.

4.6.1. Point Estimation Task

The results of graph reproduction tests were calculated across all subjects by calculating the RMSE between the estimated values with true values. The descriptive statistics of the RMSE of all participants for each condition are summarized in Table 4.14.

Conditions	Count	Mean	SD	Median	IQR
Simple 1	12	8.55	4.66	8.16	6.80
Simple 2	12	14.91	4.45	14.88	3.38
Simple 3	12	18.65	5.95	19.68	6.60
Simple 4	12	21.85	7.04	21.46	6.90
Medium 1	12	12.13	4.91	13.42	4.92
Medium 2	12	12.65	6.19	11.34	5.48
Medium 3	12	12.94	5.20	12.10	4.02
Medium 4	12	14.00	6.57	12.04	7.48
Medium 5	12	13.30	4.81	11.98	4.00
Medium 6	12	15.07	5.44	14.97	9.06
Complex 1	12	14.27	6.81	12.96	4.38
Complex 2	12	12.94	5.20	12.16	4.36
Complex 3	12	12.33	9.14	12.06	4.45
Complex 4	12	15.12	5.37	14.30	6.29

Table 4.14. Mean, Standard deviation (SD), Median, and Inter Quartile Range (IQR) of root mean squared error (RMSE) of 12 participants across 14 conditions (Simple-1-4, Medium-1-6, Complex-1-4).

To visualize the fluctuation of RMSE mean values with the addition of notes, we plotted each condition's respective values in mean and box plots, as shown in Figure 4-1. It shows that the RMSE means of the simple graph task increased linearly then had an obvious drop at the beginning of the medium series graphs. The values become more stable between medium and complex series with only minor fluctuations.

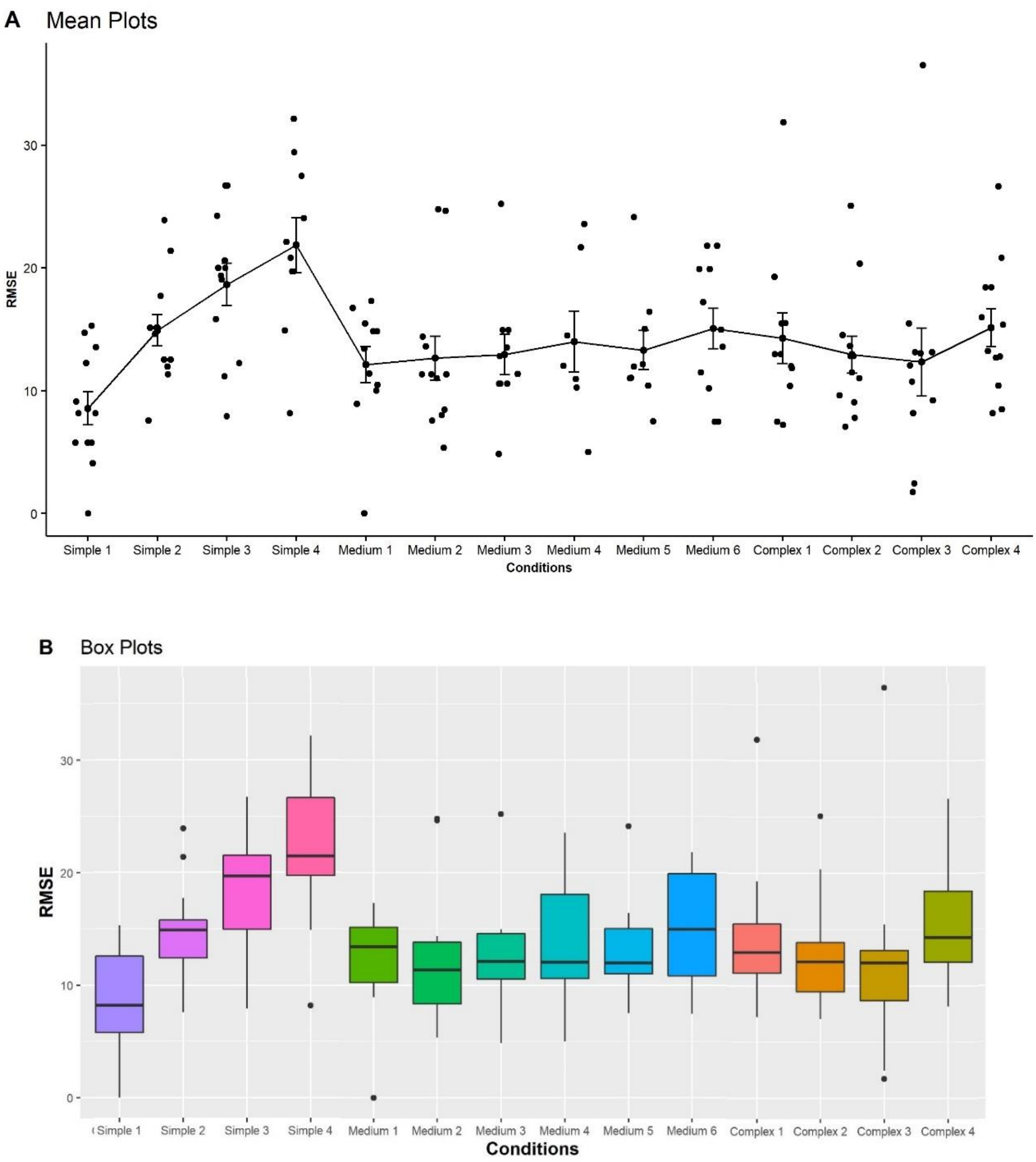


Figure 4-1. Comparison of RMSE Values in Mean Plots (A) and Boxplots (B) during Passive Listening Interaction across All Conditions

To examine whether the decline is statistically significant, we initially test the normality of RMSE means for each condition. The histogram of RMSE means in Figure 4-2 showed that six conditions were normally distributed (bell-curve shapes) while the rest were either skewed to the left or right. Therefore, we used non-parametric tests which don't assume a normal distribution.

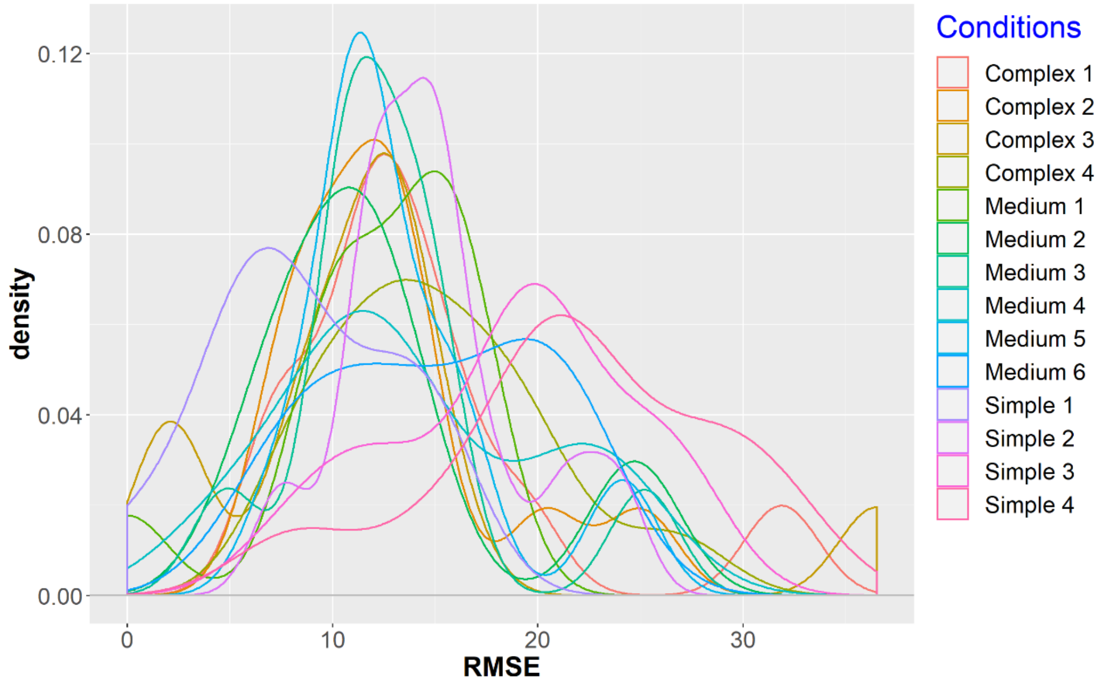


Figure 4-2. Histogram of All Conditions (Simple 1-4, Medium 1-6, Complex 1-4)

To examine the difference between all conditions, we performed a non-parametric Kruskal-Wallis test. As the calculated *p*-value was less than the significance level of 0.05, there were significant differences between the conditions (Kruskal-Wallis Chi-Squared = 31.373, $df = 13$, $p = 0.003$).

A post-hoc analysis using the Wilcoxon test was employed to determine which specific pairs of conditions were significantly different to follow up this finding. To control the familywise error rate (FER) or the false discovery rate (FDR) - rejected null hypotheses that are false (incorrect rejections) – we applied the Benjamini and Hochberg's (1995) "BH" procedure.

Table 4.15 displays the adjusted *p*-values after applying the BH correction for all pairs. Differences were found between **the simple 1 condition vs simple 3 condition ($p = 0.045$) and the simple 1 condition vs simple 4 condition ($p = 0.045$) from a total of 91 pairs.**

According to the result, hypothesis one was found to be not proven. There was a significant difference on RMSE values across all conditions, but only certain pairs of condition differed significantly. This difference indicated that participants no longer made more errors when listening to more complex graphs after the complexity became higher than a certain number of points.

	complex1	complex2	complex3	complex4	Medium1	Medium2	Medium3	Medium4	Medium5	Medium6	Simple1	Simple2	Simple3
Complex2	0.820	-	-	-	-	-	-	-	-	-	-	-	-
complex3	0.820	0.928	-	-	-	-	-	-	-	-	-	-	-
complex4	0.751	0.577	0.407	-	-	-	-	-	-	-	-	-	-
Medium1	0.948	0.871	0.699	0.640	-	-	-	-	-	-	-	-	-
Medium2	0.683	0.852	0.997	0.407	0.751	-	-	-	-	-	-	-	-
Medium3	0.843	0.928	0.843	0.640	1.000	0.843	-	-	-	-	-	-	-
Medium4	0.993	0.947	0.843	0.843	0.950	0.928	0.997	-	-	-	-	-	-
Medium5	0.944	0.937	0.843	0.620	1.000	0.843	0.989	0.992	-	-	-	-	-
Medium6	0.787	0.640	0.428	0.997	0.545	0.620	0.620	0.928	0.843	-	-	-	-
Simple1	0.161	0.262	0.599	0.114	0.192	0.428	0.288	0.400	0.243	0.114	-	-	-
Simple2	0.820	0.409	0.407	0.928	0.599	0.284	0.409	0.697	0.461	0.997	0.114	-	-
Simple3	0.198	0.159	0.140	0.400	0.114	0.159	0.159	0.407	0.159	0.407	**0.045**	0.339	-
Simple4	0.114	0.114	0.114	0.138	0.105	0.114	0.114	0.207	0.129	0.161	**0.045**	0.129	0.470

Table 4.15. Pairwise Comparison using Wilcoxon-Test. p-values adjustment shown. Bold marked shows the RMSE difference is significant after applying BH correction

4.6.2. Graph Reproduction Tasks

To investigate how closely the users' mental models matched the actual plot, we run a correlation analysis using Kendall tau non-parametric tests. Kendall's tau is a robust technique to outliers and the normality assumption and is an alternative to the Pearson correlation parametric test. The means of Kendall's tau correlation between the true and observed values for all participants are depicted in Figure 4-3. The complete values of the coefficient for each participant is provided in Appendix F.

The Kendall tau's correlation means ranged from 0.640 (medium 6) to 0.894 (simple 1), indicating a medium to a strong positive relationship. As shown in Figure 4-3 the correlation means appeared to decline continually with slight fluctuations with increasing complexity. The values fell from the simplest graph points, the Simple 1 task (6 data points) to Simple 3 and started to increase at Simple 4 (9 data points) before declining again from Medium 1 (14 data points). The correlations were then relatively stable until the most complex task (i.e. Complex

4 with 36 data points): the correlation values fluctuated between 0.84 and 0.67. This result supported the assumption that the addition of complexity may contribute to reduced ability to reproduce graphs up to a certain number of points.

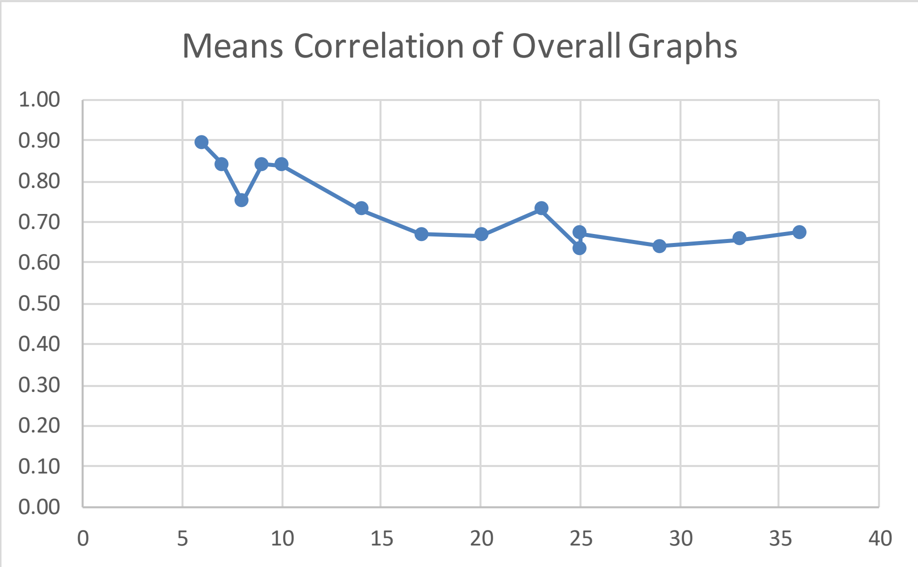


Figure 4-3. Mean Values of Kendall's Tau Correlation Coefficient for Overall Graphs. The X-axis Represents the Number of Points, the Y-axis Represents Their Corresponding CC Values

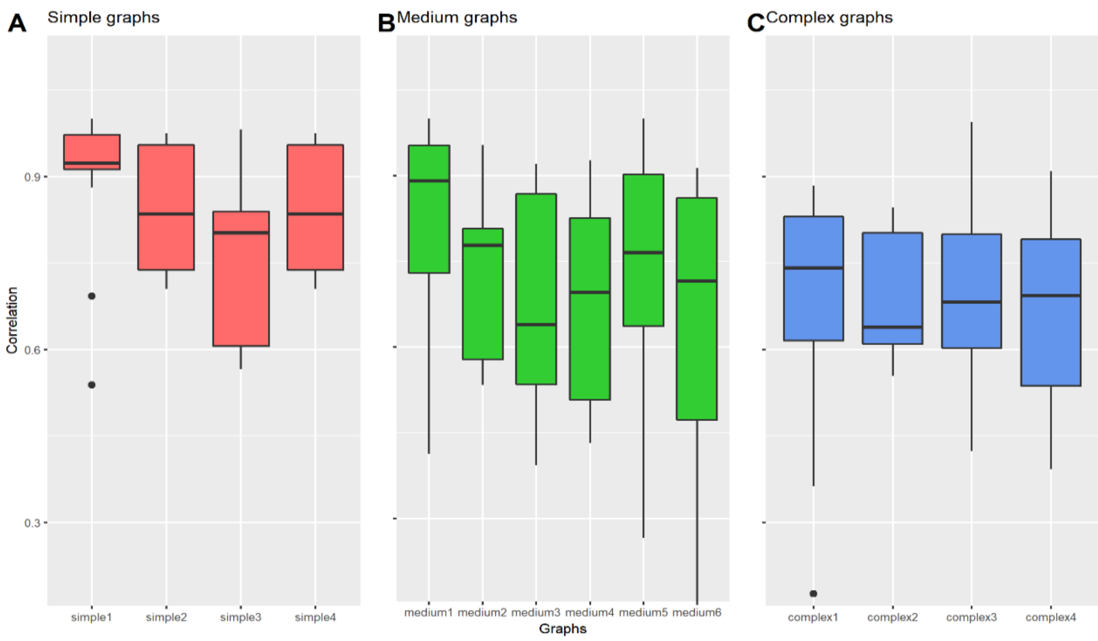


Figure 4-4. Boxplot of the Kendall correlation means and their quantile from each graph: simple, medium, complex (left to right). The ends of the whisker are set at 1.5*IQR above the third quartile (Q3) and 1.5*IQR below the first quartile (Q1). If the Minimum or Maximum values are outside this range, then they are shown as outliers. Labels indicate the type of context.

The boxplots in Figure 4-4 shows the correlation means and the respective quantiles from each group of graphs: simple, medium, and complex. Because the boxplot displayed several outliers, we transformed the data to improve its normality distribution. The inverse normal transformation (INT), a more advanced method than rank-based non-parametric solutions, was selected (Derrick et al., 2017). The correlation values before and after INT procedure is provided in Appendix F. while the histogram is shown in Figure 4-5. The pattern of boxplots for each task is relatively similar between before and after data transformation: the correlation means that the group of simple graphs is the highest compared to medium and complex categories, as shown in Figure 4-6. However, no outliers were observed in the new boxplot after data transformation.

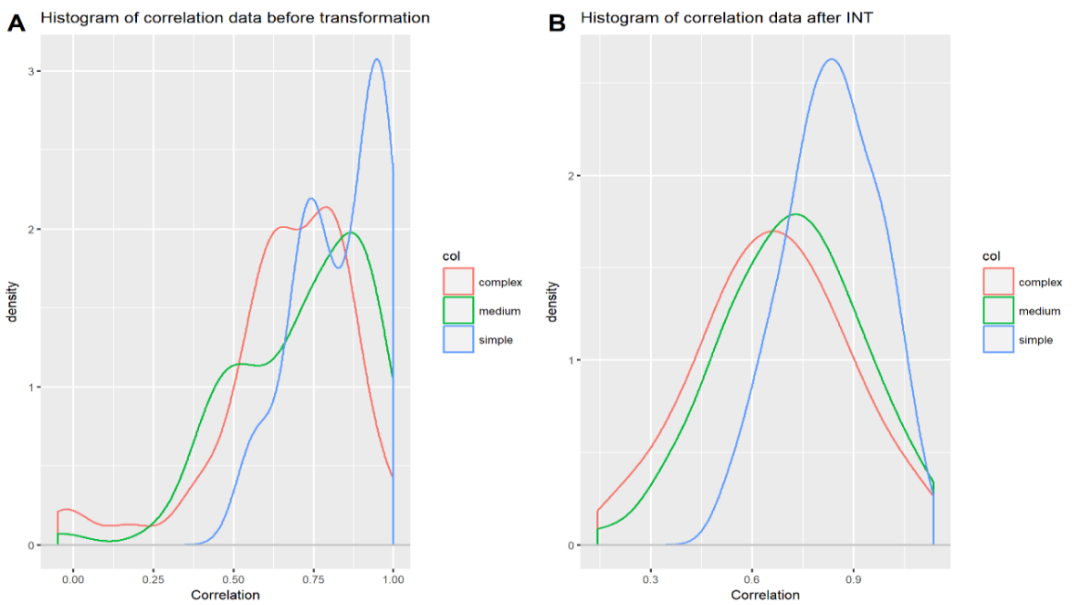


Figure 4-5. Histogram of the Correlations before Transformation (A) and After Transformation (B) for Simple, Medium, and Complex Graphs, denoted by Three Different Colors.

After data transformation, we employed two samples *t*-test to measure whether any significant difference between each group's pairs existed. We applied Bonferroni correction method to protect from making the Type I error (i.e., incorrectly rejecting a null hypothesis) because we did the multiple analysis on the same dependent variable. In this analysis, the adjusted *p*-value is obtained by dividing the original α by the number of analysis, that is 0.05/3 $\approx$ 0.0167.

The results of the t-tests showed that the significant differences were found between the simple vs medium graphs (t=3.827, $p < 0.001$) and simple vs complex graphs (t=4.6104, $p <$

0.001). However, the medium and complex graphs did not differ significantly (t=1.2756, p = 0.103).

We also asked the participants for feedback on the difficulty of the tests, and they agreed that the estimation was easier in the simple category. The difficulty was being moderate in the medium category and became very difficult in the complex categories. As predicted in hypothesis two, the simple graphs had the highest correlation compared to medium and complex graphs.

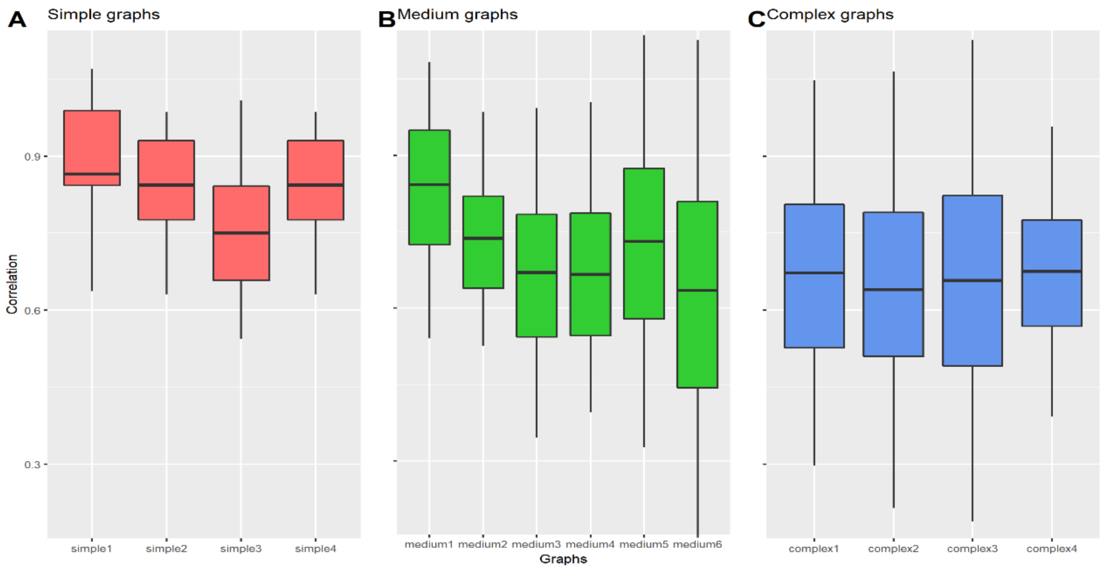


Figure 4-6 Boxplot of the Kendall correlation means and their quantile after inverse-normal transformation (INT) from each graph: simple, medium, complex (left to right). The whisker ends are set at 1.5*IQR above the third quartile (Q3) and 1.5*IQR below the first quartile (Q1).

4.6.3. Length of Pause (Time Interval) Analysis

Since each participant might have a different listening experience, we also explored the usability of adding the speed option. The participants were allowed to change the tempo of note playback and their interactions in doing this were monitored. We recorded their choice of pause length by noting the setting of the search bar they were controlling. The MAG app had five speed levels from 1 to 5 representing 500 milliseconds (level 1) to a maximum of 2500 milliseconds (level 5). The level selection by each participant for each task is shown in Table 4.16. Figure 4-7 displayed the summary of speed levels in seconds. It shows that the pause duration was increased slightly, from 0.82 s in the simple graphs to 1.09 s in the complex graphs. Further analysis using non-parametric statistics Kruskal-Wallis showed no significant differences in task groups' time interval (Kruskal Wallis, chi-square = 3.169, *df*=2, *p* = 0.205). In contrast with hypothesis three, although participants used longer pause durations when assessing more complex graphs, the differences between task groups were not significant.

	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12
Simple	2	1	5	1	1	1	1	1	1	2	1	2
	2	2	3	1	1	1	1	2	2	2	1	2
	2	2	2	1	1	1	1	2	2	2	2	2
	2	2	2	1	1	1	1	2	2	2	2	2
Medium	1	2	5	1	2	1	1	2	2	2	1	2
	1	2	3	1	2	1	1	2	2	3	1	2
	1	4	4	1	2	1	1	2	2	3	1	2
	1	3	4	1	2	1	1	2	2	3	1	2
	1	3	3	1	2	1	1	2	2	3	1	2
	1	2	5	1	2	1	1	2	2	3	1	2
Complex	1	3	5	1	2	1	1	2	2	3	1	2
	1	3	5	1	2	1	1	2	2	3	1	2
	1	3	5	4	2	1	1	2	2	4	1	2
	1	4	5	4	2	1	1	2	2	4	1	2

Table 4.16. Summary of Speed Level Selection When Listening the Auditory Graph from 12 Participants (Y1 to Y12)

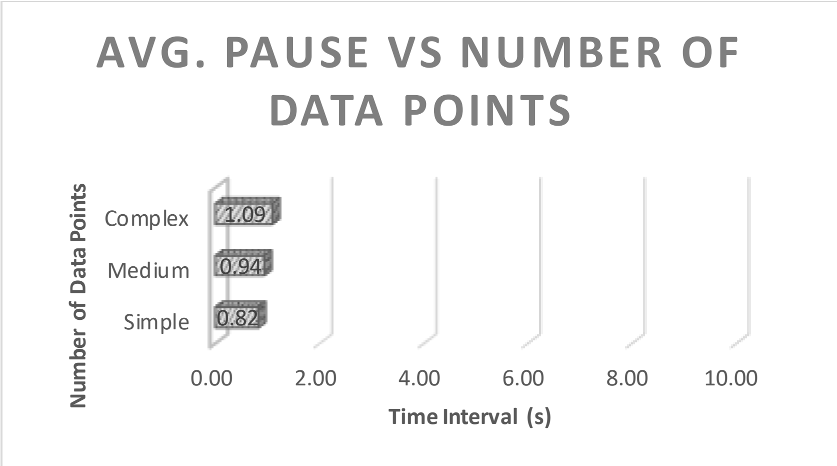


Figure 4-7. Summary of Users Average Pauses (in seconds) between Group of Data Points Compared to the Task Where the Number of Data Points Was Varied

4.7. Discussion

Our findings showed that participants made more errors in point estimation tasks with the addition of the number of data points up to a certain threshold in Simple 4 task (9 data points, M=21.85 (7.04)) before the RMSE declined from Medium 1 (10 data points, M=12.13 (4.91)) and then the RMSE is relatively stable. It was particularly observed after "simple" to "medium" five minutes break. This result indicated that the level of difficulty became stable after reaching a certain number of points. In contrast with our initial assumption and previous work, increasing the numbers of data points will increase linearly the number of errors in point estimation tasks (Nees & Walker, 2008).

This finding might be attributed to the learning effect during the experiment because participants could replay graphs as often as needed. Estimating points accurately becomes a challenging task as the complexity increases without the opportunity for multiple playback repetitions due to the limited capacity of auditory memory (Harrar & Stockman, 2007). Moreover, providing a break as we did might have influenced the participants' familiarity with the MAG app or gave them a mental break. Hence, their ability to complete more complex tasks was better than expected.

These findings have potential value for operators tasked with interpreting information relayed through auditory displays where the repetition of training tasks is an option. It is possible to provide reasonably frequent mental breaks, enabling them to maintain performance levels. It was apparent that the error rate did not increase linearly as the number of data points was increased from 10 to 25. Also, it was clear that in spite of the fact that the training included demonstrations of the pitches of the maximum and minimum Y values, there were noticeable individual differences in the ability of participants to map data points presented in audio into the scale from 0 to 100. Furthermore, the order of presentation of graphs in this exploratory study was simple to most complex in terms of the numbers of data points. This ordering almost certainly facilitated learning by participants. It is suggested, therefore, that future studies should employ randomisation of the task order.

Our study also found that the correlation between estimated and true values declined with slight fluctuations as the number of data points increased.

A significant difference between groups was found between the simple and medium graphs and between simple and complex graphs. However, the categorisation of the graphs' complexity into simple, medium and complex series was based on arbitrarily chosen boundary points to simplify the usability test of our first exploratory study using the first version of MAG. The transition points in the RMSE analysis was observed at the Medium 1 task (10 data points) while in the correlation analysis it was observed at the Medium 2 task (14 data points). The lag in the observed change in the correlation results is perhaps a result of the fact that the correlation figures continue to be influenced by the relatively good correlation scores achieved during the earlier tasks employing simpler graphs.

4.7.1. Pause duration change results

Examining the results of changes participants made to the pause duration between notes, it appears they typically took an exploratory approach, incrementing the pause duration in single steps of 500MS to hear how the resulting sonification sounded. However, the statistical analysis showed no significant pause difference between the simple, medium, and complex graphs.

4.8. Conclusions

This initial exploratory study found that using a touch screen device in the first version of MAG supported point estimation and reproduction tasks on auditory graphs.

Participants demonstrated the ability to create an approximately accurate reproduction of the graphs they auditioned. Concerning the estimation of individual points on the graphs, the number of errors made increased as the number of points in each graph increased but remained relatively stable when the number of points in each graph became larger than 10. A similar finding was also seen in the correlation values between the estimated and the true Y coordinate values. However, this was delayed by about 4 points, probably due to earlier good correlation values influencing the calculation.

These results somewhat contradicted our initial assumption that increasing the number of data points would linearly increase the number of errors in graph reproduction tasks. This issue can occur because of the learning effect, as participants tested a series of diagrams where the number of notes increased sequentially without being randomized. Further studies could address this issue by randomly assigning the order of graphs presented.

Also, each auditory graph's number of playbacks should be limited, for example, to up to three. Previous studies by Nees (2008) have shown that the mean difference for the RMS error over time with the number of auditory graphs suggests that participants do not need to hear the auditory graph's stimuli as often to maintain task performance. Furthermore, Harrar (2007) limited the number of times participants listened three times before plotting the graph. They are allowed to listen to it again because the participants concentrate on specific events and temporarily ignore the rest of the information. They were allowed to listen to the audio graph again when they changed from one feature to another or changed focus.

Notwithstanding the limitations of this exploratory study, these results demonstrate that sighted users could perform point estimation and graph reproduction tasks with a fair level of accuracy using this first version of the MAG app. For the remaining studies described in this thesis, the MAG prototype is enhanced to support multi-touch swipe gestures for interactively moving through the points in auditory graphs. This feature provides an alternative, interactive means of navigating auditory graphs compared to passive listening (playback modality).

In the following chapters, we move on to studies involving visually impaired users, as these represent the primary target population for this research.

Chapter 5. Research study two: graph reproduction tasks with additional modalities for VI users

5.1. Introduction

Our findings in Chapter 4 have explicitly highlighted that sighted participants could perceive and interpret auditory graphs presented in the MAG v1 prototype. The medium to strong correlations between the estimated and true values indicated most of the sighted participants could complete the tasks, reproduce and draw these graphs, suggesting that testing with VI participants is worth exploring. Therefore, in the current study, we will examine whether visually-impaired (VI) users can perceive and interpret auditory graphs as achieved by their sighted peers.

As used in our previous study, a similar approach for investigating the effect of the complexity of the task on point estimation task performance will also be examined but using fewer data points.

In study 1, we used 36 data points. This relatively large number of data points meant that participants required multiple playback repetitions because of the limited capacity of human auditory memory (Harrar & Stockman, 2007). Thus, in this study 2, the number of allowed playbacks will be limited to a maximum of three times to limit the possible learning effects. Further, we reduce the number of data points to a maximum of 11 which will be divided into three conditions: simple graphs (4 - 5 data points), medium graphs (7 - 8 data points), and complex graphs (10 - 11 data points). We chose the maximum number of notes at 11 because the results of our previous study showed that participants were still able to reproduce the graph without many repetitions when the number of notes was below 11.

In addition to reduced number of data points, a further refinement was also conducted by exploring multi-touch gestures (swipe modality) to the presentation of auditory graphs. This kind of modality can be an alternative tool for mental visualization and comprehension of data compared to passive listening (playback modality) only (Ferguson, Beilharz, & Calò, 2012).

Buxton (2010) define multi-touch as:

"The ability to sense simultaneously the location of multiple points of contact."

Moreover, Walker et.al (2005) proposed that adding modalities into auditory graphs resulting a better understanding of quantitative information. This encourages us to improve touch screen devices' interactivity by implementing interactive data exploration through multi-touch gestures.

VI learners usually comprehend the graphs with tactile graphs as used in the traditionally embossed graphs (Azenkot et al., 2012; Baker et al., 2014; Chew, 2014; Zou & Treviranus, 2015). Embossed graphs are the usual medium employed to convey non-textual information to VI students using tactile representations of images, maps, graphics, diagrams and other images. Therefore, we wish to explore whether adding a multi-touch gesture on a mobile device will help VI users perceive and interpret auditory graphs.

5.2. Aims

The general aim of study 2 is to examine how well the VI users can retain the memory of auditory graphs by determining the accuracy of their point estimation and graph reproduction. This study is also testing the length of the pose in between notes. We will also investigate whether interactive data exploration through multi-touch gestures can improve the interpretation of sonified data graphs compared with solely passive listening. Moreover, as previously explored for sighted users in study 1, user performance variation across the graph complexity (simple, medium, and complex) will be analyzed in this current study.

5.3. Research questions

According to the aforementioned aims, this study will examine the following research questions:

1. What is the effect of adding more points to a line graph on the VI users' graph reproduction task performance? The evaluation will be assessed by comparing the RMSEs and correlation means of the predicted and true values across all six conditions for each modality.
2. What is the effect of an additional modality using interactive data exploration through multi-touch gestures on VI users' graph reproduction task performance? Does this multi-touch gesture data exploration modality lead to better or worse interpretations of auditory graphs than passive listening? The evaluation will be assessed by

comparing the RMSEs and correlation means of the predicted and true values between passive listening and multi-touch gesture modalities.

3. Furthermore, we shall explore whether VI users can understand the trend of auditory graphs by calculating the correlation between predicted and true values. If the correlations means between the estimated and true values from VI users are relatively high (*r* > 0.7), it indicates that the VI users' interpretation of the auditory graphs is good (Ratner, 2009).

The following hypotheses fall into one of two categories. Hypotheses that compare the same interaction mode for different numbers of data points (H1 and H2), and hypotheses compare modes of interaction (H3 and H4). Finally, H5 applies to all conditions.

We formulated the following hypotheses:

H1:	VI participants will make more errors (i.e., with more complex audio graphs) and perform statistically significantly worse (i.e., lower correlation means) when listening to more data points using the passive listening modality.
H2:	VI participants will make more errors (i.e., with more complex audio graphs) and perform statistically significantly worse (i.e., lower correlation means) when listening to more data points using the multi-touch gesture modality
H3:	VI participants will make statistically significantly fewer errors (i.e., lower RMSE) using multi-touch gestures than when using the passive listening modality
H4	VI participants will perform statistically significantly better (i.e., higher correlation means) using the multi-touch gestures than when just using the passive listening modality
H5	VI participants will have good interpretation of the auditory graphs (r> 0.7) for all conditions regardless of the modality used.

5.4. Study Design

5.4.1. Participants

Fifteen VI users were recruited to participate in this study; ten males, five females, with ages ranging from 18 to 32. They comprised one high school graduate, two undergraduate students, and 12 bachelor's degree graduates. All participants were from Jakarta, Indonesia and were legally blind; seven participants reported they lost their sight since birth, while the other eight lost their eyesight before adolescence. Table 5.1 summarises the participants' age and gender.

Age range	Female	Male
18-20	0	1
21-29	4	8
30-39	0	2

Table 5.1. Distribution of Respondents Based on Age and Gender

We checked the following issues to ensure that the participants could follow the experimental task without any difficulties: their familiarity with the interaction with technologies using speech-based screen readers, their basic understanding of musical notes, and their understanding of graphical XY coordinates.

All participants used JAWS (Job Access With Speech), speech-based screen readers, on PC as their primary assistive technology. JAWS is the most popular PC screen reader which delivering voice and braille output (Freedom Scientific, 2019). Most of the participants (85%) had Android smartphones with the TalkBack screen reader, which provides the user with spoken feedback to use their device without looking at the screen (Google LLC, 2019). Moreover, they reported familiarity with basic concepts of musical notes and the ability to play at least one musical instrument, learned at their previous schools. All participants also confirmed their understanding of the presentation of the data in graphical form.

5.4.2. Apparatus

The MAG app interface was arranged into a navigation menu, graph area, and button area, respectively, as depicted in Figure 5-1. The navigation menu had a panel on the left side screen which displayed the MAG app primary navigation options. It had three menu lists, i.e., simple, medium, and complex. When a level is selected, this menu hides and the main screen displays the first graph corresponding to that level.

The MAG app 2.0 is a functional extension of the previous MAG app, making the app accessible to VI users by utilising TalkBack, a screen reader for Android devices. Talkback has been renamed 'Voice Assistant' (VA) at the beginning of launching Android 5.1.1 on certain Samsung Galaxy devices. As we used Samsung Galaxy Tab S2 for this experiment, the VA application came pre-installed. Android has developed the interaction using VA differently than in normal mode. When VA is on, the swipe function on the MAG app can only work with two-finger gestures instead of one, single finger gestures are only possible in non-VA mode.

In addition to making the app accessible, we modified the the MAG 1.0 app by removing the red layer that previously covered the diagrams to block sighted participants from seeing the data notes. The speed setting function to change the tempo of data notes was also omitted. Furthermore, the number of graphs tested was reduced to six graphs to shorten each participant's time required.

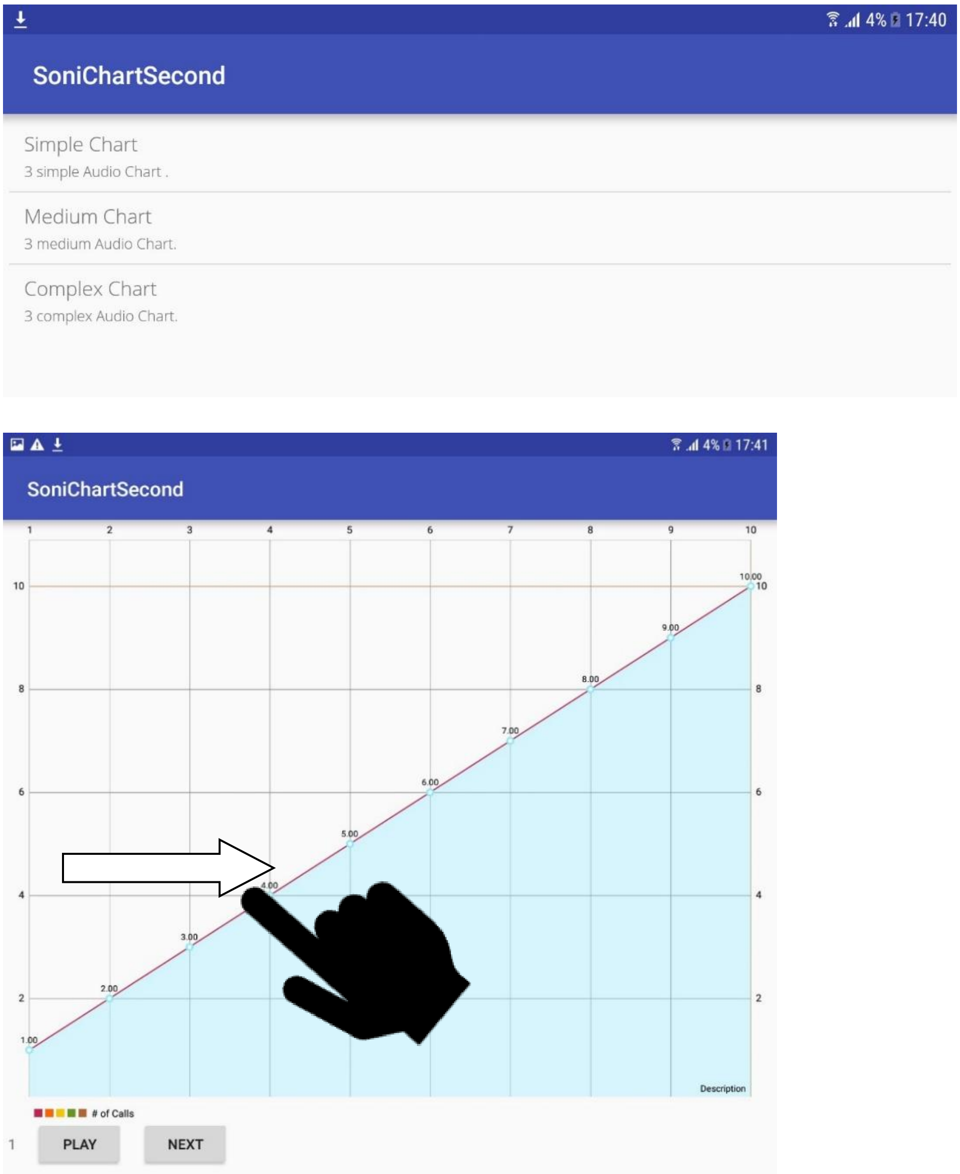


Figure 5-1 The Navigation Menu with a Panel on the Left Side Screen Showing the MAG app Primary Navigation Options with Three Menus: Simple, Medium, and Complex (Above). Illustration of User's Hand Interacting with the MAG Interface by Mobile Touch Screen Gesture (Bottom).

We created the auditory representation of each data set, using the frequency mapping used in chapter 4, by linearly scaling the Y coordinate values between 1 and 10 from each of the graphs.

The maximum number of notes was set at 11, based on study 1 regarding the limit that participants could render the graph with only a few retries. The X-axis was rendered by spacing the notes evenly in time, with a pause of 500 milliseconds between each note. The Y-points were represented as pitches with values ranging from 1 to 10 mapped into frequencies as shown in Table 3.1. The lowest frequency value was mapped into 69.8 Hz and was increased linearly to attain maximum value (10) in 698 Hz frequencies. Therefore, the higher the data point on the Y-axis, the higher the frequency.

The MAG app 2.0 was displayed on a Samsung Galaxy Tab S2 on a 9.7-inch screen running the Android 7 OS. The graphs presented in chronological order began with the two simple graphs with the least data points (4 – 5), then medium, and complex graphs. Participants moved to the next graph by tapping the "next" button and continued with the next task by navigating to the menu button to select the appropriate graph series in order. Participants performed the tasks using the touchscreen then tapping the "play" button so that the system played back the graph by default with an interval of 0.5 seconds between each point. For example, if the participant had the first simple series of graphs with four points, this graph's overall playback audio was 2 seconds (4 times 0.5s). Using the multi-touch modality, users can swipe backwards through the data points to hear the sounds, as well as being able to move forwards.

5.4.3. Self-Report Survey

Two types of questions were distributed before and after the experimental study. Before the test, the first questionnaire was administered to collect the demographic data, the degree of their visual impairment, their understanding of musical notes, and the assistive technology they normally used. The pre-study questionnaire is provided in Appendix A. After the test, a semi-structured interview consisted of several questions about the respondents' understanding when using auditory graphs using MAG app 2.0, as shown in Appendix B. We also asked their feedback about the difficulty of performing more complex tasks and the possibility of better ways to understand and interpret the auditory graphs.

5.4.4. Training

The experiment was conducted in a silent room to avoid disruption from ambient sounds so users could concentrate. The volume of the sound could be adjusted if participants requ ested. During the experiment, the researcher recorded all the subject's audio activities.

Before performing the main experimental tasks, each participant run a series of initial training, so he/she had familiarity with the MAG app and the tasks. To perform the sample of auditory graph tasks, participants initially tapped the playback button on the presented graphs, then swiped over the same samples. The participants were presented with a sample graph from the lowest frequency to the highest one on the first trial. They were trained to memorise the notes and predict all numbers in the range of 1 to 10. They were encouraged to explore the application by familiarising themselves with the application controls and get a sense of graph complexity. We informed them about the example graphs' appearance compared with the graphs used in the experiment in terms of complexity. The initial training lasted about 10 minutes.

5.4.5. Experimental Procedure

Each participant performed six sessions which consisted of three graph series: two simple, two medium, and two complexes. It took the average participant 10-15 minutes to complete each session, so each participant's total test lasted from 60-90 minutes. As there was only one device available to display the app, each participant performed the experimental tasks separately at different times.

After training, the participants took the main tests with a different set of tasks. Participants tapped the playback button, listened to the tested graphs and then attempted to estimate the graphs' points based on the pitches. Their predictions of the point positions were based on their mental visualization and comprehension of sonified data.

The first participants were permitted to listen to the first graph three times before estimating the points. After this first graph, we leave it up to the participant to choose how often they want to repeat the graph.

For the second graph, the participant was asked to swipe along the graphs, i.e., use the multi-touch gesture, thus triggering the application to generate the pitches corresponding to the second graph's points. Again, the participant was asked to estimate the points in the graph.

The participant continued to the next graph and so forth, alternatively switching between the two modalities, passive listening and swiping, until all six graphs had been completed.

The next participant performed the same task procedures, but he or she started with the multi-touch gesture for the first graph, switched to passive listening for the second graph and so forth. All graphs were rendered an equal number of times in both passive listening and multi-touch modes in random order of conditions between simple -with the least notes- to complex -with the maximum notes. The randomisation here only applied to the order in which each graph in the pair with the same level of complexity was presented; the ordering of simplest to medium to complex graphs was always maintained.

Participants spoke their estimates of all the points in each graph after listening to all the notes. The researcher audio recorded their estimates and wrote them on a sheet of paper. The paper had a table with several columns: participant, the modality of interaction (multi-touch gestures or passive listening), and the values of the estimated data points by that participant.

Following the experiment, a semi-structured interview was conducted to discuss the issues and challenges encountered using the auditory graphs. The feedback from participants would be extremely helpful to improve and refine further MAG app development. Appendix B described the questions during the interview.

5.5. Result

5.5.1. Statistical Analysis

Comparing participants' estimated point values to the true values will be utilized as a performance indicator either using passive listening or multi-touch gesture modalities. We will calculate the root-mean-square error (RMSE) measure by taking the differences between estimated values with true values. The trend of RMSE across six conditions (simple to complex graph tasks) will be analyzed as well as the RMSE difference between two modalities to examine which modality offers better respondent's performance.

Moreover, we will calculate the correlation between the estimated values (Y-values predicted by the users) with the true values (Y-values presented by pitch) using Pearson correlation (*r*) to determine the strength and direction of the linear relationship between two continuous variables (Benesty et al., 2009). The Pearson product-moment correlation *r* can be obtained by calculating the covariance of the two variables in question. Then the standard deviation of each variable must be calculated. To determine the correlation coefficient, we divide the covariance by the two variables' standard deviation product. The formula for ρ with a pair of random variables (x,y) is :

$$\boldsymbol{\rho_{xy}} = \frac{\boldsymbol{Cov(x,y)}}{\boldsymbol{\sigma_x \sigma_y}} \qquad \textbf{(4)}$$

where:

ρ_{xy}= Pearson product-moment correlation coefficient

Cov(x,y) = covariance of variables x and y

σy = standard deviation of y

If the correlation is close to "1", it indicates proper interpretation of the graphs, and shows that the estimated values are close to the true values. A lower correlation indicates that the comprehension of the plot is poor.

5.5.2. RMSEs

5.5.2.1. RMSE of Passive listening modality

The results of point estimation tasks using passive listening modality were calculated across all 15 participants by calculating their respective RMSEs as shown in Appendix F. A summary of the descriptive statistics is displayed in Table 5.2.

Conditions	Count	Mean	SD	Median	IQR
Simple-1	15	1.31	1.04	1.22	0.74
Simple-2	15	1.41	0.73	1.41	0.91
Medium-1	15	2.23	1.05	2.39	1.43
Medium-2	15	1.95	0.72	2.15	0.76
Complex-1	15	2.50	0.81	2.65	0.80
Complex-2	15	2.30	0.80	2.24	0.83

Table 5.2 Mean, Standard deviation (SD), Median, and Inter Quartile Range (IQR) of root mean squared error (RMSE) of 15 VI Participants across Six Conditions (Simple-1-2, Medium-1-2, Complex-1-2 using Passive Listening Modality.

To observe the fluctuation of RMSE mean values across six conditions, we visualized the distribution of the respective values in box plots, as shown in Figure 5-2. The participants showed better performance on point estimation tasks in the Simple 1 and 2 conditions as indicated by their lower median and quantiles RMSEs values compared with the medium and complex conditions.

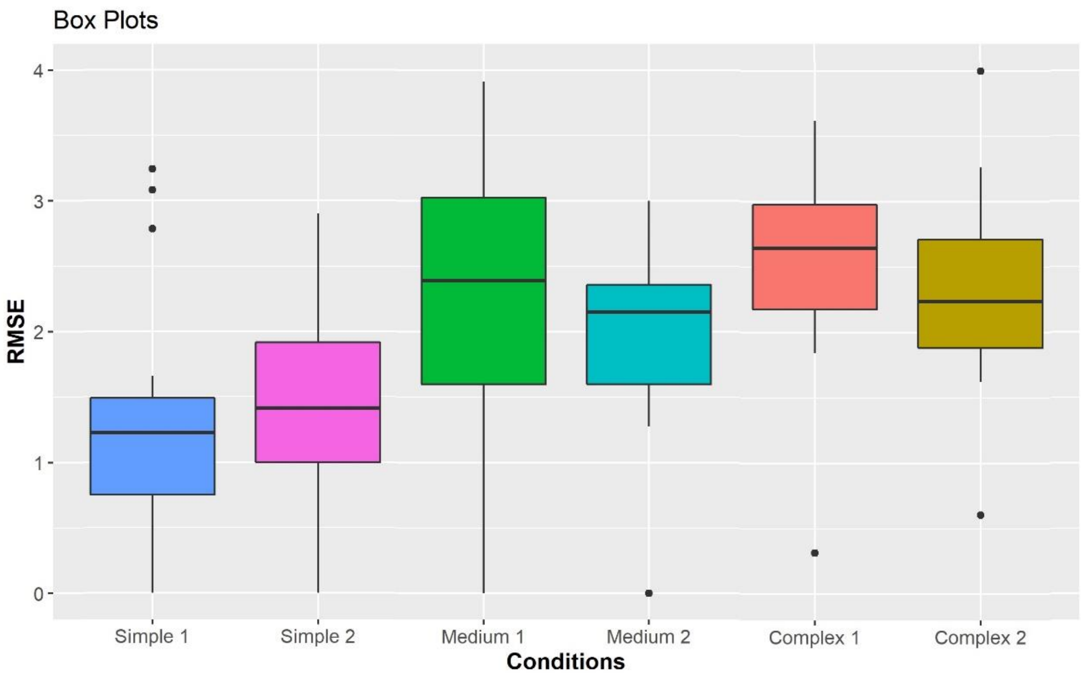


Figure 5-2 Boxplots showing the distributions of root mean squared error (RMSE) of point estimation tasks from 15 visually impaired (VI) participants using passive listening modality as displayed on the Y-axis, obtained from six conditions (Simple-1-2, Medium-1-2, Complex-1-2) displayed on the X-axis.

To evaluate the normality assumptions, we observed the histogram of RMSE means of six conditions, as shown in Figure 5-3. Most of the conditions were not normally distributed, either skewed to the left or right. Therefore, we used non-parametric tests which are more robust to normality assumption.

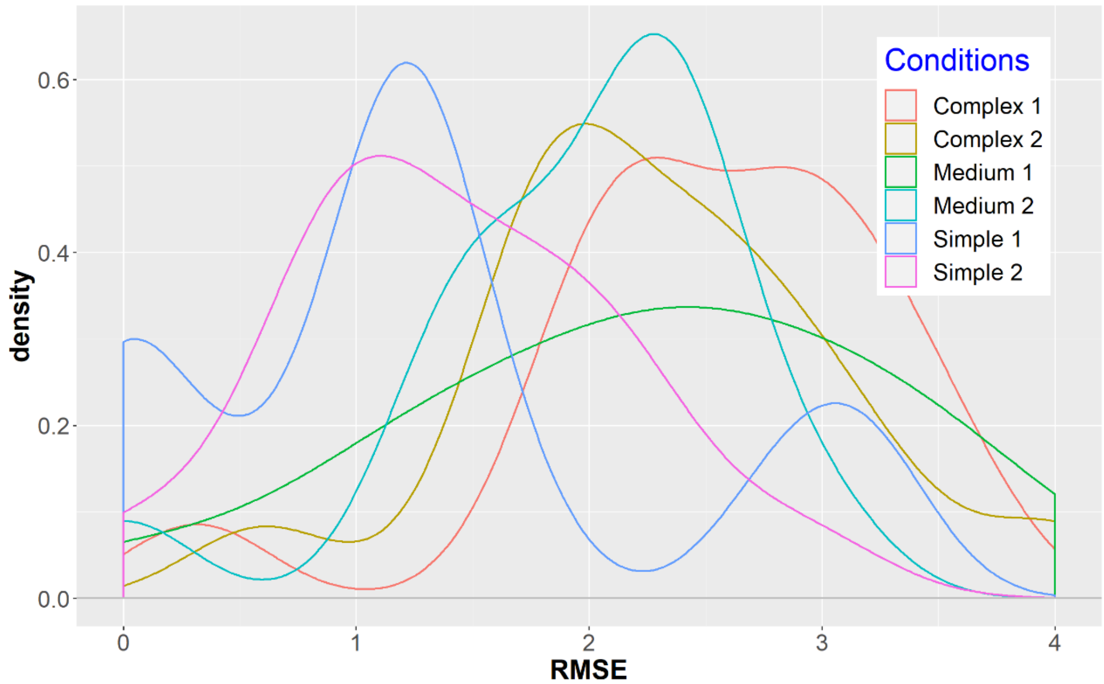


Figure 5-3 Histograms of Root Mean Square Error (RMSE) Distributions from Six Conditions (Simple-1-2, Medium-1-2, Complex-1-2) of 15 VI Participants using Passive Listening Modality.

To examine whether any difference between conditions existed, we conducted a non-parametric Kruskal-Wallis test. As the calculated *p*-value was less than the significance level of 0.05, we can conclude that **there are significant differences between the conditions** (Kruskal-Wallis chi-squared = 21.596, $df = 5$, $p < 0.001$).

Further post-hoc analysis using the Wilcoxon test was performed to determine which levels of the independent variable differ from each other level. In this case, adjustments to the *p*-values were calculated to control the family-wise error rate (FER) or control the false discovery rate (FDR).

	Complex1	*Complex2*	*Medium1*	*Medium2*	*Simple1*
Complex2	0.4382	-	-	-	-
Medium1	0.5252	0.9339	-	-	-
Medium2	0.0737	0.4351	0.4382	-	-
Simple1	**0.0230**	**0.0230**	0.0534	0.0547	-
Simple2	**0.0092**	**0.0230**	0.0533	0.0534	0.7087

Table 5.3 Pairwise Comparisons between Six Conditions (Simple 1-2, Medium 1-2, Complex 1-2) using Wilcoxon-Test. p value adjustment. Bold marked shows the RMSE difference Using Passive Listening Modality is significant after applying BH correction.

The results of all pairwise comparison showed that four pairs are significantly different: **simple 1 condition vs complex 1 condition (*p* = 0.023), simple 1 condition vs complex 2 condition (*p* = 0.023), simple 2 condition vs complex 1 condition (*p* = 0.009), simple 2 condition vs complex 2 condition (*p* = 0.023)** (see Table 5.3). **These findings supported our hypothesisone that VI participants would perform worse (i.e., higher RMSE) when listening to more data points using the passive listening modality.**

5.5.2.2. RMSE of Multi-Touch Gesture Modality

The descriptive statistics of RMSE of all VI participants while performing point estimation tasks using multi touch gestures are displayed in Table 5.4. The RMSE values for each participant across six conditions were provided in Appendix F.

Conditions	Count	Mean	SD	Median	IQR
Simple-1	15	1.12	0.77	1.22	1.27
Simple-2	15	1.02	0.87	0.77	1.44
Medium-1	15	1.07	0.76	1.07	0.70
Medium-2	15	1.52	0.93	1.50	1.00
Complex-1	15	1.41	0.88	1.55	1.34
Complex-2	15	1.67	0.96	1.73	0.91

Table 5.4 Mean, Standard deviation (SD), Median, and Inter Quartile Range (IQR) of root mean squared error (RMSE) of 15 VI Participants across Six Conditions (Simple-1-2, Medium-1-2, Complex-1-2 using Multi Touch Gesture Modality.

To visualize the error fluctuation across six conditions, we plotted the RMSE values into boxplots as shown in Figure 5-4. The distribution of RMSEs are relatively equal for all conditions, indicating the VI participant's performance did not change much when doing point estimation task from simple to complex graphs using multi-touch gesture modality as compared with passive listening.

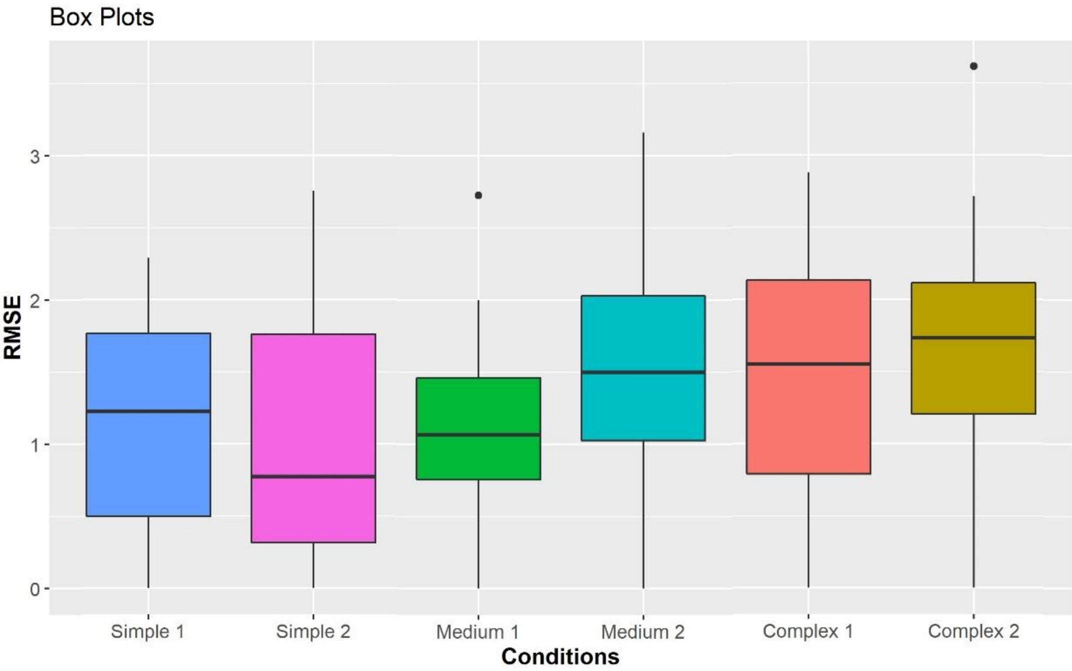


Figure 5-4 Boxplots Showing Comparison of RMSE on Point Estimation Tasks of 15 VI Participants Using Multi Touch Gestures Modality. RMSE values is displayed on Y-axis, obtained from Six Conditions (Simple-1-2, Medium-1-2, Complex-1-2) on X-axis.

As the histogram plot of all RMSE distributions were skewed to the right (see Figure 5-5), we conducted a non-parametric Kruskal-Wallis test to evaluate the significant differences between the six conditions. Contrary to our hypothesis two, the statistics showed **no significant error differences between the six conditions when using multi-touch gesture modality (Kruskal-Wallis chi-squared = 6.714, *df* = 5, *p* = 0.243).**

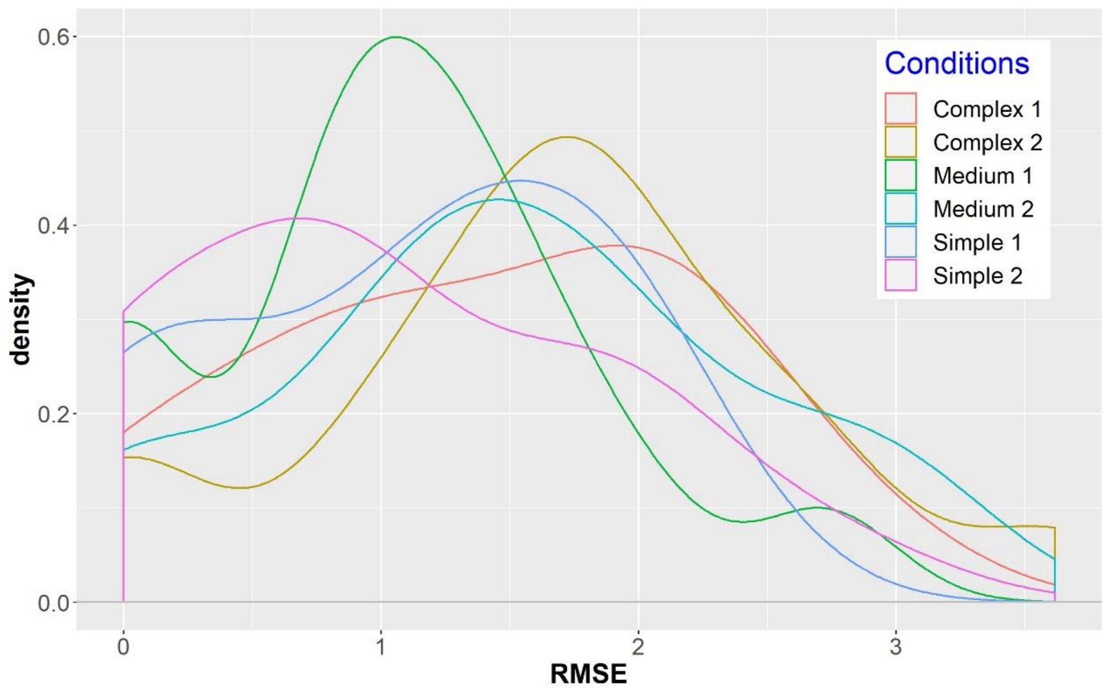


Figure 5-5 Histograms of Root Mean Square Error (RMSE) Distributions from Six Conditions (Simple-1-2, Medium-1-2, Complex-1-2) 15 VI Participants using Multi-Touch Gestures Modality.

5.5.2.3. RMSE of Passive Listening vs. RMSE of Multi-Touch Gesture Modality

To compare the point estimation task performance between two different modalities, we initially combined the RMSE values of all six conditions from 15 participants for each modality. The descriptive statistics of the combinations, including the mean, standard deviation (SD), median, and interquartile range (IQR), are revealed in Table 5.5.

Conditions	Count	Mean	SD	Median	IQR
Passive listening	90	2.27	0.77	2.27	1.18
Multi-touch gestures	90	1.81	0.86	1.80	1.17

Table 5.5 The comparison of mean, standard deviation (SD), median, and interquartile range (IQR) of root mean squared error (RMSE) values from point estimation tasks of 15 visually impaired (VI) participants between two modalities (passive listening and multi-touch gesture modalities).

To show the difference between the two modalities, we plotted the combination of RMSE values into boxplots, as shown in Figure 5-6. **The participants performed better on point estimation tasks using multi-touch modality as indicated by lower median as compared with passive listening modality**.

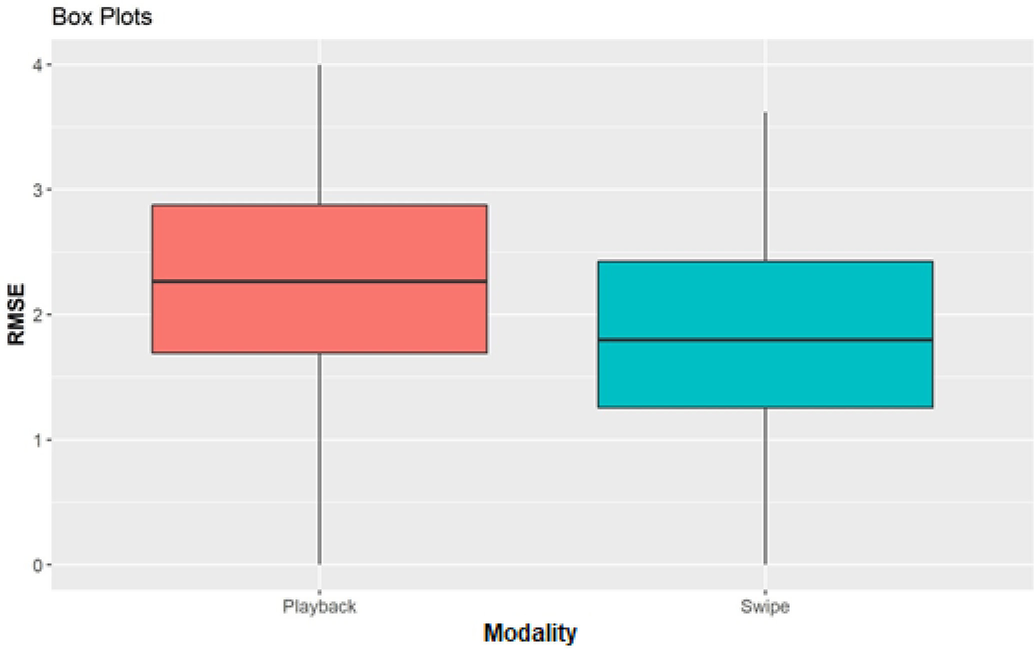


Figure 5-6 Boxplots Showing Comparison of RMSE on Point Estimation Tasks of 15 VI Participants Between Passive Listening and Multi-Touch Gesture Modalities.

Before proceeding to further parametric test, we plotted the RMSE distribution of two modalities into histogram to examine the normality assumption (see Figure 5-7). As the histogram displayed bell-curved shapes (normal distribution), we performed a two samples independent *t*-test to evaluate whether any significant difference of RMSE between the two interaction modalities existed. As predicted in hypothesis three, our *t*-test statistics showed that VI participants made significantly less error (i.e., lower RMSE) when using multi-touch gestures modality than passive listening ($t = 20.96$, $df = 290.78$, $p < 0.001$).

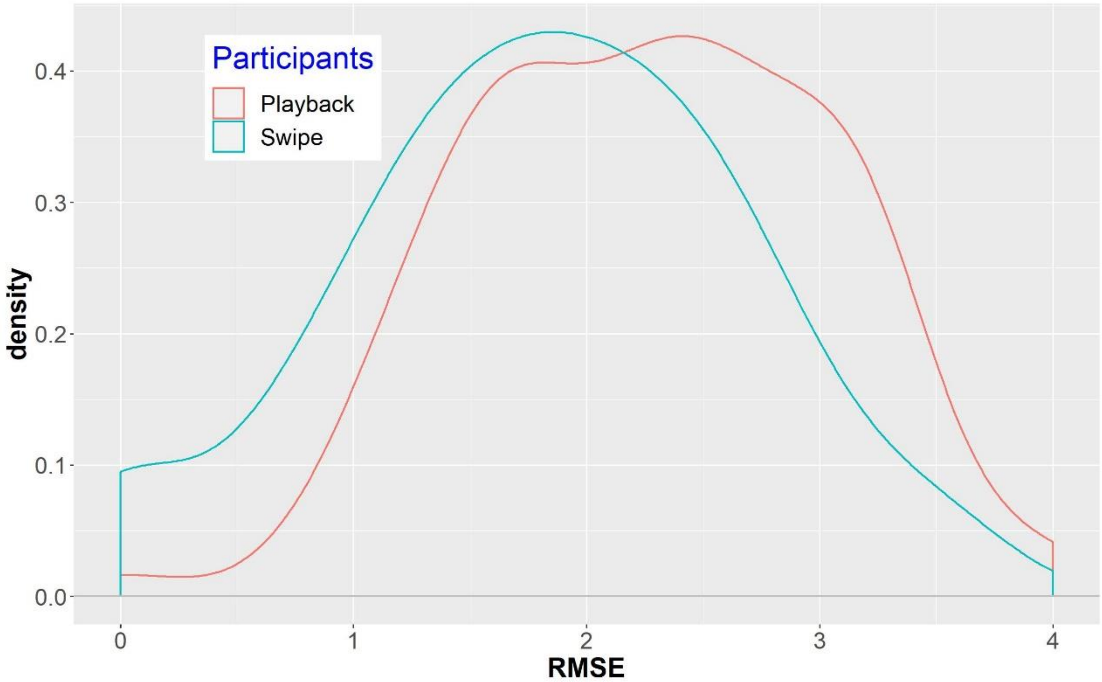


Figure 5-7 Histograms of RMSEs from Passive Listening and Multi-Touch Modalities, showing Normal Distribution.

5.5.3. Correlations

5.5.3.1. Correlation of Passive Listening Modality

To analyze the performance of graph reproduction tasks using passive listening modality, we run a Pearson *r* correlation analysis. The descriptive statistics of the Pearson correlation between the predicted and true values are displayed in Table 5.6. The complete values of the coefficient correlation for each participant are provided in Appendix F.

Conditions	Count	Mean	SD	Median	IQR
Simple-1	15	0.65	0.44	0.83	0.47
Simple-2	15	0.83	0.19	0.89	0.16
Medium-1	15	0.34	0.48	0.30	0.71
Medium-2	15	0.48	0.32	0.49	0.37
Complex-1	15	0.32	0.39	0.35	0.50
Complex-2	15	0.47	0.37	0.60	0.40

Table 5.6 The comparison of mean, standard deviation (SD), median, and interquartile range (IQR) of correlations (r) from graph reproduction tasks of 15 visually impaired (VI) participants between six conditions (Simple-1-2, Medium-1-2, Complex-1-2) using passive listening modality.

The coefficient correlation values were then plotted into six boxplots to visualise their distribution for each condition. As shown in Figure 5-8, **the correlation means across six**

conditions appeared to have unequal distribution with a tendency to decline as the number of points increased in passive listening interaction.

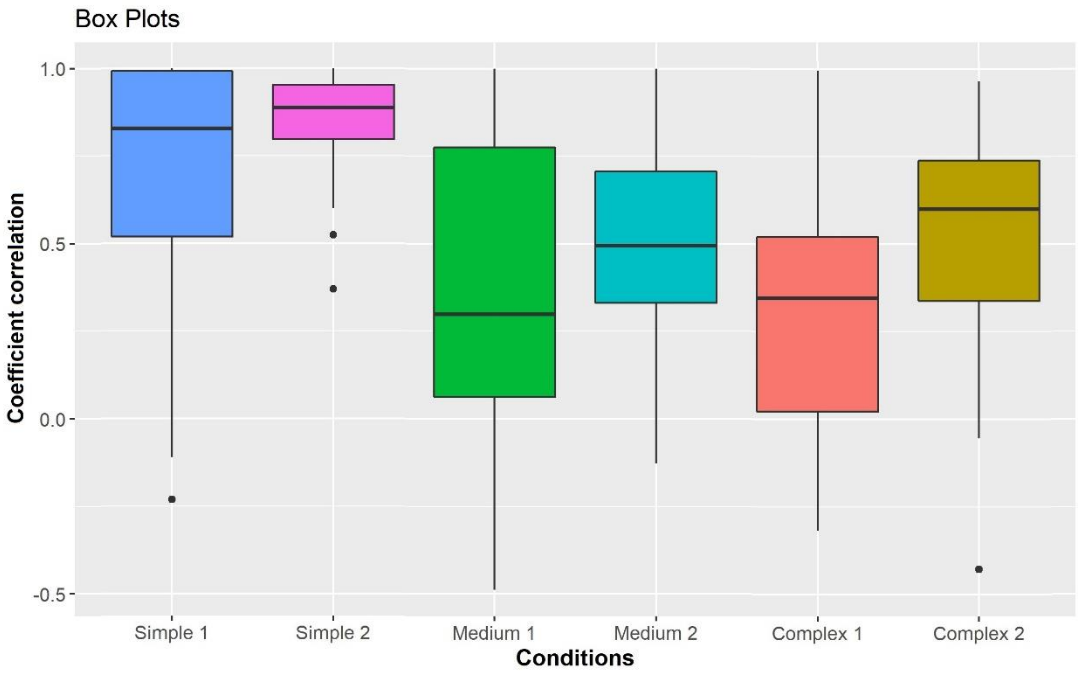


Figure 5-8 Boxplots Showing Coefficient Correlation (*r*) of Graph Reproduction Tasks of 15 VI Participants Using Passive Listening Modality as Displayed on Y-axis, Obtained from Six Conditions (Simple-1-2, Medium-1-2, Complex-1-2) on X-axis.

Before conducting further statistical analysis, we plotted the correlation for each graph into histogram to check the distributions' normality. Because all distributions were skewed to the left, as shown in Figure 5-9, we performed a non-parametric Kruskal-Wallis test. The calculated p-value of less than 0.05 led to the conclusion that **there are significant differences between the conditions (**$df = 5$, $p = 0.002$).

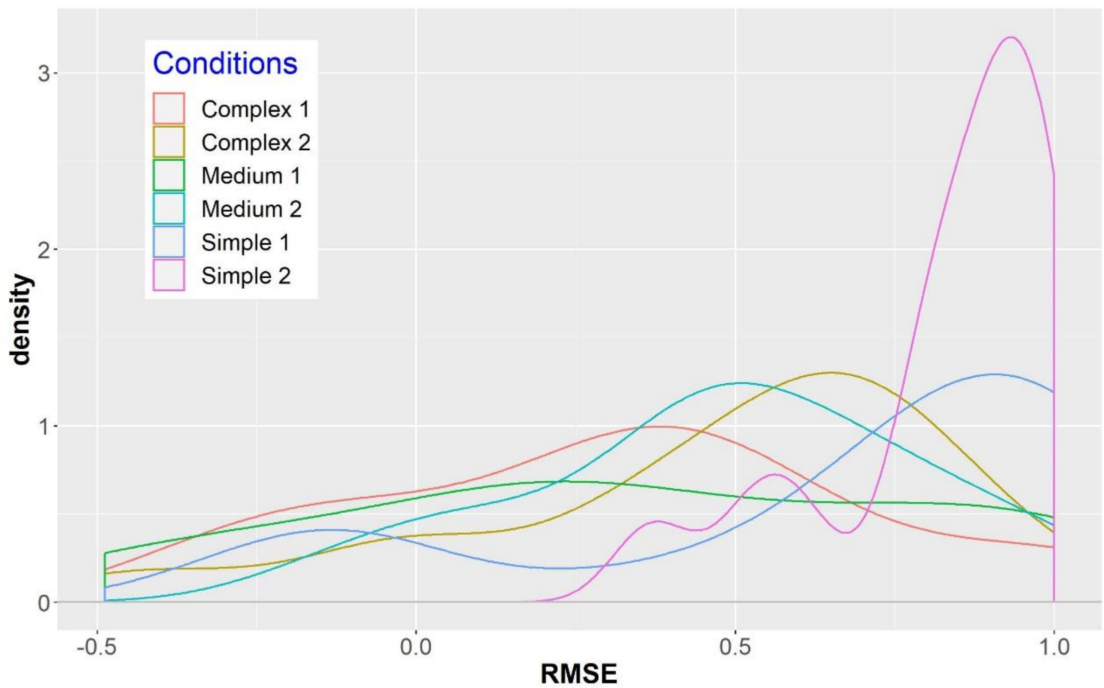


Figure 5-9 Histogram of Coefficient Correlation (*r*) from Six Conditions (Simple-1-2, Medium-1-2, Complex-1-2) Using Passive Listening Modality, Showing that All Conditions Are Not Normally Distributed.

A post hoc Wilcoxon test was then conducted to determine which pairs of conditions differ from each other.

Table 5.7 shows, after applying BH adjustment correction, that the differences were found between: **Simple 2 and Medium 1 (*p* = 0.034), Simple 2 vs Medium 2 (*p* = 0.014), Simple 2 vs. Complex 1 (*p* = 0.012), and Simple 2 vs. Complex 2 (*p* = 0.012). These results supported our hypothesis 1 that VI participants perform statistically significantly less accurately when listening to more data points using the passive listening modality.**

	Complex-1	*Complex2*	*Medium1*	*Medium2*	*Simple1*
Complex2	0.303	-	-	-	-
Medium1	1	0.667	-	-	-
Medium2	0.303	0.895	0.604	-	-
Simple1	0.083	0.132	0.132	0.181	-
Simple2	**0.012**	**0.012**	**0.034**	**0.014**	0.853

Table 5.7 Pairwise Comparisons between Six Conditions (Simple 1-2, Medium 1-2, Complex 1-2) using Wilcoxon-Test. *p value* Adjustment. Boldly marked shows the correlation coefficient difference using passive listening modality is significant after applying BH Correction.

5.5.3.2. Correlation of Multi-Touch Gesture Modality

Table 5.8 summarizes the descriptive statistics of the Pearson correlation coefficient of graph reproduction task performance using multi-touch gesture modality. The complete values of the coefficient correlation for each participant are provided in Appendix F.

Conditions	Count	Mean	SD	Median	IQR
Simple-1	15	0.79	0.28	0.92	0.42
Simple-2	15	0.91	0.18	0.99	0.11
Medium-1	15	0.85	0.25	0.95	0.14
Medium-2	15	0.67	0.36	0.77	0.28
Complex-1	15	0.76	0.28	0.85	0.29
Complex-2	15	0.70	0.39	0.85	0.20

Table 5.8 Mean, Standard deviation (SD), Median, and Inter Quartile Range (IQR) of Correlations (*r*) of 15 VI Participants across Six Conditions (Simple-1-2, Medium-1-2, Complex-1-2 using Multi-Touch Gesture Modality.

The coefficient correlation values were then plotted into six boxplots to visualise the change across all conditions. The boxplots in Figure 5-10 show the correlation means and the respective quantiles from each condition, showing that **the performance of graph reproduction tasks using multi-touch gestures interaction modality resulted in a relatively equal distribution for all conditions.**

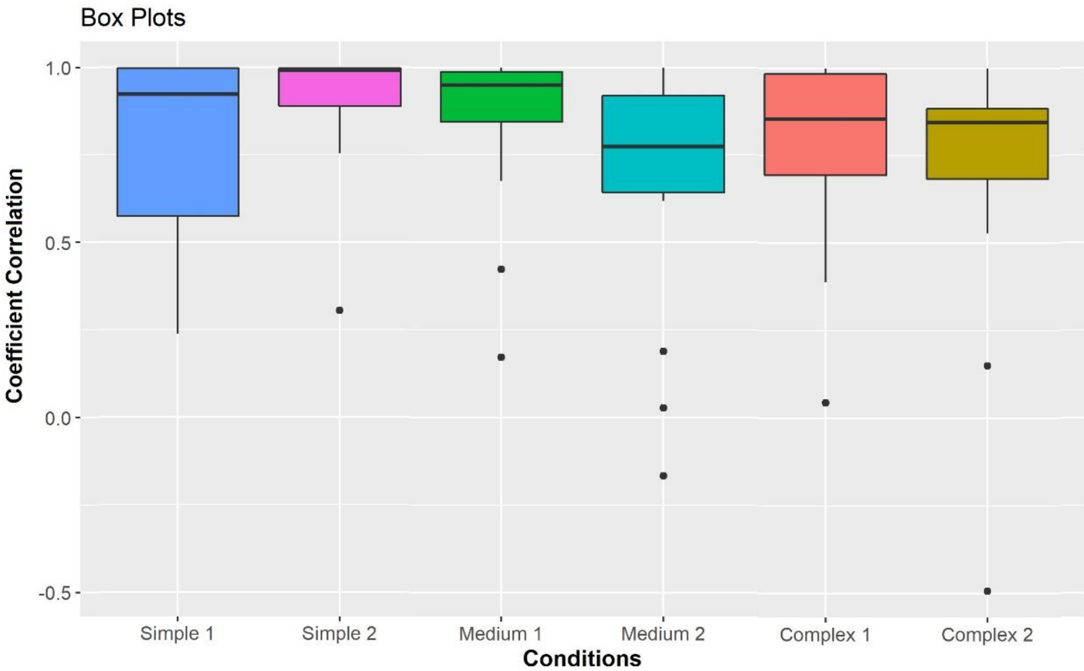


Figure 5-10. Boxplots Showing Coefficient Correlation (*r*) Distributions on Graph Reproduction Tasks of 15 VI Participants Using Multi-Touch Gesture Modality as displayed on Y-axis, obtained from Six Conditions (Simple-1-2, Medium-1-2, Complex-1-2) on X-axis.

As depicted in Figure 5-11, all conditions' distribution was skewed left, indicating non-normal distributed data. Thus, a non-parametric Kruskall Wallis test was employed to test the correlation between six conditions. The calculated *p*-value showed **no significant differences between the conditions (p = 0.068) which did not confirm our hypothesis 2.**

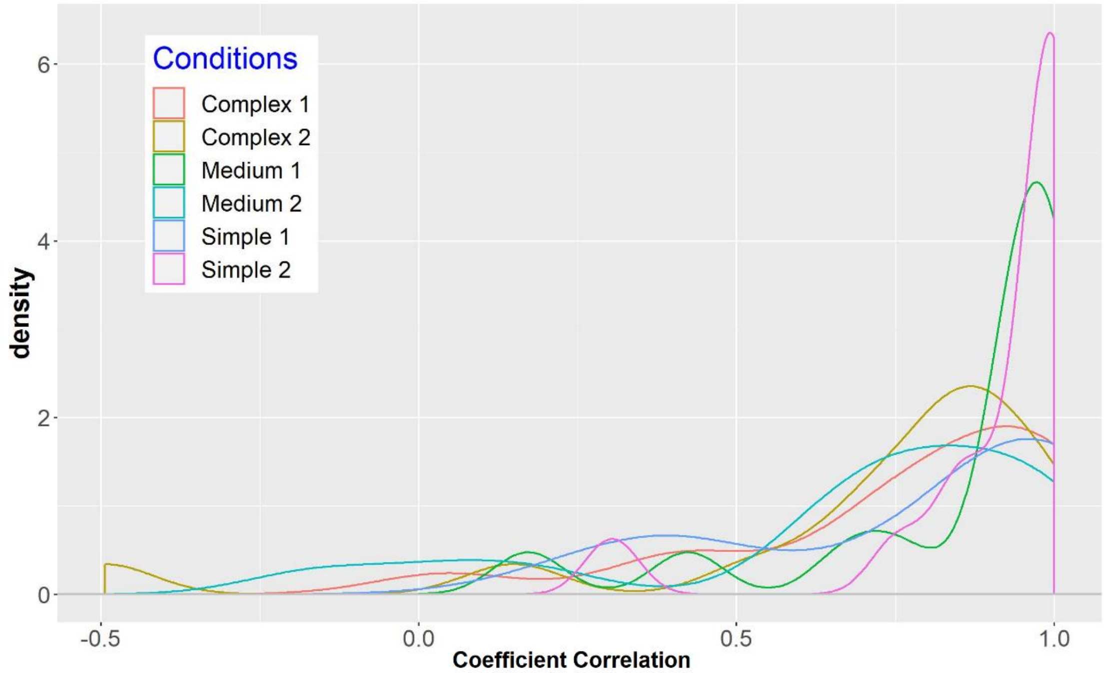


Figure 5-11. Histogram of Coefficient Correlation (*r*) from Six Conditions (Simple-1-2, Medium-1-2, Complex-1-2), Showing that All Conditions Are Not Normally Distributed.

5.5.3.3. Correlation of Passive Listening vs. Multi-Touch Gesture Modality

We combined the mean correlation values across six conditions from each modality to compare the graph-reproduction task performance between the two different modalities. Table 5.9 showed a summary of descriptive statistics after combining the data.

Modalities	Count	Mean	SD	Median	IQR
Multi-touch gesture	90	0.514147	0.408521	0.589216	0.595374
Passive listening	90	0.777951	0.301961	0.888778	0.289657

Table 5.9. Mean, Standard deviation (SD), Median, and Inter Quartile Range (IQR) of Correlation (*r*) from Graph Reproduction Tasks of 15 VI participants Between Two Modalities (Passive Listening and Multi-Touch Gestures)

The coefficient correlation values were then visualized into boxplots to examine the distribution between passive listening and multi-touch gesture modalities (see Figure 5-12). The multi-touch gesture modality resulted in a higher positive correlation (median = 0.8) than passive listening (median = 0.6)., implying that VI participants **performed more accurate estimations of true values on graph reproduction tasks using the additional modality**.

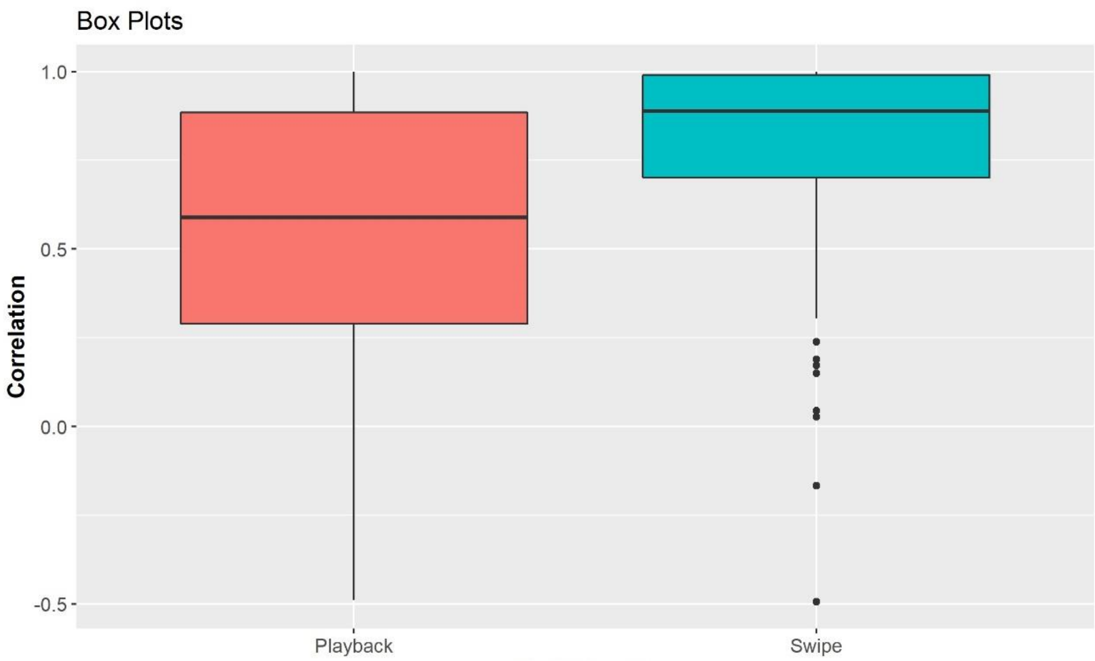


Figure 5-12 . Boxplots Showing Comparison of Coefficient Correlation (*r*) Distribution on Graph Reproduction Task of 15 VI Participants Between Passive Listening and Multi-Touch Gesture Modalities

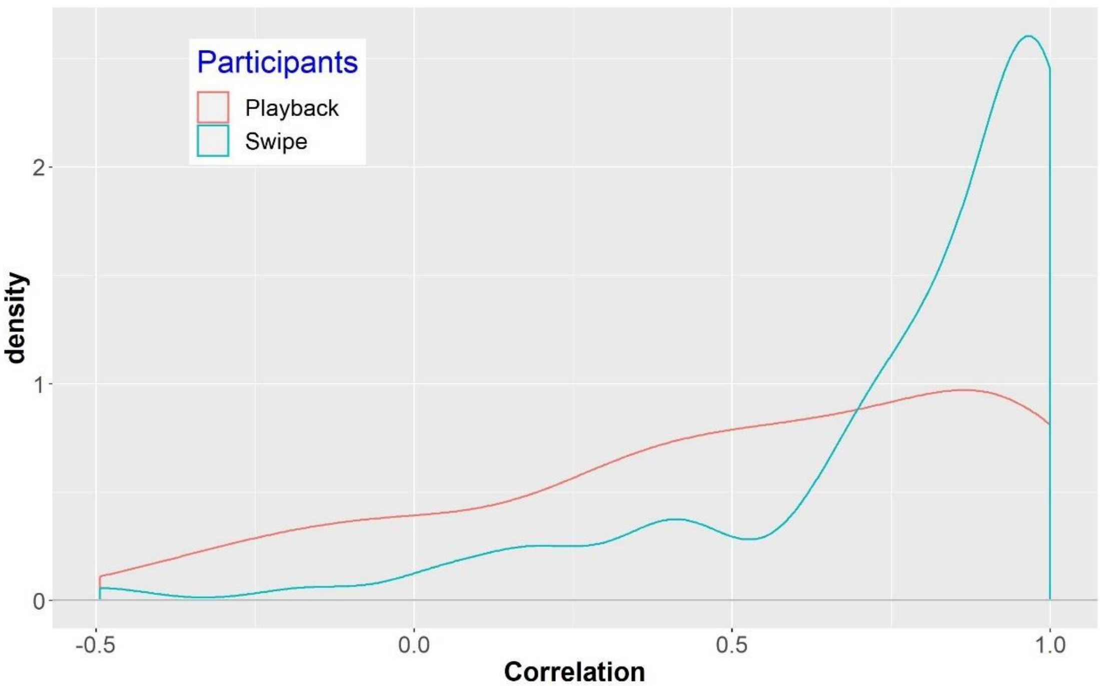
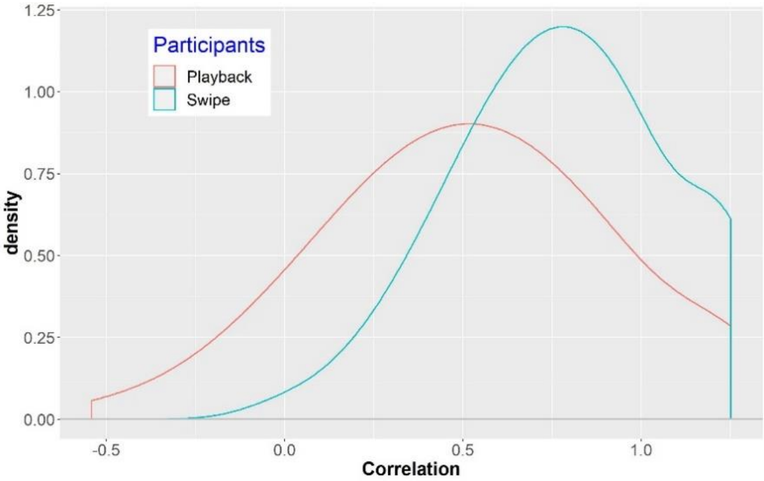


Figure 5-13. Histograms of Correlation (*r*) from Passive Listening and Multi-Touch Modalities, showing Non-Normal Distribution.

To evaluate whether the difference was statistically significant, we plotted both data into a histogram. Figure 5-13 showed that all distributions were skewed to the left which confirmed the violation of the normality assumption. Therefore, we conducted a non-parametric Mann-Whitney-Wilcoxon Test (Mann & Whitney, 1947) to evaluate the correlation means difference between both modalities. Contrary to our hypothesis 4, **the correlation between passive listening and multi-touch gestures in the graph reproduction tasks is not significantly different** (W = 15975, $p = 0.81$).

However, we noticed several outliers in the boxplot of multi-touch gesture data as shown in Figure 5-12. Even though we had employed a non-parametric test that was more robust to outliers, the boxplot's visualization showed considerably high discrepancies between the two interaction modalities. Instead of removing outliers, which should not be ignored, we decided to transform the original data set using the inverse normal transformation (INT) strategy.

The Inverse Normal Transformation (INT) is generated by log transform and arcsine for negative values (Derrick et al., 2017). It was preferred in this scenario because it is arguably a more advanced method for transforming the data than the commonly used log and arcsine transformation (Derrick et al., 2017). The correlation values before and after INT are provided in Figure 5-14. After data transformation, each modality's data set showed normal distributions, as displayed in Figure 5-14.



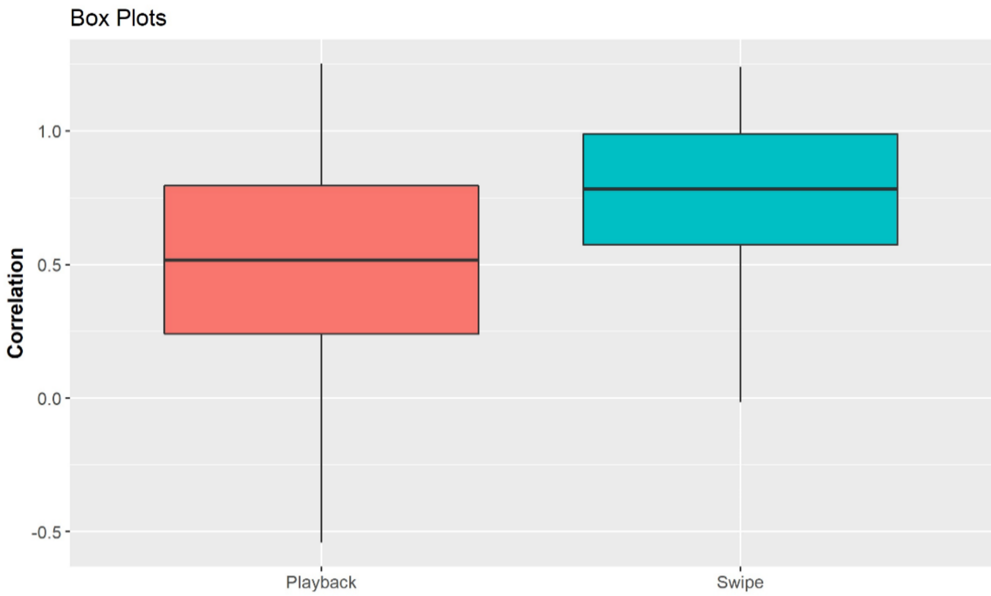


Figure 5-14. Histogram (Above) and Boxplot (Below) after applying the Inverse Normal Transformation (INT) between the Passive Listening and Multi-Touch Modalities.

Since data distribution already met the normality assumption, an independent *t*-test was used to re-analyze the significant difference between the two interaction modalities. The *t* statistics showed that **VI participants predicted data points significantly more accurately using multi-touch gesture than passive listening on graph reproduction tasks** (t = -4.9265, p < 0.001) **which supported our hypothesis four**.

5.5.4. Evaluation of Graph Reproduction Tasks across Different Graph complexity levels

This study aims to evaluate how well VI users can estimate the trend of auditory graphs. We do this by observing the pattern and direction of correlation data across all conditions. As shown in Figure 5-15, the mean correlation of using both multi-touch gesture and passive listening modalities showed a tendency to decline with the addition of notes (i.e., task complexity). The VI participants performed more accurately when using multi-touch gesture than passive listening modality as indicated by higher mean correlation values for all conditions. In general, the multi-touch gestures modality resulted in coefficient values above 0.7 which indicated a strong relationship (see Table 5.8).

In contrast, the mean correlation values of passive listening modality were much lower and fluctuated more across all conditions than those obtained with multi-touch gestures. The coefficient dropped to 0.4 as the complexity increased (see Table 5.8). Therefore, this finding partially supported hypothesis five that VI participants had a good interpretation of the auditory graph task using multi-touch gesture modality but not good enough using passive listening modality.

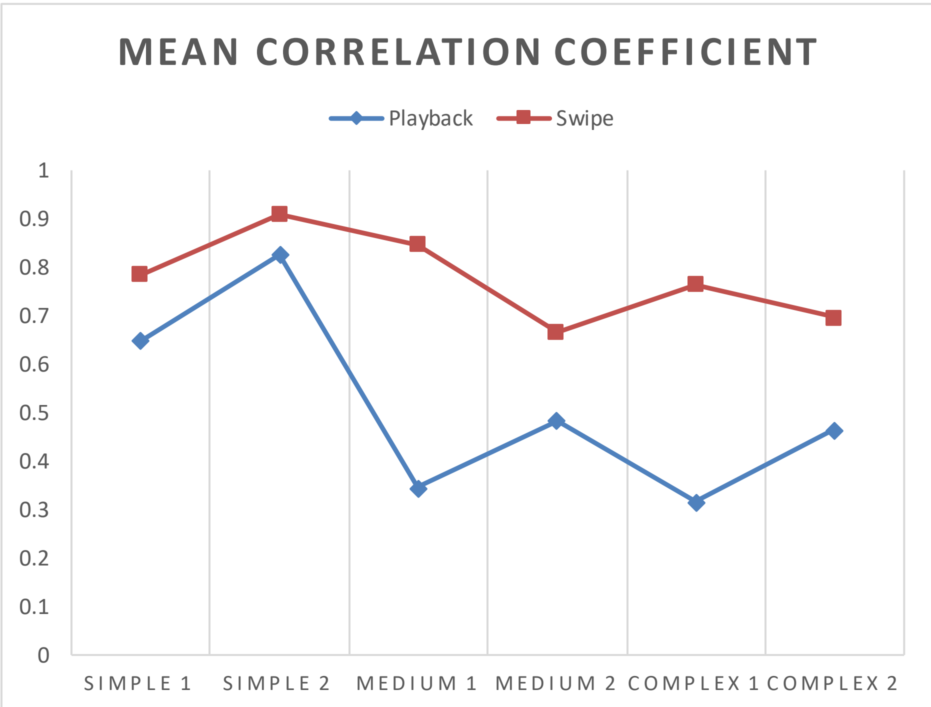


Figure 5-15. Mean Correlation Coefficient *r* Values of All Graphs (Passive Listening and Multi-touch Gesture). The X-axis represents the number of points, categorised as simple1-2 (4-5 notes), medium1-2 (7-8 notes) an and Complex-1-2 (10-11 notes) graphs, and their corresponding correlation coefficient *r* values on the Y-axis from 0 to 1 for maximum correlation.

5.6. Discussion

Part of our goal was to understand how VI participants would react to another form of interaction with their smartphones, in which gestures were performed on the touch screen of their smartphones. Our findings are following previous work by Kane et al. (2008), Duarte et al. (2017), Guerreiro et al. (2008) and Bonner et al. (2010) on the potential use of touch screen on the mobile device for VI people.

We introduce a new multimodal approach, based on multi-touch gesture interaction, aiming to lead to more accurate point estimations of the plots and improve smartphone user interfaces' accessibility, as discussed below.

5.6.1. Analysis of the point-estimation performance

The results of study 2 for VI participants showed that the point estimation performance using the multi-touch gesture modality generated more accurate results than passive listening. As shown in Figure 5-4, **the RMSEs of multi-touch gesture interaction were distributed equally for almost all conditions while those of passive listening tended to increase**.

The passive listening interaction resulted in poorer user performance on point estimation tasks as the number of data points increased (see Figure 5-2) which later was confirmed by the respective *p*-values of the pairwise comparison statistics (see Table 5.3). Further *t*-tests also supported a statistically significant difference in point estimation task performance **between passive listening and multi-touch gesture modalities**. Thus, passive listening interaction produced less accurate estimations in comparison with multi-touch gestures.

These findings answered part of our first and second research questions about the performance of point estimation tasks – as measured by RMSE – presented using both modalities for increasing numbers of data points. These results agree with Walker et.al (2005) that complementing auditory graphs with additional modalities results in an improved understanding of quantitative information. In earlier research, Geldard (1960) described how touch is the only form of human sense that engages with objects by actively manipulating them and passively perceiving them. As the body's largest organ, the skin can be considered a rich alternative touch input channel for those whose visual and auditory sensory channels are either disabled or overloaded. When a mobile device user moves his/her finger on a graph, his/her attentional resources are reserved partly for passively monitoring and reacting to the pitches of the notes and partly to the process of actively navigating along with the graph through successive data points.

Unlike passive listening mode, which transmits the auditory graph unidirectionally from the device to the user, a key feature of multi-touch interaction is the bi-directional flow of information to and from the user, allowing the user to perceive and actively engage with the system. Touch sensations combined with audio effectively close a feedback control loop between the system and the user, providing cues to the user, enabling them to actively and intuitively control the interaction.

5.6.2. Analysis of graph reproduction tasks

In this second study, we also aimed to investigate how well participants performed in interpreting auditory graphs by calculating the correlation coefficients between the estimated (predicted) values and the true values. As revealed in Table 5.7, participant performance during passive listening varied significantly for some pairs of conditions. Users' performance was worse on medium and complex graphs. In comparison, in the multi-touch condition, performance remained stable as the numbers of data points were varied. In the multi-touch condition, participants' performance was better than that in the passive listening condition, as shown by the fact that their mean correlation values for all conditions were above 0.7 (0.77 – 0.99), indicating a good correlation (Table 5.8). In contrast, the mean correlation values for the passive listening condition ranged from 0.3 to 0.89 (Table 5.6).

Furthermore, the *t*-test statistic showed a statistically significant difference between the two modalities, evidence **that the multi-touch gesture interaction provided better performance than passive listening in graph reproduction tasks**.

These results answered our third research question concerning user performance between the two modes of interaction. Participants maintained their level of performance when more notes were added using multi-touch gestures. Moreover, these findings are consistent with our previous analysis of point estimation task performance.

5.6.3. Analysis of the multi-touch gesture modality

This current study aimed to extend the interactivity provided by a touchscreen device to assist VI users in becoming more engaged with STEM and work tasks that involve graphs. Our results showed that data exploration through multi-touch gestures had significantly better performance than playback or passive listening on point estimation and graph reproduction tasks. In a semi-structured interview conducted upon completion of the experimental tasks, participants stated that they found the interactive data exploration through multi-touch gestures more interesting than passive listening. In his work on sonification on mobile touchscreen devices, Nikitenko (2014) suggests that audio playback and user interaction in combination offer an advantage over procedures that rely solely on passive listening to audio. Such a passive listening approach provides a rather unembodied series of point estimations, which may be lacking in that it comes without any sense of the position of each data point inside the physical graph area. Researchers have argued that additional modalities could be employed to overcome this issue (Bornschein, Prescher, & Weber, 2015; McDonald et al., 2014).

The results described in this chapter support this and provide a practical, specific instance of how a more embodied, multimodal form of interaction can support point estimation tasks.

5.7. Chapter summary

This chapter described an observational study in which VI performed the graph reproduction task in auditory graph. It described the results from interactive data exploration through multi-touch gestures to interpret graphs of sonified data compared with solely passive listening. It firstly examined the process as a whole and how the users interacted to perform the activities in section 5.4.5. It reported measures of the RMSE between participants' estimated point values to the true values from the actual graphs as the reference in section 5.5.2 and their correlations in section 5.5.3. It discusses the evaluation of how well VI users can estimate the trend of auditory graphs by observing the pattern and direction of correlation data between all conditions in section 5.5.4. It describes point estimation performance in graph reproduction tasks and how the distribution of RMSE for each condition in section 5.6.1. The chapter explores how well participants performed in interpreting auditory graphs by correlation in section 5.6.2. Finally, the chapter provides an overview of multi-touch gesture in graph reproduction in section 5.6.3. The following chapter will examine the non-visual point estimation task's exploratory study by integrating multiple tones as references to represent each of the notes in an auditory graph.

Chapter 6. Research study 3: point estimation tasks using multiple-reference marks

6.1. Introduction

This chapter introduces a new approach to multi-reference sonification to assist non-visual point estimation tasks based on an approach by Metatla (2016). We then describe an experimental study in which we compared the results we obtained using this new multi-reference approach to results obtained using the single point estimation approach.

A comparison should ideally be made with the results of the Metatla (2016) approach, but this is not possible for several reasons.

Differing objectives: Metatla's experiment used a visual target for sighted users, while our experiment used sound because the target audience is primarily VI users.

Further, Metatla designed a specific interface with sliders and a cursor for reaching the target as will be explained further in section 6.1.1. This means the task completion times are not comparable since our participants gave their estimates verbally. But clearly, if the duration of notes and pause length between them is the same, our approach will be faster for all cases where the distance between the point to be estimated and the reference point is greater than six because our approach never needs to render more than six notes.

In addition, the number of Metatla's units and polarities were different. Using the up and down arrow keys on the keyboard, Metatla's interface allows users to manipulate the position of a point on an axis containing a total of 30 positions (in the range -15 to 15, with 0 being the middle position). On the other hand, our approach is scalable, allowing a larger range of units (in this test, we scale up to 100 with 0 as the origin). The version of our approach used in this study only supports positive numbers. We extend the approach to negative numbers in study 4.

6.1.1. (Metatla et al., 2016) approach

The study by Metatla et al. (2016) showed that using multiple references could improve the accuracy of point estimation tasks in auditory graphs.

Their study developed a simple user interface to predict a point by providing a vertical slider that can be moved along the Y-axis with the two modes: single-point display and multi-reference display. The users can estimate the position of a point when placing it at a desired location on an axis, as shown in Figure 6.1. They can manipulate the position of a point using the up and down arrow keys on the keyboard on the Y-axis with a total of 30 positions (Metatla et al., 2016).

In the single point display, the users predict a point on an axis by mapping the pitch of a sine tone to the Y coordinate of the point using up-up polarity, i.e., increasing the pitch for point up and decrease the pitch for point down.

Using the same up-up polarity as the single point display, the point in the multi-reference display's position can be predicted relative to an origin with multi-reference tones. When the key is released, the user hears several consecutive reference tones with different pitches that match all points between the current location and the original reference, instead of just one reference point. Unlike single point display, the estimation can be done by determining both the pitch difference on that point compared to the subsequent points, the length and the number of the successive notes separating it from the origin. Thus, a greater distance provides a longer sequence of tones. An ascending set of tones will be produced for points located below the origin, and a descending set of tones for those above the origin. For example, the point on position 7 will trigger a descending sequence of tones composed of all the pitches of points 6, 5, 4, 3, 2, 1 and 0, the origin. On the contrary, the point on negative position of 7 will trigger an ascending sequence of tones composed of all the pitches of points -6, -5, -4, -3, -2, -1 and 0 (Metatla et al., 2016).

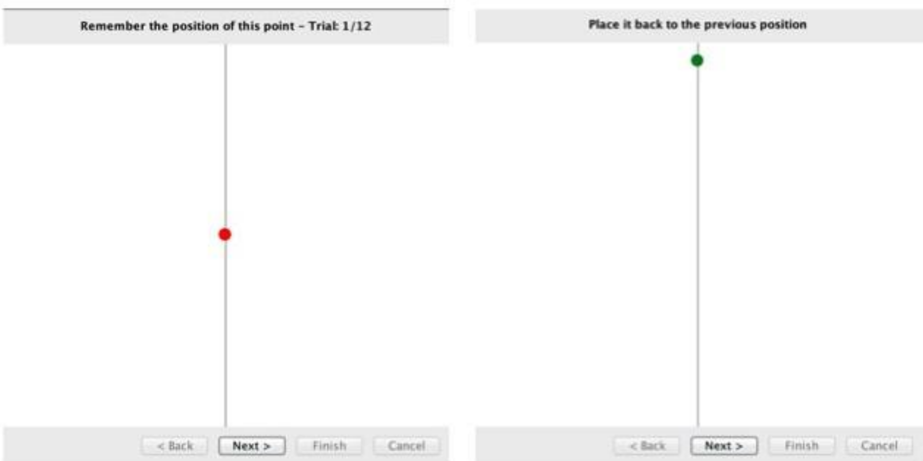


Figure 6-1 Point estimation method developed in Metatla's experiment; participants were first asked to remember the location of a target position (left) and then to move a second point (right) from a starting point generated at random on the y-axis, back to the target position remembered before (Metatla et al., 2016).

(Metatla et al., 2016) showed that his multi-reference approach enables users to make more accurate estimates and found the drawback that it takes more time than using single-point mode. In Metatla's (2016) approach, a note was played corresponding to every unit of difference between the origin and the estimated point. As the number of units (or distance) between the reference point and the point to be estimated increases, the number of reference tones increases, taking longer to present and requiring the user to keep track of an increased number of tones.

6.1.2. A modified multi-reference sonification approach

We propose further development of the approach proposed by Metatla (2016). The approach we propose uses a multiple reference scheme, but involves fewer notes than the approach of Metatla (2016).

We developed an algorithm which works for positive Y values ascending from 0 up to a maximum value (Y_{Max}) for multi-reference as discussed in section 3.11.5[4]. The idea is to play notes in multiples of 10^{ths} of the maximum Y value leading up to the value of the point the user is trying to estimate ($Y_{Estimate}$).

The approach relies on the user retaining in their memory the pitch used to represent (Y_{Max}), and understanding that the reference tones they hear represent 10ths of that value. This process can be made easier, where possible, by judicious choice of the value of (Y_{Max}). For example, if (Y_{Max}) = 100, then the user hears ascending tones will represent units of 10 ascending up to the point to be estimated. The last tone played corresponds to the estimated point, mapped to the appropriate frequency within the overall scaling.

A further refinement was made to reduce the number of notes required for each estimate. In the MAG app., Depending on whether $Y_{Estimate}$ is greater than or equal to half of Y_{Max} ($Y_{Estimate} >= 0.5 * Y_{Max}$) or below half of Y_{Max} ($Y_{Estimate} < 0.5 * Y_{Max}$), the system starts playing the sequence of reference tones from a different point.

If $Y_{Estimate}$ is greater than or equal to half of Y_{Max} ($Y_{Estimate} >= 0.5 * Y_{Max}$), the system starts playing the reference tones from a value of ($0.5 * Y_{Max}$).

[4] We deal with the case of negative numbers in chapter 7

If $Y_{Estimate}$ is less than half of Y_{Max} ($Y_{Estimate} < 0.5 * Y_{Max}$) the system starts playing the reference tones from a value of 0.

We also use different timbres to help distinguish between the two ranges. For values of ($Y_{Estimate} >= 0.5 * Y_{Max}$), the reference tones are presented using a piano sound. The reference tones are presented using a coin sound for values of $Y_{Estimate} < 0.5 * Y_{Max}$. We use two timbres in the displays since we require an approach that does not need more time to render the graph. Therefore, we split the reference to playing piano sound for any number above half of Y Maximum and using a coin for any number below half of the Y maximum. The coin sound is intuitively chosen because It is very often used to sample tones other than MIDI sound.

Using this scheme, the user will never hear more than 6 notes played, including the value of $Y_{Estimate\ in}$ the representation of a point. For example, if the y coordinate to be estimated has the value of 96, with $Y_{Max} = 100$, the listener hears the sequence of tones composed of all the pitches of points 50, 60, 70, 80, 90, and 96. Similarly, when the user hears the number 46, they hear the sequence of tones composed of all the pitches of points 0, 10, 20, 30, 40 and 46.

The value with the lowest frequency was mapped into the corresponding midi pitch. It was increased linearly to attain the values of other points up to the maximum value of Y_{Max} at 1638 Hz, corresponding to midi note G#6 as the same setting as in the previous prototype as discussed earlier in section 5.4.2.

This approach is used to reduce the number of reference tones presented while still giving the user enough information to estimate the point of interest.

In principle, the scheme we describe here could be faster than the approach proposed by Metatla et al. (2016). It employs fewer tones and has wider application because it scales more effectively to wider Y coordinates and has a lower cognitive load. In general, a smaller number of tones needs to be processed by the user.

In the remainder of this chapter, we describe an experiment in which we compare this new approach with single point estimates. Ideally, a comparison would have been made with the results of Metatla's (2016) approach, but this was not possible for the reasons given in section 6.1.

To investigate this possibility, 20 sighted participants took part in the study in March 2019. As chapter 4 tested the general feasibility of the approach done in study 5, the study here used sighted participants to test our approach's general viability before testing with visually impaired participants in chapter 7. This is due to the fact that VI participants are difficult to find. Hence, it is necessary to test the feasibility as much as possible before conducting the test with VI participants.

6.2. Objective and research questions

The study's overall objective is to investigate whether point estimation tasks can be done more accurately than with the single pitch approach while not requiring the amount of time or number of notes employed in the approach by Metatla et.al (2016).

This study tried to address two questions as follows:

1. How will the size and number of point estimation errors compare using the new multiple points reference approach described above with the size and number of errors made using the single point estimation approach?
2. How will task completion times compare between the new multiple point reference approach and the single point approach?

Metatla (2016) also investigated how point estimation errors varied in several parts of the target positions by applying the third grouping strategy. The method consists of dividing the Y_{min} and Y_{max} ranges into segments and comparing the number of errors in each segment. He claimed that it was effective in detecting differences between the sonification conditions. While this is an interesting topic, unfortunately, he did not publish the data clearly in his paper, making it difficult to compare. Furthermore, the data from his and our studies are not compatible because our study uses a larger scale than Metatla's (2016) experiment; Metatla describes the limited scale applied in his experiment in the discussion section of his paper (Metatla et al., 2016).

6.3. Participants

6.3.1. Demographics

A total of 20 sighted participants volunteered to participate in this experiment (11 men and 9 women) between 18 and 39 years old. They were a mixture of university staff (both academic and non-academic), undergraduate and postgraduate students from the Queen Mary

University of London. They were randomly assigned to two groups of ten in a within-subject experimental design. Demographic information is presented in Table 6.1.

Age	7 (18-20) 8 (21-29) 5 (30-39)
Gender	11 Male, 9 Female
Musical instrument played by participants	14 Play no instrument, 2 Piano, 2 guitar, 1 Violin, 1 Ukulele

Table 6.1 Demographic information of the participant

6.4. Study design

This study was conducted at Queen Mary University of London's Mile End campus. Before participants were introduced to the app, they were asked to complete a questionnaire about their demographic details and musical education. Fourteen participants assessed their musical education as "playing no musical instrument" or beginner, the rest stated that they played at least one musical instrument. The participants all responded that they had no experience with non-visual interaction.

6.4.1. Training

The experiments were conducted on a Samsung Galaxy Tab S2 using a 9.7-inch screen, running the Android 7 OS. Participants were given an initial demonstration of the interface to show them how they could access the MAG app's auditory graphs. Specifically, the concept of sonification was explained and the mapping of Y coordinate values to pitches was described, together with the use of the two different timbres. Participants were trained in the app and could spend as much time as they wanted to familiarize themselves with the interface before starting the experiment.

Sample graphs were used in the training that were different from those employed during the experiment. The training usually lasted about 10 minutes for both training graphs with single-point and multi-reference point sonifications. The training graphs were first presented with the participant being able to see the graph on the screen (visual mode), and then they were introduced to the non-visual mode that would be used in the actual experiment.

After implementing swiping feature in the study 2 MAG 2.0 prototype, we designed MAG 3.0 with multi-reference sonification mapping mode to address the question on how this addition could improve the performance of non-visual point estimation tasks, as shown in Figure 6-2.

Visual interaction

In this interaction mode, participants were asked to place their finger on the left-hand side of the MAG app interface at the left-hand side of the graph, then they were asked to slide their finger until the end of the right-hand side. When their finger reached each point, they heard a tone whose pitch corresponded with the Y-value that they could see on the graph. This visual interaction mode was employed during the training to enable participants to see what points were being sonified while at the same time being able to hear the sonification.

Non-visual interaction

In the second interaction mode, which we will call non-visual, participants could see the chart area. The red lines connecting the points on the plots and the information about the point values were omitted to hide the plot. Therefore, participants could see the display as a graph without points but still see each plot's borderline to help them put their finger on the right area. Further, the users had to rely on the sonifications to estimate the Y coordinates' value as they swiped their finger through the series of points. All trials in the actual experiment employed this non-visual form of interaction.

In this non-visual interaction mode, the user's interaction with the graph was essentially the same as that described for the visual mode. They started by placing their finger at the left-most part of the graph and gradually move to the right until it had passed through all the points on the graph. Unlike the previous study, instead of hearing only one pitch note, the user listens to several consecutive reference tones with different frequencies, which are all points before the present position and the source reference.

However, in the non-visual mode, after they reached each new point, they had to pause with their swiping finger and say out loud their estimate for the point they had just reached. This value was noted down by the researcher. The researcher also had to ensure that the participant did not overshoot the target or accidentally swipe to the next point. The researcher did this by alerting the participant to pay attention to the borderlines displayed on the x-axis based on the points' order, see Figure 6-2. If they accidentally move to the next point, the researcher asked them to return to the X-axis's correct position to estimate the correct point.

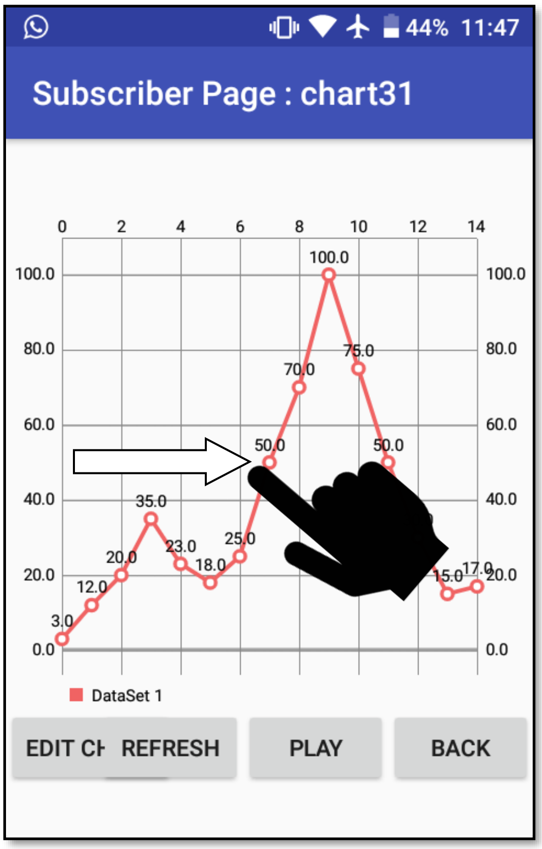


Figure 6-2 Illustration of the user's hand interacting with the MAG interface by tapping or swiping from the left part of the graph to the right. The red lines connecting each point on the plot and the value's information are omitted during the test. Therefore, participants will view the display as a graph without the data points, but still see the borderline of each graph to help them put their finger on the right area.

The first training graph was a simple linear plot with points sequentially going from Y = 0 to 100 in multiples of 10, i.e., 0, 10, 20, 30, and so forth. The purpose of displaying these values was to introduce participants to the pitches used for the lowest and highest Y coordinate values, 0 and 100, and the pitches used to represent the Y coordinates of points. Participants were then presented with the second graph which comprised randomly organised points, again with a minimum value of Y = 0 and a maximum value of Y = 100. The purpose of this second graph was not only to reinforce their memories of the range of pitch sounds, but also to familiarize them to the display of other values that were not multiples of 10, such as 18, 23, 75, , etc.

6.4.2. Experimental procedure

The experiment was designed to compare the two approaches to point estimation. During the experiment, each participant would be presented with a series of points to estimate. The two representations to be compared are single point estimation. Participants would simply hear the pitch corresponding to the Y coordinate of the point to be estimated, and multiple reference point estimation. Participants would hear the point represented using the new multiple reference scheme described in subsection 6.1.2.

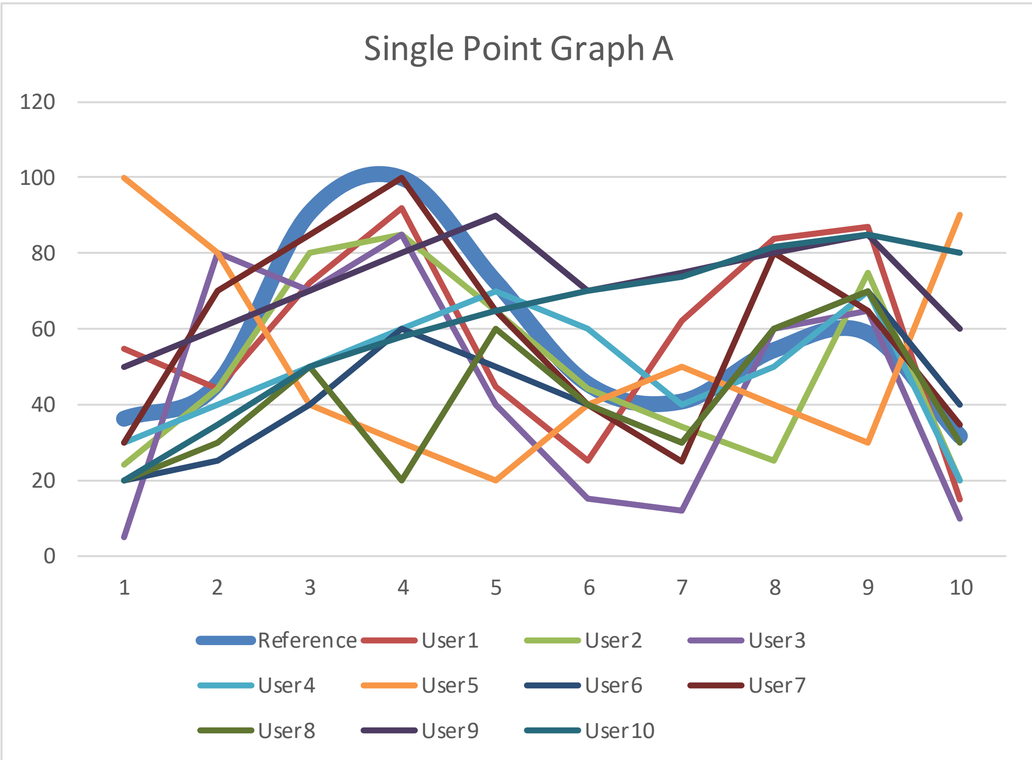
The participants were then randomly assigned to one of two groups. Each participant was asked to estimate points from two different graphs (graph A and graph B). Both graphs comprised 10 points to be estimated and participants estimated all 10 points on both graphs. The points' values were pre-generated on a computer by generating random numbers in the range 0 to 100. Before use, the distribution of points was visualised to ensure no significant clustering of points around particular Y values or ranges. Participants in group1 always started by estimating all the points on graph A using single point estimation. Then, they moved on to graph B, estimating all its points using multi-reference point estimation. Participants in group 2 did the reverse, starting with graph A using multi-reference point estimation and then progressing to graph B using single point estimation. To ensure that both graphs have the same level of complexity for the point estimation tasks, both graphs have the same number of points with the same coordinates, but the order of points in graph A is arranged to reverse the order of points in graph B. The experimental task typically lasted from 4 to 6 minutes per condition.

At the end of all trials, we conducted informal interviews and asked participants to answer several questions regarding the graphs they had just explored. This questionnaire examined the participants' difficulties in completing the trial, including the benefits and drawbacks of the different sonification conditions. The overall session, including the training and the trial, lasted between 25 minutes and 30 minutes per participant

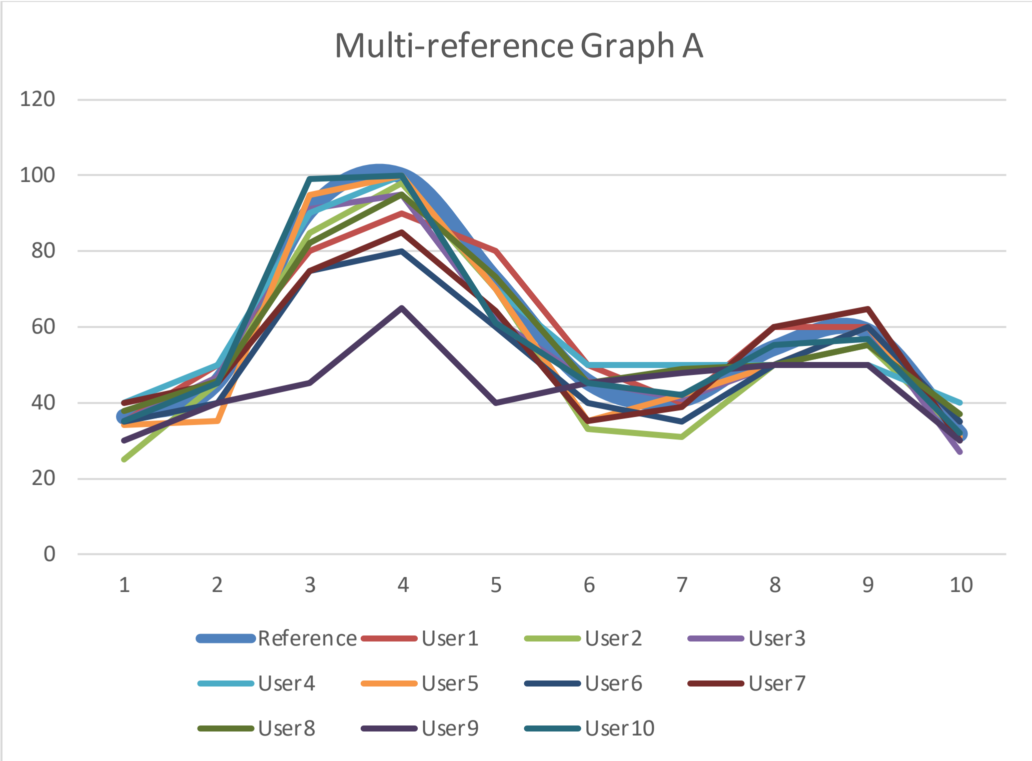
6.5. Results

6.5.1. Point estimation tasks

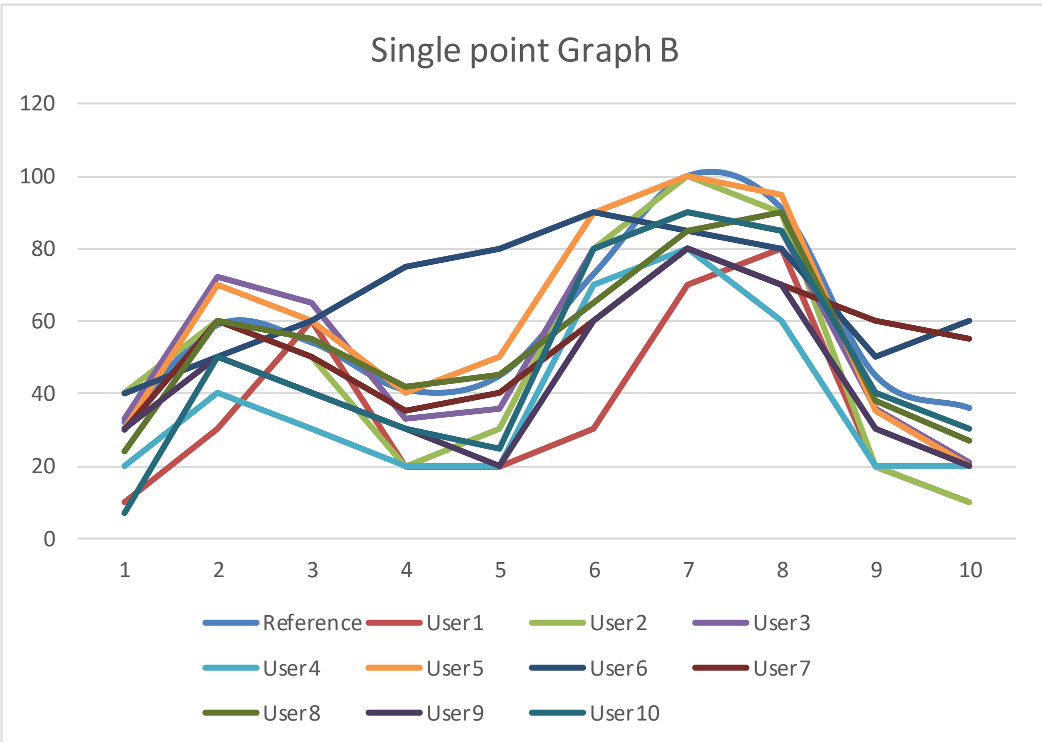
The results of two-point estimation tasks obtained from this work can be seen on the graphs in Figure 6-3, i.e. (a) The single-point mode for graph A, (b) The multi-reference mode for graph A, (c) The single-point mode for graph B, and (d) The multi-reference mode for graph B. The true value is in bold line blue colour, followed by the graphs estimated by ten participants in various colours.



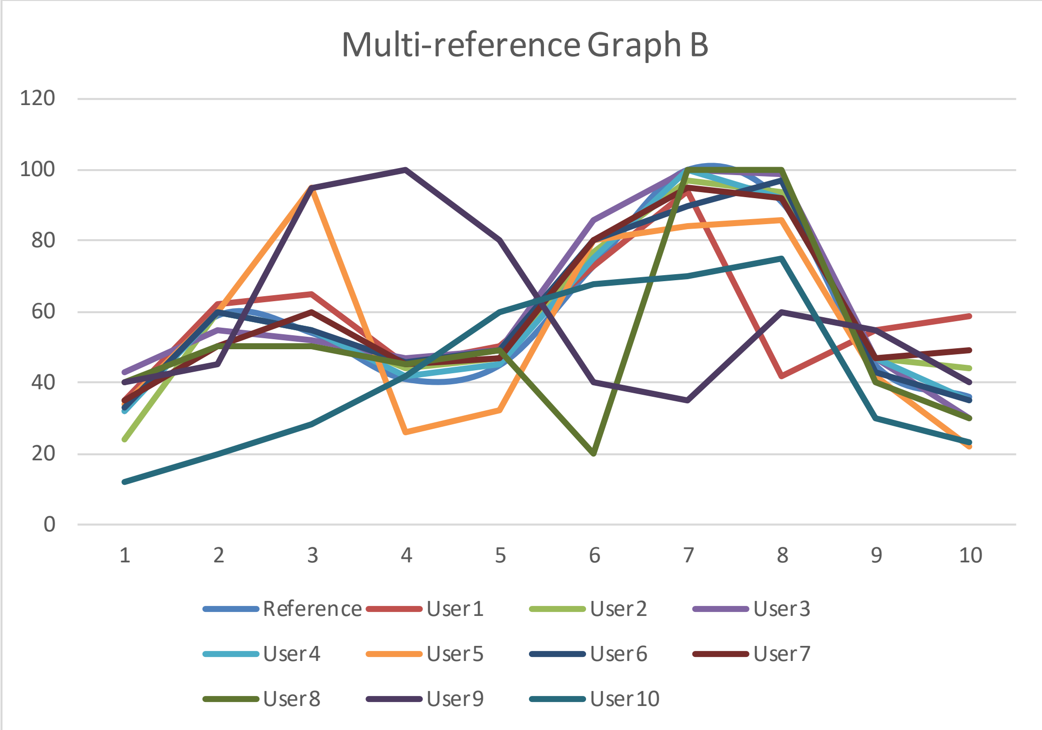
(a)


Multi-reference Graph A
120
100
80
60
40
20
0
1
2
3
4
5
6
7
8
9
10
Reference
User1
User2
User3
User4
User5
User6
User7
User8
User9
User10

(b)


Single point Graph B
120
100
80
60
40
20
0
1
2
3
4
5
6
7
8
9
10
Reference
User1
User2
User3
User4
User5
User6
User7
User8
User9
User10

(c)



(d)

Figure 6-3 Point estimations resulted by 10 users with their respective true-value graph as a reference using: (a) single-point mode in graph A, (b) multi-reference mode in graph A, (c) single-point mode in graph B, (d) multi-reference mode in graph B.

The results were then calculated across all subjects by calculating the RMSE between the estimate values with the true values; one for the single-point and the other for the multi-reference mode. This separation of using two modes was implemented as we were interested to know whether there is a relationship between the performance of point estimation tasks and the mode used to perform the tasks. After calculating the RMSE, the values were plotted into four boxplots to visualise the distribution of the error for each method and each type of graph

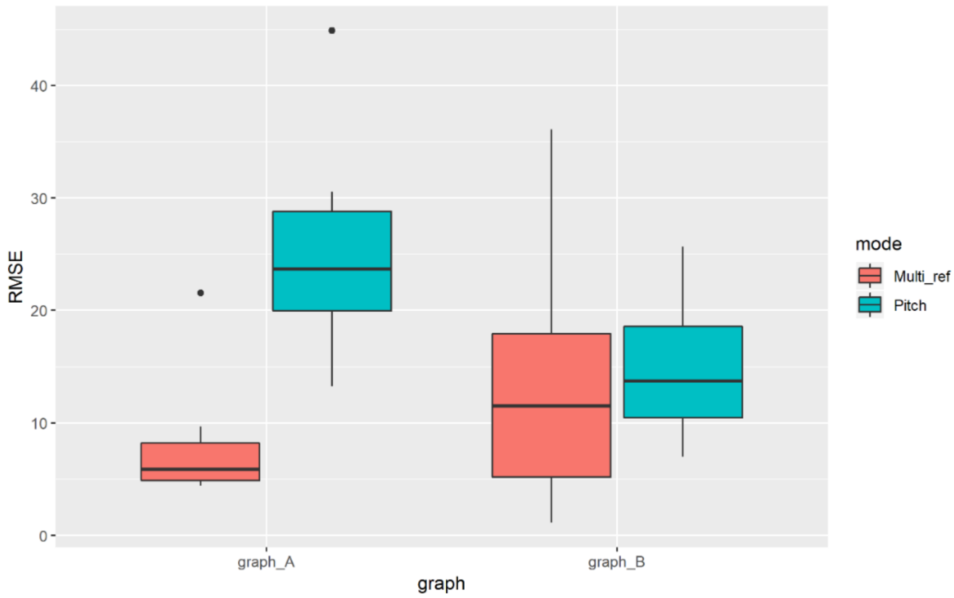


Figure 6-4 Comparison of four boxplots, representing the distributions of RMSE obtained from the multi-reference mode and the single point in graph A and graph B as displayed on the X-axis. The Y-axis shows the RMSE values between each error from 0 to 40. The legend denotes the task in the multi-reference and the single-point mode.

The boxplots from Figure 6-4 are a standardized way of displaying data distribution based on the five-number summaries: minimum, first quartile, median, third quartile, and maximum. As seen from the boxplots, the multi-reference mode used in graph A and graph B shows better performances in the point estimation tasks in terms of lower errors represented with their lower median and quantiles than those using single-point mode.

We then combined the RMSE results from graph A and graph B as both graphs were assigned with the same random values (simply in reverse order to one another). Thus they can be treated as if they were in one graph. The boxplot of these combined graphs consistently shows that using multi-reference mapping improved the performance represented by its lower RMSE values, as shown in Figure 6-5.

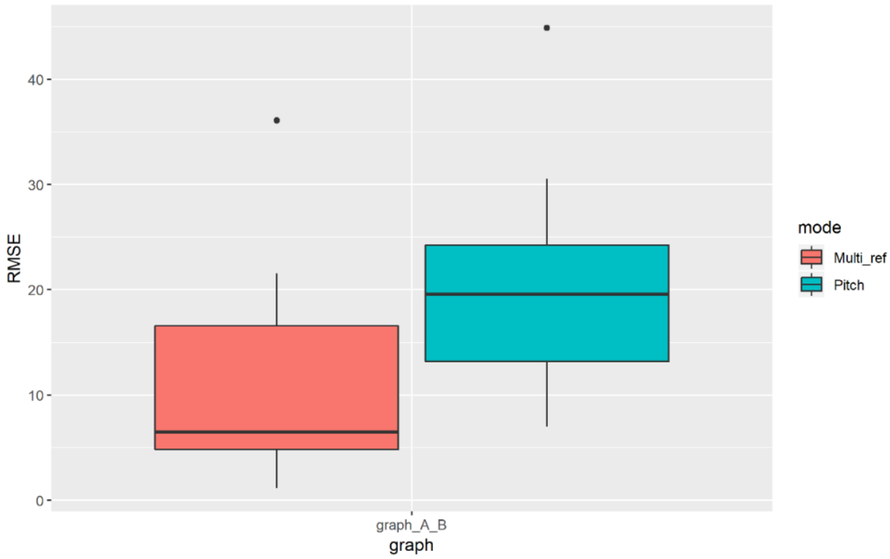


Figure 6-5 Comparison of the boxplot of RMSE values calculated from graph A and graph B combined using the multi-reference and the single-point mode.

To test whether the difference between the two modes is statistically significantly different, we performed a student t-test comparing the means of the RMSEs obtained between the two modes. We used a one-tailed t-test to compare the RMSE means with a confidence level of $\alpha = 0.05$ to analyse data between the two modes. Our null hypothesis is that the RMSE mean of the single point mode is equal to the mean of the multi-reference mode. A one-tailed test was used to test if the RMSE mean of single point mode is significantly greater than the multi-reference mode.

The t-test resulted in a significant difference in the RMSE means which implies that the two modalities are significantly different ($t = -3.30$, $p = 0.002$).

6.5.2. Completion time

The completion times of the point estimation tasks were calculated across all participants, one for the single-point and the other for the multi-reference mode. After calculating the completion time, the values were plotted into four box plots to visualise the distribution of the completion time for each method and each graph. The time is presented on Y-axis in milliseconds.

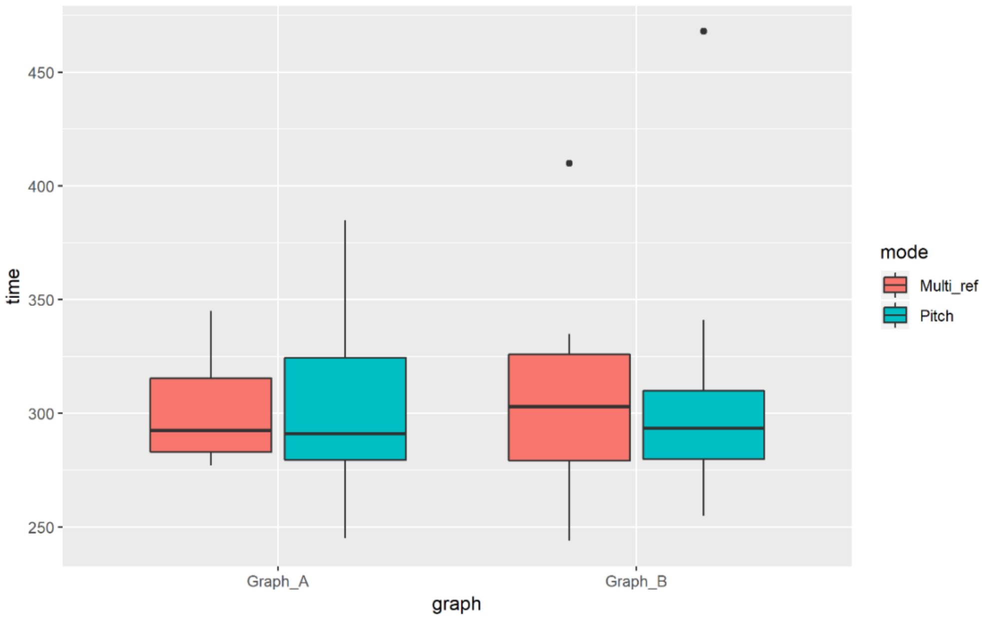


Figure 6-6 Comparison of four boxplots representing the completion time from the multi-reference and the single-point mode in graph A and graph B as displayed on the X-axis. The Y-axis shows the time in milliseconds (ms) from 250 to 450 ms. The legend denotes the multi-reference task (orange colour) and the single-point (teal colour) modality.

As seen from the boxplots in Figure 6-6, in general, the time used to complete the tasks using multi-reference mode shows similar performance as those using single-point mode in terms of their median and quantiles, although the distribution is slightly different. In graph A, the multi-reference mode has a slightly narrower distribution, while in graph B, it was the opposite.

To confirm whether the difference between the two modalities is statistically significant, we performed a student t-test comparing the completion times used between the two modalities. We used a one-tailed t-test to compare the completion time with a confidence level of α = 0.05 to analyse data between the two modalities. Our null hypothesis is that the single point modality's completion time is equal to the mean of the multi-reference mode. A one-tailed test was used to test if the single point modality's completion time mean is significantly greater or smaller than those of the multi-reference mode.

The t-test did not result in a significant difference in the completion time, which implies that the two modalities are not significantly different regarding the completion times ($t = -0.211$, $p = 0.83$). The mean completion time for single point and multi-reference are 0.304 milliseconds and 0.307 milliseconds, respectively, which are not significantly different.

When the completion time results from graph A and graph B were combined, the boxplot in Figure 6-7 consistently shows that completion time is not significantly different ($p > 0.05$).

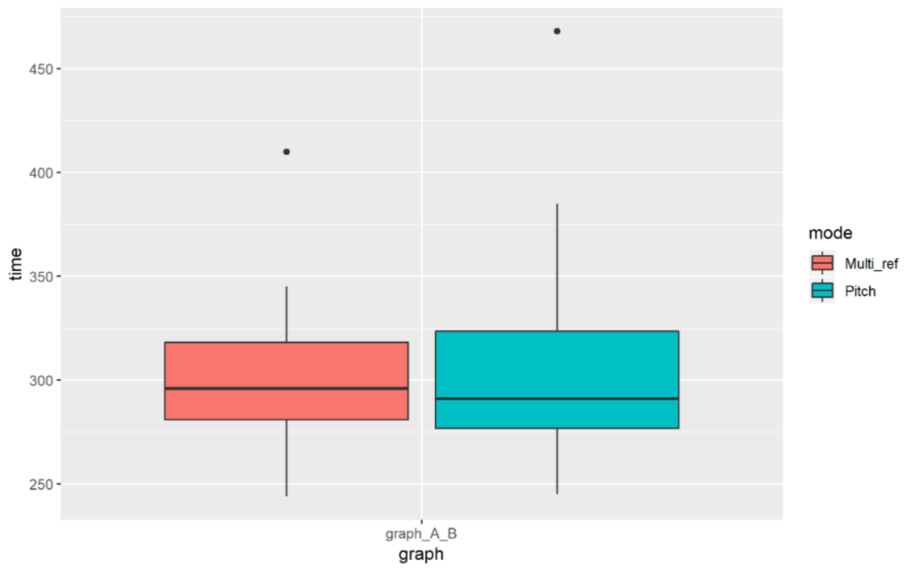


Figure 6-7 Two boxplots depicting the completion time (in milliseconds) obtained to complete point estimation tasks using the multi-reference and the single-point mode from the combined graph A and graph B.

6.6. Discussion

6.6.1. Analysis of the accuracy to estimate point-estimation tasks

In general, the results of this study showed that the multi-reference mode generated more accurate results when the participants estimated points on both graphs A and B. The *t-test* resulted in a significantly different RMSE means, which implies that the results obtained using the two sonification modes are significantly different ($M_{\text{single point}} = 20$, $M_{\text{multi-reference}} = 6$) ($t = -3.30$, $p = 0.002$). This study's first research question has been answered for this population - that the users produce higher point estimation errors when using the single point sonification mapping compared to the multi-references sonification mappings.

However, we saw a problem manifested in a few of the results for the multi-reference condition. In our experiment, we employed two different timbres, a coin-like timbre for values in the range $0 <= Y < 0.5 * Y_{Max}$, and a piano timbre for values in the range $0.5 * Y_{Max} <= Y <= Y_{Max}$.

For example, assuming $Y_{Max} = 100$. If the coin timbre is mistakenly interpreted as the piano timbre, the sonification of $Y = 10$ could be mistakenly interpreted to have the value $Y = 60$.

The opposite of this scenario could also happen if the piano timbre is mistakenly interpreted as the coin timbre. For example, the sonification of Y = 70 could be interpreted to be the sonification of Y = 20. The pitches involved are of course widely different. But if participants either ignored or didn't or couldn't hear this timbral difference, large errors will arise. If this happens, assuming the participant interprets everything else correctly, the resulting error size will be exactly 0.5 * $Y_{Max.}$

This error is high compared to the size of errors seen in calculating the number of reference steps, typically the range of 10-20%. This issue can be found in Figure 6-3 (d), in which the multi-reference mode used in graph B is predicted by participants 5,6 and 9, where they have several points that they estimated far above the true-value graph 30%-50% of Y_{Max}.

However, this only appeared in a few participants' results and rarely happened. The overall results were not influenced substantially by this phenomenon.

On the contrary, in Figure 6-3 (b), the multi-reference mode in graph A shows fewer errors as participants generally seem able to correctly map the sonification. As a result, the RMSE of point estimation tasks depicted on the boxplot for multi-reference in graph B has a higher median (M=12) compared to those in graph A (M=6), while the error distribution is wider in graph B.

The following quote from the work of Flowers (2005) may have some relevance to this problem:

> *"Listening to simultaneously plotted multiple continuous pitch mapped data streams, even when attention is given to timbre choice for different variables to reduce unwanted grouping, is probably not productive. It is possible that with levels of consistent practice that are well beyond those of most sonification evaluation studies, we might do somewhat better at listening to multiple sonified streams than is currently apparent. But it is generally the case that attending to three or more continuous streams of sonified data is extremely difficult even when care is given to selection of perceptually distinct timbres or temporal patterning."*

The situation here is rather less complicated than the ones Flowers is discussing however, because A, the different timbral streams are not being presented simultaneously; and B, there are only two of them. However, we need to be wary of the possibility that confusion might arise across the two timbres. Users might forget the relevance of them, or mentally swap their meaning. Therefore, in study 4, we examine whether this effect is eradicated using only the piano sound.

6.6.2. Analysis of the completion time to finish the tasks

In general, this study's results showed that the completion time performances for both multi-reference mode and single-point mode are similar. The t-test did not significantly differ between the completion times in the two sonification conditions, which implies that the two modes are not significantly different ($t = -0.211$, $p = 0.83$) with the mean completion time for both modes equal 0.3 milliseconds.

This study's second research question has been answered for this population that the users produce equal completion times when using the single point sonification mapping compared to the multi-references sonification mappings.

Concerning the trial duration, the users' opinions were divided between those who considered the single-point mode to be faster and those who had the opposite opinion. In general, most participants considered the multi-reference condition was faster for point estimation rather than the single-point mode. However, the single-point mode could be the most demanding in terms of memory overload; to the extent that most participants believed that the trial estimation time to complete the tasks was shorter in the multi-reference mode.

We quote a participant describing their reaction to the two modes:

> *"I think the single note condition is faster, however I believe that the multiple reference condition is more accurate. Personally, I found the pitch-supported distinction between >50 and <=50 more helpful than the reference itself."*

Another participant has said:

1. *"I think multi-reference is more accurate and faster."*
2. *"I prefer multi-reference over single reference since it is faster."*
3. *"In my opinion, the condition of the single note is more difficult to guess and it takes longer for me."*

Most participants found it was really difficult to perform the point estimation using the single-point mode. It was also observed that while estimating points in this mode, participants used the previous two or three points as their 'pseudo' pair minimum-maximum references. Even though the participants had previously been trained to become familiar with our minimum and maximum note frequency references. As a result, the participants lost focus and spent longer to estimate because the auditory memory stores information only for a short period (Harrar & Stockman, 2007; Mondor & Morin, 2004).

Thus, while the single-point mode did not display the references like in the multi-reference mode, participants could generate their reference based on the prior note, at the expense of taking a long time to complete the point estimation task.

According to some participants, the multi-reference mode problem is the time needed to perform the tasks. It took longer to get used to multi-reference mapping because it requires the user to remember more preliminary note. This is not surprising and the results are also in agreement with the literature. Meyer (1956) explained that pitch training can require considerable motivation, time and effort if it is not maintained with constant practice and reinforcement.

Although a few participants felt that the completion time used to conduct the test on multi-reference tasks was longer than the ones that used single-point estimations, there were no significant differences in the task completion times between the two modes as shown in Figure 6-7. The conclusion seems to be different from Metatla's (Metatla et al., 2016) conclusion as they found a compromise between speed and accuracy for multi-reference sonification. The development of the multi-reference point estimation scheme proposed here helps to mitigate that trade-off by requiring fewer reference points to be sonified than in Metatla's (2016) approach. Furthermore, the approach proposed here scales more effectively. Remembering that the approach proposed by Metatla (2016) sonified every unit of difference between the reference point and the point to be estimated, the results presented here show that users with no previous experience of sonified data were able to make fairly accurate estimates of points on a scale from 0 to 100.

A weakness of the approach we propose here is that it requires users to work in 10ths of Y_{Max}.. The difficulty of doing this can be reduced where it is possible to choose a value of Y_{Max} which can easily be divided by 10, such as $Y_{Max} = 100$ or $Y_{Max} = 1{,}000$. The difficulty arises if Y_{Max} has a value such as 173, which would force the user to work in multiples of 17.3. Where relatively rough estimates are required, the 10ths of Y_{Max} might be approximated to a more amenable whole number. For example, if $Y_{Max} = 190$, the user might think in terms of multiples of 20 and accept that doing this will lead to a regular but small overestimate. The approach could be modified, of course, to work for different number systems, such as binary, where the multiple reference points might be in multiples or powers of 2.

In conclusion, most participants responded that multi-references were faster in the estimation trial. However, the statistical test calculating the completion time between the two modes confirmed that the difference was not significant ($t = -0.311$, $p = 0.83$).

6.6.3. The usability of using the multi-reference mapping graphs

As these results examined the participants' performance with each sonification approach; another important aspect to be examined is to observe the ease and comfort felt by the participants while listening to the sonification sounds. For both phases, the participants' comfort levels when reproducing the graphs were reflected in the responses to the questionnaire.

We quote some participants describing their reaction below:

1. *"The standard sounds are already good, only need to be familiar with the sounds."*
2. *"Easy and very helpful."*
3. *"This app is really cool and super easy."*

The participants also considered that the multi-reference graphs were easier to predict than the single point estimation tasks that were difficult to perform.

"I think multi-reference is easier to estimate"

"In my opinion, the condition of the single note is more difficult to guess and it takes longer for me."

Using the single point approach, participants reported that the higher valued points sounded very similar to each other. For example, it was difficult to tell the difference between values 80 and 90. As discussed in section 2.2, humans can determine both the pitch and timing of a sound signal with far greater accuracy by using an exponential distribution mapping instead of linear mapping. Metatla et.al. (2016) suggested that exponential distribution mapping is more effective than linear mapping since the difference in sequential frequencies for the human auditory system is a coefficient rather than a constant term. There is relatively little difference in frequency mapping with exponential and linear mappings at relatively low frequencies, but the differences become greater at higher frequencies. Therefore, we will use the exponential mapping of Y values to frequencies in the next study to address this issue, as described in section 3.11.6.

There was a large difference in how participants rated the ease or comfort between the two modes. 16 out of 20 participants preferred the multi-reference sonification approach after performing the tasks. According to ten participants, the single-point approach was considered the most demanding in terms of the person's cognitive load doing the point estimation. We quote four participants describing their reaction below:

1. *"It is difficult to recognize notes, especially the highest and the lowest notes and always the last one to be the benchmark for the next pitch."*
2. *"I have really no idea how to guess the single point mode. I am more comfortable guessing with multiple references because it has a pattern to guess."*
3. *"The single point is the most difficult for me rather the multi-reference."*
4. *"It takes me longer to guess the single point note compare to the multi-reference."*

This was also confirmed by the performance results that showed the multi-reference presentation generated better results. The evaluation of the multi-reference results showed almost the same trend as the real-value graphs, as shown in Figure 6-3.

6.7. Conclusion

In this study, we tested a modified multi-reference sonification scheme for non-visual point estimation following the work of Metatla (2016). In general, the experiment results show that the multi-reference mode generated more accurate results compared to the single point mode. The evaluation confirms Walker's (2002) research that adding context to auditory graphs such as tick marks enhances auditory graphs' perception.

We showed that the approach led to more accurate point estimations than the single point approach even though our multi-reference scheme requires fewer reference points than Metatla (2016) approach.

We showed the task completion times using this approach are comparable with those achieved using the single-point approach.

The approach is also scalable in the sense that there will never be more than 6 reference tones (including the one corresponding to the point itself) no matter what the numbers are on the Y-axis.

In the following chapter, we move on to studies involving visually impaired users, as these represent the primary target population for this research. The study will also explore the representation of negative numbers for non-visual point estimation tasks by integrating and evaluating the representation in several multi-reference sonification schemes.

Chapter 7. Comparison of multi-reference schemes and Representation of negative numbers for point estimation tasks

7.1. Introduction

This study aims to examine whether presenting the auditory graphs using multi-reference sonification mappings improves the accuracy and efficiency of non-visual point estimation tasks compared to tasks using a single point mapping. Several alternative representations using reference tones are explored. We examine the effect of presenting more or less audio cues on the users' ability to estimate the point's location.

The addition of contextual information for data sonification improves interaction with auditory graphs (Nees & Walker, 2008). Additionally, Metatla et al. (2016) used multi-reference sonification mapping as context information. In study 3, we examined an approach that used multi-reference tones, but fewer of them and in a more scalable form than the approach of Metatla (2016). In this study, we take this approach a step further by including condition 3, in which we only provide reference tones for pitches lying at a 5^{ths} of the distance between 0 and Y_{Max}.

In this fourth study, four pitch-based sonification mappings are examined: single point, single-reference, and two multi-reference schemes.

In the research of Metatla (2016) reference mapping, both positive and negative numbers were used by setting the origin point (i.e., zero) as the note's base point. If the point is positive, the tone will decrease from the point to the original value, and vice versa; if the point is negative, the tone will increase from the point until it ends at zero. We call this a "polarity-based" mapping because it relies on the note's polarity sign to determine whether the note sequence used to represent it is ascending or descending.

Our present study included a representation of negative numbers to investigate the effect of this on the point estimation tasks for the sonification schemes included here.

7.2. Details of the sonifications used in study 4

7.2.1. Sonifications employed in the four conditions of study 4

This study explores several alternative displays of points represented in audio to support point estimation tasks including negative numbers, and three other schemes for representing Y-axis values employing reference tones.

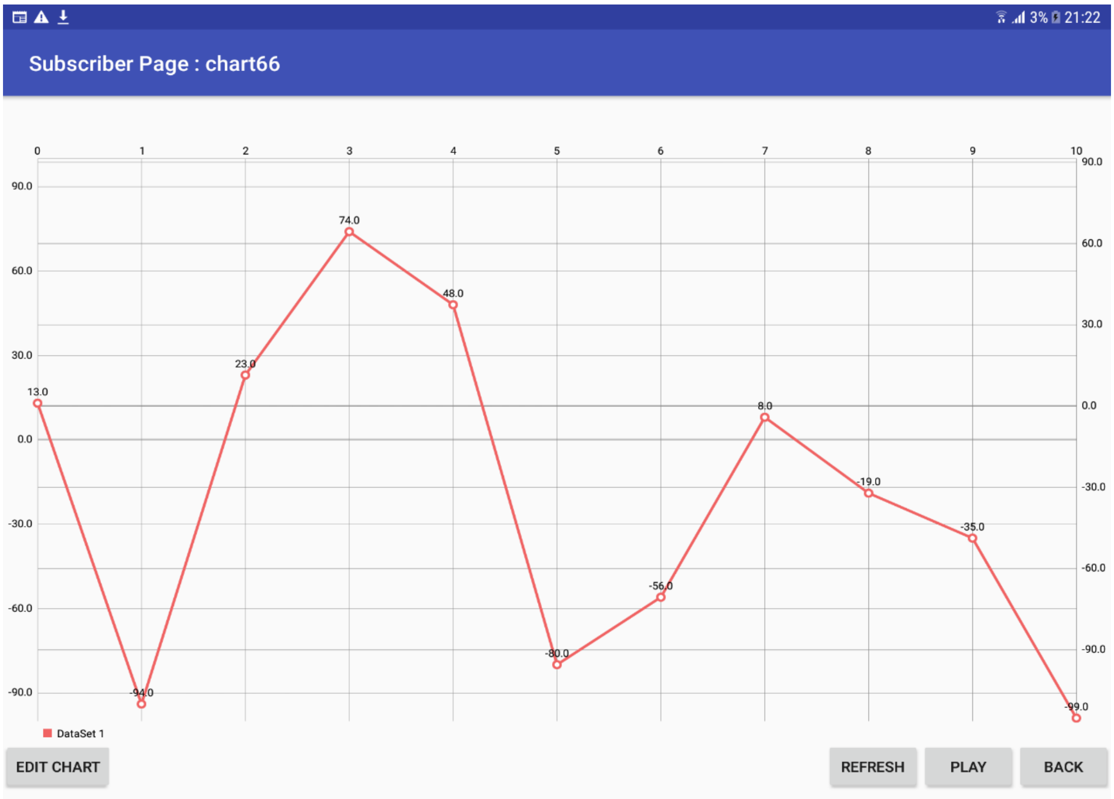


Figure 7-1 Final version of the MAG app for study 4 across the 4 conditions with the Y-axes ranged from -100 to 100.

In this study 4, the Y-axes across all four conditions were ranged from -100 to 100, which one value as one increment as shown in the graph with random notes in Figure 7-1. The MAG app 4.0 is a functional extension of the MAG app 3.0 since this study 3 prototype version is only supported positive numbers. As described in section 3.11.6, the MAG graph's sonification extends the single point mapping (condition 1) to some multi-reference conditions (condition 2, 3, and 4). Beside we wished to have a usable representation of negative numbers for use in these tasks, we also evaluated whether there was any difference across the four conditions on point estimation tasks and a polarity sign task for each group of participants. The values to frequency mappings across three multi-reference schemes are represented in an ascending order following the standard piano key frequencies shown in Table 3.1.

We have experience from previous studies 1 and 3 by starting with exploratory testing with sighted participants for further implementation with VI participants in studies 2 and 4. The reason we begin with sighted participants was that this auditory graph could benefit not only for VI users, but also for sighted users. Also, it is beneficial for recruiting sighted participants for exploratory study since it was difficult to get many VI participants.

In condition 1, the Y-axes ranged from 0 to 100 with one unit of a value corresponding to one increment. For example, a point on position 87 will trigger a pitch of point 87 only.

In condition 2, the user will never have more than two notes played, including the value of the $Y_{Estimate}$. For example, when the user hears the value 46 with Y_{Max} = 100, he or she will hear the sequence of tones comprising the pitches of points 0 and 46 as the $Y_{Estimate}$. This condition corresponds to the approached termed "one reference mapping" in Metatla's (Metatla et al., 2016) study, that is, the user hears a ***single reference corresponding to zero*** as a "reference" to an origin point, followed by the tone corresponding to the point to be estimated.

In condition 3, for Y_{Max} = 100, the user will hear the pitch for 0, and then upwards in fixed step sizes of 20 and then finally the value corresponding to $Y_{Estimate}$. For example, when the user hears the value 46, he or she will hear the sequence of tones comprising all of the pitches of points 0, 20, 40, and 46 as the $Y_{Estimate}$. This condition is a modified version of the multi-reference scheme described in chapter 6, but the user hears the reference tones increment in units of 20, for Y_{Max} = 100. More generally, in 5^{ths} of Y_{Max} up to when the 5^{th} below the $Y_{Estimate}$ is heard, and then the last pitch corresponding to $Y_{Estimate}$.

However, the user hears the $Y_{Estimate}$ value twice if the Y value is a multiple of 5^{ths}. For example, the number 40 will give the sequence 0, 20, 40, and 40 again.

In condition 4, the user will hear from 0, in fixed steps sizes of 10, including the value of the $Y_{Estimate}$. When the user hears the value 46, he or she will hear the sequence of tones comprising all the pitches of points 0, 10, 20, 30, 40, and 46 as the $Y_{Estimate}$. This condition is also a modified version of the multi-reference scheme described in chapter 6, but the user hears the reference tones increment in units of 10, for Y_{Max} = 100. More generally, in 10^{ths} of Y_{Max} up to when the 10^{th} below the $Y_{Estimate}$ is heard, and then the last pitch corresponding to $Y_{Estimate}$ itself.

The user hears the $Y_{Estimate}$ value twice when the value is a multiple of 10^{ths}. For example, the number 40 will give the sequence of pitches corresponding to points 0, 10, 20, 30, 40, and 40 again.

7.2.2. Representation of negative number

This study also explores the representation of negative numbers for non-visual point estimation. Negative numbers may be represented as *component-based representation*, i.e., as two separate components (one digit and one sign) or may be represented as *holistically*, as discussed in section 2.5.3. In contrast with Metatla's (2016) study that listeners need to hear more tones as they get further away from 0, we want to show an approach that does not rely on polarity, but uses a component-based approach, which is increasingly more efficient than Metatla's approach for numbers that are further away from 0, as explained in section 2.5.3.

For this study, we choose the *component-based representation* approach to represent the mapping by having the same positive mapping reference for the digit and adding one sign before the digit with "sonar" sound. The sonar sound is chosen because it brings to mind a submarine, whose position is below 0 meters, suggesting the idea of something below 0.

7.3. Research Questions

In this study, we aim to examine two research questions:

1. Which of the four sonification schemes, ***single point***, ***single reference***, ***step10*** or ***step20***, will prove best in terms of user performance in point estimation tasks?
2. Which of the four sonification schemes Will prove best for user performance on polarity sign selection tasks?

we formulated the following hypotheses:

H1: Participants will make significantly more point estimation errors when using the ***single point*** sonification mappings compared with the ***single reference*** and the multi-references mappings using the step20 and step10 sonification mappings.

H2: Participants will make significantly more point estimation errors when using the ***single reference*** sonification mapping as compared with the ***multi-reference mappings of 20 steps*** and ***10 steps***.

H3: Participants will make significantly more point estimation errors when using the ***multi-reference mapping of 20 steps*** compared to ***the multi-reference mapping of 10 steps***.

H4: Participants will make significantly better polarity sign selections when using the ***single point*** and ***single reference*** sonification mappings compared with the one-reference and the multi-references mappings of 20 steps and 10 steps.

H5: VI participants will perform better on point estimation tasks compared to sighted participants in all conditions

H6: VI participants will perform better on polarity sign selection tasks compared to sighted participants in all conditions

7.4. Study Design

7.4.1. Participants

A total of 40 participants volunteered to take part in this experiment, consisting of 20 with a visual impairment (11 men and 9 women) and 20 sighted participants (14 men and 6 women). Their ages ranged between 17 and 49 years (VI participants) and between 21 and 49 years (sighted participants). All of the VI participants were completely blind and used a speech-based screen reader on their mobile device.

All VI participants were from Indonesia because they were easier to recruit in that country, and the researcher is a native Indonesian speaker. The community of VI people in Indonesia is close knit, with good communications, which was of considerable help in recruiting VI participants for the study.

Both sighted and VI participants received cash incentives for their participation. As for all of the earlier studies, we obtained approval from the research ethics committee of the Queen Mary University of London.

The VI participants had more varied occupational backgrounds than their sighted peers, including athletes and musicians. The distributions of their occupational background are presented in Figure 7-2 and Figure 7-4. Table 7.1 summarizes the respondent's demographic information

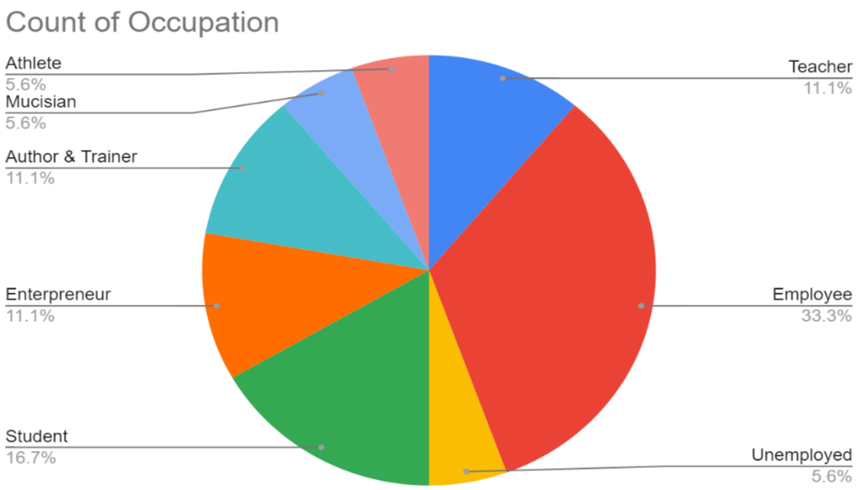


Figure 7-2 Distribution of Occupational Background of VI participants

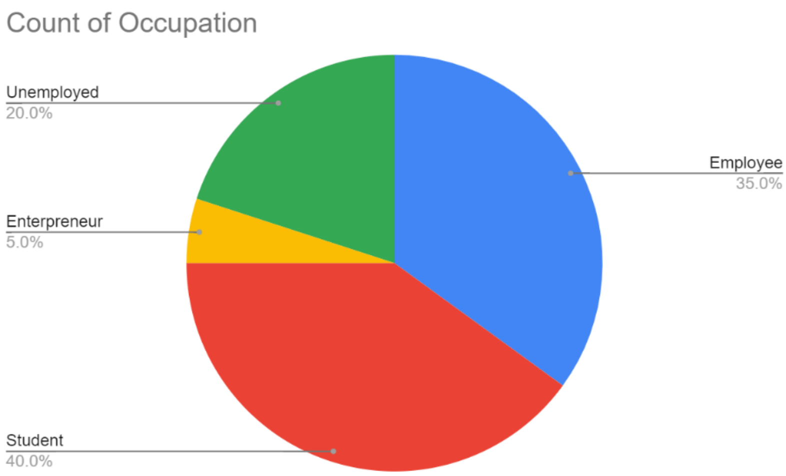


Figure 7-3 Distribution of Occupational Background of sighted participants

All participants were asked about their level of musical training and experience of playing an instrument. Two of the VI participants had formal musical training (to grades 4 and 5), three had informal training (estimated to be roughly to grade 3), while the rest had none (see Figure 7-4). The levels of musical training and experience of their sighted peers are shown in Figure 7-5.

Variable	Participants

			VI	Sighted
Age (years)	-	< 20	- 1	- 0
	-	21 – 29	- 10	- 8
	-	30 -39	- 8	- 10
	-	40 - 49	- 1	- 2
Gender	-	Male	- 11	- 14
	-	Female	- 9	- 6
Qualification	-	High School graduate	- 50%	- 20%
	-	Undergraduate	- 45%	- 60%
	-	Post graduate	- 5%	- 20%

Table 7.1 Demographic Information of the Participants

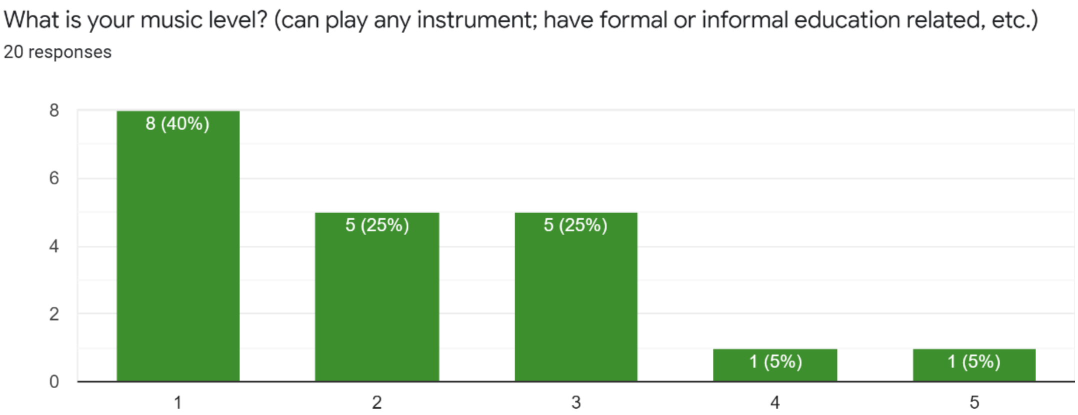


Figure 7-4 Musical Level Experience according to VI participants Questionnaire

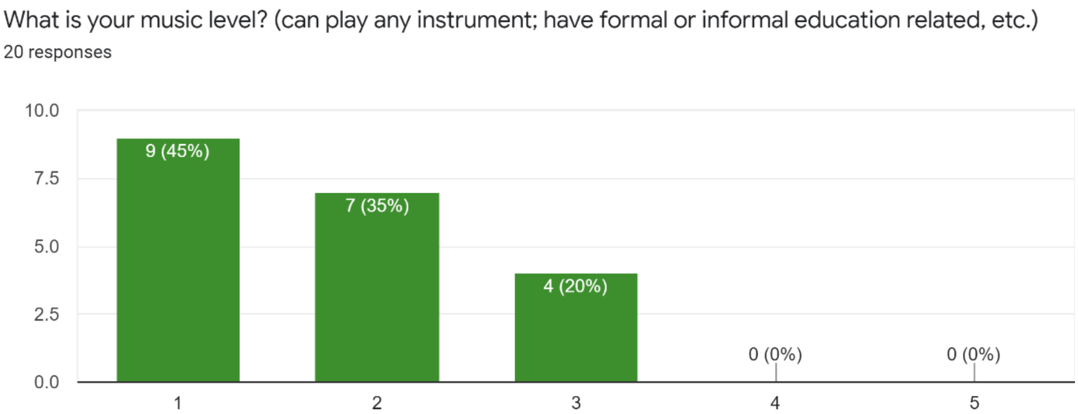


Figure 7-5. Musical Level Experience according to Sighted Participants Questionnaire

The study was conducted between October and November 2019 in three countries: the UK, Germany, and Indonesia. The experiment took place in the home of each participant. Care was taken to minimise disruptions and external noise during the experiment. Participants were given information about the study and the experimental tasks they would undertake, but not about the experiment's objectives. For the VI participants, the information was provided digitally to read it using the screen reader on their tablet device.

7.4.2. Preparation and training

After completing the demographic questionnaire, participants were introduced to the concept of sonification, both verbally and using an example, and told them to estimate the Y coordinates of points on a graph using different sonification schemes. They were then introduced to the MAG app interface and shown how it is used to display and sonify graphs. Participants were then walked through the process of estimating points on a graph in response to a sonification. Participants then listened to the respective displays for all the different sonification conditions. They could ask to listen as often as they wished to the different sonification conditions until they felt they understood them. The training period lasted roughly ten minutes for each condition.

The first sonification condition for each participant was chosen randomly. For each condition, the participant would listen to the range of possible pitches in increments of ten, going from 0 to the Y_{Max} set at 100 in this experiment. Then the range of negative pitches would be played in descending order, descending in increments of 10 from 0 down to the Y_{Min} set at -100 in this experiment. This test implements the component representation approach for negative numbers to represent the mapping by using the same positive mapping reference for the digit and adding a sign before the digit with a "sonar" sound.

These values were presented to train the participants' to differentiate between the sound representing values from the lowest to the highest points in the positive and negative ranges, respectively.

Each condition was tested twice. Firstly, with only multiples of ten, and the second time with sets of numbers randomly distributed spanning each range respectively to expose participants to plenty of values that were not multiples of 10. This procedure continued until all 4 four conditions had been completed.

7.4.3. Main Experimental Session

The main experimental session was conducted using a Samsung Galaxy Tab S2 with a 9.7-inch screen, running the Android 7 OS. The duration of each sonification condition lasted from 10 to 20 minutes. During each condition, participants listen to an auditory graph played by the researcher. The researcher put his finger on the left-hand side of the MAG app interface displaying the graph and moved his finger through the points until the graph's end.

When the researcher touched each point, the participant heard a tone representing the current condition's Y-value. The participants then estimated that point verbally so that the researcher could note it down.

Participants were asked to complete all four conditions (*single point*, *single reference*, multi-reference for steps of 20 or *condition step20*, and multi-reference for steps of 10 or *condition step10*) in random order to avoid possible learning or bias effects.

Each condition consisted of 40-point estimation trials with the same trial order applied to all four conditions. For example, if condition one began with point values of 23, 35, -9, and so forth; then conditions 2, 3, and 4 would have exactly the same number 23, 35, -9, and so forth. Point values were randomly assigned to ensure a comprehensive coverage from Y_{Min} to Y_{Max} across the 40 trials, and care was taken to ensure equal representation of positive and negative numbers.

Participants were not given any feedback about the real Y-values after each trial.

At the end of all trials, we carried out informal interviews to get participants' feedback about their experience undertaking the point estimations tasks. This interview allowed the researcher to obtain feedback from participants about the relative advantages and drawbacks of the different sonification conditions.

The overall sessions, consisting of the training and the main trials, lasted between 60-90 minutes per participant. However, most of the VI participants spent longer, approximately 90 minutes, than their sighted peers. This is because VI participants typically have to spend more time conducting training and getting used to interacting with this interface than sighted participants

Overall, it took about 60 hours to perform the study.

7.5. ANOVA statistical testing

7.5.1. Normality test

To further establish the principles that underpin all of the statistical analysis that follows this study, we follow the statistical analysis testing flowchart as discussed in 3.8. We explore the differences between the sighted and VI group results in a little more depth. Especially the phenomenon whereby sighted participants' results showed more significant differences between the four conditions than those of the VI participants. We further carry out statistical testing (e.g ANOVA) that will allow the interaction effect of participant group or condition to be evaluated.

Before we decided to use parametric or non-parametric analysis, we assessed each variable's normality (univariate normality) by inspecting the respective Shapiro-Wilk Normality test and also skewness and kurtosis values. If the p-value of the Shapiro-Wilk test is less than 0.05, then the data violate the normality assumption. However, researchers often use the skewness and kurtosis values, which are less conservative than the Shapiro-Wilk test. The skewness occurs when responses are more frequent at one part of the measurement scale and affect the variance-covariance among variables. Kurtosis reflects the flatness in the data distribution. The further the value of skewness or kurtosis is from zero, the more likely it is that the data are not normally distributed. The skewness, kurtosis values, and their respective z-scores for each condition is provided in Appendix H.

If the data meet the normality assumption, we determine to use the parametric test and non-parametric test otherwise. To evaluate whether any significant differences of the variable of interest between groups (i.e., RMSE on point estimation task), we will conduct ANOVA mixed design or Friedman test if data distribution was not normal. Both analyses compare three or more groups where the participants are the same in each group.

7.5.2. Point Estimation Task

A. Parametric Test- ANOVA Mixed Design

If we decide to use ANOVA design -in addition to sphericity assumption (Table 7.10), we need homogeneity of variance assumption as shown in Table 7.3.

	Mauchly's	Sig (p-value)	Greenhouse-Geisser
RMSE 4 conditions	0.484	<0.001	0.672

Table 7.2. Sphericity of Data Study 4

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Cond1	1.676	1	38	.203
Cond2	1.021	1	38	.319
Cond3	1.918	1	38	.174
Cond4	.939	1	38	.339

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Group2

Within Subjects Design: RMSEcond

Table 7.3. Levene's Test of Homogeneity

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.
RMSEcond	Sphericity Assumed	5933.558	3	1977.853	23.335	.000
	Greenhouse-Geisser	5933.558	2.015	2945.028	23.335	.000
	Huynh-Feldt	5933.558	2.184	2716.861	23.335	.000
	Lower-bound	5933.558	1.000	5933.558	23.335	.000
RMSEcond * Group2	Sphericity Assumed	598.911	3	199.637	2.355	.076
	Greenhouse-Geisser	598.911	2.015	297.260	2.355	.101
	Huynh-Feldt	598.911	2.184	274.230	2.355	.096
	Lower-bound	598.911	1.000	598.911	2.355	.133
Error(RMSEcond)	Sphericity Assumed	9662.466	114	84.758		
	Greenhouse-Geisser	9662.466	76.561	126.206		
	Huynh-Feldt	9662.466	82.991	116.428		
	Lower-bound	9662.466	38.000	254.275		

Table 7.4. ANOVA Mixed Design of Point Estimation Task Across Conditions for both VI and Sight Groups (Conditions x Groups)

The effect of conditions is significant ($p<0.001$), while the interaction effect of conditions and group is not significant ($p=0.101$). The insignificant interaction effect suggests that the effects of one variable (conditions) did not depend on the second variable's level (group of participants).

Table 7.13 displayed the ANOVA results for our between-groups variable and group participants (blind vs. sighted). Since the p-value is less than 0.05 (p=0.035), we can conclude that the main effect for the group is significant, so it means that the RMSE of sighted participants are much more significant than the VI participants.

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	67864.520	1	67864.520	417.573	.000
Group2	776.456	1	776.456	4.778	.035
Error	6175.808	38	162.521		

Table 7.5. Between-subjects' effects of Mixed ANOVA Design

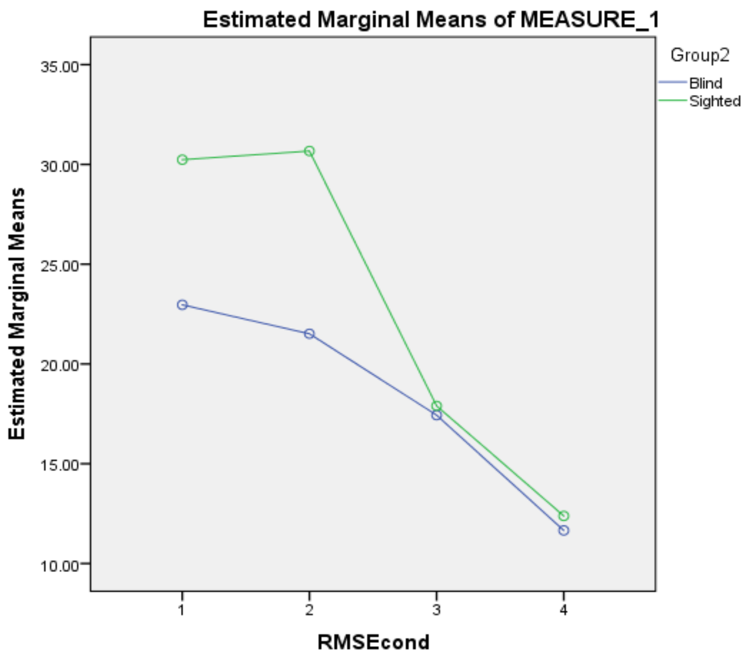


Figure 7-6. Diagram of interaction effect of RMSE condition between VI and sighted group

It can be seen in this Figure 7-6 that the effect of the conditions did not depend on the group of participants. Looking at the two lines, we see that RMSE for both groups were declined from condition 1 to condition 4. Therefore, the conditions produced much of change in RSME of both VI and sighted participants.

B. Non-Parametric ANOVA Mixed Design

When grouped the tasks into four conditions for VI group and sighted group, we found significant difference using Friedman-test (Chi-Square=20.580, df =3, p<0.001 for VI group; Chi-Square=39.420, df =3, p<0.001 for sighted group) and , as shown in Table 7.6 and Table 7.7.

Test Statistics[a]

N	20
Chi-Square	20.580
Df	3
Asymp. Sig.	.000

a. Friedman Test

Table 7.6. Friedman Test on Point Estimation Task (RMSE) for four conditions VI Groups

Test Statistics[a]

N	20
Chi-Square	39.420
Df	3
Asymp. Sig.	.000

a. Friedman Test

Table 7.7. Friedman Test on Point Estimation Task (RMSE) for four conditions VI Groups

C. Non-Parametric-Post-hoc test-Wilcoxon

To follow up on which condition is different, we conducted the Wilcoxon-paired test, as discussed in section 7.6.1.1. for VI group and section 7.6.2.1 for sighted group.

7.5.3. Polarity Sign Task

We could not conduct mixed ANOVA for polarity sign task because too many condition data violated normality assumption particularly, so we could not evaluate the interaction effect between condition and groups of participants. Therefore, we performed the Friedman test (an alternative for ANOVA). In parallel with ANOVA results, there is no significant difference both for VI and sighted groups for the polarity sign task, as shown in Table 7.8 and Table 7.9.

Test Statistics[a]

N	20
Chi-Square	4.232
Df	3
Asymp. Sig.	.237

a. Friedman Test

Table 7.8. Friedman Test on Polarity Sign Task for four conditions of VI Groups

Test Statistics[a]

N	20
Chi-Square	.638
df	3
Asymp. Sig.	.888

a. Friedman Test

Table 7.9. Friedman Test Polarity Sign Task for four conditions of Sighted Groups

7.6. Results

We will examine the results separately for each condition for each participant group. We will analyze the difference between groups after aggregating all the data for each group across all conditions (see sub-section 7.6.3).

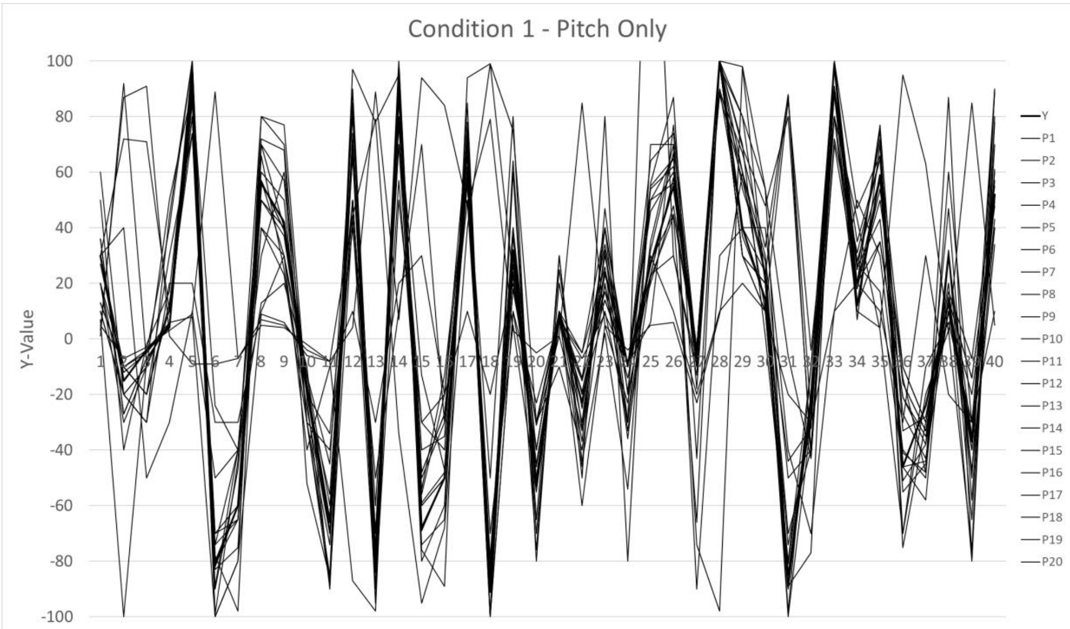
As in our previous studies, the RMSE between the estimated (observed) and true values will be used as an outcome measure of point estimation performance. The RMSEs were calculated for all participants across the four conditions. Moreover, the representation of negative numbers will be calculated by the percentage of correct polarity sign guesses.

To examine whether all data distributions meet the normality assumption, we performed the Shapiro-Wilk tests in addition to observing the skewness of the respective histograms. A significant value (p-value less than 0.05) indicates deviation from normality. When data were not normally distributed, a non-parametric test, which does not require the data to be normally distributed, was performed to analyze the data. The level of significance was set at $\alpha < 0.05$ for all of the tests.

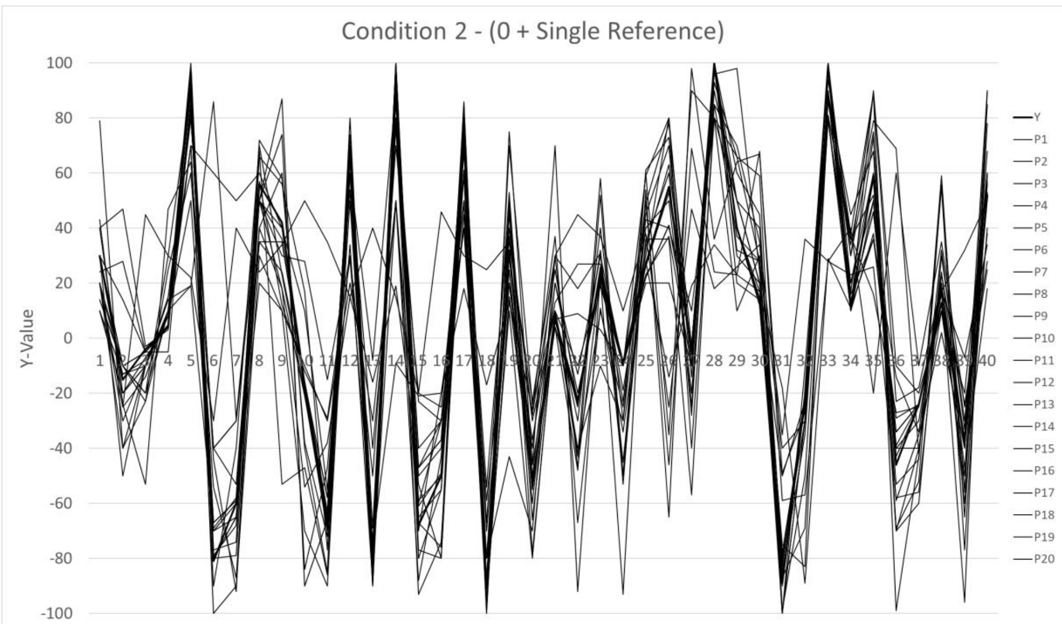
7.6.1. Results for VI Participants

7.6.1.1. Point Estimation Tasks

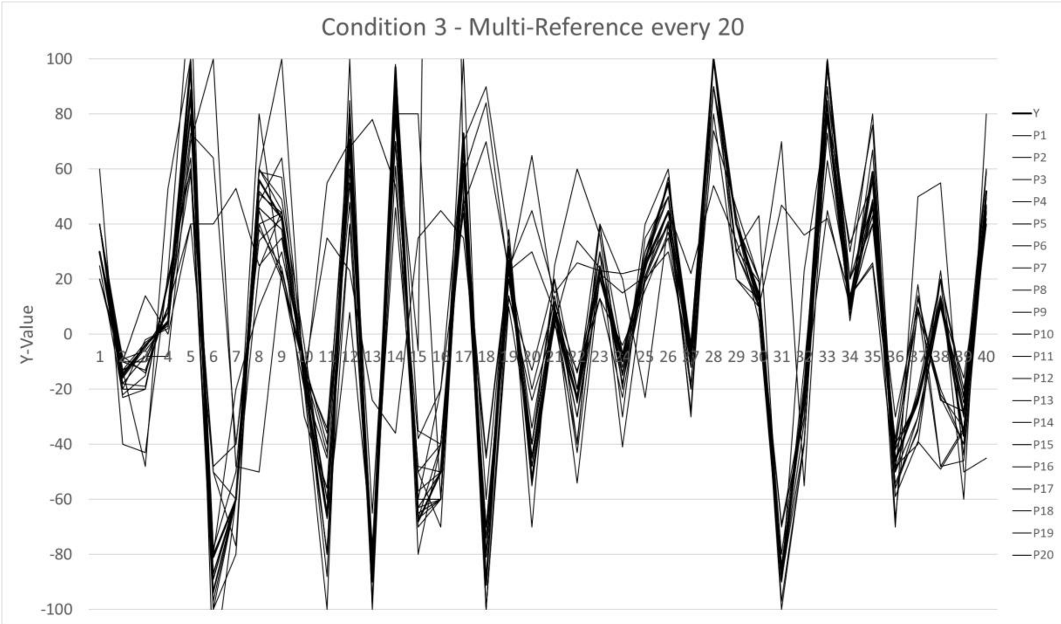
The results of the VI point estimation tasks for each condition are displayed in Figure 7-7, consisting of four sonification conditions: the *single point mode, single reference mode, multi-reference for condition step20, and multi-reference for condition step10*.



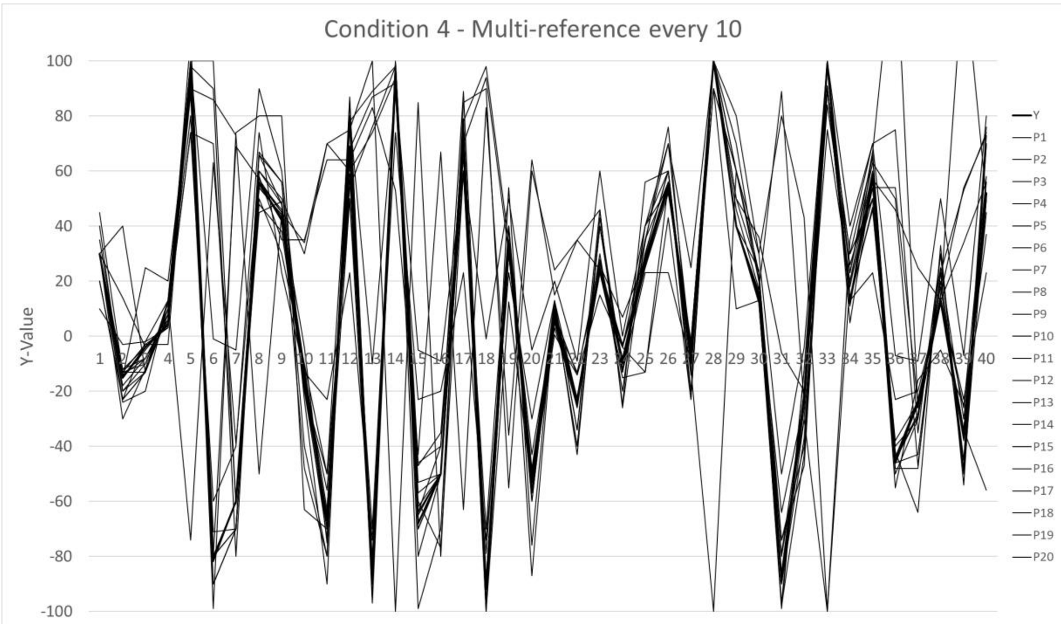
(a)



(b)



(c)



(d)

Figure 7-7 Point estimations of 20 VI participants with their respective true-value graph as a reference using: (a) single point mode (b) single reference mode (c) multi-reference for condition step20, (d) multi-reference for condition step10.

The descriptive statistics of VI participants are summarized in Table 7.10.

Conditions	Count	Mean	SD	Median	IQR
Condition 1	20	22.96	11.73	18.31	19.25
Condition 2	20	21.51	10.15	20.80	14.18
Condition 3	20	17.90	17.41	15.02	7.55
Condition 4	20	11.66	8.18	10.24	10.81

Table 7.10. Mean, Standard Deviation (SD), Median, and Interquartile Range (IQR) of RMSE on Point Estimation Tasks of 20 VI Participants across All Conditions

To visualize whether any relationship between the performance of point estimation tasks and the type of conditions existed, the distribution of RMSE values are displayed into mean and box plots as shown in Figure 7-8.

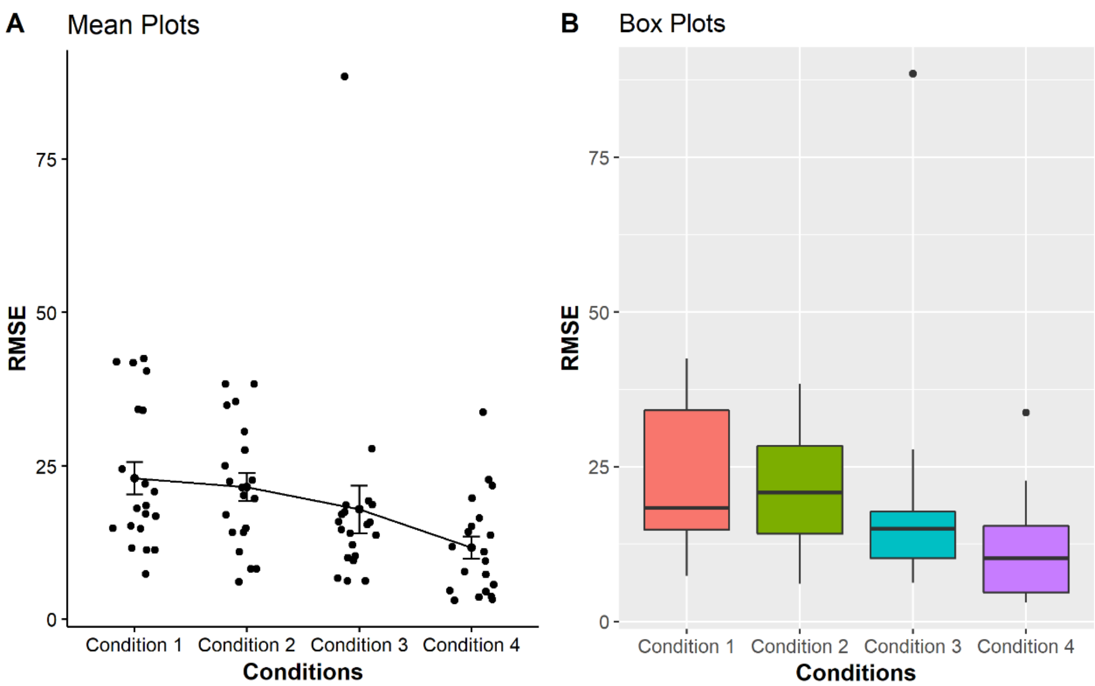


Figure 7-8. Mean Plots (a) and Boxplot (b) of 20 VI participants, representing the Distribution of RMSEs Obtained from Conditions 1 to 4

The multi-reference mode used in conditions 3 and 4 shows a better performance in the point estimation tasks than conditions 1 and 2. This was indicated by smaller RMSEs which provided lower median values and smaller distributions than those using the single point condition (condition 1) and the single reference condition (condition 2).

The mean variance of each RMSE condition tends to decrease from condition 1 to condition 4.

The Shapiro-Wilk test showed non normal distribution for condition 1 ($W = 0.86$, $p = 0.009$), condition 3 ($W = 0.50$, $p < 0.001$), and condition 4 ($W = 0.89$, $p = 0.032$). The data of condition 2 was normally distributed as indicated by p-value > 0.05 ($W = 0.94$, $p = 0.34$). The visualization of the distribution of these conditions tended to be skewed to the right, also supporting this finding, as shown in Figure 7-9.

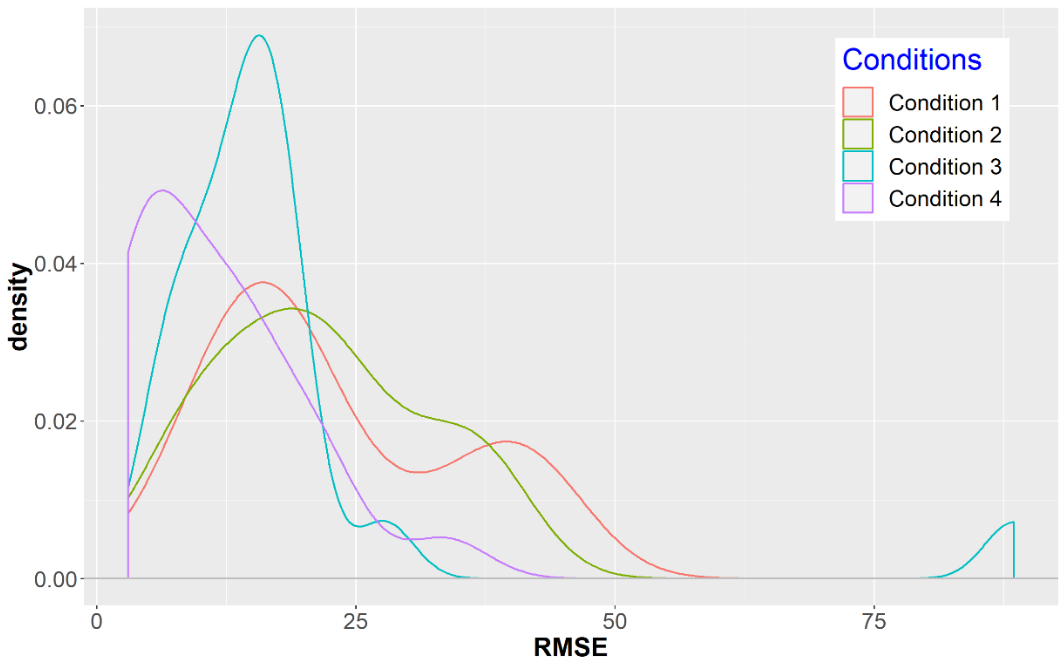


Figure 7-9. Histograms of the RMSEs from all Conditions (1-4) of 20 VI Participants on Point Estimation Tasks

To test whether any significant difference of RMSE means was found across all conditions, we ran the Kruskal-Wallis test (Kruskal & Wallis, 1952). The statistics showed a p-value less than the significance level of 0.05 ($p = 0.001$), indicating a significant difference between conditions 1 to 4.

To follow up this finding, a post-hoc analysis using the Wilcoxon test was performed to determine which independent variable levels differ from each other. In this case, we applied Benjamini-Hochberg (BH) correction method to control familywise error rate (FER) or false discovery rate (FDR) (Benjamini & Hochberg, 1995).

The pairwise comparisons results indicate that differences were found for condition 1 vs condition 4 ($p = 0.0061$) and condition 2 vs condition 4 ($p = 0.0064$) as displayed in Table 7.11.

	Condition1	**Condition2**	**Condition3**
Condition2	0.797	-	-
Condition3	0.0702	0.0702	-
Condition4	**0.0061**	**0.0064**	0.0864

Table 7.11. Pairwise Comparisons between Conditions 1, 2, 3, and 4 using the Wilcoxon Rank sum test. *p*-value-adjustment shown. Bold marked shows that a significant difference was found between RMSE means after applying the BH correction

7.6.1.2. Representation of Negative Numbers

To analyse the performance of VI participants on sign polarity estimates, their respective values across all VI participants for all conditions were calculated. We employed a stimuli detection task containing a mix between positive and negative numbers over the 40 trials. The set of positive and negative scores was divided equally between each of the 20 trials in random positions for all four conditions.

We calculated the percentage of correct polarity sign estimates. For example, if all 20 participants correctly predicted negative polarities during a trial of negative numbers, then the percentage is 100. However, if two participants falsely estimated the number as positive, then the percentage dropped to 90. The same rules apply to positive trials. The summary of descriptive statistics for 40 trials is displayed in Table 7.12 and the number of each sighted and VI participants' false polarity estimations are shown in Appendix H.

Based on the table in Appendix H, if user 1 for condition 1 got 2, for example, it means that he/she made two false polar estimations out of 40 trials. The value is positive, but the participant falsely declares it as negative or vice versa.

Conditions	Trial	Mean	SD	Median	IQR
Condition 1	40	95.87	5.76	100	5
Condition 2	40	94.75	7.33	95	5
Condition 3	40	93.62	9.47	95	10
Condition 4	40	92.12	7.67	95	6.25

Table 7.12. Mean, Standard Deviation (SD), Median, and Interquartile Range (IQR)of Percentage of Correct Polarity Sign estimates of 20 VI Participants' across All Conditions

To investigate the relationship between the negative number representation task performance and the type of conditions, we visualized the percentage of correct polarity sign estimates as four boxplots and mean plots as shown in Figure 7-10.

The boxplots of the sonification mappings used in conditions 1 and 2 showed better performance in the polarity sign estimation tasks, as indicated by their higher percentage of correct polarity sign estimates. Their higher means, medians, and smaller distributions also supported this finding as compared with those using multi-reference for condition step20 and 10 (see Table 7.12). The mean-variance of each condition tends to decrease from condition 1 to condition 4, as shown by the mean plot.

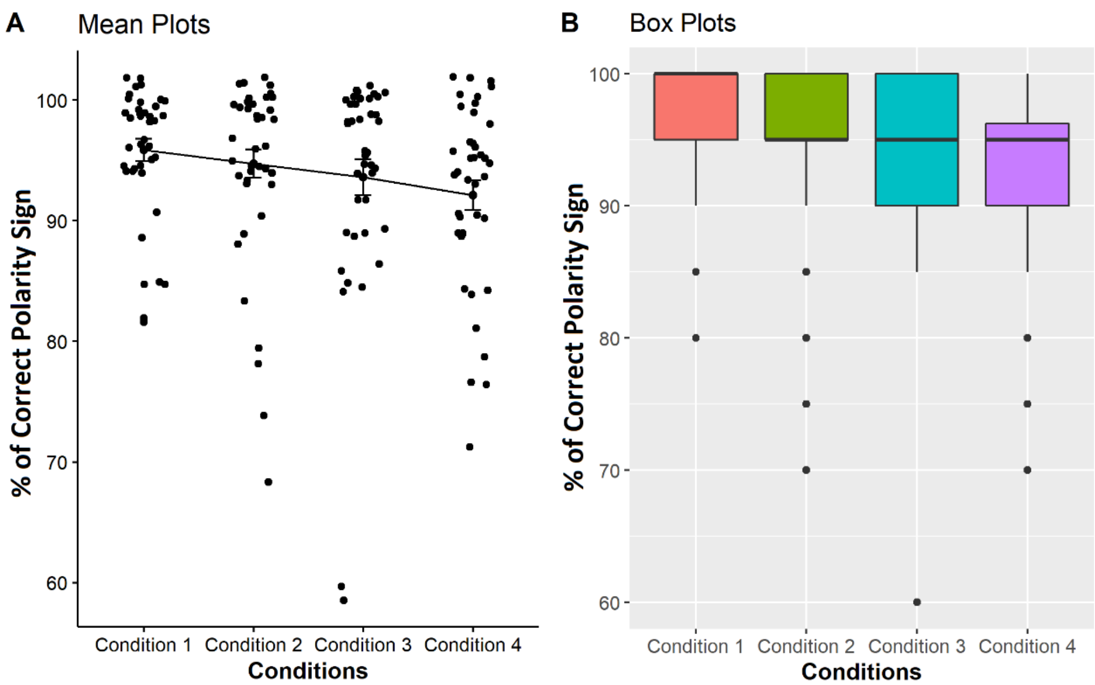


Figure 7-10. Mean Plots (a) and Boxplots (b) for VI participants, representing the Percentage of Correct Polarity Sign estimations in 40 trials obtained from Conditions 1 to 4.

The result of Shapiro-Wilk test showed all conditions data were not normally distributed (p-values< 0.05): condition 1 (W = 0.72, p < 0.001), condition 2 (W = 0.71, p < 0.001), condition 3 (W = 0.67, p < 0.001) and condition 4 (W = 0.85, p < 0.001). The histogram of four conditions were all skewed left as shown in Figure 7-11. The non-parametric Kruskal–Wallis test (1952) showed that all four conditions did not differ significantly in the percentage of correct polarity sign estimates as indicated by its p-value being larger than 0.05 (Kruskal-Wallis Chi-squared = 7.5951, p= 0.052).

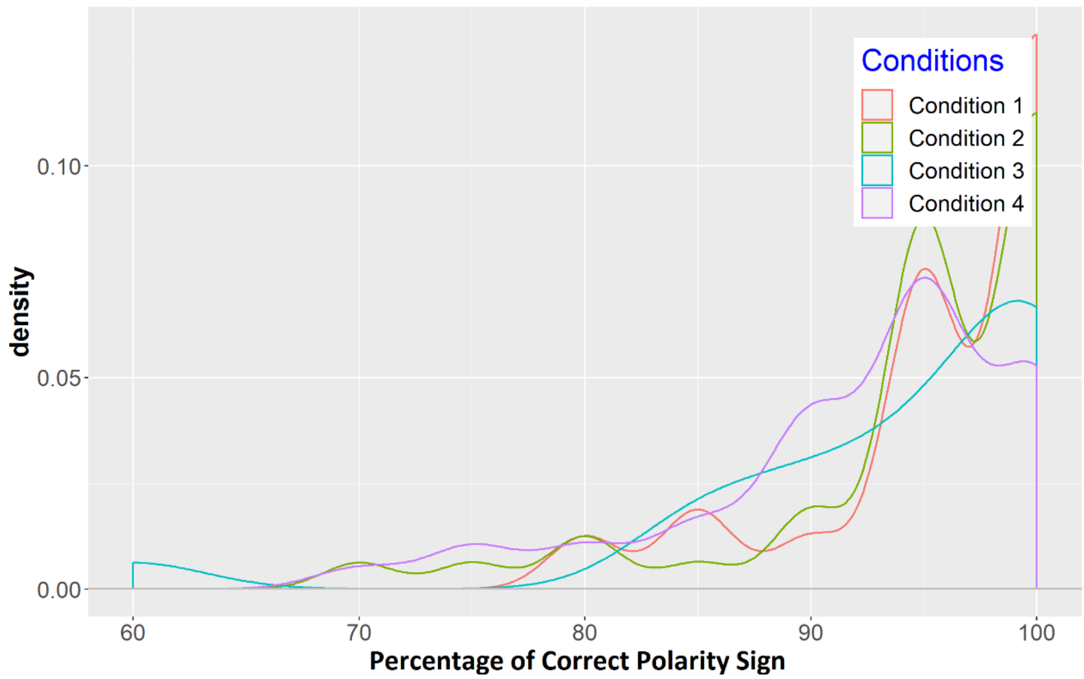
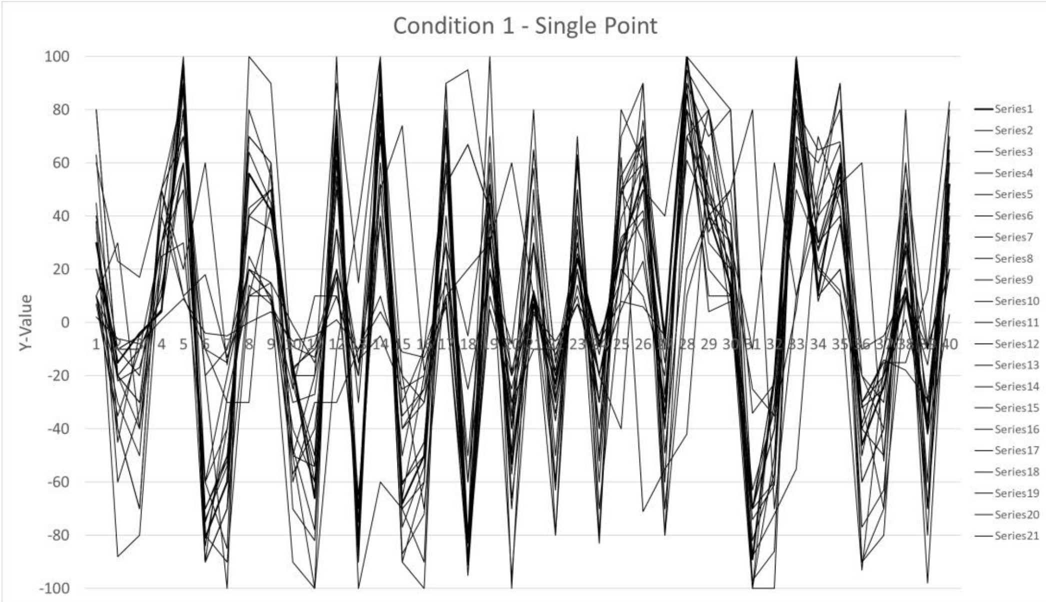


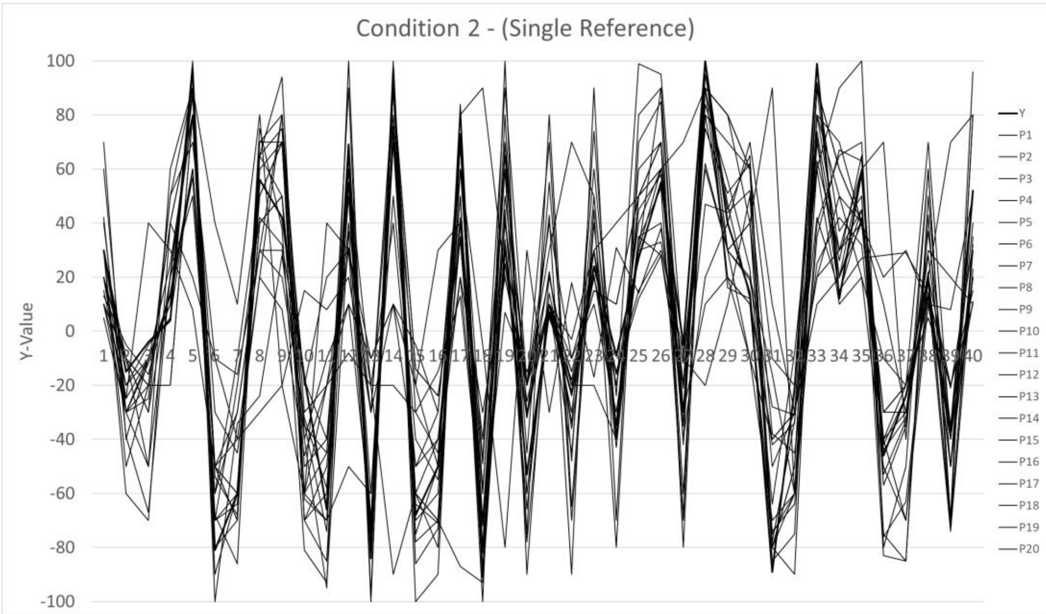
Figure 7-11. Histograms of the Percentage of Correct Polarity Sign estimates of 20 VI Participants across All Conditions

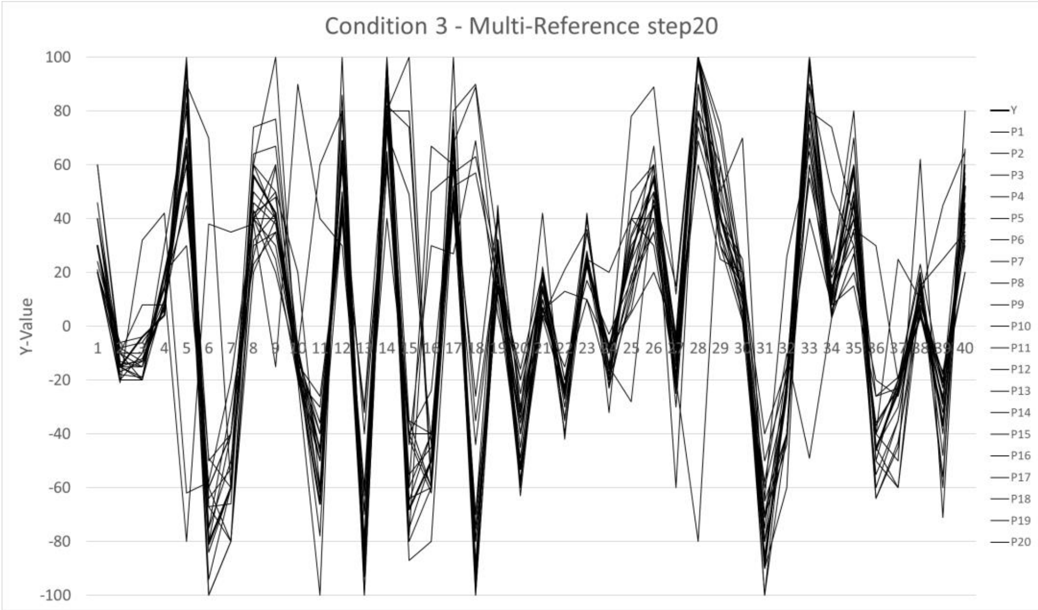
7.6.2. Results for Sighted Participants

7.6.2.1. Point Estimation Tasks

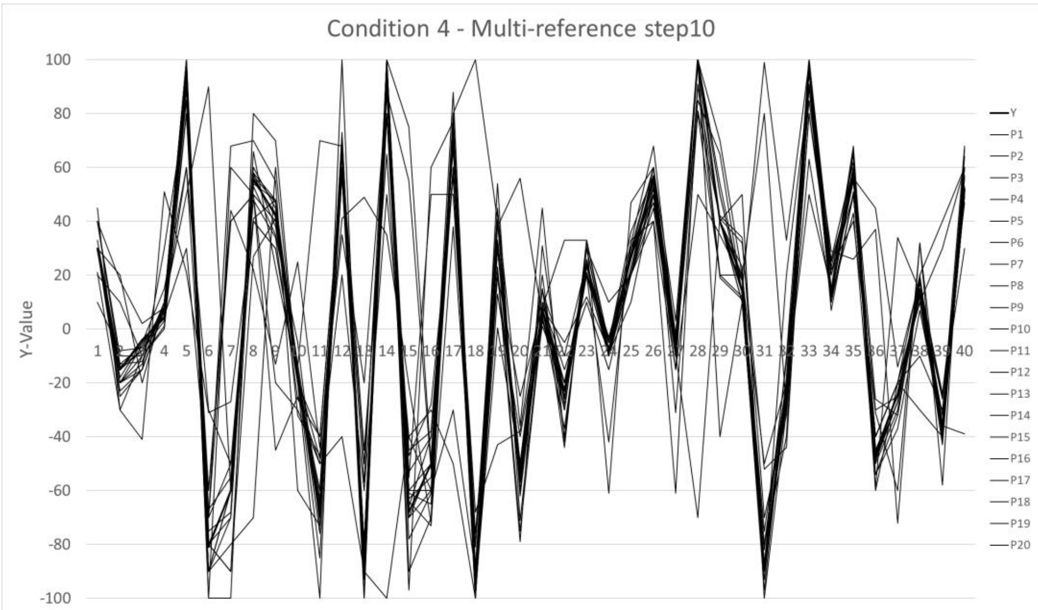
We followed the same data analysis procedure as had been conducted for VI participants. The point estimation task results for each condition: the single point mode, single reference mode, multi-reference for condition step20, and multi-reference for condition step10 were shown in Figure 7-12. The bold line showed the true values to be estimated, followed by the graphs produced by each participant.


Condition 1 - Single Point
Y-Value
Series1
Series2
Series3
Series4
Series5
Series6
Series7
Series8
Series9
Series10
Series11
Series12
Series13
Series14
Series15
Series16
Series17
Series18
Series19
Series20
Series21
(a)


Condition 2 - (Single Reference)
Y-Value
Y
P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
P11
P12
P13
P14
P15
P16
P17
P18
P19
P20
(b)



(c)



(d)

Figure 7-12 Point estimations of 20 sighted participants with their respective true-value graphs as a reference using: (a) single point mode (b) single reference mode (c) multi-reference for condition step20, (d) multi-reference for condition step10.

We then calculated the RMSE between the estimated and true values across all participants, as shown in Table 7.13.

The summary of the descriptive statistics of sighted participants is displayed in Table 7.13.

Conditions	Count	Mean	SD	Median	IQR
Condition 1	20	30.24	9.86	30.01	10.46
Condition 2	20	30.67	7.95	30.28	11.62
Condition 3	20	17.44	4.76	16.98	4.90
Condition 4	20	12.39	6.31	10.15	9.21

Table 7.13 Mean, Standard Deviation (SD), Median, and Interquartile Range (IQR) of RMSE on Point Estimation Tasks of 20 Sighted participants across All Conditions

As we did for the VI participants, we split the RMSE data into four groups (i.e., conditions) to evaluate the point estimation performance's relationships across all conditions. The mean and boxplots of the RMSE distribution values are displayed in Figure 7-13.

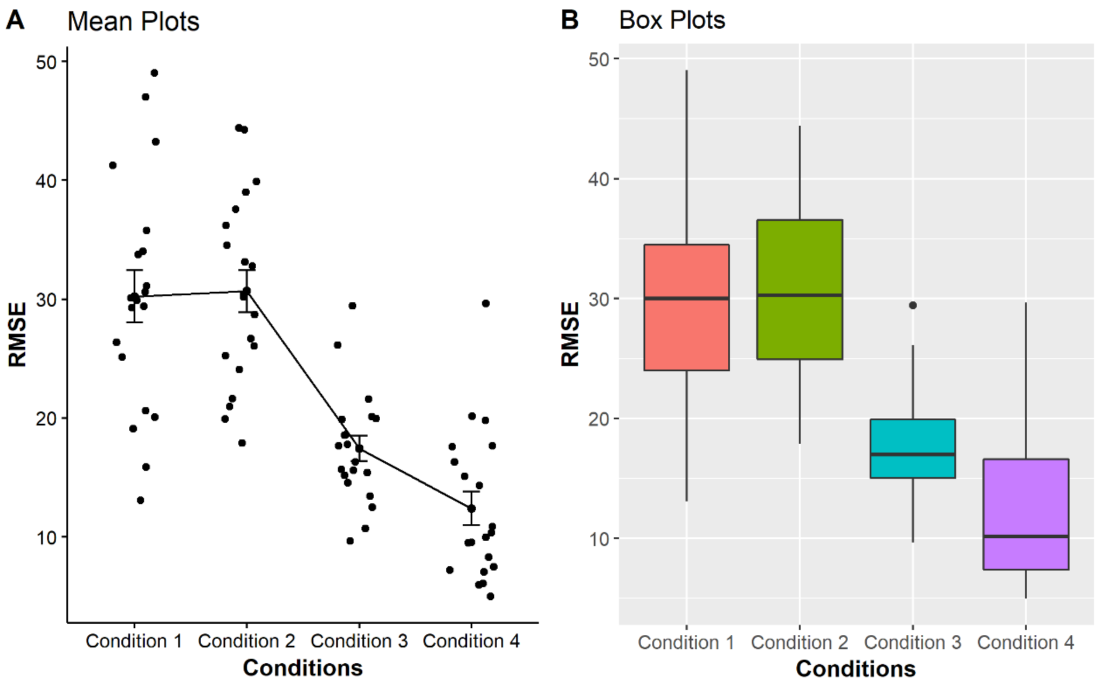


Figure 7-13 Mean Plots (a) and Boxplot (b) of Sighted Participants, representing the Distribution of RMSE values Obtained from Conditions 1 to 4.

The point estimation tasks using multi-reference mode in conditions 3 and 4 shows better performance as indicated by their smaller errors. Both conditions showed lower medians and quantiles as compared with conditions 1 and 2. According to the mean plot, the mean-variance of each RMSE condition was stable between conditions 1 and 2 and tended to decrease from condition 2 to condition 4.

We ran a Shapiro-Wilk test to check whether the data met the normality assumption. The results showed that data in conditions 1, 2 and 3 were normally distributed (W_1 = 0.97, p_1 = 0.72, W_2 = 0.97, p_2 = 0.82, and W_3 = 0.95, p_3 = 0.35, respectively) while condition 4 violated the normality assumption (W = 0.89, p = 0.03).

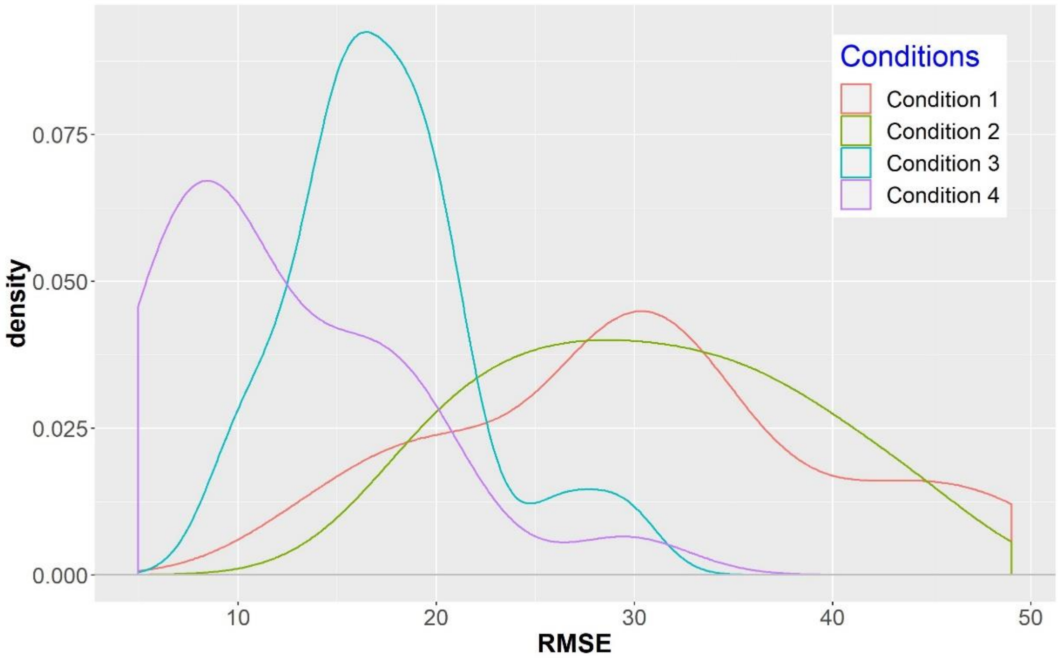


Figure 7-14 Histograms of the RMSE values for all Conditions (1-4) of 20 Sighted Participants on Point Estimation Tasks.

The histograms of conditions 3 and 4 were skewed right which supported the non-normality distribution, as shown in Figure 7-14. Therefore, we performed a Kruskal-Wallis test to examine the significant difference of RMSE means among sighted participants between the four conditions. We found the p-value is less than 0.05 ($p < 0.001$), so there were significant differences between the conditions.

Post hoc analysis using a Wilcoxon test with Benjamini-Hochberg correction procedure was applied to determine which pairs of conditions differed significantly. Table 7.14 revealed that the pairwise comparison of conditions 1 and 2 was not significantly different (p=0.82) while other pairs differed significantly.

	Condition1	Condition2	Condition3
Condition2	0.82	-	-
Condition3	**< 0.001**	**< 0.001**	-
Condition4	**< 0.001**	**< 0.001**	**< 0.001**

Table 7.14 Pairwise Comparison between Conditions 1, 2, 3, and 4 using the Wilcoxon Rank sum test. *p*-value-adjustment shown. Boldly marked shows a significant difference between RMSE means after applying the BH correction

7.6.2.2. Representation of Negative Numbers

The descriptive statistics of the percentage of correct polarity sign estimates for all conditions for all sighted participants are summarized in Table 7.15.

Conditions	Trial	Mean	SD	Median	IQR
Condition 1	40	95.75	4.17	95	5
Condition 2	40	94.13	5.17	95	10
Condition 3	40	95.38	6.03	97.5	10
Condition 4	40	94.25	5.13	95	6.25

Table 7.15. Mean, Standard Deviation (SD), Median, and Interquartile Range (IQR) of RMSE on Point Estimation Tasks of 20 Sighted Participants across All Conditions

To examine whether the performances of point estimation tasks among sighted participants were different across four conditions, we plotted the percentage of correct polarity sign values into boxplots and mean plots, as displayed in Figure 7-15.

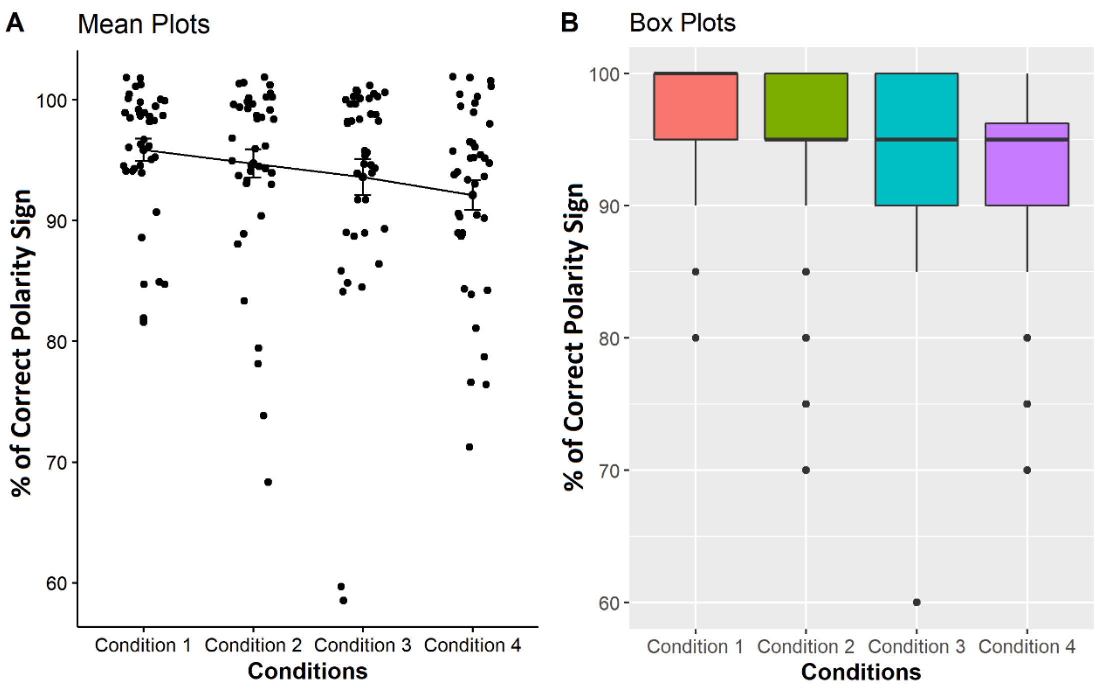


Figure 7-15. Mean Plots (a) and Boxplot (b) of 20 Sighted Participants, representing the Distribution of RMSEs Obtained from Conditions 1 to 4.

The sighted participants performed the polarity sign task better with the sonification mappings used in conditions 1 and 2 compared with those using the multi-reference sonification mappings for condition step20 and 10. The higher percentage of correct polarity sign estimates in conditions 1 and 2 were supported by higher mean, median and lower error distributions in the boxplot, as shown in Figure 7-15. Furthermore, the mean plot showed that each condition tended to decrease from condition 1 to condition 4.

After checking the normality assumption by means of a Shapiro-Wilk test, all conditions were not normally distributed: condition 1 ($W = 0.82$, $p < 0.001$), condition 2 ($W = 0.87$, $p < 0.001$), condition 3 ($W = 0.76$, $p < 0.001$), and condition 4 ($W = 0.81$, $p < 0.001$).

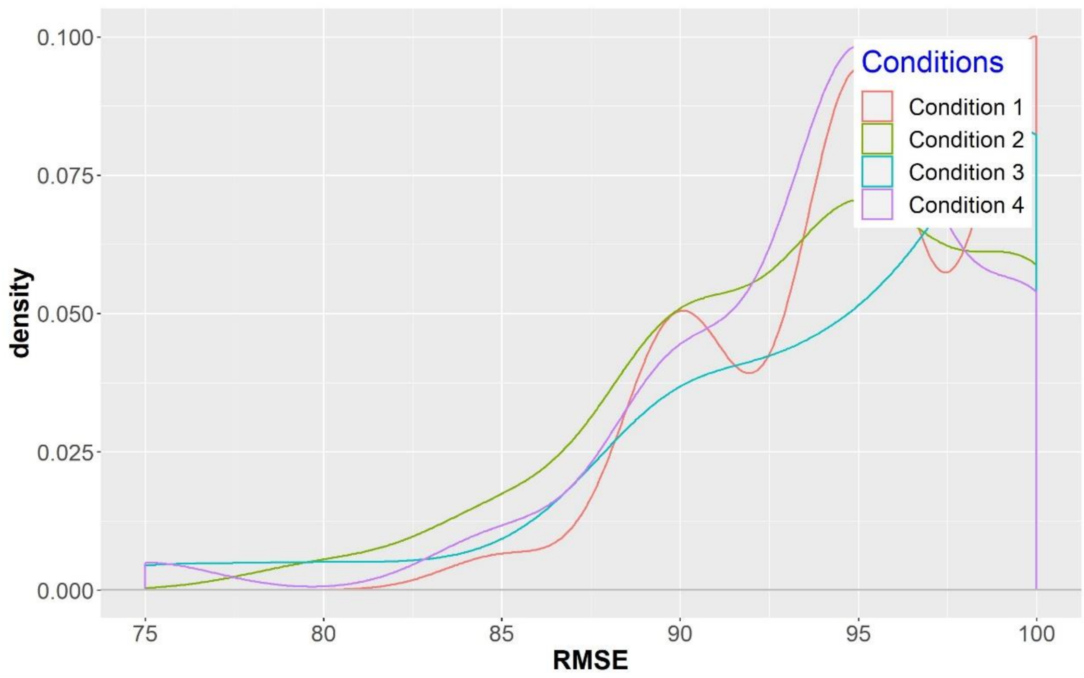


Figure 7-16 Histograms of the Percentage of Correct Polarity Sign estimates of 20 Sighted Participants across All Conditions.

The histograms of all conditions were skewed left, showing non-normal distributions, as shown in Figure 7-16. Therefore, we performed Kruskal-Wallis non-parametric tests to investigate the difference in performance in the polarity sign estimation tasks between the four conditions. As the *p*-value was larger than 0.05, we could conclude no significant difference between the four conditions (Kruskal-Wallis chi-squared = 3.942, p = 0.27).

7.6.3. Results for VI Participants vs. Sighted Participants

7.6.3.1. Point Estimation Tasks

To evaluate the performance difference in point estimation tasks between sighted and VI participants, we aggregated all data across all four conditions for each group of participants, as summarized in Table 7.16.

Conditions	Count	Mean	SD	Median	IQR
Sighted Participant	80	22.68	10.86	20.08	14.98
VI Participant	80	18.51	12.89	15.62	11.36

Table 7.16. Comparison of Mean, Standard Deviation (SD), Median, and Interquartile Range (IQR) of RMSE after Point estimation Data Aggregated for Sighted and VI Participants.

The distribution of RMSEs for each group was transformed into boxplots and mean plots, as shown in Figure 7-17.

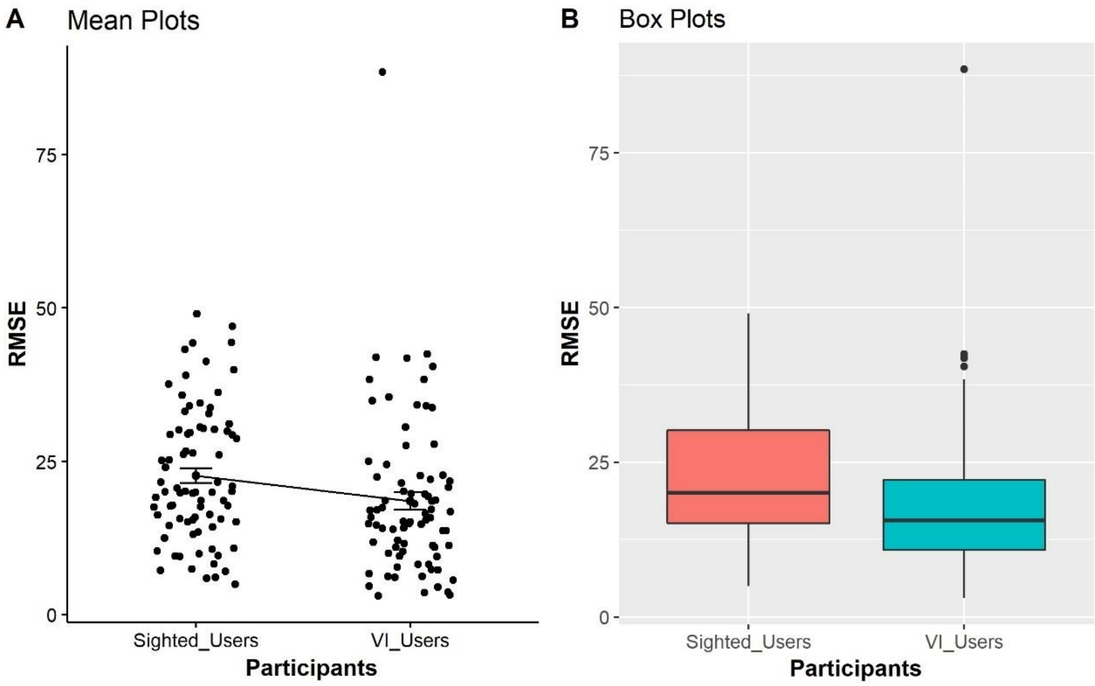


Figure 7-17. Comparison of RMSE Means after Data Aggregation between Sighted and VI Participants.

The descriptive statistics revealed that the VI participants made fewer errors on the point estimation tasks as represented by their lower mean and median values compared with sighted participants. It was also worth noting that the standard deviation (SD) of the VI participants were higher than those of sighted participants, indicating their point estimation errors are more spread out from the mean value (see also Figure 7-17)

Furthermore, it was important to compare the performance between sighted and VI participants to investigate the performance differences between the two groups under different conditions. Examining the descriptive statistics for VI participants (Table 7.10) and sighted participants (Table 7.12), the SD of sighted participants (10.86) is smaller than the SD of VI participants (12.89). While for the IQR, the sighted participant (14.98) is bigger than the IQR of VI participants (11.36). Therefore, the mean RMSE for sighted participants is still bigger than VI participants, as expected.

The SD of VI participants was bigger than sighted participants because we have several long outliers for the VI participant results. The existence of outliers implicitly shows the diversity of the abilities of the VI participants.

We did a small study with a statistical calculation to remove the outlier, but the data remains non-normally distributed. It is difficult to control for outliers in the population due to the difficulty of recruiting VI participants. Therefore, we choose the non-parametric test for analysis.

We should choose the "preferred" measures of center and spread when outliers are present in a set of data. This is because our central tendency measure is higher than most of the data points (RMSE) and because our data is skewed. Using the Mean makes more sense if we have a symmetric data set and the data distribution isn't substantially skewed in one direction. So, in our case here, the median is a much better measure of central tendency. Two VI participants are active in music and in playing musical instruments, three had informal training, but the rest are not, as shown in Figure 7-4.

Meanwhile, sighted participants had no better musical level than the VI participants, as shown in Figure 7-5. Moreover, half of the VI participants only received education up to the upper secondary level. In contrast to sighted participants, 80% of them can study at a higher level, as shown in the demographic in Table 7.1.

Concerning the spread, the standard deviation is based on the mean value. Since it is based on the mean value, which is not a good measure of central tendency in this situation, this will also skew the SD. The SD is going to be larger than if we look at the actual values. Therefore, the interquartile range (IQR) is more appropriate, especially when the data is skewed in one direction.

Before proceeding to further statistical analysis, we checked the normality assumption of the data by observing the histogram in Figure 7-18. The RMSE distributions are skewed right, indicating a non-normal distribution due to several outliers, particularly in the VI participants' data.

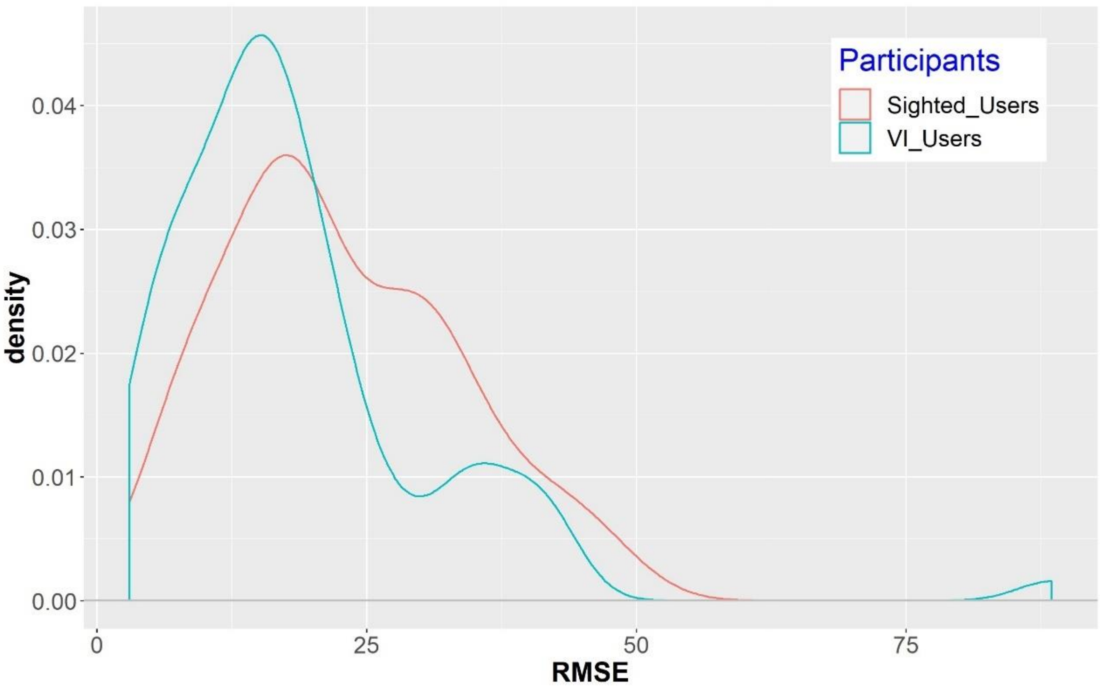


Figure 7-18. Histogram of the RMSE Means for Sighted and VI Participants.

We then performed a Mann-Whitney non-parametric tests to examine whether the RMSE means were different between VI and sighted participants. The result showed that the performance difference on point estimation tasks between sighted and VI participants were statistically significant (W = 25600, $p < 0.001$).

7.6.3.2. Representation of Negative Numbers

After calculating the number of correct polarity sign estimates across all group participants, the values were plotted into two boxplots and mean plots to visualise the distribution, as shown in Figure 7-19.

By examining the boxplots and mean plots, both groups apparently had no performance differences on the correct polarity sign task represented by the means, medians, and data distributions.

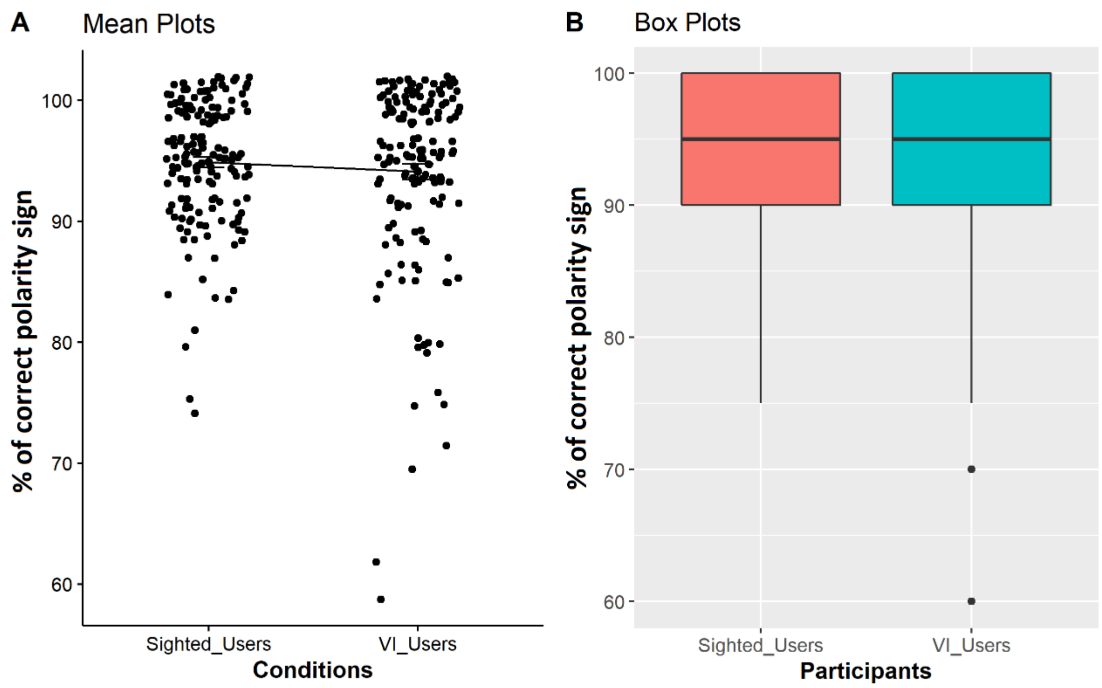


Figure 7-19. Comparison of Percentage of Correct Polarity Sign estimates after Data Aggregation for Sighted and VI Participants.

7.7. Discussion

7.7.1. Analysis of the Point Estimation Tasks

Our first research question on the point estimation performance task across four conditions was directed to address the following hypotheses:

H1: Participants will make significantly more point estimation errors when using the ***single point*** and ***single reference sonification*** mappings as compared with the one-reference and the multi-references mappings using 20 steps and 10 steps

H2: Participants will make significantly more point estimation errors when using the ***single reference*** sonification mapping as compared with the ***multi-references mappings of 20 steps*** and ***10 steps***.

H3: Participants will make significantly more point estimation errors when using the ***multi-references mapping of 20 steps*** as compared to ***the multi-references mapping of 10 steps***

In general, the VI participants produced more accurate results or fewer errors in conditions 3 and 4 which used the multi-reference sonification mappings. This was indicated by its respective lower mean, median and narrower distribution compared with conditions 1 and 2 as shown in Figure 7-8. Nevertheless, further post-hoc analysis revealed that only condition 1 vs condition 4 ($p = 0.006$) and condition 2 vs condition 4 ($p = 0.0064$) differed significantly. The other pairs were not found to be significantly different ($p > 0.05$) (see Table 7.11).

The results for sighted participants showed even greater differences when compared with VI participants. The accuracy of the multi-reference modes at condition 3 (Median = 16.98) and condition 4 (Median = 10.15) were much smaller, being almost less than or equal to half of the RMSE means for condition 1 (Median = 30.01) and condition 2 (Median = 30.28) (see Table 7.13). Their respective boxplots also showed narrower distributions, as depicted in Figure 7-13. This finding was supported by the results of pairwise comparisons which revealed that all pairs were significantly different except between condition 1 vs. condition 2 (p= 0.82) (see Table 7.14).

Therefore, as predicted in hypotheses one and two, both VI and sighted participants produced higher point estimation errors when using sonification mappings with fewer reference tones (conditions 1 and 2) compared to the multi-reference mappings of 20 steps and 10 steps. These results align with Metatla's finding (Metatla et al., 2016) who had similar results using narrower scales than the one used in our experiment. It suggested that the multi-reference approaches provide more "anchor" points or more guidance to assist in the point estimation tasks.

Furthermore, hypothesis five predicted that VI participants would have better performance on point estimation than the sighted group. As displayed in Figure 7-17, the VI participants performed more accurately on point estimation tasks, represented by a lower mean and median of RMSE values than sighted participants. The result of the Mann-Whitney Test (1947) statistics also showed that the differences were statistically significant (W = 25600, $p < 0.001$).

Condition	Count	Mean (μ)		SD		Median ($\tilde{x}$)		IQR	
		VI	Sighted	VI	Sighted	VI	Sighted	VI	Sighted
1	20	22.96	30.24	11.73	9.86	18.31	30.01	19.25	10.46
2	20	21.51	30.67	10.15	7.95	20.80	30.28	14.18	11.62
3	20	17.90	17.44	17.41	4.76	15.02	16.98	7.55	4.90
4	20	11.66	12.39	8.18	6.31	10.24	10.15	10.81	9.21

Table 7.17 . Mean, Standard Deviation (SD), Median, and Interquartile Range (IQR) of RMSE values on Point Estimation Tasks of VI vs Sighted Participants across All Conditions

However, when we compared data for each condition separately (see Table 7.17), it seems that the RMSE values of conditions 3 and 4 for both groups of participants were not different. Meanwhile, in conditions 1 and 2, sighted participants made twice the number of errors than

VI participants. This shows that VI participants performed more accurately on point estimation tasks with fewer references (i.e., conditions 1 and 2).

When the reference was added for every multiple of 20 or 10, both groups showed somewhat equal performance ($\tilde{x}_{cond.3}$ = 15.02, $\tilde{x}_{cond.4}$ = 10.24 for VI; and $\tilde{x}_{cond.3}$ = 16.98, $\tilde{x}_{cond.4}$ = 10.15 for sighted participant, respectively). So, both VI and sighted groups had similar performance in the multi-reference conditions. Both groups performed better in condition 4 (step10 sonification mapping) than condition 3 (step20 sonification mapping). These results tend to show that all participants seemed to perform better as the number of reference points increased, in this case, when the range from 0 to $Y_{Estimate}$ was split into 10^{ths} rather than 5^{ths}.

Interestingly, by comparing the standard deviation and IQR of each condition, we observed that sighted participants' data variability was far lower than VI participants across all conditions.

The participants' demographic characteristics (see sub-section 7.4) revealed that several VI participants had a better level of musical training than the other VI participants, as discussed previously in sub-section 7.6.3.1. Our data analysis revealed that the more varied data and outliers came from VI participants with less musical training. Meanwhile, our sighted participants generally had no better musical levels than VI participants whose musical level distribution was more varied. Further studies are needed to investigate the possible relationship between *musicianship level* for effective point estimation in this context.

We felt the log / exponential mapping made a noticeable difference to the accuracy of point estimates in this 4th study, compared to the accuracy of point estimates in the three previous studies. While it didn't make a big difference for the result, this mapping solution minimizes the usability problem that many participants had complained about in the previous study, as discussed in section 6.7.3.

We also found no issue in this study as that happens in different timbral streams issue in study 3. In this issue, the participant mistakenly guesses the coin as piano and vice versa, resulting, for example, 70 turns to 20 and vice versa. In other words, whether $Y_{Estimate}$ is greater than or equal to half of Y_{Max} ($Y_{Estimate} >= 0.5 * Y_{Max}$), it could mistakenly perceive as below half of Y_{Max} ($Y_{Estimate} < 0.5 * Y_{Max}$) and vice versa. While it rarely happens in study 3, the solution to use only one type of sound, however, minimizes this potential issue.

7.7.2. Analysis of Negative Number Reference

Regarding negative number reference, our hypothesis is stated as follow:

H4: Participants will make significantly better polarity sign selections when using the ***single point*** and ***single reference*** sonification mappings as compared with the one-reference and the multi-references mappings of 20 steps and 10 steps.

The results of the polarity sign task revealed no significant difference across all conditions both in the VI participants (Kruskal-Wallis chi-squared = 7.5951, *p*= 0.051) and sighted group ((Kruskal-Wallis chi-squared = 3.942, *P*-value = 0.267). Table 7.18 summarized the descriptive statistics of the percentage of correct polarity sign. Despite the conditions used, both VI and sighted participants demonstrated high accuracy with all mean values and medians being above 90%.

Condition	Trial	Mean (μ)		SD		Median ($\tilde{x}$)		IQR	
		BVI	Sighted	BVI	Sighted	BVI	Sighted	BVI	Sighted
1	40	95.87	95.75	5.76	4.17	100	95	5	5
2	40	94.75	94.13	7.33	5.17	95	95	5	10
3	40	93.62	95.38	9.47	6.03	95	97.5	10	10
4	40	92.12	94.25	7.67	5.13	95	95	6.25	6.25

Table 7.18. Mean, Standard Deviation (SD), Median, and Interquartile Range (IQR) of RMSE on Polarity Sign Task of VI vs. Sighted Participants across All Conditions.

Metatla's (2016) used a *polarity-based approach* using exponential function -instead of linear mapping- from the smallest negative number to the largest positive number.

In this research, however, we attempted to validate that **negative numbers are represented mentally in the form of component representation values, not a holistic representation**, as discussed in section 2.5.3 and 7.2.1.

According to Kong et. al. (2012):

> *"negative numbers in the auditory modality are generated from the set of positive numbers, that the absence of a semantic congruity effect for negative numbers was the result of the minus polarity sign being perceived as indicating "small" and the number itself as indicating "large", and that they offset each other. Thus, supporting a components representation."*

This is our main reason to implement the *component representation* approach for negative numbers using the same positive mapping reference for the digit -as described in Table 3.1- and adding a sign before the digit with a "sonar" sound.

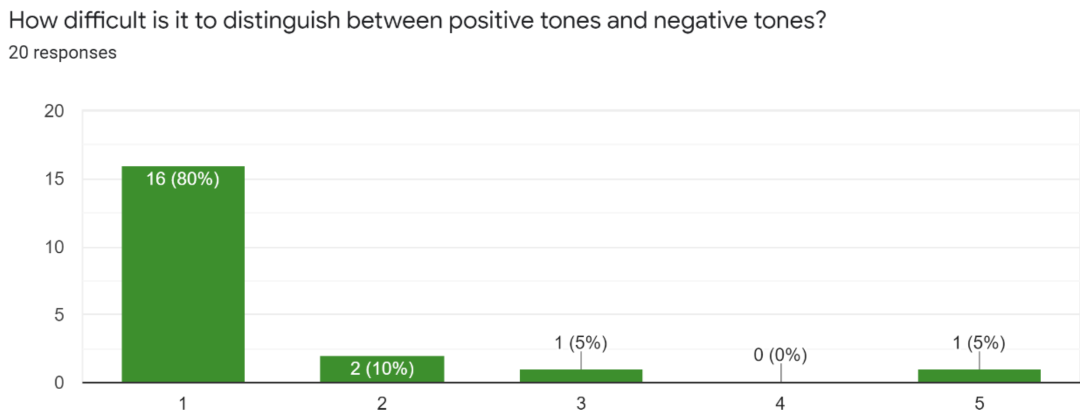
Unfortunately, we cannot compare our results with Metatla because he did not describe the polarity data results in his paper. However, it is clear concerning a weakness in Metatla's approach that point estimation's task becomes inefficient as the numbers to be represented get further from 0.

7.7.3. Self-Perceived Usability

In this study, we also evaluated the user experience about several issues while performing the point estimation and polarity sign tasks across all conditions. To examine the perceived ease of use of the respective sonification conditions, a questionnaire was developed using a five-point Likert scale that ranged from 'very easy' (=1) to 'very difficult' (=5). We aimed to obtain participants' feedback about: (1) the perceived ease of use of the polarity sign estimation approach, (2) Participants' preference between the sonification conditions and (3) the perceived ease of use of each of the individual sonification conditions.

7.7.3.1. Perceived Ease of Use of the Polarity Sign estimation Approach

The polarity sign estimation approach was rated as 1 (very easy) by 80% of VI participants and 60% of sighted participants, as shown in Figure 7-20. Only one VI participant and none of the sighted participants perceived it was very difficult to distinguish between positive and negative tones.



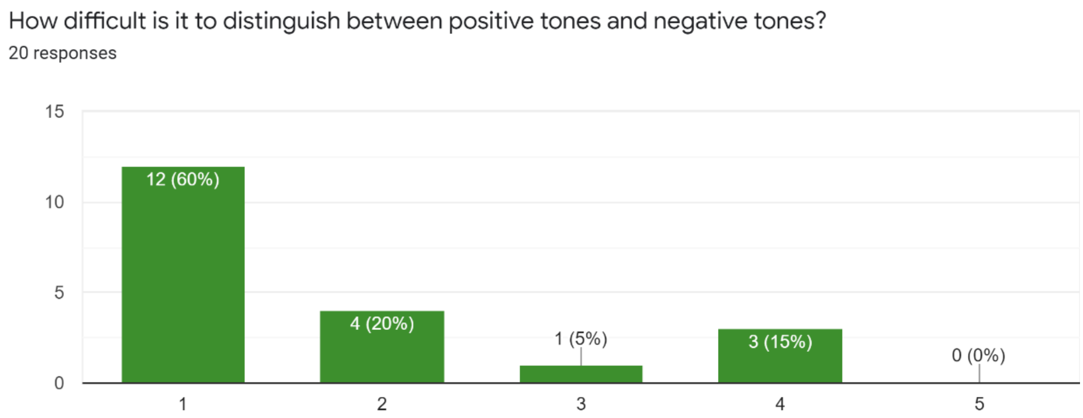
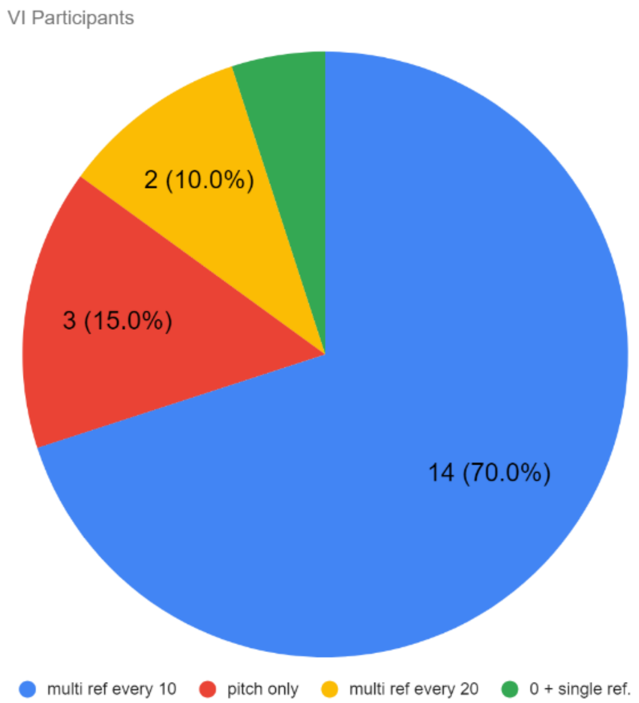


Figure 7-20. Perceived Ease of Use of the Polarity Sign Task for VI Participants (above) and Sighted Participants (bottom). Likert-Scale Ranges from Very Easy (Level 1) to Very Difficult (Level 5)

7.7.3.2. Preference between sonification conditions

Participants were also asked to rank which sonification condition was easiest to use while performing the point estimation task. Figure 7-21 showed that 70% of the VI participants and 65% of the sighted participants perceived that the multi-reference approach using step10 (condition 4) was the easiest condition to use. The single point condition was their second preference, selected by 15% of VI participants and 20% of their sighted peers.



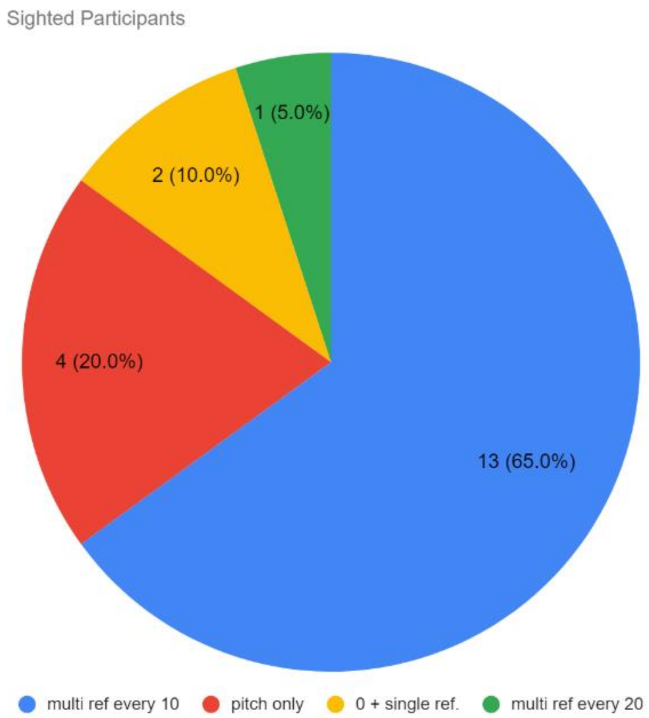
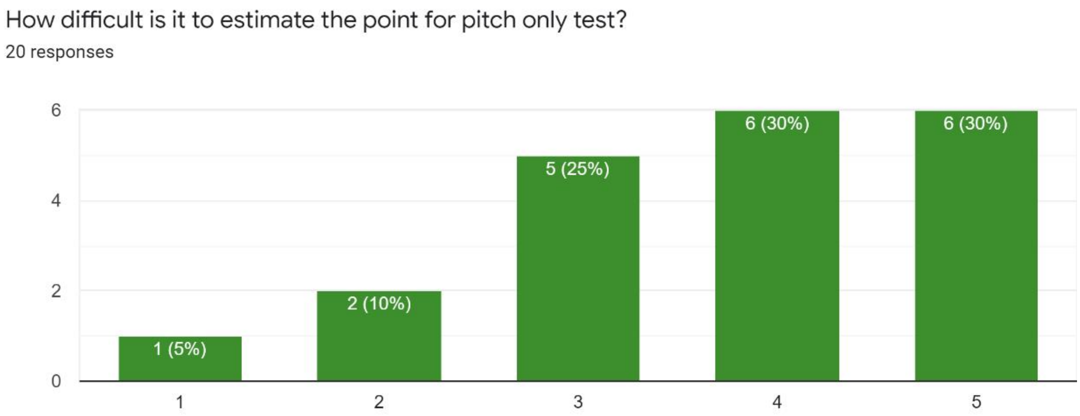


Figure 7-21. Perceived Ease of Use for Each sonification Condition of the VI Participants (above) and the Sighted Participants (Below). Coloring Scheme: Single Point Mode (Red), Single Reference (Green), Multi-reference for Condition Step20 (Yellow), Multi-reference for Condition step10 (Blue).

It is interesting to note that condition 3 (multi-reference step20) was the least preferred approach for both groups. They considered this condition was the most challenging point estimation, see Figure 7-24. This might be explained by the fact that when compared with condition 4, multiples of 10 are more natural to use than multiples of 20. These findings are in spite of the fact that step20 (condition 3) results were generally better than the results for conditions 1 and 2.

7.7.3.3. Perceived Ease of Use of the Single Point Condition



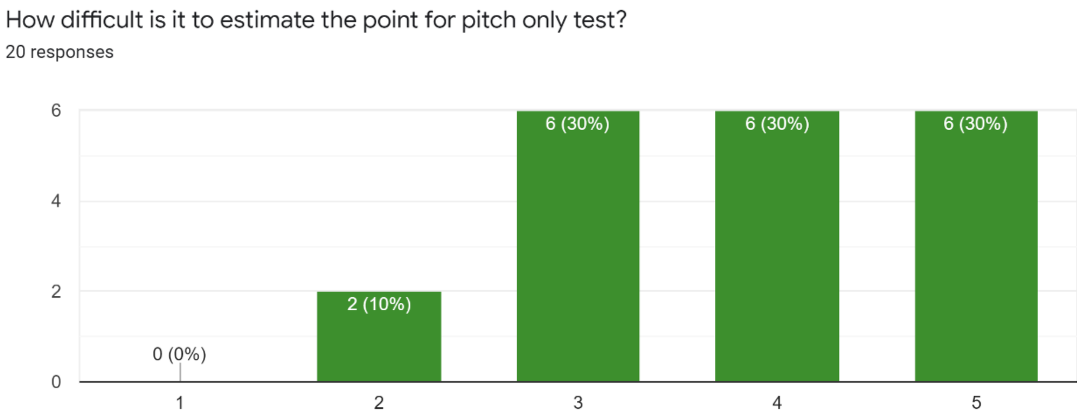
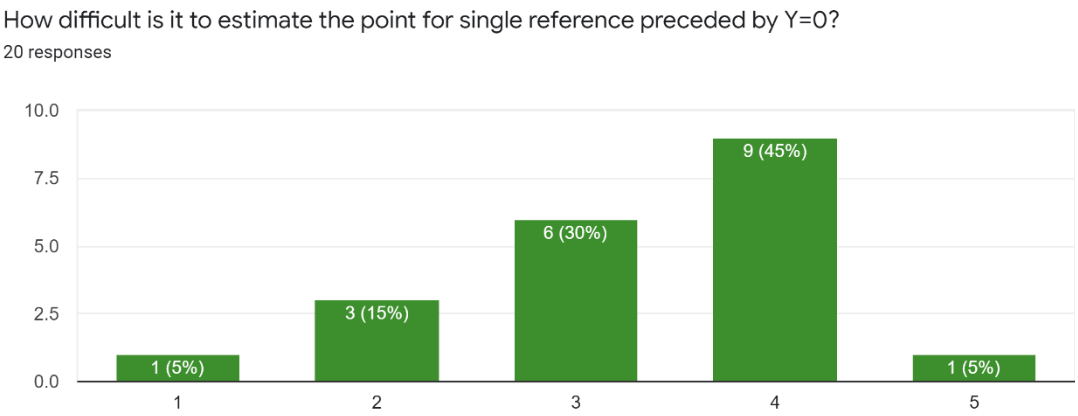


Figure 7-22. Perceived Ease of Use for the Single Point (pitch only) Condition among VI Participants (above) and Sighted Participants (below)

Both VI and sighted participants rated the difficulty of using the single point condition in the range 3 to 5 equal to 85% to 90% of participants, as shown in Figure 7-22. This indicates that it was relatively difficult to perform point estimations in this condition. Only 2 to 3 (10% - 15%) participants rate single point condition are relatively "easy".

While the approach is very immediate and intuitive, it probably imposes the most cognitive load.

7.7.3.4. Perceived Ease of Use of Zero as Single Reference Condition



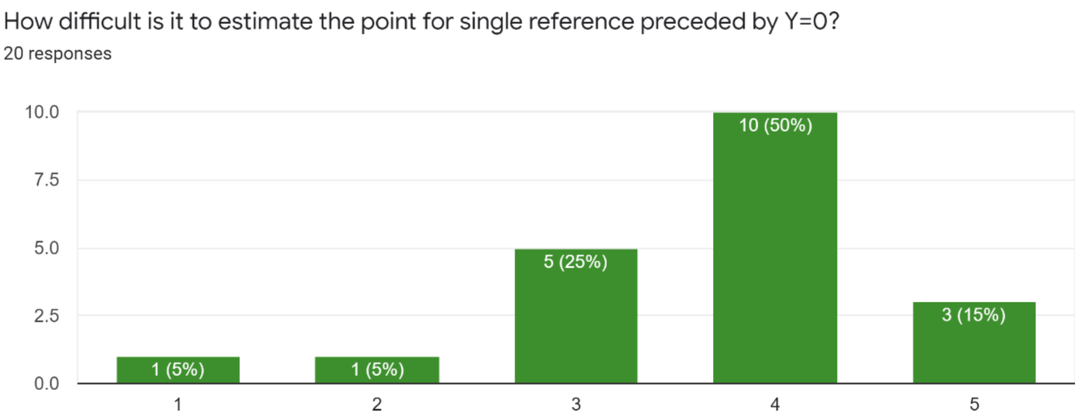
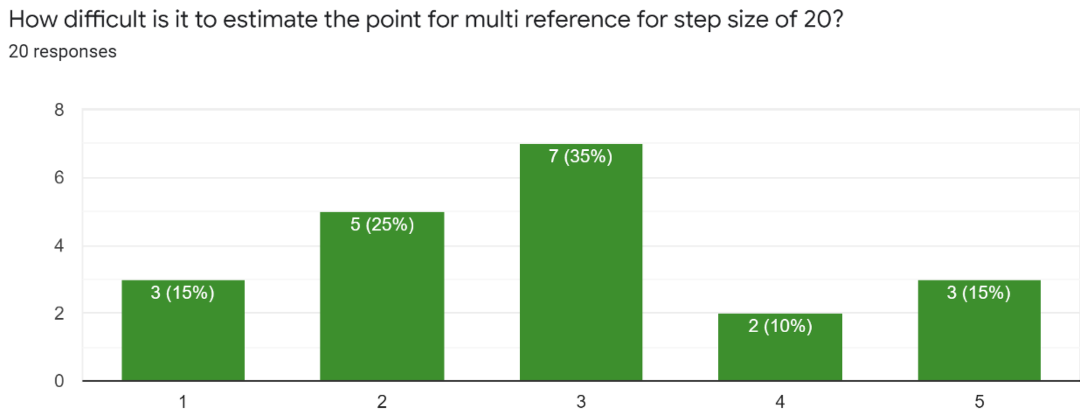


Figure 7-23. Perceived Ease of Use for the Single Reference Y=0 Condition for VI Participants (above) and Sighted Participants (below)

On average, both VI and sighted participants considered the single reference Y = 0 condition to be difficult to perform, rating the condition 2 to the levels 3 (30% and 25%, respectively), level 4 (45% and 50%) and level 5 (5% and 15%). Only one participant (5%) in each group perceived the condition to be very easy (level 1) and another three VI participants and one sighted participant that rate it in level 2.

7.7.3.5. Perceived Ease of Use of the Multi-reference step20



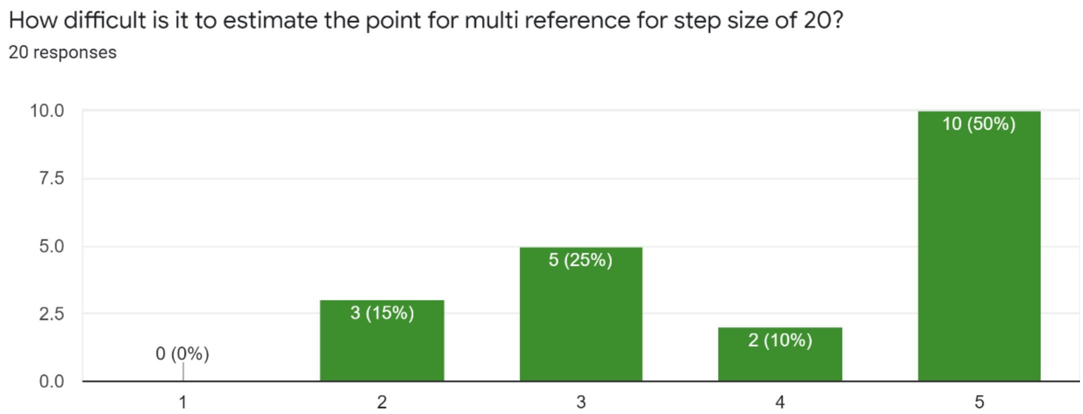
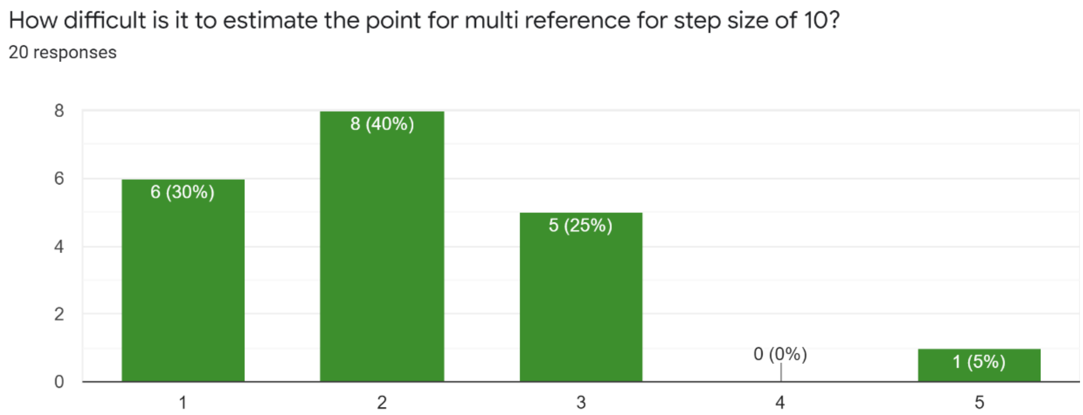


Figure 7-24. Perceived Ease of Use for the Multi-Reference step20 Condition for VI Participants (above) and Sighted Participants (below)

The response of both groups of participants varied concerning condition 3, as shown in Figure 7-24. While 35% of the VI participants rated it neutrally at level 3, 40% of them (15% for level 1 and 25% for level 2) rated it as easy.

On the contrary, a substantial proportion of the sighted participants (50%) rated this condition at level 5 and 10% at level 4, indicating the condition was very difficult. None of them giving a rating of level 1 and only 15% rating it at level 2. A longer and less intuitive approach to the other condition, it probably represents the greatest cognitive load.

7.7.3.6. Perceived Ease of Use of Multi-reference for Condition Step10



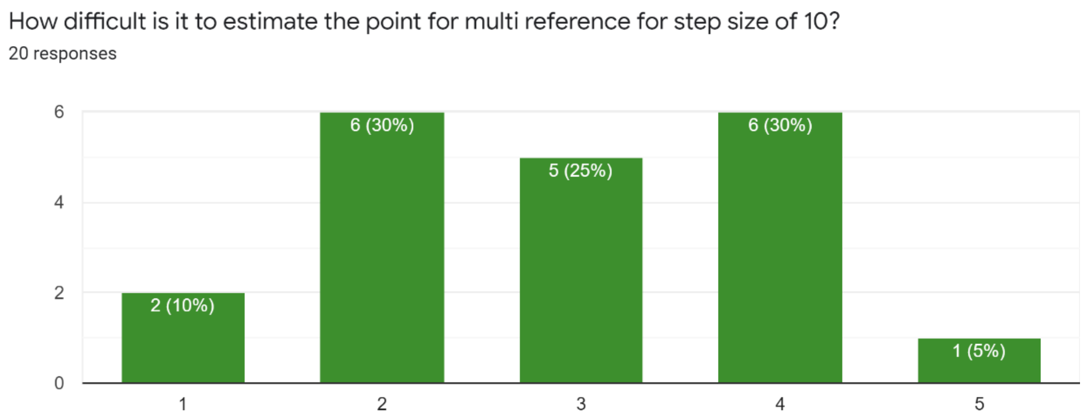


Figure 7-25. Perceived Ease of Use of the Multi-Reference step10 Condition by VI Participants (above) and Sighted Participants (below)

Both groups agreed that the point estimation task using condition 4 was relatively easy compared to the other three conditions. While most of the VI participants rated it in the range

1 to 3 (30%, 40%, and 50%, respectively), the sighted participants rate it mostly in level 2 to 4 (30%, 25%, and 30%, respectively).

This suggests that the VI participants got the biggest advantage of using the multi-reference for condition step10. Moreover, these results are consistent with their response regarding the conditions most preferred for perceived ease of use.

In general, all of these findings suggest that not only were the VI participants better at point estimation using sonification, as discussed in section 7.6.1. They also were more comfortable with the process of using sonification for the point estimation task. Further, their generally lower rating scores across all conditions suggest adapting to the different sonification conditions easier than their sighted counterparts.

7.8. Conclusion

In this study, we tested modified multi-reference sonification schemes for non-visual point estimation for larger scales with fewer references following the work of Metatla (2016). In general, the experiment results show that the step10 condition proved more effective and was preferred to other conditions. The study further showed that VI participants performed more accurately on point estimation tasks with fewer references than their sighted peers. Both VI and sighted groups had similar performance and seemed to perform better as the number of reference points increased, in this case, when the range from 0 to $Y_{Estimate}$ was split into 10^{ths} rather than 5^{ths}.

This work implements the component representation approach for note of negative numbers to represent the mapping using the same positive mapping reference for the digit and adding a sign before the digit. The polarity sign task results revealed no significant difference across all conditions, both in the VI participants and the sighted group. Therefore, this approach is more efficient than Metatla's approach for numbers further from 0.

In summary, the above results contribute to the study on non-visual interaction by extending relevant research which is also scalable in the sense that there will never be more than 10^{ths} reference tones no matter what the numbers are on the Y-axis with graphs.

Chapter 8. Conclusion

8.1. Overview of the thesis

This work examines the performance of point estimation and graph reproduction tasks by sighted and VI users employing auditory graphs and multi-touch gestures. It offers a framework related to improving performance in point estimation and graph reproduction tasks. The introductory chapter motivates the work, defines the objectives, the coverage and main research questions. The following chapter addresses related research covering a range of topics from available guidelines for designing auditory displays to HCI methodologies. The concept of multi-reference sonifications and the representation of negative numbers are explored.

Chapter 2 aimed to examine the existing practice in the field of auditory display design and provides a literature study of relevant research.

Chapter 3 discusses the choice of research methods employed in each study, the selection of participants, and ethical considerations. It details the changes made to each successive version of the Mobile Audio Graphs (MAG) prototype.

Chapter 4 presents an exploratory study to test the usability of the MAG app for sighted users. This chapter describes our investigation into whether the complexity of audiographs affects the ability of sighted users to perform charting tasks.

Chapter 5 presents an exploratory study to test the usability of the MAG App for VI users. The investigation in this chapter involves studying the impact of adding complexity or additional modalities to the auditory graphs to improve the performance of point estimation and graph reproduction tasks.

Chapter 6 describes the evaluation of the MAG app for the multi-reference sonification point estimation task and relevant problems to be solved in this multi-reference sonification setting. The evaluation is performed by comparing the size of errors made and task completion times with those for the single point sonification condition.

Chapter 7 describes an in-depth analysis of the evaluation of multiple references for different conditions and the presentation of negative numbers.

Sighted and VI users contributed to this research and we gratefully acknowledge their assistance in being able to undertake these studies.

This work is aimed at three target groups: the sonification community, the wider community of HCI researchers and professionals, and the visually impaired community. Each of these disciplines is contributed to by the work in various ways. To the best of the knowledge we have, this is the first in-depth study of the design practice in the area of auditory graph presentation on mobile devices. The mobile auditory graph was specifically designed to give these communities the opportunity to create a common base of design knowledge that can be used to leverage previous work. In the long term, this thesis is intended to have an implication on the sharing of best practices and make effective mobile auditory graphing more widely used in everyday technology.

The rest of this chapter is structured as follows. In section 8.2, we revisit the three main research questions posed in chapter 1. We then go on to describe the main contributions of this work. Section 8.5 concludes the chapter with some reflections and some ideas and perspectives for future work.

8.2. Research questions and Contributions

8.2.1. Research questions

In this subsection, we shall return to each of the three main research questions described in chapter 1, drawing conclusions based on the results of the work achieved.

Research question 1) How accurately can visually impaired users estimate the values of data points rendered in auditory graphs presented on a mobile device?

The studies in this thesis yield the unsurprising result that the answer to this question is not simple. We have seen that it is influenced at least by individual differences, the complexity of the auditory graphs involved and what other interaction modes might be available. We have gathered some evidence that in some circumstances, the accuracy of VI users can surpass that of sighted users. This seems to be particularly the case when there is sparser (less contextual) information available in the sonification approach used to render the auditory graphs (see contribution 8 in section 8.2.3). Similarly, the ability of VI users appears to be more robust to changes in the sonification approach employed (contribution 8).

Research question 2) Are there modes of interaction which can improve the ability of visually impaired people to perform point estimation tasks presented on a mobile device?

Study 2 in chapter 5 yields clear evidence that adding multi-touch gestures as a means of interacting with the sonified graphs improves the accuracy of point estimation and graph reproduction tasks. The chapter describes how the introduction of multi-touch gestures in this situation can be conceived as closing a feedback loop involving the user, the multi-touch interaction, the device and the perceived sonification. Participants report the additional interaction mode as making the task more interesting. The improvement in the accuracy of the results obtained is clear for point estimation tasks (contribution 4) and graph reproduction tasks (contribution 5).

In many ways, these findings might be considered the most exciting of those reported here because of the potential to further build on them. As haptic displays improve and the technology concerned enables a wider range of interaction possibilities, the potential for more sophisticated forms of multi-touch interaction will increase (see suggestions for further work in section 8.5).

Research question 3) What format should the auditory display take to enable accurate understanding and efficient processing of auditory graphs?

In a single person project of this length, the range of sonification possibilities that can be explored is only a fraction of the possible options available. For example, we have limited ourselves only to parameter mapping approaches, these being the most widely reported in the auditory display literature over the last 35 years.

Two of the key attributes of auditory displays we reported in chapter 2 are that they are transitory and need time to be rendered. In a sense, these qualities are conflicting, because the longer an auditory display takes to be rendered, the harder it is likely to remember the parts of the display rendered earlier in the temporal sequence. Techniques might be employed to overcome this, such as continuing to render those parts of the auditory display rendered early, but perhaps at a steadily decreasing amplitude, signifying their historical property. Still, such approaches run the risk of creating a display which is cognitively demanding to interpret, and in the case of repeating values, may have severe problems due to masking.

For these reasons, in the multi-reference sonification approach described in chapter 6, we aimed to put forward an approach that will not require more time to render each point as their values get further from zero.

The initial evaluation of the multi-sonification approach we propose showed that it induced smaller errors and had similar task completion times compared with the single point approach (contribution 6). Furthermore, the two variants of the approach evaluated in chapter 7, as conditions 3 and 4, performed better for both sighted and VI populations than the simpler sonification approach evaluated as conditions 1 and 2 (contribution 7).

Of course, the point could be made that the sonification will take longer to render employing our proposed approach as the number of points in the auditory graph increases, but this would be the case for any of the sonification schemes evaluated in the present work. In a sense, that is a different problem, though under some circumstances, for example, in a situation where few or no abrupt turning points occur in a graph, a similar multi-sonification approach might be employed to render an entire graph, sampling it at 10^{th} or 20^{ths} of its entire length. Clearly as well, in such an approach, each sampled value might be weighted in some way to take into account the values of points in its vicinity.

8.2.2. Contributions

In this section, we shall discuss the original contributions to research that have resulted from this work. While these studies have led to several results in answer to the 3 posed research questions, they have also left others open and posed new questions. As well as reporting contributions, in the following section, we will try to identify several important questions that have arisen from this work. The set of contributions of the work are summarised together in section 8.2.3. Section 8.5 considers future lines of research.

8.2.2.1. Study 1: Exploratory

The experiment tried to investigate the research questions on how well sighted users can estimate points on the Y-axis in a graph along the X-axis? We aim to evaluate how well the sighted users can estimate points for varied numbers of data points by calculating the root mean squared errors (RMSE) between the estimated (predicted) values to the true values. We also aim to evaluate how well the sighted users can perform graph reproduction tasks by calculating the correlation test between the estimated values to the true values. In addition, we want to assess the performance of point estimation and graph reproduction tasks on the auditory graphs, whether they have similar or worsened performances between simple, medium, and complex graphs. From the results of this exploratory study, it can be concluded that the MAG app helped the sighted users obtain more information better to identify the auditory graph shape in a mobile device. Based on this finding, we assume that further experiments with VI people would be possible and expected to have a better result. There are limitations to these findings, however. The results may differ if the order of conditions was randomized instead of performed as a sequence of simple, medium, and complex conditions as used in this experiment. In particular, the study had a maximum of 36 data points, thus making the activity of graph reproduction tasks was very difficult to do without multiple repetitions for this large number of points. For further study, we limit the trial of playback to a maximum of three times to anticipate the possible learning effect. As compensation, we need to reduce the number of data points as Metatla (2016) found that if a lot of notes were played, people lost track of them, and they became less useful.

8.2.2.2. Study 2: Going Multi-Modal

Part of our goal was to understand how VI participants would react to another form of interaction with their smartphones, in which gestures were performed on the touch screen of their smartphones. We introduce a new multimodal approach based on multi-touch gesture interaction, aiming to have a more accurate mental model of the plots and improve smartphone user interfaces' accessibility, as discussed below. The main goal of this thesis was to ensure that auditory display design can be incorporated into the overall domain of interaction design. From our point of view, this issue is critical for the auditory display to fit into the broader HCI design space. The mobile auditory graph has been designed with this aspect in focus. Although we have developed all concepts and methods to facilitate auditory display design, they are open for enhancements and can be flexibly used in a wider context. The experiment in this study tried to investigate the user perceptual skills, listening experience, and any mental model of how the data values with multiple points on the Y-axis can vary and how changes in sound have used to represent changes in data. On MAG app 2.0, an additional modality using swipe interaction was implemented to help the users to locate the points on the X-Y coordinates.

The results of study 2 for VI participants showed that the point estimation performance for the multi-touch gesture generated more accurate results in which **there are no significant differences between the conditions for the multiple number of notes.** Further, having additional modalities could improve the performances as there are no significant differences between the conditions since users can plot shapes of graphs with higher correlations.

However, the results for passive listening interaction showed a far greater difference in which the estimation performances are worsened by higher error if the number of notes was increasing. Under the categories of medium and complex graphs, adding complexity to the graphs also worsened their correlation performances (rho < 0.7) in passive listening. The research question of this study has been answered that the users produce better point estimation performance and graph reproduction performance when using additional modalities compared to listen to the audio passively. This is aligned with Walker et.al. (2005) suggestion that is adding modalities into auditory graphs resulting in a better understanding of quantitative information. Nikitenko (2014), in his work on sonification on mobile touchscreen devices, suggests that audio playback and user interaction combined procedures offer an advantage over procedures that rely solely on audio. Such a passive listening approach can only stand for one interpretation, which may be insufficient, incomprehensible, incomplete, or even wrong. Researchers have argued that additional modality alternatives to the interface could be implemented to overcome this issue (Bornschein et al., 2015; McDonald et al., 2014).

The above results support the investigation of non-visual interaction with graphs by adding active rather than passive point estimation tasks to the relevant research. However, these findings are limited. The results may vary if the sonification of notes is applied to support the two-dimensional motion. In addition, it would be interesting to investigate several kinds of sonification, using a musical scale, and to compare them to linear sonification in the present experiment.

8.2.2.3. Study 3: Multi-reference in auditory graph

The review of design practice both inside and outside the community has revealed that it is still challenging to produce accurate point estimation in auditory graphs. Among the elements detected are:

- Gaps identified in the documentation of research studies in this area, especially with regard to the rationale of design decisions,
- The innovative and cross-disciplinary nature of the process

We tried to address the question in this exploratory study on how adding multi-reference sonification mapping in auditory graphs could improve the performance of non-visual point estimation tasks. The first research question of this study has been answered for this population that the users produce higher point estimation errors when using the single point sonification mapping compared to the multi-references sonification mappings. Concerning the duration of the trial, the users' opinions were divided between those who considered the single point mode to be faster and those who had the opposite opinion.

In general, most participants considered the multi-reference was faster for point estimation trial rather than single point. Their feedbacks were in the opposite, as stated in the literature, as usually the completion time to conduct the multi-reference mapping is longer. However, the single point mode could be the most demanding, not of course due to the number of points involved, but because it is simply a mentally difficult task to perform, to the extent that most participants believed that the trial estimation time to complete the tasks was shorter in the multi-reference mode. In conclusion, most of the participants responded that multi-references were faster in the estimation trial, although the statistical test calculating the completion time between the two modalities confirmed that the difference was not significant.

In general, the results of the experiment show that the multi-reference mode generated more accurate results compared to the single point modality. The evaluation confirms previous researches that adding context to auditory graphs such as tick marks could enhance the perception of auditory graphs.

Due to the small number of participants, the quantitative results of this study have limited scope for generalization. However, the qualitative results of this study are relevant as they provide initial guidance for the implementation and future development of applications based on multi-reference marks, especially for visually impaired users.

8.2.2.4. Study 4: comparison of 4 sonification schemes and representation of negative numbers

The work has demonstrated a representation of negative numbers for non-visual point estimation tasks using another form of sonification by integrating multiple tones as references to represent a note, following on from study 3. The study in this chapter investigates whether employing multiple tones combining with an audio component as a representation of negative numbers can assist point estimation tasks for gaining a better perception and interpretation in auditory graphs. The prototype, therefore, has four conditions, i.e., the single point (condition 1), single reference preceded by zero pitch (condition 2), multiple references with step size of 20 (condition 3), and multiple references with step size of 10 (condition 4).

The results of this study 4 for VI participants showed that the point estimation task **for condition 1 vs. condition 4 (**$p = 6.10 \times 10^{-3}$**) and condition 2 vs. condition 4 (**$p = 6.40 \times 10^{-3}$**) are significantly different**. In contrast, the result for sighted participants showed that only **the condition 1 vs. condition 2 (p = 0.82) that are not significantly different** ($p > 0.05$). Comparing the overall result of VI participants and sighted participants, we found that differences between sighted and VI participants in the point estimation task are significant ($W = 25600$, $p = 2.20 \times 10^{-16}$).

Interestingly, by comparing each condition result of VI participants vs. sighted participants, it can be seen that the condition 1 and 2 from sighted participants that doubled the error in point estimation task. But if the reference is added for every multiple of twenty or multiples of ten, both groups have equal performance. We can conclude that the VI participants have better performances in the point estimation tasks only if the reference note is minimum because of their advantage in melody listening experience (in general, they had a higher level of musical training than their sighted counterparts). While if the reference is added for every multiple of twenty or multiples of ten, both groups have equal performance because they have implemented a similar listening strategy for multi-reference task estimation.

This work also implements the component representation approach for negative numbers to represent the mapping by using the same positive mapping reference for the digit and adding a sign before the digit with a "sonar" sound. This approach leads to better accuracy of the polarity sign. *Our null hypothesis is that the correct polarity sign guesses between conditions are no different*. As the *P*-value for sighted participants and VI participants are more than the significance level 0.05, we can conclude that there are **no differences between the conditions for both types of users.**

In terms of usability, both groups seemed to perform better as the number of reference points increased, in this case, when the range from 0 to $Y_{Estimate}$ was split into 10^{ths} rather than 5^{ths}. These results are consistent with their answers when asked which conditions are most preferred.

In general, VI participants were not only better at point estimation tasks, but they were also easier to adapt to various reference conditions in the auditory graph compared to sighted participants.

8.2.3. Summary of contributions

The main contributions of this thesis can be summarised as follows.

1) An accessible and functional prototype demonstrating multimodal interaction with auditory graphs.
2) The results of study 1 demonstrate that sighted users were able to use this prototype to perform point estimation and graph reproduction tasks with a fair level of accuracy.
3) A new multimodal approach based on multi-touch gesture interaction for performing point estimation and graph reproduction tasks.
4) The multi-gesture interaction approach performed better than passive listening for point estimation tasks.

   The results of study 2 show that the RMSEs of multi-touch gesture interaction were distributed equally for almost all conditions, while the results obtained for passive listening tended to increase. The passive listening interaction resulted in poorer user performance on point estimation tasks as the number of data points increased, which later was confirmed by the respective p-values of the pairwise comparison statistics. Further t-tests also supported that there was a statistically significant difference in point estimation task performance between passive listening and multi-touch gesture modalities. Thus, passive listening interaction produced less accurate estimations in comparison with multi-touch gestures.

   Unlike passive listening mode, which transmits the auditory graph unidirectionally from the device to the user, a key feature of multi-touch interaction is the bi-directional flow of information to and from the user, allowing the user to perceive and actively engage with the system. Touch sensations combined with audio effectively close a feedback control loop between the system and the user, providing cues to the user, enabling them to actively and intuitively control the interaction.

5) The multi-gesture interaction approach performed better than passive listening for graph reproduction tasks.

   We calculated the correlation coefficients between the estimated (predicted) values and true values.

   Participant performance during passive listening varied significantly for some pairs of conditions. Users' performance was worse on medium and complex graphs. In comparison, in the multi-touch condition, performance remained stable as the numbers of data points were varied. In the multi-touch condition, participants' performance was better than that achieved in the passive listening condition, as shown by the fact that their mean correlation values for all conditions were above 0.7 (0.77 – 0.99), indicating a good correlation. In contrast, the mean correlation values for the passive listening condition ranged from 0.3 to 0.89. Furthermore, the t-test statistic showed a statistically significant difference between the two modalities, evidence that the multi-touch gesture interaction provided better performance than passive listening in graph reproduction tasks.

   These results were supported by answers given during the semi-structured interviews that took place at the end of study 2, which confirmed that participants found the multi-touch interaction approach more interesting and engaging.

6) A new approach to multi-reference sonification based on dividing the numeric space between $y = 0$ and Y_{Max} into 10^{ths} to sonify points between $Y = 0$ and $Y_{Estimate}$. The results of study 3 show that participants produced lower RMSE values using the multi-reference sonification than the single point sonification approach.

   A possible problem with the multi-reference sonification approach could be that it might take longer. However, the results of study 3 on task completion times show that this was not the case.

   Metatla et. al. (2016) found that using their approach, there was a compromise between speed and accuracy for multi-reference sonification. The development of the multi-reference point estimation scheme proposed here helps to mitigate that trade-off by requiring fewer reference points to be sonified than in Metatla's (2016) approach. Furthermore, the approach proposed here scales more effectively.

Remembering that the approach proposed by Metatla (2016) sonified every unit of difference between the reference point and the point to be estimated, the results presented here show that users with no previous experience of sonified data were able to make fairly accurate estimates of points on a scale from 0 to 100.

A weakness of the approach we propose here is that it requires users to work in 10ths of Y_{Max}. The difficulty of doing this can be reduced where it is possible to choose a value of Y_{Max}, which can easily be divided by 10, such as $Y_{Max} = 100$ or $Y_{Max} = 1{,}000$. The difficulty arises if Y_{Max} has a value such as 173, which would force the user to work in multiples of 17.3. Where relatively rough estimates are required, the 10^{ths} of Y_{Max} might be approximated to a more amenable whole number; for example, if $Y_{Max} = 190$, the user might think in terms of multiples of 20 and accept that doing this will lead to a regular but small overestimate.

The results of the semi-structured interviews conducted at the end of study 3 confirmed that the multi-reference approach was preferred and found generally to be easier to use by participants than the single point sonification approach.

7) Multi-reference approaches work better than pitch only or single reference sonification approaches.

   The results of study 4 showed that both VI and sighted users performed better on point estimation tasks with the two multi-reference conditions (conditions 3 and 4) than with the pitch only and single point reference sonification (conditions 1 and 2). The difference in performance was more marked for sighted participants than for VI participants. We consider that this results from the fact that the multi-reference sonification approaches provide more anchor points (or more guidance than the pitch only and single reference conditions).

8) VI participants generally perform better on point estimation tasks across all sonification approaches reported here and are more adaptable to different sonification approaches.

   VI participants performed more accurately on point estimation tasks, represented by a lower mean and median of RMSE values when compared with sighted participants. The result of a Mann-Whitney Test also showed that the differences were statistically significant ($W = 25600$, $p < 0.001$).

However, when we compared data for each condition separately (see Table 7.9), it seems that the RMSE values of conditions 3 and 4 for both groups of participants were not different. Meanwhile, in conditions 1 and 2, sighted participants made twice the number of errors than VI participants. This shows that VI participants performed more accurately on point estimation tasks with fewer references (i.e., conditions 1 and 2).

Both VI and sighted groups had similar performance in the multi-reference conditions, with both groups performing rather better in condition 4 (step10 sonification mapping) than condition 3 (step20 sonification mapping). These results tend to show that all participants seemed to perform better as the number of reference points increased, in this case, when the range from 0 to $Y_{Estimate}$ was split into 10^{ths} rather than 5^{ths}.

Comparing the standard deviation and IQR values for each condition, we see that the variability of point estimations by sighted participants was far lower than VI participants across all conditions.

Examining the descriptive statistics for VI participants and sighted participants (Table 7.16), the SD of sighted participants (10.86) is smaller than the SD of VI participants (12.89). While for the IQR, the sighted participant (14.98) is bigger than the IQR of VI participants (11.36). The IQR result suggested that the mean RMSE for sighted participants is still bigger than VI participants, as expected.

The reason why the SD of VI participants was bigger than sighted participants is due to the fact that we have several long outliers for the VI participant results. The existence of outliers implicitly shows the diversity of the abilities of the VI participants.

9) The component-based approach to representing negative numbers shows a success rate of around 90% in polarity estimates performed by both VI and sighted users.

The component-based approach was implemented by preceding the representation of the absolute number of $Y_{Estimate}$ by a sonar ping sound reminiscent of a submarine, associated with being below ground level.

The results observed for both sighted and VI participants had a mean of over 90% for all four conditions. However, for both sighted and VI participants, there was some falling off of accuracy going from condition 1 to 4. We speculate that this might be due to an inverse recency effect, where for those sonification conditions involving more notes (the multi-reference conditions), there was a higher number of instances of incorrect polarity estimates.

There was no statistically significant difference between the results achieved by sighted and VI participants on the polarity estimation tasks.

10) We don't have any evidence in our study that the result is affected much because of differences in cultural background, education, linguistic, musical experience etc. But in our study 4, the participants' demographic characteristics revealed that several VI participants had better level of musical training than the other VI participants. We observed that the data variability of sighted participants was far lower than VI participants across all conditions. Our analysis on the data revealed that some varied data and outliers came from VI participants with less musical training. Meanwhile, our sighted participants generally had no better musical levels than VI participants whose their musical level distribution was more varied. Further studies are needed to investigate possible relationship of the level of musicianship for effective point estimation in this context.

8.3. MAG App version 5

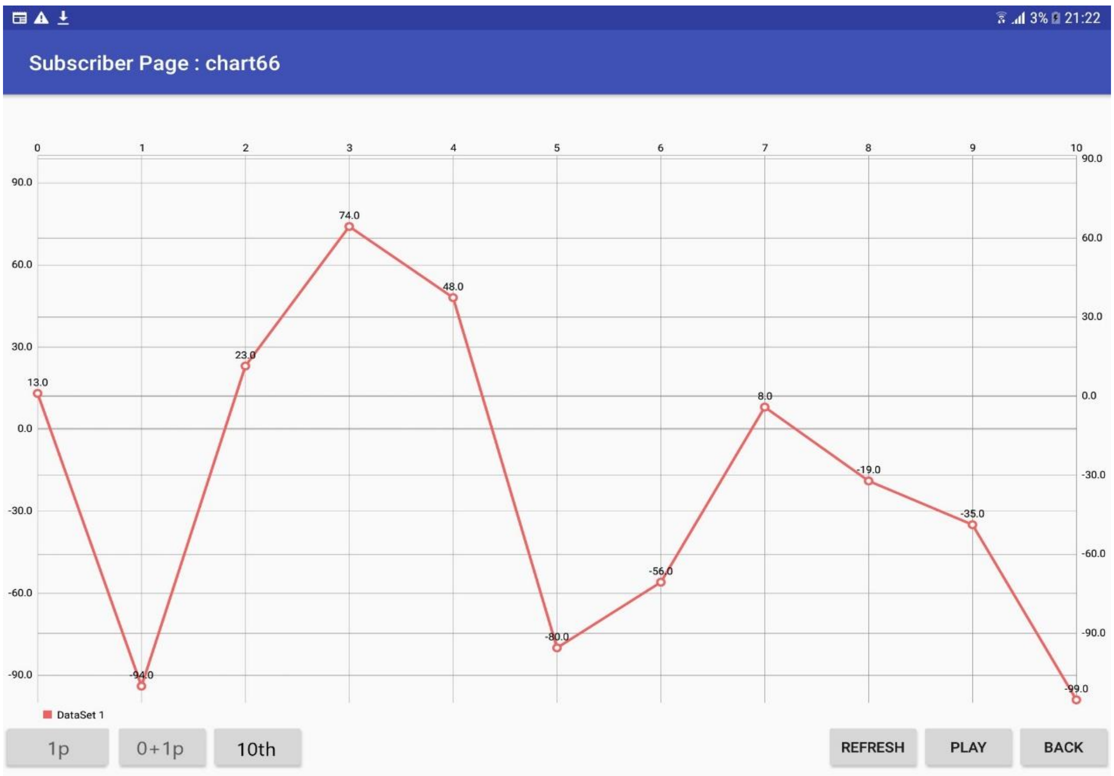


Figure 8-1. MAG App version 5 (final design) with three additional buttons for auditory reference mode options

In the final version of the MAG App (version 5), the Y-axes across all four conditions ranges from -100 to 100, with one value corresponding to one increment as shown in the graph with random notes in Figure 8-1. The final MAG app is a functional extension of MAG app version 4.0 with the addition of three buttons on the left side below the graph for the selection of auditory reference mode options. The first button is labelled "1p" corresponding to single point reference mode, followed by the button labelled "0+1p" for single reference mode, and "10th" for the multi-reference mode corresponding to 10 steps. The design does not include the steps of every 5th (the steps of 20 units) since there is probably not enough evidence to include that feature in the final design. While for the condition using step20 (condition 3) results were generally better than the results for conditions 1 and 2, the mode was the least liked by participants and considered the most difficult one in usability testing.

The "10th" button (corresponding to condition step10) is active by default since it was selected as participants' most preferred mode in usability testing and was the most accurate mode for point estimation. When this mode is active, the other two buttons are inactive and their button labels are grayed out. Activating any of the 3 buttons automatically deactivates the other two.

8.4. Theoretical implications

In this section, we discuss theoretical implications that reflects on what we learnt through the PhD research about visually impaired and use of sonified graph.

8.4.1. Multi-reference scheme

In general, VI participants' performance across all conditions was better than that of sighted people, and that their performance was more robust to changes in the sonification condition. As shown in Table 7.16, the descriptive statistics revealed that the VI participants (M=18.51 (12.89), median= 15.62) made fewer errors on the point estimation tasks, as represented by their lower mean and median values, compared with sighted participants (M=22.68 (10,86), median= 20.08).

The study suggested that users' performance is better using a step size of 10 integers rather than a step size of 20 integers. It is by no means obvious that would be the case, because, for example, the sonification using a step size of 10 integers means auditioning more sounds.

One possible question might be raised that if we reduce the step to 1 integer, the subjects will probably be able to perform the task with a null error (if they are able to count up to 100). However, using a step size of one integer entirely misses the point that we are aiming to produce a system that is useful and usable, not one where users will be likely to take upwards of 30 seconds (assuming a pause length of 0.5 seconds) in order to estimate on each individual point. It becomes lengthy in duration and it also becomes increasingly cognitively demanding when numbers far from 0 are involved.

In this research, we found that negative numbers were represented mentally in the form of component representation values instead of a holistic representation. The component-based approach showed a success rate of around 90% in polarity estimates performed by both VI and sighted users.

8.4.2. Differences in cultural background, education, musical experience

Based on our demographic analysis, we can assume there are correlations between the level of musical training and the variability in the results. The participants' demographic characteristics (see sub-section 7.4) revealed that several VI participants had better levels of musical training than the other VI participants, as discussed in section 7.6.3.1. The existence of outliers from VI participants with less musical training implicitly shows the diverse abilities of the VI participants. Meanwhile, our sighted participants generally had no better musical levels than VI participants. Earlier in study 1 to 3, we have no evidence that the result is affected much because of this different background issue.

The most challenging is to attract the visually impaired participants to the experiment. Therefore, I need to go back to my home country to get VI participants from Indonesia as I cannot get sufficient number in the UK. However, we need to control the variable by recruiting them based on several filtering criteria, such as the participants should understand the concept of the graph and also have understanding about musical note. Control of subjects with this limitation will be more restricted and cases related to outliers will be more difficult to avoid.

Moreover, we would like our study generalisable and easier to replicate by any researchers in the future so that it could be used by VI users in all countries. Thing that help this study is the concept of basic music education in Indonesia is similar to the concept of music in any other country in the West.

8.4.3. Contextual cues role

Earlier in chapter 5, we have pointed out the importance of contextual clues, the fact that it seems to work better when it is part of an interactive control loop with the human at the centre of the interaction. Unlike passive listening mode, which transmits the auditory graph unidirectionally from the device to the user, a key feature of multi-touch interaction is the bi-directional flow of information to and from the user, allowing the user to perceive and actively engage with the system. Touch sensations combined with audio effectively close a feedback control loop between the system and the user, providing cues to the user, enabling them to control the interaction actively and intuitively. This is confirmed by Smith and Walker (2002) that contextual cues (i.e, axes, labels and reference marks) increases perceptions and readability by providing the means to estimate point at any position.

8.4.4. Training and design implications

The need to benefit from additional training was also evident in the results reported in Study 2. A number of instances occurred in which VI participants were at a loss to identify the graph boundary inside active screen. Further, each mobile device has different frame width to the active screen. Likewise, there were a small number of incidents in which a VI user was unable to reach or pass a specific note. During training, participants either located the specific notes by going back to the previous note, or simply navigated the entire graph again to find the notes. Confronting these problems during the study led many VI participants to emphasize the need to use the interface frequently so they could become familiar with its structure before experiment. Given that sound is a temporal medium and all VI participants utilized speech-based screen readers in this study, this implies that VI users do not have a running picture of the screen they are working with in front of their eyes compared to their sighted counterparts. Potentially, this suggests that more training is needed to allow VI users to memorize more of the interface structure. The designer should also consider using a mobile device with a wider screen instead of a pocket-sized device, and also get a device where the active screen boundary is short on the edge of the device. Reflecting these difficulties, findings in Study 2 likely represent an underestimate regarding how much could be achieved in terms of task performance if participants had received more intensive training prior to performing the tasks.

8.4.5. Implication of mobile screen readers to improve user's experience

The user is left with adapting the access tool to use it as effectively as possible in cases where applications are not designed from the outset to support different access modes. Multiple access tools provide different functionality with scripting and setting up in other ways. Android screen readers such as TalkBack and Voice Assistant let the user and designer customize settings and create scripts that can be either global or specific to a mobile application. As discussed in Section 3.11.1, the Voice Assistant (VA) screen reader was chosen here because it was built into our device set. Android has scripting capabilities that can be used to further set how the screen reader should respond or sound when it receives an interaction. We were able to identify barriers that users would encounter when accessing the MAG app with screen readers after conducting the accessibility review. In the MAG app, we used scripts with multi-touch capabilities to get two-finger input when accessibility is enabled. We carefully labelled each part so that the reader could read it effectively. From the results and findings in Study 2, these features were well accepted and positively impacted VI users' experience and performance in the tasks.

Enhancing the user experience in interfaces that were not designed with accessibility in mind requires the designer to rigorously assess the interface using an analytical usability technique. The result of this process allows the designer in identifying the constraints preventing the target users from achieving a goal when using the app in the specific context of use. Following the identification of the barriers, the designer needs to modify the accessibility tool settings and develop scripts to increase accessibility.

8.5. Future work

In this final section, we will discuss potential ways to advance research on specific issues related to multimodal graph interaction derived from the thesis and to address specific areas of auditory graphs in different contexts and using mobile devices.

The design and analysis of **mobile auditory device** discussed in this thesis have demonstrated that such multimodal interactions have positive influences on auditory display design. This work has not only demonstrated the ability of sighted and VI participants' to perform point estimation and graph reproduction tasks but has also provided important findings on the auditory graph design process in the context of interaction design.

8.5.1. Extended studies

The studies described here could be extended, for example, to apply to different types of plots other than line graphs such as pie charts and bar charts. It is likely that entirely new, or at least substantially revised sonification approaches than those reported here may be required to render these effectively.

Different sonification approaches might be explored for different types of relatively straightforward 2D plots, for example, the efficacy of granular synthesis in representing scatter plots.

Similarly, the line of research investigated in studies 2-4 here on a more active approach to interaction through the use of multi-touch gestures could be taken further. As the hardware develops, techniques to enable the easy comparison of individual point values or comparison of ranges of values using gestures from two hands might be examined.

Networking of devices can extend interaction to collaborative work between users with different abilities, jointly examining data and capturing their findings in a shared workspace.

The exploration of multimodal graphs in collaborative work should certainly be seen as a fruitful path for future research. This should include the identification of behavioral patterns that occur during creation, reading, updating, and deletion (CRUD) and user communication. Exploring ways in which users can use social networks to perform collaborative graphing and other tasks, investigating the advantages and possible costs of such communication, and the resulting user behavior advances the development of support for multimodal collaborative interaction, and thus the integration of VI users in educational and work environments.

In Study 4, we tested the representation of negative number with component-based only due to time limitation. We cannot push our participants to take another test because the main test itself take about 1 hour per participants. But we can see the result that the percentage of correct polarity sign is high, which is all above 90% in all condition.

While for the holistic-based representation, we assume the result would be obvious since if we map all the value from the most minimum negative number to the most maximum number, user will need to remember the all mapping twice more than component-based representation.

A further extension would be to examine approaches to sonifying more complex graphs such as stacked bar graphs and 3D presentation such as 3D spectral plots. This might be done by keeping the present scheme for of using pitch to represent the value of Y coordinates. Position along the X-axis might be indicated through lateral spacing, that is the sonification might move from left to right as its coordinate on the X-axis increases. A number of parameters might be explored for representing changes of location along the Z-axis. A strong candidate for investigation is amplitude, where increases in amplitude might be associated with increasing values of location on the Z-axis.

Care would need to be taken in such studies to take into account psychoacoustic effects between whichever parameters are explored. For example, Fletcher and Munsen (1933) have shown relationships between perceived amplitude and frequency. According to Fletcher-Munsen curve, at low listening volume, the mid frequencies sound more prominent, yet the low and high frequency ranges seem to fade into the background. Further at high listening volumes, the lows and highs sound more prominent, whilst the midrange seems relatively softer. The interaction between the perception of amplitude and frequency could lead to misjudgements on the part of the listener of the values of amplitude and frequency and so misjudgements of values in the 3D plot.

References

Albinsson, P.-A., & Zhai, S. (2003). High precision touch screen interaction. In *Proceedings of the conference on Human factors in computing systems - CHI '03* (p. 105). https://doi.org/10.1145/642611.642631

Aliakseyeu, D. … Subramanian, S. (2008). Multi-Flick - An Evaluation of Flick-Based Scrolling Techniques for Pen Interfaces. In *Proceedings of the International Conference on Human Factors in Computing Systems (CHI'08)* (pp. 1689–1698). https://doi.org/10.1145/1357054.1357319

Amar, R. … Sellers, C. (2003). Mobile ADVICE: An Accessible Device for Visually Impaired Capability Enhancement. *CHI '03 Extended Abstracts on Human Factors in Computing Systems - CHI '03*, 918. https://doi.org/10.1145/765891.766069

Anderson, J. (2005). Creating an Empirical Framework for Sonification Design. *Proceedings of ICAD 2005*.

Anderson, J., & Sanderson, P. (2004). Designing Sonification for Effective Attentional Control in Complex Work Domains. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. https://doi.org/10.1177/154193120404801606

Apitz, G., & Guimbretière, F. (2004). CrossY: A Crossing-Based Drawing Application. *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology - UIST '04*, *6*(2), 3. https://doi.org/10.1145/1029632.1029635

Azenkot, S. … Ladner, R. E. (2012). Input finger detection for nonvisual touch screen text entry in Perkinput. *Proceedings of Graphics Interface 2012*, (d), 208. Retrieved from http://dl.acm.org/citation.cfm?id=2305297

Baddeley, A. (1997). *Human Memory: Theory and Practice*. Psychology Press.

Baker, C. M. … Ladner, R. E. (2014). Tactile Graphics with a Voice Demonstration. *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*, 321–322. https://doi.org/10.1145/2661334.2661349

Balik, S. P. … Sigler, V. J. (2014). Including Blind People in Computing Through Access to Graphs. *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*, 91–98. https://doi.org/10.1145/2661334.2661364

Banks, W. P. … Kayra-Stuart, F. (1976). Semantic congruity effects in comparative judgments of magnitudes of digits. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(3), 435–447. https://doi.org/10.1037/0096-1523.2.3.435

Barkhuus, L., & Rode, J. A. (2007). From Mice to Men - 24 Years of Evaluation in CHI. https://doi.org/10.1145/1240624.2180963

Barnston, A. G. (1992). Correspondence among the Correlation, RMSE, and Heidke Forecast Verification Measures; Refinement of the Heidke Score. *Weather and Forecasting*, *7*(4), 699–709. https://doi.org/10.1175/1520-0434(1992)007<0699:CATCRA>2.0.CO;2

Batusic, M., & Urban, F. (2002). Preparing Tactile Graphics for Traditional Braille Printers with

BrlGraphEditor. *Proceeding ICCHP '02 Proceedings of the 8th International Conference on Computers Helping People with Special Needs*, 535–536.

Benesty, J. … Cohen, I. (2009). Pearson Correlation Coefficient (pp. 1–4). https://doi.org/10.1007/978-3-642-00296-0_5

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blandford, A. (2013). Semi-Structured Qualitative Studies. *The Encyclopedia of Human-Computer Intreraction*.

Blattner, M. … Greenberg, R. (1989). Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction*, *4*(1), 11–44. https://doi.org/10.1207/s15327051hci0401_1

Bly, S. (1982). Presenting information in sound. *Proceedings of the 1st Conference on Human Factors in Computing Systems (CHI '82)*, 371–375. https://doi.org/10.1145/800049.801814

Bonebright, T. L. (2005). a Suggested Agenda for Auditory Graph Research. *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland, July 6-9, 2005*, 1–4.

Bonebright, T. L. … McCain, G. R. (2001). Testing the Effectiveness of Sonified Graphs for Education: A Programmatic Research Project. *2001 International Conference on Auditory Display*, 62–66. Retrieved from http://www.acoustics.hut.fi/icad2001/proceedings/papers/bonebri1.pdf

Bonner, M. … Edwards, W. (2010). No-look notes: accessible eyes-free multi-touch text entry. *Pervasive Computing*. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-12654-3_24

Bordman, G. N. (1876). Notes and Diatonics Scale. *New England Journal of Education*, *3*(19), 224. Retrieved from https://www.jstor.org/stable/44764823%0D

Bornschein, J. … Weber, G. (2015). Collaborative Creation of Digital Tactile Graphics. *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, 117–126. https://doi.org/10.1145/2700648.2809869

Bowen, G. M., & Roth, W.-M. (1998). Lecturing graphing: What features of lectures contribute to student difficulties in learning to interpret graph? *Research in Science Education*, *28*(1), 77–90. https://doi.org/10.1007/BF02461643

Bowen, G. M., & Roth, W.-M. (2005). Data and graph interpretation practices among preservice science teachers. *Journal of Research in Science Teaching*, *42*(10), 1063–1088. https://doi.org/10.1002/tea.20086

Bregman, A. (1990). Auditory Scene Analysis. In *Auditory Scene Analysis: The Perceptual Organization of Sounds*. London: MIT Press, Cambridge.

Brewster, S. A. (2002). NonSpeech Auditory Output. In *The human-computer interaction* *handbook: fundamentals, evolving technologies and emerging applications* (pp. 220–239). Hillsdale: L. Erlbaum Associates Inc. Retrieved from http://portal.acm.org/citation.cfm?id=772089

Brewster, S. A. … Edwards, A. D. N. (1994). A detailed investigation into the effectiveness of earcons. *Auditory Display: Sonification, Audification, and Auditory Interfaces*.

Brown, L. M., & Brewster, S. A. (2003). Drawing By Ear : Interpreting Sonified Line Graphs. *International Conference on Auditory Display*, *35*(July), 6–9.

Brown, L. M. … Riedel, B. (2003). Design Guidelines for Audio Presentation of Graphs and Tables. *Psychology*, 1–5.

Buxton, B. (2010). A Touching Story: A Personal Perspective on the History of Touch Interfaces Past and Future. *SID Symposium Digest of Technical Papers*, *41*(1), 444. https://doi.org/10.1889/1.3500488

Chew, Y. C. (2014). *Assessing the Use of Auditory Graphs for Middle School Mathematics*.

Clough, J. (1979). Aspects of Diatonic Sets. *Journal of Music Theory*. https://doi.org/10.2307/843693

Cockburn, A., & Savage, J. (2003). Comparing Speed-Dependent Automatic Zooming with Traditional Scroll, Pan, and Zoom Methods. *Bcs Hci*, 87–102. https://doi.org/http://doi.acm.org/10.1145/1456650.1456652

Cohen, R. F. … Skaff, J. (2005). PLUMB: Displaying Graphs to the Blind Using an Active Auditory Interface. *Assets'05*, 182–183. https://doi.org/10.1145/1090785.1090820

Conrad, R., & Hull, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science*. https://doi.org/10.3758/BF03331446

Davison, B. K. (2013). Universal Graph Literacy : Understanding How Blind and Low Vision Students Can Satisfy the Common Core Standards With Accessible Auditory Graphs, (June).

Davison, B. K., & Walker, B. N. (2007). Sonification Sandbox Reconstruction: Software Standard for Auditory Graphs. *Proceedings of the 13th International Conference on Auditory Display (ICAD2007)*, 509–512. Retrieved from Proceedings/2007/DavisonWalker2007.pdf

de Campo, A. (2007). A DATA SONIFICATION DESIGN SPACE MAP. *Proceedings of the 2nd International Workshop on Interactive Sonification (2007) 1-4*, 342–347. Retrieved from http://sonify.psych.gatech.edu/~ben/references/decampo_toward_a_data_sonification_design_space_map.pdf

Denzin, N. K., & Lincoln, Y. S. (2000). Introduction: The discipline and practice of qualitative research. In *Handbook of qualitative research (2nd edition)*.

Derrick, B. … Toher, D. (2017). An Inverse Normal Transformation Solution for the comparison of two samples that contain both paired observations and independent observations. *Journal of Modern Applied Statistical Methods*, *16*(1), 137–157. https://doi.org/10.22237/jmasm/1493597280

Dombois, F. (2001). Using audification in planetary seismology. *Proceedings of the 7th International Conference on Auditory Display (ICAD2001)*, 227–230.

Dowling, W. J., & Hollombe, A. W. (1977). The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception & Psychophysics*. https://doi.org/10.3758/BF03199469

Duarte, C. … Dumas, B. (2017). Designing Multimodal Mobile Interaction for a Text Messaging Application for Visually Impaired Users. *Frontiers in ICT*, *4*(December). https://doi.org/10.3389/fict.2017.00026

Ferguson, S. … Calò, C. A. (2012). Navigation of interactive sonifications and visualisations of time-series data using multi-touch computing. *Journal on Multimodal User Interfaces*, *5*(3–4), 97–109. https://doi.org/10.1007/s12193-011-0075-3

Field, A. P. … Field, Z. (2012). *Discovering Statistics using R*. London, UK: SAGE Publications.

Fischer, M. H. (2000). Do irrelevant depth cues affect the comprehension of bar graphs? *Applied Cognitive Psychology*, *14*, 151–162. https://doi.org/10.1002/(SICI)1099-0720(200003/04)14:2<151::AID-ACP629>3.0.CO;2-Z

Fischer, M. H. (2003). Cognitive Representation of Negative Numbers. *Psychological Science*, *14*(3), 278–282. https://doi.org/10.1111/1467-9280.03435

Fischer, M. H., & Rottmann, J. (2005). Do negative numbers have a place on the mental number line ? *Psychology Science*.

Fletcher, H., & Munson, W. A. (1933). Loudness, Its Definition, Measurement and Calculation. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1915637

Flowers, J. H. (2005). Thirteen Years of Reflection on Auditory Graphing : Promises , Pitfalls , and Potential New Directions. *International Conference on Auditory Display*, 1–5.

Flowers, J. H. … Turnage, K. D. (1997). Cross-Modal Equivalence of Visual and Auditory Scatterplots for Exploring Bivariate Data Samples. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(3), 341–351. https://doi.org/10.1518/001872097778827151

Foreit, K. G. (1976). Short-lived auditory memory for pitch. *Perception & Psychophysics*. https://doi.org/10.3758/BF03204245

Forlines, C., & Balakrishnan, R. (2008). Evaluating tactile feedback and direct vs. indirect stylus input in pointing and crossing selection tasks. *Proceedings of ACM CHI 2008 Conference on Human Factors in Computing Systems*, *1*, 1563–1572. https://doi.org/10.1145/1357054.1357299

Freedom Scientific. (2019). JAWS. Retrieved November 12, 2019, from https://www.freedomscientific.com/products/software/jaws/

Friesen, J. "JavaJeff." (2010). *Learn Java for Android Development*. *Learn Java for Android Development*. Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4302-3157-8

Frysinger, S. P. (2005). A brief history of auditory data representation to the 1980s.

Proceedings of the 11th International Conference on Auditory Display (ICAD2005), 410–413.

Ganor-Stern, D., & Tzelgov, J. (2008). Negative numbers are generated in the mind. *EXPERIMENTAL Psychology*, *55*(3).

Gardner, J. a. … Sahyun, S. (1996). Triangle: a practical application of non-speech audio for imparting information. *Proceedings of the 3rd International Conference on Auditory Display (ICAD 1996)*, 59–60.

Gaver, W. (1986). Auditory Icons: Using Sound in Computer Interfaces. *Human-Computer Interaction*, *2*(2), 167–177. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/s15327051hci0202_3

Geldard, F. A. (1960). Some neglected possibilities of communication. *Science*. https://doi.org/10.1126/science.131.3413.1583

Goncu, C., & Marriott, K. (2015). GraCALC: An Accessible Graphing Calculator. *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, 311–312. https://doi.org/10.1145/2700648.2811353

Google LLC. (2019). Android Accessibility Help. Retrieved November 20, 2019, from https://support.google.com/talkback

Greene, R. L., & Samuel, A. G. (1986). Recency and Suffix Effects in Serial Recall of Musical Stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/0278-7393.12.4.517

Grussenmeyer, W., & Folmer, E. (2017). Accessible Touchscreen Technology for People with Visual Impairments. *ACM Transactions on Accessible Computing*, *9*(2), 1–31. https://doi.org/10.1145/3022701

Guerreiro, T. … Jorge, J. A. (2008). From Tapping to Touching: Making Touch Screens Accessible to Blind Users. *IEEE Multimedia*, *15*(4), 48–50. https://doi.org/10.1109/MMUL.2008.88

Guest, G. … Mitchell, M. L. (2017). *Collecting Qualitative Data: A Field Manual for Applied Research*. *Collecting Qualitative Data: A Field Manual for Applied Research*. https://doi.org/10.4135/9781506374680

Gutwin, C., & Fedak, C. (2004). Interacting with big interfaces on small screens: a comparison of fisheye, zoom, and panning techniques. *Proceedings of Graphics Interface 2004*, 145–152. Retrieved from http://dl.acm.org/citation.cfm?id=1006076

Hanson, W. E. … Creswell, J. D. (2005). Mixed methods research designs in counseling psychology. *Journal of Counseling Psychology*. https://doi.org/10.1037/0022-0167.52.2.224

Harrar, L., & Stockman, T. (2007). Designing Auditory Graph Overviews: An Examination of Discrete vs. Continuous Sound and the Influence of Presentation Speed. *Proceedings of the 13th International Conference on Auditory Display (ICAD2007)*, 299–305. Retrieved from Proceedings/2007/HarrarStockman2007.pdf

Hawkins, H. L., & Presson, J. C. (1986). Auditory information processing. In *Handbook of* *perception and human performance*. New York, New York, USA.

Hermann, T. (2002). *Sonification for Exploratory Data Analysis*. University of Bielefeld.

Hermann, T., & Ritter, H. (2004). Sound and meaning in auditory data display. *Proceedings of the IEEE*, *92*(4), 730–741. https://doi.org/10.1109/JPROC.2004.825904

Howe, A. (1997). Refusal of videorecording: What factors may influence patient consent? *Family Practice*. https://doi.org/10.1093/fampra/14.3.233

Igarashi, T., & Hinckley, K. (2000). Speed-dependent automatic zooming for browsing large documents. *Uist '00*, *2*, 139–148. https://doi.org/10.1145/354401.354435

Irani, P. … Yang, X. D. (2006). Improving selection of off-screen targets with hopping. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI 06*, (January), 299. https://doi.org/10.1145/1124772.1124818

Johnsen, A. D. … Bygstad, B. (2012). Making Touch-Based Mobile Phones Accessible for the Visually Impaired. *Norsk Informatikkonferanse*, *2012*, 177–188. Retrieved from http://www.detgarbra.no/publish/fil/vis/1053

Johnson, E. A. (1965). Touch display—a novel input/output device for computers. *Electronics Letters*, *1*(8), 219. https://doi.org/10.1049/el:19650200

Kane, S. K. … Wobbrock, J. O. (2008). Slide Rule: Making Mobile Touch Screens Accessible to Blind People Using Multi-Touch Interaction Techniques. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility - Assets '08* (p. 73). New York, New York, USA: ACM Press. https://doi.org/10.1145/1414471.1414487

Kane, S. K. … Ringel Morris, M. (2011). Access Overlays: Improving Non-Visual Access to Large Touch Screens for Blind Users. *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology - UIST '11*, 273–282. https://doi.org/10.1145/2047196.2047232

Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*. https://doi.org/10.1561/1500000012

Kennel, A. R. (1996). Audiograf : A Diagram-Reader For The Blind, 51–56.

King, A. … Wood, J. (2004). Presenting UML Software Engineering Diagrams to Blind People. *International Conference on Computers Helping People with Special Needs*, 522–529. https://doi.org/10.1007/978-3-540-27817-7_76

Kong, F. … You, X. (2012). Components representation of negative numbers: Evidence from auditory stimuli detection and number classification tasks. *Quarterly Journal of Experimental Psychology*, *65*(4), 691–701. https://doi.org/10.1080/17470218.2011.622048

Krajcsi, A., & Igács, J. (2010). Processing negative numbers by transforming negatives to positive range and by sign shortcut. *European Journal of Cognitive Psychology*. https://doi.org/10.1080/09541440903211113

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal* *of the American Statistical Association*. https://doi.org/10.1080/01621459.1952.10483441

Lahiri, A. … Basu, A. (2005). Sparsha: a comprehensive indian language toolset for the blind. *Assets '05 Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*, 114–120.

Larkin, J. H., & Simon, H. A. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science*. https://doi.org/10.1016/S0364-0213(87)80026-5

Li, K. a. … Hinckley, K. (2008). BlindSight: Eyes-Free Access to Mobile Phones. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '08*, 1389–1398. https://doi.org/10.1145/1357054.1357273

LicNSc, kka-L. H. (2001). Video Recording as a Method of Data Collection in Nursing Research. *Nordic Journal of Nursing Research*, *21*(2), 21–26. https://doi.org/10.1177/010740830102100204

Lin, G. G., & Scott, J. G. (2012). A mobile phone system to find crosswalks for visually impaired pedestrians, *100*(2), 130–134. https://doi.org/10.1016/j.pestbp.2011.02.012.Investigations

Liu, Y. (2012). *Multimodal Interaction : Developing an Interaction Concept for a Touchscreen*. Ludwig-Maximilians-Universität München.

Ljubic, S. … Kukec, M. (2015). Finger-based pointing performance on mobile touchscreen devices: Fitts' law fits. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-20678-3_31

Lopes, A. G. (2016). Using Research Methods in Human Computer Interaction to Design Technology for Resilience. *Journal of Information Systems and Technology Management*, *13*(3). https://doi.org/10.4301/S1807-17752016000300001

Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*. https://doi.org/10.1214/aoms/1177730491

Mansur, D. L. … Joy, K. I. (1985). Sound graphs: A numerical data analysis method for the blind. *Journal of Medical Systems*, *9*(3), 163–174. https://doi.org/10.1007/BF00996201

Mascetti, S. … Bernareggi, C. (2016). Sonification of guidance data during road crossing for people with visual impairments or blindness. *International Journal of Human-Computer Studies*, *85*, 16–26. https://doi.org/10.1016/j.ijhcs.2015.08.003

McDonald, S. … Hurst, A. (2014). Tactile aids for visually impaired graphical design education. *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility - ASSETS '14*, 275–276. https://doi.org/10.1145/2661334.2661392

Mcgee, R. (2009). Auditory Displays and Sonification : Introduction and Overview.

Metatla, O. … Martin, F. (2016). Sonification of Reference Markers for Auditory Graphs: Effects on Non-Visual Point Estimation Tasks. *Peer Journal Computer Science*, *2*(April).

https://doi.org/10.7717/peerj-cs.51

Metatla, O. … Bryan-Kinns, N. (2014). Non-Visual Menu Navigation: the Effect of an Audio-Tactile Display. In *Proceedings of the 28th International BCS Human Computer Interaction Conference: Sand, Sea and Sky - Holiday HCI, HCI 2014*. https://doi.org/10.14236/ewic/HCI2014.33

Meyer, J. (2000). Performance with tables and graphs: Effects of training and a visual search model. *Ergonomics*, *43*(11), 1840–1865. https://doi.org/10.1080/00140130050174509

Meyer, M. F. (1956). On Memorizing Absolute Pitch. *The Journal of the Acoustical Society of America*, *28*(4), 718–719. https://doi.org/10.1121/1.1908465

Mezrich, J. J. … Slivjanovski, R. (1984). Dynamic Representation of Multivariate Time Series Data. *Journal of the American Statistical Association*, *79*(385), 34–40. https://doi.org/10.1080/01621459.1984.10477059

Mondor, T. A., & Morin, S. R. (2004). Primacy, recency, and suffix effects in auditory short-term memory for pure tones: Evidence from a probe recognition paradigm. *Canadian Journal of Experimental Psychology*. https://doi.org/10.1037/h0087445

Morse, J. M. (1994). Emerging from the data : the cognitive processes of analysis in qualitative inquiry. Critical issues in qualitative research methods. *Critical Issues in Qualitative Research Methods*.

Moscovich, T., & Hughes, J. (2004). Navigating documents with the virtual scroll ring. *Proceedings of the 17th Annual ACM …*, *6*(2), 57. https://doi.org/10.1145/1029632.1029642

Murdock, B. B., & Walker, K. D. (1969). Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior*. https://doi.org/10.1016/S0022-5371(69)80120-9

Mynatt, E. D., & Weber, G. (1994). Nonvisual presentation of graphical user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94*, 166–172. https://doi.org/10.1145/191666.191732

Nees, M. A., & Walker, B. N. (2007). Listener, task and auditory graph: Toward a conceptual model of auditory graph comprehension. *International Conference on Auditory Display*, 266–273. Retrieved from http://sonify.psych.gatech.edu/~walkerb/publications/pdfs/2007ICAD-NeesWalker.pdf

Nees, M. A., & Walker, B. N. (2008). Data density and trend reversals in auditory graphs. *ACM Transactions on Applied Perception*, *5*(3), 1–24. https://doi.org/10.1145/1402236.1402237

Nesbitt, K. V., & Barrass, S. (2004). Finding trading patterns in stock market data. *IEEE Computer Graphics and Applications*, *24*(5), 45–55. https://doi.org/10.1109/MCG.2004.28

Newson, R. (2002). Parameters behind "nonparametric" statistics: Kendall's tau, Somers' D and median differences. *Stata Journal*, *2*(1), 45–64. https://doi.org/The Stata Journal

Nikitenko, D., & Gillis, D. (2014). Touching the data: Exploring data sonification on mobile touchscreen devices. *Procedia Computer Science*, *34*, 360–367. https://doi.org/10.1016/j.procs.2014.07.038

Olivan, J. … Roessen, M. (2004). Easy listening to sleep recordings: Tools and examples. *Sleep Medicine*, *5*(6), 601–603. https://doi.org/10.1016/j.sleep.2004.07.010

Oviatt, S. … Kruger, A. (2017). *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 1*. *The handbook of multimodal-multisensor interfaces. Volume 1, Foundations, user modeling, and common modality combinations*. https://doi.org/10.1145/3015783

Pauletto, S., & Hunt, A. (2005). A comparison of audio & visual analysis of complex time-series data sets. *Proceedings of the 11th International Conference on Auditory Display (ICAD2005)*, 175–181.

Peres, S. C., & Lane, D. (2003). Sonification of Statistical Graphs. *International Conference on Auditory Display, Boston …*, (July), 6–9. Retrieved from http://icad.org/Proceedings/2003/PeresLane2003.pdf

PictureBraille. (n.d.). Retrieved November 10, 2016, from http://www.pentronics.com.au/index_files/PictureBraille.htm

Pinhas, M., & Tzelgov, J. (2012). Expanding on the mental number line: Zero is perceived as the “smallest”. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1187–1205. https://doi.org/10.1037/a0027390

Pinker, S. (1990). A Theory of Graph Comprehension. In *Artificial intelligence and the future of testing* (pp. 73–126). https://doi.org/10.1145/2046684.2046699

Pollack, I. (1952). The Information of Elementary Auditory Displays. *The Journal of the Acoustical Society of America*, *24*(6), 745–749. https://doi.org/10.1121/1.1906969

Potter, R. L. … Shneiderman, B. (1988). Improving the accuracy of touch screens: an experimental evaluation of three strategies. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 27–32. https://doi.org/10.1145/57167.57171

Ratner, B. (2009). The correlation coefficient: Its values range between 1/1, or do they. *Journal of Targeting, Measurement and Analysis for Marketing*. https://doi.org/10.1057/jt.2009.5

Rossing, T. (2014). *Springer Handbook of Acoustics*. (T. D. Rossing, Ed.) (Vol. 45). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-0755-7

Sahyun, S. (1999). *A comparison of auditory and visual graphs for use in physics and mathematics. Ph.D. Thesis*. Oregon State University. Retrieved from http://academics.uww.edu/physics/scs/phd/index.html

Sánchez, J., & Maureira, E. (2007). Subway mobility assistance tools for blind users. *9th ERCIM Workshop on User Interfaces for All, ERCIM 2006*, *4397 LNCS*, 386–404. https://doi.org/10.1007/978-3-540-71025-7

Schiffman, H. R. (1977). Fundamental auditory functions and phenomena. In *Sensation and*

Perception: An Integrated Approach. The American Journal of Psychology. https://doi.org/10.2307/1421748

Sears, A., & Shneiderman, B. (1991). High precision touchscreens: design strategies and comparisons with a mouse. *International Journal of Man-Machine Studies*, *34*(4), 593–613. https://doi.org/10.1016/0020-7373(91)90037-8

Sengpiel, E. (n.d.). Note names of Musical Notes Keyboards Piano Frequencies. Retrieved from http://www.sengpielaudio.com/calculator-notenames.htm

Shah, P. (1997). A Model of the Cognitive and Perceptual Processes in Graphical Display Comprehension. *AAAI Technical Report*, 94–101. Retrieved from http://aaaipress.org/Papers/Symposia/Fall/1997/FS-97-03/FS97-03-012.pdf

Shah, P., & Freedman, E. G. (2011). Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science*, *3*(3), 560–578. https://doi.org/10.1111/j.1756-8765.2009.01066.x

Shaki, S., & Petrusic, W. M. (2005). On the mental representation of negative numbers: Context-dependent SNARC effects with comparative judgments. *Psychonomic Bulletin & Review*, *12*(5), 931–937. https://doi.org/10.3758/BF03196788

Shin, H. … Kyung, K. U. (2013). Background display for visually impaired people in mobile touch devices. *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, 464–465. https://doi.org/10.1109/ICCE.2013.6486977

Smith, D. R., & Walker, B. N. (2002). Tick-Marks, Axes, And Labels : The Effects Of Adding Context To Auditory Graphs. *Proceedings of the 2002 International Conference on Auditory Display*, 1–6.

Smith, D. R., & Walker, B. N. (2005). Effects of auditory context cues and training on performance of a point estimation sonification task. *Applied Cognitive Psychology*, *19*(8), 1065–1087. https://doi.org/10.1002/acp.1146

Song, H. J., & Beilharz, K. (2007). Spatialization and timbre for effective auditory graphing. *Proceedings of the 8th WSEAS International Conference on Acoustics & Music: Theory & Applications*, (June 2007), 18–26. Retrieved from http://portal.acm.org/citation.cfm?id=1361371.1361375

Stevens, S. S. … Newman, E. B. (1937). A Scale for the Measurement of the Psychological Magnitude Pitch. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1915893

Stigler, S. M. (1978). Some forgotten work on memory. *Journal of Experimental Psychology: Human Learning and Memory*. https://doi.org/10.1037/0278-7393.4.1.1

Strumillo, P. (2010). Electronic interfaces aiding the visually impaired in environmental access, mobility and navigation. *3rd International Conference on Human System Interaction, HSI'2010 - Conference Proceedings*, 17–24. https://doi.org/10.1109/HSI.2010.5514595

Su, J. … Truong, K. N. (2010). Timbremap: : Enabling the Visually-Impaired to Use Maps on Touch-Enabled Devices. *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '10*, 17–26.

https://doi.org/10.1145/1851600.1851606

Suits, B. H. (1998). Physics of Music - Notes. Retrieved from http://www.phy.mtu.edu/~suits/notefreqs.html

Surprenant, A. M. … Crowder, R. G. (1993). Auditory Recency in Immediate Memory. *The Quarterly Journal of Experimental Psychology Section A*, *46*(2), 193–223. https://doi.org/10.1080/14640749308401044

Tu, H. … Wang, F. (2014). Evaluation of Flick and Ring Scrolling on Touch-Based Smartphones. *International Journal of Human-Computer Interaction*, *30*(8), 643–653. https://doi.org/10.1080/10447318.2014.907017

Tufte, E. R. (1990). Color and Information. In *Envisioning Information* (Vol. I, pp. 81–96).

Tzelgov, J. … Maymon-Schreiber, K. (2009). The Representation of Negative Numbers: Exploring the Effects of Mode of Processing and Notation. *Quarterly Journal of Experimental Psychology*, *62*(3), 605–624. https://doi.org/10.1080/17470210802034751

Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied*, *8*(4), 211–221. https://doi.org/10.1037/1076-898X.8.4.211

Walker, B. N. (2010). Career: Fundamental Research, Design and Evaluation of Auditory and Multimodal Graphs, 1–15.

Walker, B. N. … Lane, D. M. (2000). Psychophysical scaling of sonification mappings. *Proceedings of ICAD 2000*, 99–104. Retrieved from http://dev.icad.org/Proceedings/2000/WalkerKramer2000.pdf

Walker, B. N., & Mauney, L. M. (2010). Universal design of auditory graphs: A comparison of sonification mappings for visually impaired and sighted listeners. *ACM Transactions on Accessible Computing*, *2*(3), 1–16. https://doi.org/10.1145/1714458.1714459.http

Walker, B. N., & Nees, M. A. (2011). Theory of Sonification. *Principles of Sonification: An Introduction to Auditory Display and Sonification*, 1–32. Retrieved from http://sonify.psych.gatech.edu/publications/pdfs/2006preprint-WalkerNees-SonificationChapter_v09.pdf

Walker, B. N. … Nw, C. S. (2005). AN AGENDA FOR RESEARCH AND DEVELOPMENT OF MULTIMODAL GRAPHS School of Psychology Georgia Institute of Technology. *Audio*, *1*, 1–5. Retrieved from http://sonify.psych.gatech.edu/ags2005/pdf/AGS05_WalkerNees.pdf

Weingarten, M. A. … Margalit, A. (2001). A comparison of videotape and audiotape assessment of patient-centredness in family physicians' consultations. *Patient Education and Counseling*, *45*(2), 107–110. https://doi.org/10.1016/S0738-3991(00)00199-3

Wherry, E. (2003). Scroll ring performance evaluation. *CHI '03 Extended Abstracts on Human Factors in Computing Systems - CHI '03*, 758. https://doi.org/10.1145/765891.765973

Williams, K. … Bontempo, D. (2013). Comparing Audio and Video Data for Rating

Communication. *Western Journal of Nursing Research*, *35*(8), 1060–1073. https://doi.org/10.1177/0193945913484813

Worrall, D. (2009). *SONIFICATION AND INFORMATION CONCEPTS , INSTRUMENTS AND TECHNIQUES*. University of Canberra.

Zacks, J. … Schiano, D. (2002). Graphs in print. *Diagrammatic Representation and Reasoning*, (August), 187–206. https://doi.org/10.1007/978-1-4471-0109-3_11

Zhao, S. … Baudisch, P. (2007). Earpod. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07* (p. 1395). New York, New York, USA: ACM Press. https://doi.org/10.1145/1240624.1240836

Zou, H., & Treviranus, J. (2015). ChartMaster : a Tool for Interacting with Stock Market Charts using a Screen Reader. *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, (Figure 1), 107–116. https://doi.org/10.1145/2700648.2809862

Zwicker, E. (1938). On Peripheral Processing in Human Hearing. In *HEARING — Physiological Bases and Psychophysics* (pp. 104–110). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-69257-4_16

APPENDIX A NORMALITY ASSESSMENT OF RMSE, CORRELATION COEFFICIENT, AND LENGTH OF PAUSE FOR EACH CONDITION IN STUDY 1

Criteria	Condition	Shapiro Wilk	Skewness	SE	Z skew	Kurtosis	SE	Z Kurto
RMSE	Simple 1	0.126	0.657	0.637	1.031	-0.941	1.23	-0.764
	Simple 2	0.23	0.981	0.637	1.540	1.069	1.23	0.868
	Simple 3	0.357	-0.507	0.637	-0.796	-0.704	1.23	-0.571
	Simple 4	0.329	-0.803	0.637	-1.261	0.344	1.23	0.279
	Medium 1	0.094	-1.237	0.637	-1.942	1.823	1.23	1.480
	Medium 2	0.02	1.287	0.637	2.020	0.943	1.23	0.765
	Medium 3	0.23	0.836	0.637	1.312	2.334	1.23	1.894
	Medium 4	0.189	-0.144	0.637	-0.226	-1.392	1.23	-1.130
	Medium 5	0.007	1.386	0.637	2.176	0.801	1.23	0.650
	Medium 6	0.702	0.108	0.637	0.170	-0.982	1.23	-0.797
	Complex 1	0.061	1.43	0.637	2.245	2.167	1.23	1.759
	Complex 2	0.094	1.256	0.637	1.972	1.433	1.23	1.163
	Complex 3	0.045	1.516	0.637	2.380	3.525	1.23	2.861
	Complex 4	0.625	0.738	0.637	1.159	0.487	1.23	0.395
Aggregate	Simple	0.064	0.168	0.343	0.490	-1.109	0.67	-1.645
	Medium	0	1.286	0.283	4.544	2.839	0.56	5.079
	Complex	0.001	1.243	0.343	3.624	2.341	0.67	3.473
Coefficient Correlation	Simple 1	0.001	-1.969	0.637	-3.09	3.505	1.23	2.845
	Simple 2	0.011	0.057	0.637	0.09	-2.171	1.23	-1.762
	Simple 3	0.199	-0.064	0.637	-0.10	-1.305	1.23	-1.059
	Simple 4	0.237	-0.631	0.637	-0.99	-0.952	1.23	-0.773
	Medium 1	0.028	-1.484	0.637	-2.33	2.227	1.23	1.808
	Medium 2	0.162	-0.241	0.637	-0.38	-1.188	1.23	-0.964
	Medium 3	0.386	-0.492	0.637	-0.77	-0.221	1.23	-0.179
	Medium 4	0.133	0.125	0.637	0.20	-1.601	1.23	-1.300
	Medium 5	0.119	-0.996	0.637	-1.56	0.168	1.23	0.136
	Medium 6	0.095	-1.148	0.637	-1.80	1.063	1.23	0.863
	Complex 1	0.031	-1.286	0.637	-2.02	0.946	1.23	0.768
	Complex 2	0.002	-2.284	0.637	-3.59	6.478	1.23	5.258
	Complex 3	0.161	-1.198	0.637	-1.88	2.49	1.23	2.021
	Complex 4	0.743	-0.182	0.637	-0.29	-0.967	1.23	-0.785
Aggregate	Simple	0.346	-0.833	0.637	-1.31	0.315	1.23	0.256
	Medium	0.412	-0.863	0.637	-1.35	0.347	1.23	0.282
	Complex	0.043	-1.381	0.637	-2.17	2.351	1.23	1.908
Length of Pause	Simple 1	0	2.661	0.637	4.18	7.731	1.23	6.275

Criteria	Condition	Shapiro Wilk	Skewness	SE	Z skew	Kurtosis	SE	Z Kurto
	Simple 2	0.006	0.439	0.637	0.69	-0.337	1.23	-0.274
	Simple 3	0	-0.812	0.637	-1.27	-1.65	1.23	-1.339
	Simple 4	0	-0.812	0.637	-1.27	-1.65	1.23	-1.339
	Medium 1	0	2.276	0.637	3.57	6.478	1.23	5.258
	Medium 2	0.011	0.478	0.637	0.75	-0.868	1.23	-0.705
	Medium 3	0.011	0.912	0.637	1.43	-0.337	1.23	-0.274
	Medium 4	0.028	0.854	0.637	1.34	-0.014	1.23	-0.011
	Medium 5	0.01	0.354	0.637	0.56	-1.447	1.23	-1.175
	Medium 6	0.003	1.847	0.637	2.90	4.132	1.23	3.354
	Complex 1	0.009	1.492	0.637	2.34	2.521	1.23	2.046
	Complex 2	0.009	1.492	0.637	2.34	2.521	1.23	2.046
	Complex 3	0.045	0.8	0.637	1.26	-0.512	1.23	-0.416
	Complex 4	0.022	0.634	0.637	1.00	-1.143	1.23	-0.928
Aggregate	Simple	0.056	0.786	0.637	1.23	1.03	1.23	0.836
	Medium	0.029	0.967	0.637	1.52	0.695	1.23	0.564
	Complex	0.052	1.115	0.637	1.75	1.116	1.23	0.906

APPENDIX B QUESTIONAIRE FOR STUDY 2 PRIOR THE EXPERIMENT

<u>Section 1: General Questions (1 of 3)</u>

1. Age Range:
   - o 17 or younger
   - o 18-20
   - o 21-29
   - o 30-39
   - o 40-49
   - o 50-59
   - o 60 or older
2. Gender
   - o Male
   - o Female
3. Occupation
   - o Student
   - o Employed
   - o Unemployed
4. Qualification
   - o GCSE's and A-levels obtained

- Undergraduate
- Postgraduate or higher Degree

5. What assistive technology you use

- None.
- Screen reader
- Braille display
- Screen magnifier software
- Others, please Specify____

6. What musical instrument you play

- None
- Guitar
- piano
- violin
- keyboard
- Others, Please Specify____

APPENDIX C SEMI-STRUCTURED INTERVIEW AFTER THE EXPERIMENT

1. What sort of information do you know about graphs?

2. Have you ever learnt about graphs in school? What tools did you use to understand graphs?

3. At what stage you felt that it was hard to finish the task?

4. Do you think that there are better methods to understand the tone change and range?

5. How do you describe the process of graph reproduction tasks?

6. Did you find difficulties in understanding the audio graphs?

7. Did you find difficulties in understanding math?

8. Did you find the haptic feedback is helpful for the graph reproduction in line with the audio feedback?

APPENDIX D RESULTS STUDY 2

- Playback

Task	X Value	Y Reference	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
Simple 1	1	2	4	2	2	3	1	1	2	1	2	1	1	2	2	4	1
	2	6	5	6	6	5	5	1	5	7	5	8	6	1	6	6	5
	3	3	9	3	3	6	4	4	4	5	3	4	5	5	3	4	2
	4	4	5	4	4	4	3	6	6	3	4	5	3	7	4	3	3
Simple 2	1	5	2	5	5	5	3	5	7	6	5	4	4	5	5	5	5
	2	2	7	2	2	2	1	2	2	1	2	1	2	2	2	4	3
	3	3	3	3	4	7	2	1	3	5	3	2	3	1	6	3	1
	4	8	10	8	10	6	5	6	8	9	7	7	7	5	8	7	4
	5	6	8	6	6	4	6	8	6	10	8	5	6	6	5	6	5
Medium 1	1	2	3	2	2	2	1	2	3	1	1	2	1	1	2	3	1
	2	6	5	6	8	7	3	1	5	1	4	1	4	1	2	4	5
	3	4	8	4	3	4	4	5	4	6	3	5	3	3	2	2	1
	4	2	4	2	2	3	1	7	6	5	1	1	2	2	6	6	6
	5	8	6	8	8	5	3	2	3	6	8	8	5	3	4	6	4
	6	3	8	3	4	6	5	5	4	7	2	3	3	4	8	4	2
	7	4	9	4	5	3	4	8	5	8	3	4	4	5	5	3	3
Medium 2	1	6	2	6	6	4	6	4	5	6	6	5	4	6	6	5	4
	2	3	3	3	3	1	3	3	4	1	3	2	2	3	2	4	3
	3	2	5	2	2	2	1	1	3	2	1	1	1	2	3	3	2
	4	5	4	5	5	4	5	2	5	5	4	9	6	7	8	7	5
	5	8	3	8	8	5	9	5	3	6	6	2	5	6	6	6	8
	6	3	1	3	6	6	1	4	5	8	5	7	4	8	5	8	9
	7	6	6	6	4	3	7	6	3	6	4	5	3	7	4	6	6
	8	5	5	5	5	2	8	4	5	5	3	5	1	9	5	4	7
Complex 1	1	3	4	3	3	2	3	2	3	1	3	1	2	3	3	4	2
	2	6	6	6	6	7	4	4	6	1	3	4	4	2	3	6	3
	3	5	5	5	5	3	5	5	5	4	8	3	3	4	6	5	5
	4	4	8	4	4	8	8	6	4	5	5	2	2	6	5	6	8
	5	9	7	9	9	7	10	5	9	10	10	8	1	5	7	5	7
	6	3	3	3	7	4	3	8	7	2	7	1	7	7	9	7	9
	7	7	7	6	3	3	4	7	6	7	8	6	5	6	8	8	8
	8	4	9	4	6	5	5	10	8	8	6	2	3	7	3	9	3
	9	3	8	3	8	4	6	9	5	9	5	1	1	8	4	6	2
	10	4	5	4	9	2	7	7	4	2	3	2	2	9	5	8	3
Complex 2	1	4	1	4	4	5	4	4	4	5	4	3	4	3	2	4	5
	2	3	3	3	3	4	3	1	3	4	3	1	2	3	2	3	3
	3	2	4	2	2	2	2	2	2	3	2	5	1	2	3	6	1
	4	7	3	7	7	3	10	5	8	7	8	2	5	1	4	7	4
	5	3	2	4	5	6	4	3	6	8	6	9	7	6	4	6	6
	6	9	1	3	9	4	5	7	8	9	8	3	5	5	6	4	8
	7	4	3	4	4	2	3	4	9	10	9	4	2	4	5	2	2
	8	5	5	5	5	5	5	5	7	6	7	5	6	9	6	3	4
	9	6	3	6	6	6	6	6	6	7	6	7	7	7	8	4	6
	10	8	2	8	8	7	7	4	9	8	9	6	6	9	7	2	6
	11	7	1	7	7	4	8	5	8	9	8	6	2	10	4	8	2

- Swipe

Task	X Value	Y Reference	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
Simple 1	1	2	1	2	2	3	2	1	2	1	2	1	1	2	2	5	1
	2	6	2	6	6	5	6	5	5	6	5	8	7	7	6	7	3
	3	3	3	3	3	6	5	2	3	5	3	2	4	5	3	4	4
	4	4	2	4	4	5	7	3	4	4	4	4	6	6	4	2	2
Simple 2	1	5	2	5	5	4	3	4	5	6	3	5	4	6	5	5	5
	2	2	4	2	2	2	2	1	3	1	1	2	1	3	2	3	1
	3	3	3	3	3	3	5	2	4	2	2	3	6	2	3	4	3
	4	8	9	8	8	7	10	8	8	9	5	8	10	9	8	7	2
	5	6	7	6	6	5	9	6	6	7	4	6	7	5	6	6	5
Medium 1	1	2	1	2	2	2	2	1	2	2	1	1	1	2	2	1	2
	2	6	3	6	6	5	7	6	6	9	4	6	7	5	6	4	5
	3	4	4	4	4	4	5	3	3	4	3	3	6	4	4	6	3
	4	2	3	2	2	2	1	1	2	2	1	1	1	2	2	2	2
	5	8	2	8	8	7	9	10	3	6	8	10	9	7	8	7	9
	6	3	1	3	3	4	3	1	4	5	2	2	4	4	3	6	3
	7	4	3	4	4	5	4	4	5	4	3	4	6	3	4	4	2
Medium 2	1	6	3	6	6	5	6	6	6	6	5	6	6	6	6	5	5
	2	3	4	3	3	3	5	2	3	5	2	2	7	3	3	3	2
	3	2	5	2	2	2	2	1	2	2	1	1	4	2	2	2	1
	4	5	6	5	5	5	4	5	4	10	3	9	1	5	5	5	4
	5	8	4	8	8	7	9	7	8	7	8	2	6	9	8	9	5
	6	3	3	3	3	3	5	5	6	4	2	6	9	7	6	4	3
	7	6	5	6	6	4	7	8	5	7	5	5	4	3	3	6	2
	8	5	9	5	5	6	8	7	4	6	4	5	5	6	5	5	3
Complex 1	1	3	2	3	3	2	3	2	3	5	2	1	1	3	3	5	2
	2	6	3	6	6	5	5	5	6	7	5	5	5	6	6	6	5
	3	5	5	5	5	4	6	3	5	8	4	3	4	5	5	4	3
	4	4	4	4	4	3	7	4	4	2	3	2	2	4	4	6	2
	5	9	3	9	9	8	5	9	9	10	10	10	8	10	9	10	10
	6	3	1	3	3	2	6	2	3	6	2	1	6	3	3	8	4
	7	7	2	7	7	8	10	6	7	8	7	6	3	8	7	6	3
	8	4	6	4	4	3	5	4	3	7	3	2	2	7	4	4	2
	9	3	3	3	3	4	4	5	2	5	2	1	1	6	5	7	3
	10	4	2	4	4	6	5	4	3	3	3	3	2	2	4	6	2
Complex 2	1	4	3	4	4	4	5	3	4	6	3	3	6	4	4	5	5
	2	3	5	3	3	3	4	2	3	5	2	2	4	3	3	4	3
	3	2	8	2	2	2	2	1	2	2	1	1	1	2	2	2	1
	4	7	3	7	7	7	9	5	6	10	8	5	8	6	7	8	8
	5	3	4	3	3	2	5	2	2	2	2	3	6	5	8	4	6
	6	9	3	9	9	8	10	8	8	10	9	8	10	7	9	5	8
	7	4	5	4	4	2	8	3	9	6	3	4	6	4	3	9	10
	8	5	9	5	5	3	8	8	7	7	4	5	4	5	2	6	9
	9	6	5	6	6	2	7	6	2	8	5	7	5	6	7	5	7
	10	8	4	8	8	8	8	7	5	9	7	5	6	9	8	3	10
	11	7	3	7	7	3	8	5	7	6	6	5	7	8	5	4	9

APPENDIX E RESULTS STUDY 2 DISPLAYING THE DIFFERENCE BETWEEN THE TRUE VALUES AND THE PREDICTED VALUES WITH THEIR AVERAGES

	Group A (Play Button)						Group B (Swipe)					
	ΔA1	ΔA2	ΔA3	ΔA4	ΔA5	ΔA6	ΔB1	ΔB2	ΔB3	ΔB4	ΔB5	ΔB6
Simple1	2	0	1	0	0	0	1	1	1	0	1	1
	1	6	2	1	1	0	2	1	1	0	3	1
	3	2	3	2	2	2	3	3	2	2	3	2
	1	5	3	5	2	2	2	1	1	6	2	1
Average	1.75	3.25	2.25	2	1.25	1	2	1.5	1.25	2	2.25	1.25
	0	1	1	1	1	1	0	0	0	0	0	0
	2	2	0	2	0	1	1	1	4	2	1	1
Simple2	0	1	1	1	0	3	0	4	1	1	1	0
	2	2	1	2	0	0	1	1	2	3	0	1
	2	3	0	3	0	0	1	2	1	4	1	0
Average	1.2	1.8	0.6	1.8	0.2	1	0.6	1.6	1.6	2	0.6	0.4
Medium1	1	0	1	0	1	0	1	4	1	1	1	2
	1	1	1	1	1	0	1	5	1	2	2	2
	2	1	1	2	1	1	2	2	3	2	4	2
	0	4	2	1	1	1	1	3	2	1	2	5
	3	1	4	5	1	1	1	3	4	4	0	3
	6	1	2	0	1	1	1	3	3	7	1	1
	5	2	3	1	3	0	1	3	4	6	2	1
Average	2.57	1.43	2.00	1.43	1.29	0.57	1.14	3.29	2.57	3.29	1.71	2.29
	0	2	1	1	1	1	1	0	1	0	1	1
	0	3	1	0	0	1	1	0	1	0	1	1
	0	3	2	0	0	1	1	3	2	1	1	2
Medium2	0	5	1	0	0	3	1	5	5	1	1	0
	0	1	1	2	1	2	1	2	4	3	1	0
	0	9	0	2	2	1	3	3	3	4	0	1
	2	1	2	2	4	0	5	1	4	3	0	1
	1	3	2	5	1	0	3	4	2	4	2	2
Average	0.38	3.38	1.25	1.50	1.13	1.13	2.00	2.25	2.75	2.00	0.88	1.00
	1	1	2	0	1	1	1	1	1	1	0	1
	2	1	2	0	1	1	1	0	1	1	1	2
	1	2	3	1	1	0	1	1	1	1	2	1
	3	4	1	1	1	2	1	2	1	0	2	0
	3	2	1	3	0	0	1	0	3	0	1	2
Complex1	1	1	5	3	3	3	1	0	0	1	1	2
	4	1	1	2	2	2	1	1	7	2	0	5
	2	4	2	5	1	1	1	3	4	4	1	2
	4	0	3	1	4	1	1	0	0	6	1	2
	2	6	2	1	1	0	1	1	2	5	0	1
Average	2.3	2.2	2.2	1.7	1.5	1.1	1	0.9	2	2.1	0.9	1.8
	1	1	1	1	1	0	1	1	0	0	1	0
	1	1	1	2	1	0	1	2	1	2	0	1
	1	2	3	0	2	1	3	2	0	2	0	2
	1	2	1	1	2	2	1	4	3	0	0	2
	1	2	2	1	1	1	1	2	1	0	0	1
Complex2	1	0	1	1	2	0	1	4	7	0	0	1
	1	4	2	3	2	0	1	1	3	0	0	2
	1	3	4	3	1	1	3	2	4	1	0	1
	1	1	0	2	4	3	1	2	3	2	1	4
	1	0	2	1	2	1	1	4	0	0	1	2
	1	0	3	0	1	0	1	2	1	0	1	1
Average	1.00	1.45	1.82	1.36	1.73	0.82	1.36	2.36	2.09	0.64	0.36	1.55

APPENDIX F CORRELATION COEFFICIENT *R* AND INVERSE NORMAL TRANSFORMATION (INT) INVERSE NORMAL TRANSFORMATION CORRELATION IN STUDY 1

Task	Correlation	INT_Correlation
Simple-1-1	1	1.069761
Simple-1-2	0.963624112	0.961717
Simple-1-3	0.880704846	0.775102
Simple-1-4	0.692307692	0.724335
Simple-1-5	0.963624112	0.961717
Simple-1-6	1	1.069761
Simple-1-7	1	1.069761
Simple-1-8	0.538461538	0.637057
Simple-1-9	0.923076923	0.865455
Simple-1-10	0.923076923	0.865455
Simple-1-11	0.923076923	0.865455
Simple-1-12	0.923076923	0.865455
Simple-2-1	0.948683298	0.911622
Simple-2-2	0.9	0.856912
Simple-2-3	0.705023988	0.630586
Simple-2-4	0.974679434	0.985798
Simple-2-5	0.737864787	0.77641
Simple-2-6	0.737864787	0.77641
Simple-2-7	0.974679434	0.985798

Task	Correlation	INT_Correlation
Simple-2-8	0.769483764	0.83112
Simple-2-9	0.923380517	0.883289
Simple-2-10	0.737864787	0.77641
Simple-2-11	0.974679434	0.985798
Simple-2-12	0.718184846	0.702234
Simple 3-1	0.846780395	0.871281
Simple 3-2	0.981306763	1.008312
Simple 3-3	0.566138517	0.544116
Simple 3-4	0.692820323	0.702743
Simple 3-5	0.577350269	0.629182
Simple 3-6	0.808290377	0.79772
Simple 3-7	0.923760431	0.921675
Simple 3-8	0.566138517	0.544116
Simple 3-9	0.802572354	0.750232
Simple 3-10	0.615840287	0.668482
Simple 3-11	0.802572354	0.750232
Simple 3-12	0.836501913	0.831981
Simple 4-1	0.948683298	0.911622
Simple 4-2	0.9	0.856912
Simple 4-3	0.705023988	0.630586
Simple 4-4	0.974679434	0.985798
Simple 4-5	0.737864787	0.77641
Simple 4-6	0.737864787	0.77641
Simple 4-7	0.974679434	0.985798
Simple 4-8	0.769483764	0.83112

Task	Correlation	INT_Correlation
Simple 4-9	0.923380517	0.883289
Simple 4-10	0.737864787	0.77641
Simple 4-11	0.974679434	0.985798
Simple 4-12	0.718184846	0.702234
Medium-1-1	1	1.082571
Medium-1-2	0.949967907	0.937296
Medium-1-3	0.72972973	0.724378
Medium-1-4	1	1.082571
Medium-1-5	0.413377017	0.540332
Medium-1-6	0.722185381	0.641535
Medium-1-7	0.891891892	0.860018
Medium-1-8	0.885807893	0.786329
Medium-1-9	0.889232023	0.823586
Medium-1-10	0.72972973	0.724378
Medium-1-11	0.909781679	0.897275
Medium-1-12	0.961769203	0.983203
Medium-2-1	0.794719414	0.810304
Medium-2-2	0.844227042	0.849224
Medium-2-3	0.533380747	0.525226
Medium-2-4	0.953663297	0.984928
Medium-2-5	0.566177381	0.609469
Medium-2-6	0.775959197	0.713903
Medium-2-7	0.7792865	0.760404
Medium-2-8	0.582794237	0.648388
Medium-2-9	0.883191367	0.89913

Task	Correlation	INT_Correlation
Medium-2-10	0.533380747	0.525226
Medium-2-11	0.742858624	0.682317
Medium-2-12	0.7792865	0.760404
Medium 3-1	0.906581649	0.884216
Medium 3-2	0.862421616	0.771746
Medium 3-3	0.534567888	0.543189
Medium 3-4	0.920279077	0.992852
Medium 3-5	0.392884984	0.345624
Medium 3-6	0.44103144	0.45426
Medium 3-7	0.883205515	0.821025
Medium 3-8	0.648462672	0.688792
Medium 3-9	0.630196118	0.649684
Medium 3-10	0.534567888	0.543189
Medium 3-11	0.593857446	0.60969
Medium 3-12	0.68499578	0.728785
Medium 4-1	0.856176041	0.824492
Medium 4-2	0.925237406	0.890633
Medium 4-3	0.431929748	0.395088
Medium 4-4	0.815713692	0.77291
Medium 4-5	0.532220734	0.603286
Medium 4-6	0.926208307	1.004344
Medium 4-7	0.70763037	0.686082
Medium 4-8	0.683559529	0.645147
Medium 4-9	0.725078838	0.727943
Medium 4-10	0.431929748	0.395088

Task	Correlation	INT_Correlation
Medium 4-11	0.44598667	0.506737
Medium 4-12	0.528395949	0.558318
Medium 5-1	1	1.135528
Medium 5-2	0.938343117	0.999738
Medium 5-3	0.450533313	0.505043
Medium 5-4	0.934211537	0.920753
Medium 5-5	0.265967056	0.326522
Medium 5-6	0.776527636	0.755467
Medium 5-7	0.883399167	0.805457
Medium 5-8	0.753300166	0.706584
Medium 5-9	0.699234275	0.602895
Medium 5-10	0.450533313	0.505043
Medium 5-11	0.735597447	0.656594
Medium 5-12	0.891132789	0.859156
Medium 6-1	0.860655738	0.790568
Medium 6-2	0.660190162	0.60512
Medium 6-3	0.472135995	0.443324
Medium 6-4	0.913254329	1.126491
Medium 6-5	-0.038655457	0.143166
Medium 6-6	0.77071859	0.664537
Medium 6-7	0.862424816	0.865437
Medium 6-8	0.366676486	0.308216
Medium 6-9	0.800080004	0.725298
Medium 6-10	0.472135995	0.443324
Medium 6-11	0.602812628	0.544359

Task	Correlation	INT_Correlation
Medium 6-12	0.878851955	0.961441
Complex-1-1	0.840242025	0.849019
Complex-1-2	0.740665218	0.649998
Complex-1-3	0.362690538	0.423005
Complex-1-4	0.826920219	0.791778
Complex-1-5	0.176026048	0.29682
Complex-1-6	0.850420064	0.922417
Complex-1-7	0.884254702	1.048602
Complex-1-8	0.619711669	0.603544
Complex-1-9	0.615691009	0.5263
Complex-1-10	0.742765352	0.695424
Complex-1-11	0.615691009	0.5263
Complex-1-12	0.800006152	0.741878
Complex-2-1	0.554301528	0.35728
Complex-2-2	0.778862156	0.718083
Complex-2-3	0.56843214	0.440329
Complex-2-4	0.826589998	0.839313
Complex-2-5	-0.050201208	0.214501
Complex-2-6	0.794216324	0.774545
Complex-2-7	0.846945977	1.065141
Complex-2-8	0.640434955	0.66552
Complex-2-9	0.622838634	0.534028
Complex-2-10	0.636820379	0.614122
Complex-2-11	0.622838634	0.534028
Complex-2-12	0.837173523	0.922362

Task	Correlation	INT_Correlation
Complex 3-1	0.995121951	1.126629
Complex 3-2	0.811643106	0.877505
Complex 3-3	0.423240133	0.345748
Complex 3-4	0.796542016	0.806057
Complex 3-5	0	0.188241
Complex 3-6	0.730004996	0.74377
Complex 3-7	0.995112999	0.969122
Complex 3-8	0.499426296	0.437365
Complex 3-9	0.682003769	0.657435
Complex 3-10	0.636076786	0.508813
Complex 3-11	0.682003769	0.657435
Complex 3-12	0.638044581	0.5711
Complex 4-1	0.909883826	0.957809
Complex 4-2	0.808505034	0.807846
Complex 4-3	0.670911561	0.658306
Complex 4-4	0.877406732	0.862996
Complex 4-5	0.588155774	0.623401
Complex 4-6	0.764074883	0.727342
Complex 4-7	0.715405778	0.692438
Complex 4-8	0.540086203	0.585907
Complex 4-9	0.528602061	0.517584
Complex 4-10	0.39242562	0.392934
Complex 4-11	0.528602061	0.517584
Complex 4-12	0.784923549	0.764837

APPENDIX G CORRELATION COEFFICIENT *R* AND INVERSE NORMAL TRANSFORMATION (INT) CORRELATION IN STUDY 2

Task	Playback		Swipe	
	Correlation	INT_Correlation	Correlation	INT_Correlation
Simple-1-1	-0.11003	-0.03038	0.239046	0.348547
Simple-1-2	1	1.25216	1	1.241054
Simple-1-3	1	1.25216	1	1.241054
Simple-1-4	0.377964	0.344	0.426562	0.457365
Simple-1-5	0.828571	0.755928	0.722806	0.592485
Simple-1-6	-0.07968	0.020481	1	1.241054
Simple-1-7	0.714286	0.61675	0.982708	0.927705
Simple-1-8	0.831522	0.769738	0.813157	0.680445
Simple-1-9	0.982708	1.013824	0.982708	0.927705
Simple-1-10	0.980379	0.990963	0.993019	1.017097
Simple-1-11	0.770208	0.690305	0.922139	0.813456
Simple-1-12	-0.23035	-0.09105	0.903508	0.795921
Simple-1-13	1	1.25216	1	1.241054
Simple-1-14	0.659232	0.593013	0.421927	0.425801
Simple-1-15	1	1.25216	0.377964	0.390145
Simple-2-1	0.524931	0.453064	0.754245	0.613102
Simple-2-2	1	1.25216	1	1.241054
Simple-2-3	0.96719	0.929679	1	1.241054
Simple-2-4	0.370178	0.318353	0.990771	0.989127
Simple-2-5	0.878653	0.783831	0.846239	0.716726

Task	Playback		Swipe	
	Correlation	INT_Correlation	Correlation	INT_Correlation
Simple-2-6	0.814163	0.742376	0.994637	1.053347
Simple-2-7	0.938541	0.893354	0.990771	0.989127
Simple-2-8	0.887376	0.798236	0.988106	0.962707
Simple-2-9	0.903584	0.828114	0.993399	1.034802
Simple-2-10	1	1.25216	1	1.241054
Simple-2-11	0.979648	0.969446	0.859751	0.734511
Simple-2-12	0.811452	0.729058	0.917663	0.80468
Simple-2-13	0.782472	0.703041	1	1.241054
Simple-2-14	0.927173	0.859684	0.993399	1.034802
Simple-2-15	0.600751	0.534568	0.30439	0.370251
Medium-1-1	0.196282	0.190641	0.172345	0.297517
Medium-1-2	1	1.25216	1	1.241054
Medium-1-3	0.928377	0.876226	1	1.241054
Medium-1-4	0.579197	0.511349	0.936977	0.839957
Medium-1-5	0.298511	0.271506	0.985578	0.942461
Medium-1-6	-0.48828	-0.54101	0.981287	0.913263
Medium-1-7	-0.29271	-0.21785	0.423683	0.442014
Medium-1-8	0.058275	0.123163	0.76094	0.623137
Medium-1-9	0.972933	0.949071	0.972933	0.875973
Medium-1-10	0.653882	0.581241	0.98951	0.973113
Medium-1-11	0.895861	0.812985	0.933538	0.831087
Medium-1-12	0.065313	0.140952	0.949225	0.857841
Medium-1-13	-0.26444	-0.16832	1	1.241054
Medium-1-14	0.298511	0.271506	0.676503	0.559901

Task	Playback		Swipe	
	Correlation	INT_Correlation	Correlation	INT_Correlation
Medium-1-15	0.266996	0.22132	0.9275	0.822256
Medium-2-1	0.053468	0.104626	0.027588	0.181656
Medium-2-2	1	1.093495	1	1.086564
Medium-2-3	0.755126	0.677729	1	1.086564
Medium-2-4	0.374273	0.331265	0.887561	0.778426
Medium-2-5	0.915432	0.843665	0.816131	0.689625
Medium-2-6	0.687829	0.604846	0.822716	0.698726
Medium-2-7	-0.12727	-0.05919	0.774136	0.64274
Medium-2-8	0.494593	0.441293	0.665208	0.512251
Medium-2-9	0.724882	0.628735	0.976156	0.885153
Medium-2-10	0.105777	0.15808	0.188776	0.324544
Medium-2-11	0.531048	0.464785	-0.16688	0.114606
Medium-2-12	0.452267	0.417556	0.619332	0.499346
Medium-2-13	0.563955	0.488113	0.672727	0.536785
Medium-2-14	0.288725	0.236069	0.952342	0.866872
Medium-2-15	0.441367	0.40557	0.754006	0.602891
Complex-1-1	0.242876	0.206191	0.043993	0.228999
Complex-1-2	0.988027	1.038287	1	1.241054
Complex-1-3	0.123557	0.174622	1	1.241054
Complex-1-4	0.385631	0.369008	0.854228	0.725641
Complex-1-5	0.466984	0.42946	0.388666	0.408576
Complex-1-6	-0.24148	-0.12692	0.860075	0.743342
Complex-1-7	0.573884	0.499738	0.97985	0.903787
Complex-1-8	0.422189	0.393492	0.718955	0.581867

Task	Playback		Swipe	
	Correlation	INT_Correlation	Correlation	INT_Correlation
Complex-1-9	0.60894	0.557841	0.992221	1.005698
Complex-1-10	0.995495	1.06469	0.98708	0.952496
Complex-1-11	-0.08888	-0.00398	0.669726	0.524715
Complex-1-12	-0.318	-0.28069	0.798948	0.671177
Complex-1-13	0.325465	0.29193	0.948091	0.848872
Complex-1-14	-0.07791	0.043342	0.4436	0.471976
Complex-1-15	0.345517	0.305248	0.777144	0.652335
Complex-2-1	-0.42771	-0.36968	-0.49411	-0.01449
Complex-2-2	0.621565	0.56952	1	1.086564
Complex-2-3	0.964703	0.911142	1	1.086564
Complex-2-4	0.293112	0.250474	0.786353	0.661811
Complex-2-5	0.737665	0.640814	0.86784	0.752142
Complex-2-6	0.799855	0.715952	0.845339	0.707757
Complex-2-7	0.74014	0.659132	0.528239	0.485943
Complex-2-8	0.599235	0.522956	0.879274	0.769677
Complex-2-9	0.74014	0.659132	0.977164	0.894421
Complex-2-10	-0.05453	0.06486	0.868228	0.760918
Complex-2-11	0.38255	0.356576	0.767657	0.633011
Complex-2-12	0.41158	0.381308	0.889996	0.787172
Complex-2-13	0.604056	0.546193	0.673652	0.548501
Complex-2-14	0.013804	0.085235	0.149525	0.266325
Complex-2-15	0.553711	0.476465	0.695166	0.571013

APPENDIX H NORMALITY ASSESSMENT OF RMSE, CORRELATION COEFFICIENT, AND LENGTH OF PAUSE FOR EACH CONDITION IN STUDY 4

Criteria	Condition	Shapiro Wilk	Skewness	SE	Z skew	Kurtosis	SE	Z Kurto
RMSE Blind	1	.011	.670	.512	1.309	-1.037	.992	-1.045
	2	.339	.277	.512	0.541	-.957	.992	-0.965
	3	.364	.810	.512	1.582	1.182	.992	1.192
	4	.031	1.122	.512	2.191	1.195	.992	1.205
RMSE Normal	1	.731	.222	.512	0.434	-.379	.992	-0.382
	2	.692	.167	.512	0.326	-.930	.992	-0.938
	3	<.001	3.833	.512	7.486	16.01	.992	16.139
	4	.031	1.147	.512	2.240	1.311	.992	1.322
Polarity Sign Blind	1	<.001	2.188	.512	4.273	4.800	.992	4.839
	2	<.001	3.339	.512	6.521	13.116	.992	13.222
	3	.001	2.082	.512	4.066	4.411	.992	4.447
	4	.001	2.325	.512	4.541	4.898	.992	4.938
Polarity Sign Normal	1	<.001	2.191	.512	4.279	6.469	.992	6.521
	2	.001	1.210	.512	2.363	.182	.992	0.183
	3	.017	.899	.512	1.756	3.15	.992	3.175
	4	.015	.760	.512	1.484	-.510	.992	-0.514

APPENDIX I RMSE OF 20 VI PARTICIPANTS AND 20 SIGHTED PARTICIPANTS BETWEEN FOUR CONDITION

No.	RMSE VI	RMSE Sighted	Conditions
1	14.73686	13.08816	Condition 1
2	17.1508	20.61735	Condition 1
3	11.57044	29.26901	Condition 1
4	22.09864	30.60433	Condition 1
5	34.03381	34.03087	Condition 1
6	41.95742	26.36238	Condition 1
7	14.86102	41.25561	Condition 1
8	15.21348	30.11063	Condition 1
9	41.82613	19.09974	Condition 1
10	11.27054	20.06489	Condition 1
11	11.27054	31.11149	Condition 1
12	7.394255	47.00319	Condition 1
13	18.54791	15.85402	Condition 1
14	20.7804	49.03494	Condition 1
15	18.07692	29.90694	Condition 1
16	42.51088	29.40111	Condition 1
17	16.77945	43.24118	Condition 1
18	24.51377	25.1421	Condition 1
19	40.45182	35.77779	Condition 1
20	34.20234	33.745	Condition 1
1	14.11737	21.61423	Condition 2
2	20.14572	30.16994	Condition 2

No.	RMSE VI	RMSE Sighted	Conditions
3	11.00114	24.06086	Condition 2
4	25.03897	19.92988	Condition 2
5	21.44586	17.87736	Condition 2
6	27.59121	33.14853	Condition 2
7	19.62014	28.68711	Condition 2
8	22.4544	20.96306	Condition 2
9	38.35851	26.05379	Condition 2
10	8.20975	39.91209	Condition 2
11	8.20975	26.66271	Condition 2
12	6.090977	44.4089	Condition 2
13	14.15715	30.38215	Condition 2
14	22.68039	32.77766	Condition 2
15	17.03232	36.21533	Condition 2
16	38.32362	25.22747	Condition 2
17	14.82481	38.99006	Condition 2
18	30.54832	44.26426	Condition 2
19	35.51936	37.55895	Condition 2
20	34.87908	34.52427	Condition 2
1	13.96245	9.654015	Condition 3
2	15.40941	15.40941	Condition 3
3	18.6125	18.6125	Condition 3
4	9.578622	15.59407	Condition 3
5	15.82719	13.43875	Condition 3
6	27.79613	20.101	Condition 3
7	13.7104	19.9374	Condition 3

No.	RMSE VI	RMSE Sighted	Conditions
8	10.0075	14.5396	Condition 3
9	18.6715	12.49	Condition 3
10	6.254998	10.70397	Condition 3
11	6.254998	17.77076	Condition 3
12	6.69328	26.13762	Condition 3
13	12.12023	29.44444	Condition 3
14	19.25292	21.60671	Condition 3
15	88.46751	18.57283	Condition 3
16	17.08801	19.89724	Condition 3
17	14.62703	15.67721	Condition 3
18	10.29684	16.30414	Condition 3
19	15.86033	17.64724	Condition 3
20	17.48928	15.16328	Condition 3
1	4.527693	9.944848	Condition 4
2	3.697972	10.35374	Condition 4
3	7.730621	15.10132	Condition 4
4	22.72554	8.271941	Condition 4
5	21.74569	10.84435	Condition 4
6	13.67845	4.982469	Condition 4
7	4.69308	7.20243	Condition 4
8	5.612486	7.478118	Condition 4
9	15.12944	9.526279	Condition 4
10	3.567212	5.943484	Condition 4
11	3.070016	16.2957	Condition 4
12	3.232646	17.6614	Condition 4

No.	RMSE VI	RMSE Sighted	Conditions
13	7.321202	29.65257	Condition 4
14	11.85011	17.57484	Condition 4
15	16.46739	14.31084	Condition 4
16	14.21003	9.488151	Condition 4
17	11.01363	7.039176	Condition 4
18	33.72981	6.093029	Condition 4
19	9.456479	19.80215	Condition 4
20	19.7085	20.13827	Condition 4

APPENDIX J SIGHTED AND VISUALLY IMPAIRED FALSE POLARITY ESTIMATION

Number of Sighted (S) false polarity estimation from 20 sighted participants

User	Condition 1	Condition 2	Condition 3	Condition 4
S1	1	4	2	0
S2	1	1	0	5
S3	3	3	1	0
S4	0	0	1	1
S5	1	0	1	0
S6	0	2	3	2
S7	0	0	5	1
S8	0	0	2	0
S9	2	1	2	1
S10	1	1	0	2
S11	1	1	1	2
S12	2	1	6	5
S13	1	8	2	1
S14	3	1	1	2
S15	2	7	2	6
S16	1	1	0	3
S17	7	6	3	6
S18	1	2	4	2
S19	1	6	0	4
S20	3	1	1	2

Number of Visually Impaired (VI) false polarity estimation from 20 VI participants

User	Condition 1	Condition 2	Condition 3	Condition 4
VI1	0	1	1	1
VI2	1	1	0	3
VI3	1	2	1	3
VI4	2	2	1	3
VI5	0	0	2	0
VI6	3	0	13	15
VI7	0	3	0	5
VI8	0	1	1	1
VI9	4	14	6	18
VI10	0	0	2	0
VI11	0	0	2	0
VI12	0	3	2	0
VI13	2	0	4	1
VI14	0	2	4	7
VI15	0	2	9	2
VI16	7	4	0	1
VI17	0	0	2	0
VI18	1	3	0	1
VI19	10	2	0	1
VI20	2	2	0	1

APPENDIX K LINK TO CODES AND VIDEO EXAMPLES OF MAG APP

1. Condition 1: single point mode (Study 1, 2, 3, and 4)

Code: https://github.com/zicohasan/chartMTenLog

Video: https://youtu.be/zVq2bbjTXUE

2. Condition 2: single reference mode (Study 4)

Code: https://github.com/zicohasan/chartMTenLogDouble

Video: https://youtu.be/w7m6jn5I4Uw

3. Condition 3: multi-reference mapping of 20 steps (Study 4)

Code: https://github.com/zicohasan/chartMTenLog30

Video: https://youtu.be/oBrRak1ZsEw

4. Condition 4: multi-reference mapping of 10 steps (Study 4)

Code: https://github.com/zicohasan/chartMTenLog01

Video: https://youtu.be/SJQNM0LBMQI