IMPROVING ACCESSIBILITY OF SPATIAL INFORMATION: A TECHNIQUE USING PARAMETERIZED AUDIO TO SYMBOLIZE LINES

by

MEGEN E. BRITTELL

A THESIS

Presented to the Department of Computer and Information Science
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Master of Science

December 2011

THESIS APPROVAL PAGE

Student: Megen E. Brittell

Title: Improving Accessibility of Spatial Information: A Technique Using Parametrized Audio to Symbolize Lines

This thesis has been accepted and approved in partial fulfillment of the requirements for the Master of Science degree in the Department of Computer and Information Science by:

Dr. Michal Young	Chair
and	
Kimberly Andrews Espy	Vice President for Research and Graduate Studies/ Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded December 2011

THESIS ABSTRACT

Megen E. Brittell

Master of Science

Department of Computer and Information Science

December 2011

Title: Improving Accessibility of Spatial Information: A Technique Using Parametrized Audio to Symbolize Lines

Graphics provide a rich display medium that facilitates identification of spatial patterns but are inaccessible to people who are blind or low vision. Audio provides an alternative medium through which to display information. Prior research has explored audio display of lines representing functions and location of screen objects within a graphical user interface; however, presentation of spatial attributes of lines (angle, number of segments, etc.) of geographic data has received limited attention.

This thesis explores a theoretical foundation for designing audio displays and presents an experimental evaluation of line symbology. Sighted users who were blindfolded and blind users performed a line following task and a matching task to evaluate the line symbology. Observed differences between the conditions did not reach statistical significance. User preferences and observed strategies are discussed.

CURRICULUM VITAE

NAME OF AUTHOR: Megen E. Brittell

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
University of Maryland, Baltimore County, Baltimore
Johns Hopkins, Engineering Program for Professionals, Elkridge, Maryland
Washington State University, Pullman
Willamette University, Salem, Oregon

DEGREES AWARDED:

Master of Science in Computer and Information Science, 2011, University of Oregon
Bachelor of Arts in Mathematics and Russian Studies, 2002, Willamette University

AREAS OF SPECIAL INTEREST:

Human Computer Interaction
Software Engineering
Assistive Technology

GRANTS, AWARDS AND HONORS:

J. Donald Hubbard Family Scholarship, University of Oregon Department of Computer and Information Science, 2011

PUBLICATIONS:

Megen E. Brittell. Line following: a path to spatial thinking skills. In *Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 1753-1758, Vancouver, BC, Canada, 2011.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Michal Young, for his support and encouragement throughout my time as a masters student and in particular during my involvement in the Soundscape Map project. My work has benefited and continues to benefit from his insight, ideas, and open discussion of high level ideas, subtle details, and general project direction.

I would like to thank the members of the Spatial and Map Cognition Research Lab in the Department of Geography. Their insight and suggestions were influential in the development of this work.

This thesis supported in part by the National Institutes of Health (NIH NEI 1RC1EY020316). Its contents are solely the responsibility of the author and do not necessarily represent the official views of the NIH.

The Bamboo tablet used in the experiment was generously donated by Wacom.

To M & GJ

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

Spatial thinking skills are influential across multiple disciplines [29, 40, 20, 17] and are essential for independent mobility [27]. These skills are gained through experience moving in the immediate environment and can be refined with the aid of geographic maps [33, 34]. Vision supports development of spatial skills by providing access to information both in the environment and in graphic representations of maps, but the map based learning materials available to people who are blind or low vision are much more limited. Alternative modalities (e.g. sound) have the potential to increase access to maps. In addition to facilitating development of procedural knowledge to understand and process spatial information, an accessible display also provides access to the information contained in maps.

With the increasing popularity and availability of digital information, computers and graphical user interfaces have facilitated information storage, retrieval, and manipulation. Despite the variety of output options that technology provides, vision so readily communicates spatial information that visual displays have remained the primary medium for presenting such information. This is particularly apparent for geographic data. Many maps have shifted from print to geographic information systems (GIS) where map data is rendered to a computer screen in a two dimensional layout. Without non-visual displays, users who are blind or low vision cannot access the geographic information or actively participate in the design of maps. The work reported in this thesis supports development of a minimal Geographic Information System (mGIS) which will be a multimodal educational tool for teaching spatial thinking skills to students who are blind or low vision. With access to information, individuals can practice and apply their spatial thinking skills to form their own interpretation of patterns and trends in data and participate in the design of maps and other artifacts.

Common approaches to creating accessible maps include tactile graphics, use of haptic devices, and audio displays. Static tactile maps effectively convey spatial information but are relatively inflexible [71]. Collections of printed tactile graphics can be bulky, and any change to the content or map layout requires production of a new physical graphic. Dynamic tactile displays (e.g. MudPad a surface controlled by electromagnet actuators [38]) and haptic devices (e.g. Phantom[1]) are more flexible in their ability to represent changing content. These specialized hardware devices provide dynamic cutaneous and kinesthetic feedback to the user as they explore and experience information in the display. As hardware, however, cost and portability remain a concern. Some devices are expensive due to the strength and precision required of the mechanical actuators (e.g. "linkage systems" [4]). At the other end of the spectrum, inexpensive and portable vibro-tactile devices are under development (e.g. fingertip haptic device [16]).

Audio displays also constitute an inexpensive option for accessible displays and will be discussed in more detail in the following chapters. As inexpensive hardware options develop, audio techniques are revised, and usability is verified, multimodal interfaces that can present both haptic and audio feedback (e.g. haptic audio display by Brewster [12]) are likely to become more common. Such hybrid displays can take advantage of each modality to either increase the capacity of the display or provide redundant information. In either audio-only displays or multimodal displays, effective audio symbology is critical to the success of the interface. The study described in this thesis focuses on audio symbology, but also relies on proprioceptive and kinesthetic feedback through active use of a stylus (see also Section 5.2.3).

Design of an interface with audio output requires an interdisciplinary approach. A rich history in behavioral geography (Section 2.2) and map design (Section 2.1) provide perspective on how humans understand and use spatial information. The fields of psychology (Section 2.3), psychoacoustics (Section 2.3.3), and music theory (Section 2.3.3.2) share insight into human perception and information processing and provide a vocabulary with which to discuss audio events. These diverse topics have previously been brought together by existing information sonification applications and guidelines (Section 3.1) and several applications that target spatial information (Section 3.4).

While strongly interdisciplinary, this thesis looks at how computers are (or could be) involved with the ways that people access, manipulate, and process information. From the perspective of human computer interaction, this work does not attempt to redefine theories from other fields, but draws on them to inform design choices. Computer applications must provide sufficient control over the processing to the user, be able to parse and interpret input commands from the user, and

[1] Sensable, PHANTOM haptic device, http://sensable.com/products-haptic-devices.htm

format responses in a way that is usable. To create such applications, the interface designer must understand what constitutes sufficient control and what makes information perceivable and usable. Research in user interfaces (Section 2.4) highlights some the successful approaches to the design process.

The digital environment also offers the opportunity to rapidly prototype and iteratively test new systems. Prototypes and pilot testing augmented the theoretical design of the audio interface described in this thesis. Without experiencing the audio it is hard to tell how sounds will interact and how difficult or easy they will be to interpret. While watching a collaborator interact with an early prototype, colleagues observed that following lines and interpreting their shape were difficult tasks based on the combined kinesthetic feedback from the stylus and audio feedback. Insufficient detail in the feedback was identified as a possible explanation of the difficulty. As the next iteration of development, the continuous sounds described in this thesis were designed to augment the initial click symbol.

An experiment was conducted to investigate the effectiveness of three auditory symbols for line features on a map (see Chapter V). Participants were asked to follow lines and make judgements about the information contained in the display. Two audio symbols were continuous when the distance between the cursor and the line was below a given threshold or buffer width. The feedback was either constant ("on-off") or modulated based on the distance from the line ("shape"). A third technique used a short duration audio event to indicate when the cursor had crossed from one side of the line to the other ("click"). Speed and accuracy in a following task and a matching task were collected for analysis. Results and observations from the experiment and their interpretation are presented in Chapters VI and VII, respectively.

Chapter VII presents a discussion of the ways that my understanding of the problem evolved over the course of the project along with several suggestions for future investigation. This thesis concludes with a summary of the experimental findings and their implications on future tool development.

CHAPTER II

BACKGROUND

Computer displays present information through physical stimuli that invoke perceptual responses. They are communication channels that benefit from alignment between interface characteristics and user expectations. Design of these communication channels is informed by and draws background material from multiple disciplines. The work described in this thesis focuses on design of audio displays to present spatial information. The following sections introduce choropleth maps as an application domain and highlight several facets of map design and spatial cognition that relate to computer interfaces. Several topics in psychology, including psychoacoustics (the study of audio perception), are briefly presented. Exploration of interface design is divided between the vast experience in graphical user interfaces and the more recent introduction of audio interfaces. This overview presents a collection of topics that represent selected aspects of their respective fields of research which extend far beyond the scope of this work.

The research discussed in this section is supported by experimental testing with a wide variety of populations. Their findings are neither specific to people who are blind or low vision, nor do they exclude those populations from their arguments or conclusions (with the exception of visual displays).

2.1. Choropleth Maps

Display of geographic data is a natural domain in which to explore the efficacy of audio displays for presenting spatial information. Research and advances in technology have provided increasingly efficient ways to store and search spatial data (e.g. locations of objects and relationships between their positions) and increasingly flexible ways to synthesize and play audio. While maps can convey many types of information, thematic maps present trends in data that are related to spatial location [24]. Choropleth maps are a subclassification within thematic maps. They use varying color or patterns over geographically defined area (in contrast to areas defined by similar attribute values) to represent information. For example, population density can be displayed in a choropleth map to illustrate the distribution patterns (Figure 2.1).

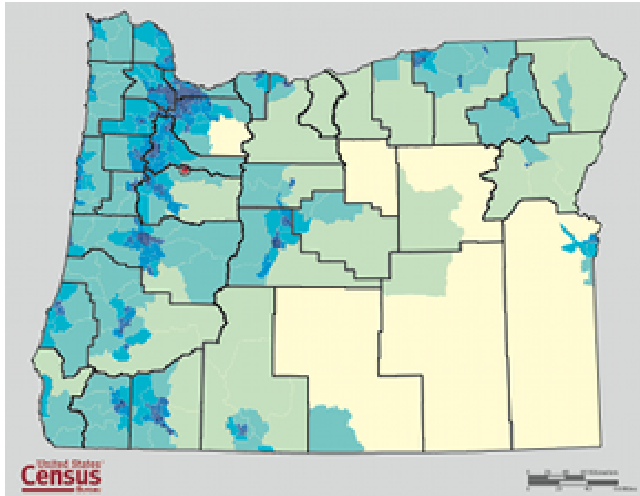

FIGURE 2.1. Choropleth maps such as this map of population density show trends in data over geographic space. This example shows higher population density on the western side of the state of Oregon and in particular along the Willamette River. (Map derived from U.S. Census Bureau, *2010 Census: Oregon Profile, Population Density by Census Tract*, U.S. Census Bureau: Public Domain, 2010)

2.1.1. Figure Ground Relationship

Thematic maps are made up of geographic context and information. Dent [23] details the relationship between these two aspects of the display as a figure-ground relationship. Geographic context helps the map user interpret information relative to its location in space or proximity to other bits of information. Context helps form the structure of a semantic hierarchy through which to organize information. The context is the ground in which information is placed. The information, or figure, is made up of the facts or variable attributes of the map elements. In tandem context and information accomplish the task of communicating a message to the map user. As part of cartographic design, the figure-ground relationship helps map makers organize map content and consistently convey information through maps based on characteristics human perception. Like the perceptual effects from vision that Dent described, audio perception possesses similar perceptual traits that may help convey the figure-ground relationship between display elements (see also Section 2.3.5).

Over many centuries, cartographers have gained extensive experience in design of visual artifacts to convey spatial information. From this experience, map designers have developed a technique of varying intensity or saturation to help direct focus to the primary map features (i.e.

the figure) while leaving background or reference markers available. Similarly parameters of sound can be adjusted to make context more or less prominant.

2.1.2. Lines in Choropleth Maps

In choropleth maps, lines are often part of the context[1]. Lines may mark the edges of two dimensional features (e.g. boundaries of counties) or represent linear map features (e.g. rivers) that help the user place the display in a larger scene. Depiction of lines in audio displays of quantitative data graphics has been studied previously (see also Section 3.3.1), but sonification techniques cannot necessarily be generalized across applications. Presentation of lines as context in an interactive audio map display is less well understood.

Location and shape of lines are expected to play important roles in understanding the information conveyed in choropleth maps. The project described in this thesis uses proprioceptive feedback (see also Section 5.2.3) to provide some location information (e.g. location of the cursor relative to the user or relative to the edges of the active tablet surface) but also relies on audio to convey context (e.g. location of the cursor relative to a line). In combination, the proprioceptive and audio feedback are intended to convey the shape of the line.

2.2. Spatial Thinking

Computers access data from both memory and input channels to process that data through operations on available hardware registers. Similarly, people combine information from memory and new sensory input to process and understand spatial information. With the help of working memory processes, people can find patterns in data and make judgements about the spatial relationships between objects in space. Long term memory (stored data), audio or visual input (digital input), working memory (registers) are all essential components of a functional system. Where computer science has explored the interaction between hardware components to facilitate digital processing and algorithms to manipulate spatial data, behavioral geographers have studied the processes through which humans understand, process, and describe space.

Through interaction with our environment, humans have necessarily developed ways to understand and communicate information about space [43, 56]. By aligning the way that people

[1] This is in contrast to, for example, lines that create a hatch pattern as a data symbol on a visual representation of the map or lines in an isopleth map (another type of thematic map) that represent information.

naturally think about spatial information with the way that the interface presents that information, interfaces become more intuitive and usable. The following sections will highlight aspects of spatial thinking and spatial cognition that influence interface design.

2.2.1. Egocentric and Allocentric Perspectives

Spatial data is understood in some frame of reference that may be relative to the observer (egocentric) or external to the observer (allocentric)[2]. The egocentric frame places the observer in the center of the frame and provides spatial cues relative to that central position. This perspective is common in navigation and is developed at an early stage when building spatial thinking skills [63, 44]. The observer can reason about the space around them in a way that directly reflects their experience in that space.

An allocentric frame of reference is external to the viewer. Features in the environment are understood in relationship to a fixed frame of reference (e.g. a telephone pole does not move relative to the buildings around it even through it does move within the egocentric perspective of someone passing by in a car).

Information in a choropleth map (and more generally in thematic maps) does not have an inherent relationship with the observer. Instead, the important relationships are between the map elements (independent of the viewer). This suggests that an allocentric perspective may be well suited to address tasks that require interpretation of data from a choropleth map. This does not, however, imply than an egocentric perspective could not result in successful performance of the same tasks. There may simply be a difference in the speed and accuracy with which tasks are completed (e.g. judging distance may take longer from an egocentric, or route, perspective than it would from an allocentric, or survey, perspective [62]).

In addition to the natural or preferred frame of reference for the display content, interface modalities may also suggest different representations. Kinesthetic feedback is typically thought to produce an egocentric representation, but does not preclude alternative representations of spatial relationships such as distance and angle [45]. Spatialized audio (see also Section 2.3.3.3) also suggests an egocentric representation which may make it less well suited (although not unusable) to presentation of maps that target a survey representation [36]. Although it may not be ideal, the

[2] Brewer and Pears [11] provide a more thorough discussion of reference frames.

stylus input (providing kinesthetic feedback from an egocentric perspective) was used in the current implementation of the interface used as the test instrument for this thesis. Spatialized audio was not included in the current implementation.

In many studies, subjects may choose for themselves which perspective or representation to use and may be influenced by either personal preference or the given task. Level of visual experience may further influence selection of a frame of reference (e.g. people who are cogentially and early blind tend to adopt a frame of reference relative to their body coordinates despite the fact that it may not be the most efficient choice [75]). When comparing visual and verbal representations of spatial data Péruch et al. observed that "the perspective from which [information] is presented (survey or route), [...] may affect the structure and accuracy of the mental representations" [62]. The potential influence of the reference frame may not always be explicitly controlled, but needs to be taken into consideration when making any inferences based on performance.

2.2.2. Mental Maps

The terms 'cognitive map' and 'mental map' refer to peoples' internal representation of spatial data that includes locations and attributes of objects along with relationships between features. Mental maps are not, however, an exact copy of the space that they represent. Ungar proposes that users "integrate experience from different senses to build up abstract (amodal) concepts" [75]. As an artifact of creating and manipulating mental maps, spatial data and spatial relationships they contain are often simplified or innacurate [56].

Distortions that occur in mental maps may be due to a variety of influences. These include mistaken judgements of euclidean distance due to the number of features that reside in the space [32] or the paths that connect points [49]. Modality and perspective may influence the choices people make in reducing complexity or quantity of details between the physical stimulus and the internal representation [62].

In contrast the to a behavioral approach of modeling observed performance, Kuipers [47] approaches the problem from the perspective of artificial intelligence and builds up a model that would meet task requirements. He argues a that mental maps must necessarily possess multiple frames of reference to overcome resource limitations and form a reliable representation of space. This perspective may help reason about how people are able to build a coherent overview based on isolated observations, which is a skill that will be critical to the success of audio display proposed in this thesis.

Using knowledge of how people understand and represent spatial information, interface designer can match display design choices to users expectations making geographic information systems easier to use [30, 56]. The audio (and tactile) modalities have lower resolution than vision, and as a result the data must be simplified to maintain usability when designing for non-visual displays. Katz et al. [41] simplify the data for presentation in an audio only display by reducing the number of variables displayed at a single time. Analogously the data used in this thesis is limited to a single line in each display and focuses on declarative attributes of lines (e.g. number of segments) and simple relational characteristics (e.g. linearity of segments, concavity of lines). The design errs on the side of simplicity and in the future will increase complexity as user interaction with the display is better understood.

2.3. Shared Concepts from Psychology

In addition to the preceding cognitive aspects of map use that are specific to understanding spatial data, more general cognitive theories are also applicable to interface design. Among these, the efficacy of audio interfaces are heavily influenced by working memory, encoding strategies, and perception. Perception will primarily be discussed in the context of audio; some parallel discussion of visual perception is included to show how decisions from visual cartographic design may transfer to an audio display.

2.3.1. Working Memory

The term 'working memory' describes a set of cognitive resources that support manipulation of data [3]. It has been divided into to parts: the phonological loop that handles speech and non-speech audio and visuospatial sketchpad for visual input. The use of these resources is closely tied with the input modality. It is important to note that it is a limited resource. Human working memory can hold only a small number of items at any one time.

Working memory has been highlighted as a challenge in general design of artifacts [61], in implementing computer interfaces [68] and especially in audio interfaces (e.g. [14, 2]). Rather than requiring people to memorize all the details of their environment, Norman [61] advocates making information available in the environment, distinguishing between "knowledge in the head and knowledge in the world." People remember a few salient details about the things around them and refer back to the environment to fill in any gaps in their knowledge. In computer interface design, limiting the number of items that users are required to remember is among Schneiderman and Plaisant's "eight golden rules" [69]. The rule is defined in the context of information availability from one page (or screen) to another in a visual display. Its interpretation can also carry over to an audio display where, unlike in a visual display, the information is not readily available for frequent rapid refresh. The audio interface must minimize the number of distinct items that the user must hold in memory. This may take the form of a limited symbol set or the explicit synthesis of multiple pieces of information into a single item.

In audio interfaces, sound's inherent dependence on time compounds the strain on working memory [65]. While vision provides information about a broad two dimensional area, audio must exist over time and the listener must remember and process those audio events to extract their meaning. This leads to serial presentation of information which may overwhelm the limited working memory resources (e.g. as noted by Brewster [12]), concurrent presentation which may impede perception of the intended information [55, 14, 9], or a combination of both serial and parallel presentations (e.g. as proposed by Brown et al. [15]).

Users of audio displays have had difficulty retaining enough details in memory to synthesize an overall interpretation of the display contents. These difficulties can be lessened by providing context. For example, Alty and Rigas [2] found that users were more successful in identifying the unified shape of display components (e.g. the letter 'E' made up of four rectangles) when they were primed to expect a certain type of figure (e.g. letter of the English alphabet). Providing structured organization of the information may also help reduce the demand on working memory [55, 14]. The approach in this thesis was to simplify the display content and limit the number of features that users are required to remember between exploration of a display and making judgements about the features in that display. When the display becomes more complex, context and hierarchical structure of information will gain importance.

2.3.2. Encoding Strategies

Information from the environment is encoded into an internal representation for mental manipulation or storage in memory. As mentioned above in reference to mental maps, this encoding is not necessarily a precise copy of the information it represents. There are multiple encoding strategies (e.g. auditory, linguistic, visuospatial) that may emphasize different characteristics of the information and make various manipulations of the data more or less difficult. Even when data comes from the same stimulus, different features of that data (e.g. the "what" and the "where" of spatial information) may each be encoded differently in working memory [54].

Selection of a strategy may be based on previous experience, comfort with a specific strategy, or personal preference, however, the chosen strategy may not be the most effective strategy for the task [54, 40]. Similar to the the way that perspective may influence task performace, Klatzky and Lederman [45] and Afonso et al. [1] noted that translation between different internal representations takes time and may introduce error. Rather than arbitrary choices in design, interfaces should reflect and accommodate trends in how people naturally understand and interact with information.

The cognitive processes used to store and analyze information varies greatly between individuals and also from one task to another (e.g. comfort with strategy [54], attention based encoding [60], representation of space [45]). Some studies have tried to control for encoding strategy by directing attention toward an encoding scheme [60] or asking participants to perform a secondary task that would preclude use of a specific encoding (e.g. counting backwards during a navigation task to avoid a linguistic representation [74]). The study described in this thesis does not try to constrain or capture the mental processes involved.

2.3.3. Psychoacoustics and Perception

With the purpose of conveying information, a usable display must not simply produce sound waves, but must produce sound waves that can be perceived and understood by listeners. This section describes psychoacoustics, the study of how humans perceive sound.

2.3.3.1. Physical Sound and Human Perception

Sounds travel as longitudinal pressure waves [35] that interact with one another (e.g. additive volume) as they travel through a transmission medium (e.g. the air). These waves can be described by their physical characteristics. Their interaction with objects in the environment and with other sound waves can be predicted by laws of physics. The ear receives sound waves as fluctuations in pressure that cause the eardrum to oscillate. As an audio stimulus reaches and passes through perceptual processing, the interaction between sounds from different sources must be interpreted by the listener. Although physics may be able to model the sound waves, influence of interaction between concurrent sounds on perception is complex.

2.3.3.2. Music Theory

Music theory provides both a vocabulary with which to discuss audio events (notation) and cultural framework in which to discuss perceptual qualities (e.g. dissonance). The structural components that distinguish music from noise follows a hierarchy (e.g. movements and phrases) which can be leveraged to convey the hierarchically structured data in a map display.

Although this work will not extensively discuss the music theory aspects of audio design, one way to describe the differences among sounds is of particular interest. Regarding frequency, the interval between two notes is a measure of the difference between the fundamental frequencies of the notes on a logarithmic scale. One example of a common interval in western music is the octave in which one frequency (f_2) is twice that of anther (f_1).

$$f_1 = \frac{1}{2} f_2$$

Frequency is a continuous variable, but western music breaks up the octave into just twelve tones. Rather than arbitrarily selecting intervals, the audio symbols in this thesis are separated by at least one step on the twelve step chromatic scale of western music (see also 2.3.3.3).

2.3.3.3. Dimensions of Sound

Sound can be described in terms of both physical properties and perceptual qualities. Levitin [52] states that "the basic elements of any sound are loudness, pitch, contour, duration (or rhythm), tempo, timbre, spatial location, and reverberation." All dimensions of sound are present in any audio event and can be modified or parameterized in synthesized audio.

This thesis describes an experiment that selected, modified, and evaluated symbols that controlled parametric settings for loudness, pitch, and timbre (Table 2.1). The other dimensions

Dimensions of Sound used as Parameters

Physical	Perceptual
sound pressure level (SPL)	loudness
frequency	pitch
spectral envelope	timbre

TABLE 2.1. The dimensions of sound can be labeled in two domains. These pairs of labels are not redundant but instead match a physical property that can be configured (left) with the desired perceptual characteristics (right).

of sound were held constant across all conditions, effectively limiting the vast number of possible combinations of parameters.

Loudness (Volume): Loudness is the perceptual quality that relates most closely with the sound pressure level (SPL) and physical amplitude of the sound wave. SPL is measured in decibel (dB) and loudness is measured in sones.

$$1\ sone = \text{loudness of a } 1000\ Hz \text{ tone at } 40\ dB$$

The loudness level in sones increases (or decreases) as intensity increases (or decreases) based on the 1 sone reference level. Phons are the units used to label the lines of constant loudness when frequency (Hz) is plotted against intensity (dB) (see Figure 2.2; see also [35]).

There is a positive correlation between SPL and loudness (i.e. when SPL increases, loudness increases), but SPL is not the only contributing factor to loudness. The curves shown in Figure 2.2 show how the SPL must change to maintain constant perceived loudness as frequency changes. For example, a tone at 10 Hz must be played at above 90 SPL to achieve the same loudness as a 1000 Hz tone played at 40 SPL. Loudness is one example of how two physical dimensions (frequency and SPL) interact to influence a single perceptual quality of sound.

Loudness may be modulated by the transmission medium through which it passes. The sound is perceived as louder near the source and softer as distance from the source increases. As a parameter of sound, amplitude (the physical characteristic heavily influencing loudness) can be varied to modify the perceived loudness.

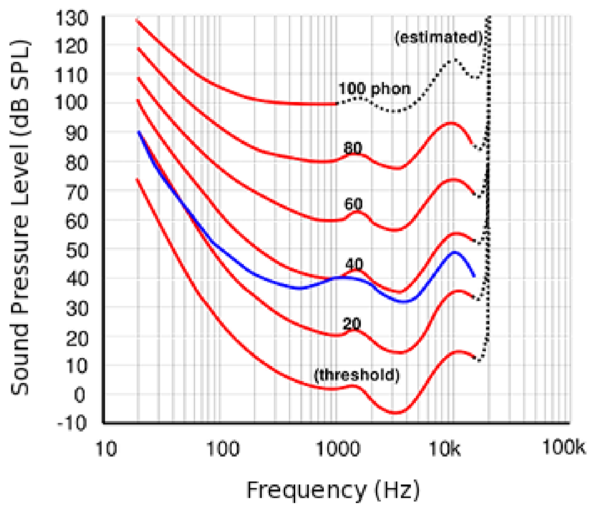

FIGURE 2.2. Equal loudness contours from ISO226:2003 revision (red) show the relationship between the physical sound pressure level (SPL) and the perceived loudness as a function of the frequency of a sound. The original ISO standard is shown in blue for 40-phons. (Image by Lindosland, *Lindos1.svg*, Wikimedia Commons: Public Domain, 2005)

Pitch (Frequency): Pitch is the perceptual counterpart to frequency. In a pure tone, pitch is directly related to frequency. In a note that is composed of concurrent integer multiples of a frequency (see also 2.3.3.3), the lowest harmonic, or fundamental frequency, determines the perceived pitch.

The interval between the harmonics is the most prominent characteristic for perception of pitch. The "missing fundamental" illustrates the binding between the frequency interval that separated components of a sound (integer multiples of the interval) and pitch. Two notes that contain the same overtones but in one of which the fundamental frequency has been removed are perceived as having the same pitch. This phenomenon of perception means that filters may be applied to a note removing a subset of the harmonics without changing the perceived pitch.

A subset of the population can identify absolute pitch values and are described as having perfect pitch. For a much larger majority, absolute pitch is difficult to identify, but relative pitches and changes in pitch can be perceived. Most people can reliably discriminate between frequencies that are separated by at least 25 cents [18]. A cent is a measure of similarity between frequencies. Cents are computed based on the equation:

$$log_2 \frac{frequency1}{frequency2} \times 1200$$

As an example, one octave (an interval over which frequency doubles) contains 1200 cents. A semitone (or half step) in musical notation is a change of 100 cents. The experimental symbology in this project used frequencies that differed by at least 200 cents (one whole step).

Timbre (Spectral Envelope): Timbre describes the overall quality of a sound and encompasses multiple physical attributes of the sound wave. There is no universal definition of timbre, but the spectral envelope of sound is widely accepted as a contributing factor [48, 52, 81, 55].

Natural sounds (i.e. those produced through vibration of physical objects as opposed to computer synthesized tones) are composed of a collection of harmonics. The term harmonics refers to the fundamental frequency and any overtones (i.e. integer multiples of the fundamental frequency) [35]. Overtones occur in different proportions based on the resonant cavity of the instrument (e.g. vocal tract or tube shaped body of a woodwind instrument). The term spectral envelope refers to the "shape of the power spectrum of sound" and is a characteristic of sound that aids identification of sound sources [81]. For comparison, Figure 2.3 presents examples of the harmonics that describe simple synthesized instruments. The 'woodwind-type' instrument is a collection of harmonics to approximate a simplified clarinet timbre in which the odd number harmonics are stronger than the even numbered harmonics [35]. For physical instruments, the harmonics for a single instrument vary slightly as frequency changes (i.e. when the instrument plays different notes) [58]. The spectral envelopes of the two notes are not identical, but they are more similar to one another than they are to the envelope of a note played on another instrument.

Musical Instruments typically have a fixed shape (i.e. the physical body of the instrument), and frequency is modified by manipulating the volume of the resonant cavity and frequency at which a source of vibration oscillates [35]. For example the trombone is stretched longer by extending the slide and on a woodwind instrument additional keys are pressed, effectively closing additional holes in the body of the instrument, to make the tube longer. Changing the tension in the source of vibration further modulates the frequency of the produced notes. The association between spectral envelope and the physical body of the instrument encourages the listener to associate sounds that display the same (or very similar) spectral envelopes as having come from the same source (see also 2.3.5).

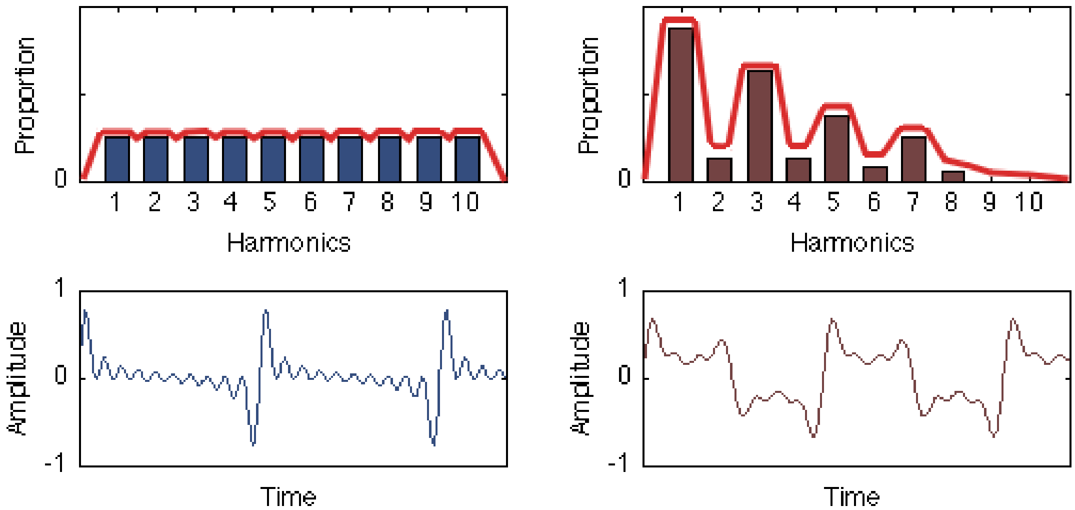

FIGURE 2.3. The spectral envelope (top) and the time domain waveform produced by those harmonics (bottom) of two simple synthesized instruments. Harmonics that have even proportions (left) are compared with a woodwind-type instrument (right). The fundamental frequency of both instruments is middle C (C4 ≈ 261 Hz).

There is not a natural ordering of different timbres so the spectral envelope is not well suited to representing ordered or continuous variable. Timbre had been recommend and used to represent distinct objects (nominal data) in audio displays [13, 46].

Spatial Location As sounds propagate from their sources they are modulated by the environment. In addition to the impact on loudness mentioned above, reflections and environmental filtering modulate the characteristics of sound in such a way that reveals the direction from which they arrived. For example, people can determine the relative location of a sound source by processing input from both the right and left ears. Further, cognitive processing that tracks pitch over time can determine the direction of a sound source in motion [5]. Spatialized audio is synthesized audio that simulates position and modulation from the environment to evoke perception of source location. As observed by Katz et al. [41], spatialized audio is more common in virtual reality applications than in data sonification.

2.3.4. Spatialized Audio

The last of these dimensions of sound, spatialized audio, has received substantial attention from researchers of audio interfaces and virtual environments. Spatialized audio places the listener in the center of the audio scene so the listener receives information from an egocentric perspective (see also Section 2.2.1). People's ability to separate audio streams based on their spatial location can be used to convey multiple discrete input streams [14]. Conveying allocentric or survey representation of spatial information through spatialized audio places strong emphasis on distance cues [36] and requires the user to mentally transform the perspective of the perceived information. Consider for example, standing in the middle of a small park and hearing sound from a source that moved around the perimeter of the park. Without vision, perceived distance from the sound source would be the primary cue by which to determine the shape of the perimeter (e.g. circular or square). With its focus on survey representations of spatial information, this work does not incorporate spatialized audio.

Spatialized audio and pan setting embody an egocentric perspective, but Brown et al. [14] also use this dimension of sound as a tool to facilitate audio streaming. In a display of data series and tables, different objects were encoded as distinct spatial locations. As sound pans from left to right, for example, users can take advantage of the spatial cues to maintain awareness of the position in the table [14]. Although spatial audio holds potential for conveying spatial information, it may not be well suited to conveying an allocentric perspective [36] (see also Section 2.2.1). Further, the need for specialized hardware (speaker array) or headphones may be prohibitive or undesirable (e.g. headphones may isolate the user form the surrounding environment) [55].

2.3.5. Streaming and Segmentation

As the environment continuously feeds our perceptual system with input, grouping and filtering sounds based on their characteristics allow us to "associate [incoming information] with discrete objects belonging to the external world [70]." Streaming describes the decision to group together a collection of audio characteristics into a unit. Bregman [10] and Winkler et al. [82] have both drawn a parallel between a stream in auditory perception and an object in visual perception.

With generated or synthesized stimuli, interface designers must carefully consider the desired perceptual qualities of the display and identify the combination(s) of physical characteristics that will achieve that result. A poorly planned display may produce ambiguous feedback or interact in ways that negatively impact perception causing communication from the display to the user to fail. Evaluation of the way that combinations of sounds influence perception is called auditory scene anlaysis (ASA) [10].

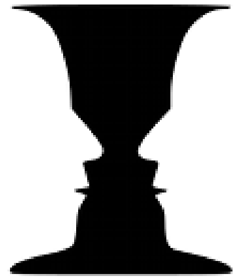

FIGURE 2.4. Rubin's Vase illustrates the visual illusion created by an ambiguous boundary. (Derived from image by Alan De Smet, *Multistability.svg*, Wikimedia Commons: Public Domain, 2007)

An example from vision illustrates how we can produce stimuli that cause ambiguity in perception. Physical objects are typically defined by the extent of their component parts (e.g. a book made up of pages and a cover). The boundary between an object (e.g. a book) and the rest of the world (e.g. other books, the shelf on which a book sits, or the air around a book) can be described as lying at the extent of the molecules that make up the object. The boundary can also be described as belonging to the object. Through interaction with the physical world we are trained to distinguish between a physical object and the rest of the world. There are examples, however, of ambiguity in the assignment of a boundary to an object. In the case of Rubin's Vase (see Figure 2.4), the viewer may see either two faces, assigning the boundary to the white portion of the image, or a vase, assigning the boundary to the black portion of the image. The assignment may shift between the two objects, but follows the principle of 'exclusive allocation' belonging to only one object at a time [10]. Similarly, there are boundaries between audio objects and Winker et al. [82] determined that attention can influence choices in assigning boundary.

Many factors can influence streaming including time of onset and similarity of the sounds (e.g. similar frequency or similar pattern of notes over time). When audio feedback is designed within a semantic hierarchy of earcons, that structure simplifies learning [7, 55], but also increases the probability that sounds will share characteristics [55]. After an audio scene has been segmented into one or more streams, the listener must interpret their meaning. Even through humans are good at picking out a single spoken stream from background noise (often called the "coctail party effect" [19]), non-speech sounds can be more difficult to separate and often only one of the streams can be attended at once.

In the display of spatial data it is common that multiple features would need to be conveyed at the same time. For example if the stylus is touching the representation of a river that flows through a park, the stylus is both in the river and in the park at the same time. The audio event that symbolizes the river and that of the park may be presented serially (i.e. one after the other, separated in time) or in parallel (i.e. concurrently, separated in a dimension of sound other than time). Serial presentation requires more time and is thus more taxing on working memory (see also Section 2.3.1). Another example would be a choropleth map that shows bivariate data (e.g. population density classified into a discrete number of levels and education levels). The ability to explore such overlapping data allows users to draw conclusions about trends in spatial data. Although a table filled with numbers may represent the same information, it can be more difficult to interpret the trends in that format.

Segregating audio into distinct streams will be critical to the success of the final minimal geographic information system. The experiment described in this thesis takes an initial step toward evaluating simple audio symbols that are presented concurrently; a background noise will always be present when the user is listening for the line symbols. While the dimensions of sound can be measured and parameterized in isolation, they occur in combination. The interdependence and interaction between the various dimensions produces a complex audio result. The perceived whole (overall scene) must be understood as more than the sum of its parts - both for usability and for aesthetic qualities.

In addition to consideration of the psychoacoustic properties of sound, interface design and the application domain further shape development. The next section turns to discussion of the user interface design.

2.4. User Interfaces

As the computer has evolved, so too have the methods of user interaction. User interfaces have been designed to convey information through all of the human senses. Implementations vary widely based on influences from both application context and available hardware, but there are several common threads that inform design.

One influential contribution to user interface design were made in the form of guidelines by Schneiderman and Plaisant [69]. These guidelines are not dependent on interface modality, however special consideration is required when the modality constrains design decisions to a space where two (or more) of the guidelines conflict. There are places in the design of audio interfaces when a balance must be struck between two of the golden rules. For example, Rigas and Alty [66] found that in their interface, consistency was more important than explicit feedback.

2.4.1. Graphical User Interfaces

The primary display in modern software applications is a graphical user interface (GUI). Users focus their point of gaze and fovea on an area of interest for high resolution details, review the screen for reminders of system state when necessary, and use peripheral vision to aid search.

2.4.2. Speech Interfaces

Within the audio display research, several classes of audio feedback have emerged. At the top level, audio feedback can be categorized as either speech or non-speech audio. Verbal (speech) feedback delivered through text-to-speech applications is a commonly commonly used mechanism to improve accessibility of information by providing an auditory representation of natural language. This is a practical approach to accessibility when the original data is natural language (i.e. text or other narrative), but has also been explored as an accessible alternative to images and graphics. An example of this is the W3C guidelines [77] that encourage the originators or preparers of the content to specify alternative text to accompany images.

One drawback of speech feedback is the time it takes to express complete words and phrases. Text-to-speech applications typically allow users to set the rate or speed of the voice. The approach of speeding up spoken text was taken even further with the introduction of Spearcons [79]. Spearcons increase the rate of speech to the point where the sound becomes an icon. Presentation is much shorter but recognition is still higher than with non-speech audio events (see also Section 3.1.1).

Speech efficiently conveys facts and nominal data, however, natural language phrases have difficulty describing spatial patterns and relationships [55, 40] and conveying procedural information [26]. In several publications, researchers have argued that natural language is ineffective for spatial knowledge [62, 56, 40]. For example a map may have a title and description that can easily be converted to verbal feedback, but the distances between cities and distribution of data would rely heavily on interpretation by the designer creating the display. Similarly, when describing a painting, text can easily provide the name of the painting and the name of the artist (i.e. nominal data), but has difficulty capturing the nuances or relevance of relative positions between elements of the image. A wealth of possible descriptions forces a choice between long narrative that tries to capture multiple aspects of the spatial layout and efficient summaries that necessarily exclude some details. Due to overlapping challenges between signification of images and signification of maps, applications that address both of these display needs will be discussed in the related work (Section 3.4).

Speech audio will likely be incorporated into future iterations of this project to present elements of the application interface (e.g. menus) and for presentation of nominal data (e.g. state names). At the present time, however, this thesis addresses concerns about display of elements (i.e. geographic data and spatial relationships) for which speech is less suitable and focuses on systems that use non-speech audio.

2.4.3. Accessible Interfaces

Graphical user interface elements have been presented to users through text-to-speech applications (e.g. menus), with non-speech audio feedback to indicate cursor position relative to display features, or with non-speech guidance to locations within the display (e.g. buttons) [64, 57, 71, 51]. Within the user interface, a target element can be specified through keyboard input or determined based on system state. Given a known target, audio feedback guides the user as they navigate through the application interface or move a cursor to a position in the display coincident with that target element. Locating a known object in the display is a slightly different task than exploring and gaining an overview of the layout of spatial information. There is no specific display feature to which to direct the stylus.

Accessibility, however, does not necessarily imply usability. Beyond the surface pointing and selection tasks, conceptual design of the interface is essential to usability of the audio interface. The Mercator Project [57] used audio feedback to guide the user through a hierarchical model of the user interface linking interface widgets to one another and using keyboard input for navigation.

W3C developed Web Content Accessibility Guidelines (WCAG) [77] to encourage techniques that make website content accessible, but the conceptual design of usable websites is underspecified [51]. The most common accommodation is compatibility with screen readers. In practice, this means that images are tagged or labeled with narrative text and guidelines recommend a hierarchical organization of graphic elements [26]. Alternate text is a step toward accessibility, but still leaves users who are blind or low vision dependent on designers to interpret and/or describe graphic information.

2.4.4. Dual Interfaces

Dual user interfaces go beyond a one-to-one mapping from a primary modality to an alternative modality. Instead, salient elements and structure of the display are identified and matched to the most suitable interaction metaphor for each modality [67]. The goal is to create simultaneous displays in more than one modality that are equivalent in content and usability.

Due to the rich information content in maps and their multiple uses, there is not a single 'correct' hierarchy of map elements. When searching for medium size cities, a natural organization of city data would be in a structure organized by size. In contrast the same cities may be organized by their respective latitudes to find the northernmost city on the map. As described by Brown et al. [15], within the audio modality there may be a need for multiple modes or configurations based on the task. This would be in addition to considerations for other modalities.

2.4.5. Implications on Design of an Audio Display

Ultimately this research will contribute to an application that includes interface navigation, but the current focus is on the display of spatial content. In contrast to an application interface where the goal is to acquire specific target positions (e.g. buttons) or select actions from among available options (e.g. menu), an interactive map display conveys the existence and characteristics of multiple objects and the spatial relationships between them by allowing the user to determine the speed and order of exploration.

As the project develops, the design will incorporate an increasing number of techniques that leverage the strengths of the audio modality (and auditory perception) and not restrict development to a potentially ill-fitting extension of an existing visual display. In the current stage, however, the interface design strongly reflects the layout of a visual display. A graphical version of the display is rendered to the screen and feedback is based on stylus position that has a one to one mapping to position in the visual display. As the project matures, users will be able to specify which attributes are salient customizing their interaction based on their current task.

Content in audio displays is often simplified to improve usability. Mental representations also include simplifications (see also 2.2.2) and design choices that mirror the choices a human would make may improve efficiency and usability of the resulting application. Choices that break expectations of the map reader may lead to misinterpretation or confusion. When making choices in simplification or distortion, the map maker and the map user must share a common schema. Rather than preserving accuracy of all relationships depicted in the map, the schema can provide a consistent interpretation of the inaccuracy.

Similar to the guidelines for user interfaces, cartographic principles also encourage consistency and hierarchical structure. The attributes of design are important for learning interfaces and helps minimize the load on working memory. In cartography, however, there is a greater focus on the balance between simplicity and completeness. As a representation of a larger space, the meaning of a map and interpretation of its component features are dependent on the context. While minimizing clutter, the interface must supply sufficient context to place the components of the display in a broader space. The context helps establish a shared schema between the map designer and the map user, but the display still needs to be clear about what assumptions and simplifications (i.e. sacrifices in terms of accuracy of the depiction) have been made.

The original data selected for this study explicitly limited complexity. An effort was made to create a data set that was representative of the data that would be found in a true map while eliminating many potential confounds (e.g. complexity of intersections and existing expectations of the shape of classes of map features). As complexity is reintroduced, methods of simplification and sonification will need to be revisited.

CHAPTER III

RELATED WORK

Audio interfaces have become popular for augmenting visual displays and for sensory substitution. Their design considers and tries to complement the natural ways that people gather and manipulate data (including but not limited to spatial data) and explores how computers can help mediate this interaction with data. In an interactive application, the computer fills one end of a bidirectional communication in which the user gives instructions and the computer provides responses. This thesis looks at audio as a display (output) option; implicit haptic feedback (in contrast to explicit force feedback) through active manipulation of a stylus is essential to the success of the display, but is not the primary focus. This chapter provides an overview of previous research in audio display and specifically audio display of spatial information.

For many years, researchers have explored computerized audio and its potential in computer displays (see also Section 3.1). Implementations have targeted both sighted users and users who are blind or low vision and many of the results are generalizable. There are, however, observable differences in strategies and preferences between these groups [51, 40, 78, 36]. Among the references cited in this chapter, all of the studies evaluated their hypotheses through experimental testing with users who were blind or low vision or with sighted users whose sight was temporarily restricted (e.g. participants were blindfolded).

3.1. Auditory Computer Interfaces

Researchers have explored audio as a potential display format to either augment or replace visual displays. In many existing applications, audio is used to enrich or extend displays in other modalities. For example, in a molecular biology application, audio presents one attribute of the data while the remaining attributes are displayed on the screen [31]. Hybrid haptic-audio displays often represent one or more attributes of the data in audio while a haptic (force feedback) display reflects shape or texture of a display element (e.g. the system designed by Brewster [12]). The audio may display information that either complements or reiterates that of other modalities. As mobile devices have become popular, sound has been explored as a way to supplement limited screen space. In addition to increasing the display space, multimodal displays also allow application developers to make the most of human abilities in information processing [33]. Although there are many promising uses of audio to augment display, this thesis focuses on audio for sensory substitution where the audio channel needs to convey all facets of the information not just the ones unavailable in or replicated by the visual or haptic display.

As noted in Section 2.4.2, text-to-speech is a common approach to accessibility, but may be inefficient or ineffective in conveying spatial information. Incorporation of text-to-speech capability to aid menu navigation in future versions of a minimal geographic information system is expected, but has not been implemented at this time.

3.1.1. Non-speech Audio Interfaces

As anticipated by Krygier in 1994 [46], the availability of MIDI opened new opportunities for integrating sound into applications. As inexpensive sound synthesis became widely available, researchers began exploring the potential and challenges of presenting information through audio. Aspects of spatial information can be mapped to non-speech audio events in a way that reflects real world events (e.g. the sound of running water) or that symbolizes items or events (e.g. a change in pitch may represent a location within a given reference frame [66] or reflect a value of the display content [84]). In cases where there is not a direct mapping between natural sounds in the environment and the display information, perceptual qualities must be observed through behavioral experiments (e.g. [78]) or trained (e.g. [36]). The wide variety of possible non-speech feedback can be subdivided into three categories: auditory icons, earcons, and soundscapes.

Auditory Icons: Auditory icons are recordings of natural events or synthesized audio events. These clips of sound represent objects or actions from the physical world, and can be associated with display elements or interface actions (see also [7]). Although they can be quite effective in conveying things and actions, auditory icons have difficulty representing abstract concepts and items that do not have an associated sound.

Earcons: The term earcon has been used to describe non-speech audio feedback in computer interfaces [7]. Earcons are event based elements of the display. In a study[1] that tested a hybrid symbol that encoded size information in earcons representing nominal data, Dingler et al. [25] found that earcons were more difficult to learn than spearcons (see also Section 2.4.2). Part of the

[1] Note that participants in the study described by Dingler et al. [25] were sighted.

challenge may have been that the assignment of earcons to the information that they represent is largely an arbitrary matching. Unlike auditory icons that represent a natural sound, earcons are only metaphorically related to the item or concepts that they represent.

Soundscapes: A soundscape is a collection of sounds that when heard together represent a space that may be real or virtual. The definition of soundscape varies across different fields. In the context of this thesis, the term is used to refer to collections of sounds presented in an audio display to convey information relative to space. For additional discussion of soundscapes, see [76].

3.1.2. Challenge: Time Dependence and Interaction of Concurrent Audio

Time plays a critical role in audio and heavily influences implementation of tools that have audio interfaces. Investigating spaces without vision requires strategies that use sequential exploration and increases exploration times [37, 12, 74, 66] (see also Section 2.3.1).

Compared to the broadband input provided by the visual system, exploration of a space with a stylus restricts input to a single point at a time. As observed in research on haptic interfaces, exploring a space with a cursor is "like foveation without peripheral vision" [75]. Memory is required for things outside the foveal area. With serial input that is inherent to both audio and tactile displays, cognitive processes must collect and synthesize small pieces of information to gain an understanding of the overall layout [6, 27]. Although the stylus (i.e. cursor) is constrained to a single point, audio feedback can be enriched with information about the surrounding area to help users build an understanding of display. Providing context prior to experiencing images in an audio display has also alleviated some difficulty in interpreting shape [2].

3.2. Guidelines for the Use of Non-Speech Audio

An audio display with non-speech feedback must map information to the various dimensions of sound. Dimensions of sound are mapped to classes of information (e.g. frequency may be mapped to position in a grid) and values are mapped to the associated variables (e.g. specific ordered frequencies may be assigned to low, medium, and high data values). Some guidelines have been developed, but there are still unanswered questions about how to perform this mapping [72]. This section presents examples of current applications using the various dimensions of sound to convey information; for definitions and descriptions of the dimensions of sound, see section 2.3.3.3.

3.2.1. Frequency (Pitch)

Frequency is the most commonly adjusted sound parameter, but is not necessarily well suited for all applications [78]. In particular, Talbot [73] noted that frequency is often used in navigation assistants to convey distance but is difficult for users to interpret. In a separate experiment by Nees and Walker [60], frequency was mapped to metric distance and was found to be effective only for the condition in which a visuospatial encoding strategy was encouraged. If frequency is used to convey information, the polarity of the mapping (i.e. as frequency increases does the attribute it is describing increase (positive polarity) or decrease (negative polarity)) must be considered and may be context dependent [78].

3.2.2. Volume (Loudness)

Loudness naturally maps to distance and can be used to convey distance representations in an audio display. The relationship mirrors the effect on sound of traveling over distance. Travel across physical distance depletes the strength of sound waves. The distance represented in audio displays may be euclidean distance between display elements [64, 36] or contrast between layers in the hierarchy of the display contents [59] (see also 2.1.1).

3.2.3. Spectral Envelope (Timbre)

Existing guidelines recommend that selected timbres be distinct from one another [13], but beyond exclusion of pure tones no quantitative description of what makes two timbres distinct was provided. Timbre has been used to discriminate between elements of the display that represent abstract contextual elements (e.g. grid position within a display [66]) or distinct objects in the display. In agreement with the guidelines suggesting consistency in interface design [69], consistency in assignment of timbre was found to be more important than embedding additional information into feedback events [66].

Paired with reduction of the perceived loudness (see also 3.2.2) caused by travel through the physical environment, there are also systematic changes in the spectral envelope over distance. The propagation of high frequencies is more affected by distance than lower frequencies. This results in diminished high frequency components in the sound when it reaches a distant listener.

3.3. Displays of Graphic Information

Several audio interfaces have tackled the challenge of mapping information to an audio display. These include signification of data graphics and spatial information.

3.3.1. Sonified Line Graphs: Mathematical Functions

To display data from mathematical functions in an audio display, the dependent variable can be symbolized with frequency while the independent variable is modified over time or advanced (or reversed) by the user. Recent work by Brown et al. [14, 15] has extended this approach to include the presentation of two lines on the same graph. In the case of these mathematical functions, possible feedback is constrained by the nature of functions: there is only one value of the dependent variable for each input value (for each line).

Line features in a geographic context, however, differ in regularity from line graphs. Line graphs commonly present one or more functions that have a single output for a given input value and the input values have absolute ordering. For example, rainfall (dependent variable) can be plotted for a series of days (independent variable). In contrast, lines on a map do not necessarily conform to a mathematical function. Techniques that map the independent values to a single dependent variable cannot generalize to include such cases. Instead of considering the line as a value associated with a single independent variable, images look at two dimensions where the output value is derived from a pair of independent variables.

3.4. Displays of Spatial Information

Maps display a rich combination of landmarks, spatial characteristics, and spatial relationships. They both support navigation and display information (e.g. thematic and choropleth maps; see also Section 2.1). This work focuses on thematic maps that convey survey information from an allocentric perspective (see also Section 2.2.1).

An audio display can present sufficient feedback to represent univariate characteristics of two dimensional map features [86, 28]. The iSonic system [86] symbolizes data attributes with continuous audio and boundaries between two areas with discrete short duration audio events. The boundaries indicate transition from one enumeration unit to its neighbor (i.e. movement of the cursor from one state to the next) but do not themselves have width or area. Initially the information display emphasized attribute data for two dimensional features. A later study evaluated participant's success in identifying shapes of the two dimensional features [22].

Details about the shapes of two-dimensional features (i.e. polygons) have been difficult for participants to interpret [22]. The boundary is key to defining and understanding the shape, but there has been limited investigation into the presentation of one dimensional features (i.e. lines) in an audio display. In the Omero display, haptic "walls" are used in combination with audio feedback to indicate boundaries between regions [28]. Similar to the short duration audio events used in iSonic, the Omero system displays haptic "walls" in combination with audio feedback to indicate transition across boundaries. While both Omero and iSonic provide feedback for transition across a boundary line, neither of these projects emphasize following the borders. This work presents and evaluates three techniques for creating an audio line symbology that facilitates line following.

Although not explored in this work, the combination of explicit egocentric radial direction along with the location and position of display elements relative to the cursor may alleviate some of the difficulties observed in line following such as perceiving corners. In one study, vowels sounds were mapped to directions [36]. Specific vowel sounds were arbitrarily mapped to the cardinal and intermediate directions and with four hours of training participants were able to reliably interpret a path displayed through audio.

3.4.1. Implications on Design of an Audio Display for Spatial Information

Presentation of irregularly shaped lines remains an unanswered question. Most of the current research in sonification of information focuses on a subset of the types of features that may be contained in a display. Display techniques have been developed for mathematical functions that relate an independent and dependent variable in a Cartesian plane (e.g. [15]). Some projects explore simple shapes using a passive interface to describe their perimeter (or the perimeter of their component parts) [2]. Projects that evaluate audio displays for geographic and spatial data typically attend to two dimensional elements (i.e. polygons) (e.g. [85, 28]). There has been little evaluation of one dimensional features (i.e. lines) in a geographic context where the shape of the features is not constrained. This thesis explores the presentation of irregular line shapes.

It may be desirable for users to be able to follow lines that convey context in the display of thematic maps. Unlike haptic feedback which can 'urge' the user toward (or away from) a feature, the stylus input device leaves the user in full control of movement. To encourage the user to move the stylus in any specific direction, it the responsibility of the audio feedback to ‘nudge’ them in the right direction.

CHAPTER IV

IMPLEMENTATION

Supported by the theories described in the previous chapters, a Java based application that uses audio to present spatial information was developed for use in experiments to test audio symbology and to provide the foundation of a minimal geographic information system. Beyond standard libraries, specialized functionality was required for audio synthesis and managing spatial data.

4.1. Audio Synthesis

Computer applications produce sound by streaming amplitude measurements to the speakers where the amplitudes are converted into fluctuations in pressure (i.e. sound waves). Digital representations of sound waves are records of amplitude measurements taken at regular intervals (sample rate) and recorded with finite precision (dynamic range). These amplitudes may originate from files containing previously recorded streams or be generated by a custom software application. The implementation presented in this thesis generates audio streams at run time through either preset configurations (MIDI) or as combinations of periodic functions (sampled audio).

Changes to the design are expected, but early design choices have favored flexibility and scalability. A MIDI library instrument is used for one of the tested line symbols while all other sounds are dynamically generated. During development flexibility is essential and the quality of the sound is less critical. When the interface design is more refined, implementation choices can be revisited to optimize sound generation and smooth out any undesired irregularities in the feedback.

4.1.1. MIDI

The Musical Instrument Digital Interface (MIDI) is a widely available standard for audio synthesis. MIDI was developed for instrumental performance of music and is based on events (e.g. notes and voices) rather than sampled wave forms [53]. This approach supports communication of MIDI events across different systems and platforms while leaving responsibility for the sound synthesis to those respective platforms (i.e. MIDI devices or computer software and hardware). Changes to sounds produced by MIDI systems are given in the form of new events. With the exception of control changes to the channels (e.g. gain or volume), individual internal parameters

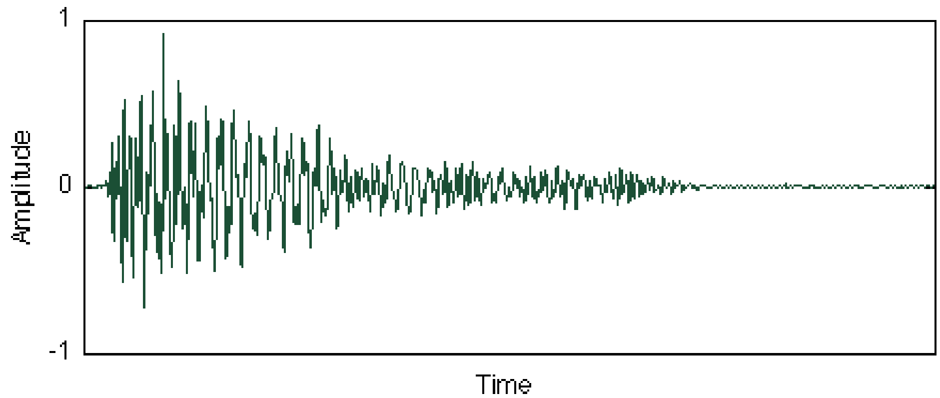

FIGURE 4.1. The "click" line symbol (short duration "woodblock" note) can be illustrated in the time domain.

of the synthesis cannot be interactively modified. The MIDI system manages the output stream(s) using pre-configured settings to provide uninterrupted output but sacrifice some responsiveness and flexibility.

One example of a MIDI note is the "woodblock" sound used in one of the audio symbols (see also Section 5.4). A formal description of the "woodblock" was available in the Java Soundbank. This includes the strength of attack, length of decay, the number and relative strength of the frequency components. A time domain representation of the "woodblock" note is presented in Figure 4.1 The experimental software configures a MIDI channel by specifing the instrument by name. Each time it is required by the interface, the software simply sends a message to the MIDI system to play a note.

4.1.2. Sampled Audio

Generation of sampled audio provides fine grained control over synthesis parameters. This control allows for precision adjustment but also requires that the developer handle all subtleties of the sound (e.g. ensuring that the stream of amplitude data is uninterrupted).

Synthesizing sampled audio places a demand on the processor and takes time to effect changes. Applications must balance responsiveness with the production of an uninterrupted stream of sampled data. Samples can be buffered to avoid interruptions in the audio stream, but the size of the buffer must be limited to maintain a maximum response time. If a buffer is small, it may drain too quickly causing in an interruption in the stream and resulting in audible clicks. Large buffers help avoid

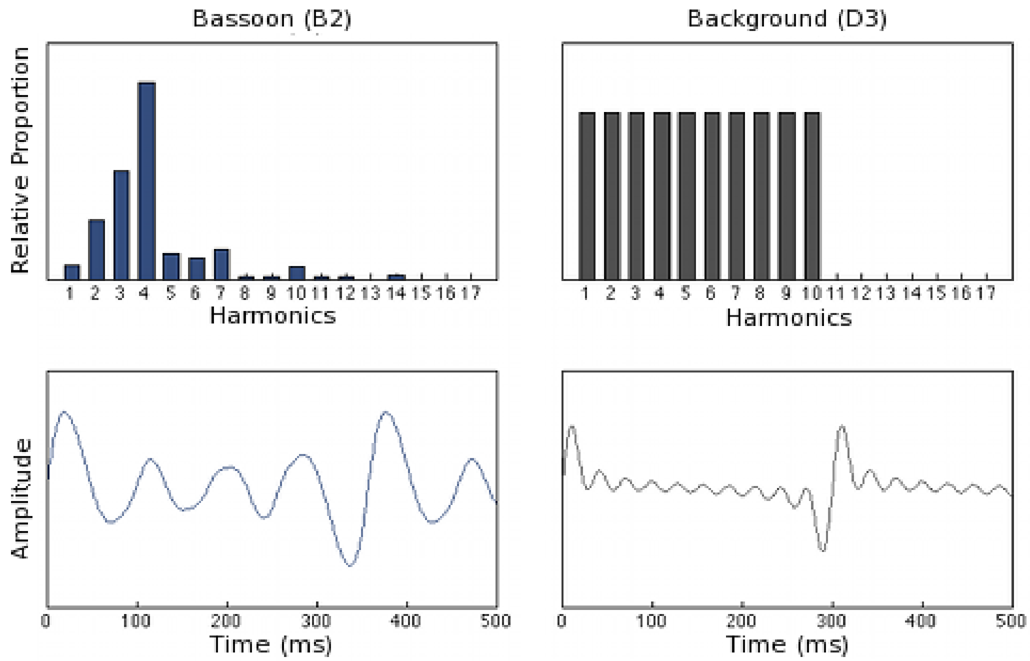

FIGURE 4.2. The continuous audio for the "on-off" and "shape" line symbols based on the Bassoon description by Nave [58] (left) and the background symbols (right) can be represented in the frequency (top) and time (bottom) domains. The height of the bar representing each harmonic reflects the relative strength of the harmonic.

interruption but may send stale audio samples to the speakers (i.e. data that was already buffered but had not yet played through the speakers when a change was specified) and lead the user to misinterpret a real time interactive display. Improvements in the experimental software to address issues of latency and responsiveness are ongoing.

Audio symbols evaluated in the experiment (see Section 5.4) controlled and modified frequency and spectral envelope. The contribution of each harmonic (see Figure 4.2) to the sample data is computed independently ($proportion$ = strength of the harmonic, i = index of the harmonic, θ = phase that changes over time, $sample\ rate$ = 44,100 samples per second).

$$amplitude_i = proportion_i * sin\left(\frac{2.0 * \pi * frequency_i * \theta}{sample\ rate}\right)$$

When generating the "shape" symbol, individual harmonics are suppressed as needed at a rate of 0.85 ($amplitude$ = original amplitude, n = number of harmonics to be filtered, i = index of the harmonic).

$$amplitude'_i = \begin{cases} amplitude_i * 0.85^{(n-i)} & i < n \\ amplitude_i & i \geq n \end{cases}$$

This rate was selected based on perceivable difference in informal pilot tests but was not formally evaluated. The value sent to the buffer is the sum of all the component harmonics.

$$amplitude_{sample} = \sum_i amplitude'_i$$

4.2. GeoTools Framework

Geographic information systems (GIS) provide ways to interact with, display, and store digitized spatial data. The nature of the data requires efficient geometric computations and special consideration for query efficiency compared to non-spatial data sets [80]. Many tools and libraries have been developed to accommodate these needs. Rather than reimplementing algorithms to compute and display spatial data, this project uses existing library capabilities.

ArcMap, a popular commercial GIS, was developed by Environment Systems Research Institute (ESRI)[1]. The ESRI tool suite provides an application programming interface (API), but the source code is closed and developers are limited to the exposed functions and interfaces. The open source community has also developed a number of libraries that specialize in processing geographic and spatial information[2]. GeoTools is one such open source project that provides cross platform compatibility and is distributed free of charge.

This project uses functionality from the GeoTools libraries to store, search, filter, query, and graphically render data. In the current minimal system, standard Java graphical components (i.e. Swing) could be sufficient to display and process distance queries regarding the single line feature in each display of the line following experiment (see also Chapter V). Looking forward, however, a more rigorous treatment of the spatial data will be required. The GeoTools framework has been employed in anticipation of the increased complexity and functionality of future versions.

GeoTools uses a `Style` object to specify symbology for the graphic display. Line strokes are characterized by color and width. Area fill is specified by color, pattern, and opacity. The audio

[1]ESRI, http://www.esri.com

[2]OSGeo, http://www.osgeo.org

symbology is currently specified outside the existing `Style` construct, and audio synthesis is handled independently from the graphic rendering. The separation of audio synthesis and graphic rendering allows each display to target different display hardware and respond to unique refresh needs (i.e. audio may update with any move of the cursor while graphics need only refresh when the window contents have been changed or revealed from behind other application windows).

In practice, the separation between mechanisms that implement non-speech audio symbols and their graphic counterparts means that developers may be less likely to incorporate audio display options. Analogously, published web pages tend to omit accessibility features for which implementation is not streamlined with the primary features. Despite guidelines that recommend them, accessibility features are implemented relatively infrequently [51]. Integration of the audio into the existing `Style` structure would allow for more natural inclusion of sound in applications built on the GeoTools framework. At the same time, this integration would impose a strict alignment between graphic symbology and audio symbology. Consistency between the two display modes may limit modality specific optimization but facilitates collaboration between users who are sighted and users who are blind [57] (see also 2.4.4). A balance must be struck between optimal approaches for each modality, consistency, and ease of implementation. The work described in this thesis moves toward a better understanding of the key factors that influence the balance.

CHAPTER V

EXPERIMENTAL METHODS

The line symbology designed for this study was evaluated through a behavioral experiment that consisted of two phases: training, during which participants were shown examples of the symbology that would be presented, and behavioral testing, in which participants explored a series of lines, traced the lines, and made judgements related to shape. Real time audio feedback reflected the object (line or background) located under the cursor, distance between the cursor and the line (up to a maximum fixed maximum distance), and/or traversal across a line by the cursor. Following an irregularly shaped line and determining its shape based solely on audio feedback was a difficult task, but overall participants' performance shows that a high level of accuracy is possible in a simple display.

5.1. Research Question

This study investigated the extent to which parameterized audio feedback can influence time and accuracy when following a line and making judgements about its shape.

> **Hypothesis 1:** Line symbology that indicates distance from the line improves speed and accuracy of line tracing.

> **Hypothesis 2:** Line symbology that indicates distance from the line during exploration and following tasks improves accuracy in making judgements about line shape.

In order to follow a line, users must first be able to perceive the line and discriminate between the line and background symbology. Although not explicitly tested, participants ability to follow lines with good accuracy implies that they were able to distinguish the line symbology from the concurrent background symbol(s). The position of the cursor relative to the line was expected to inform adjustments to the path of the cursor while tracing a line.

5.2. Considerations that Influenced Experimental Design

A number of factors influenced design of the experiment to evaluate audio symbology. This section briefly outlines several factors that were influential in the experimental design and should be noted when interpreting the results.

5.2.1. Target Population: Users who are Blind or Low Vision

Users who are blind and low vision are the target user group in this study. Due to the widely varying causes of blindness and visual impairment and the uniqueness of each individual's experiences, blindness cannot be classified into a single category [21]. This variability makes large scale studies desired but large participant numbers are often not feasible. People who are blind are a minority in the local population and it is difficult to control for experience. In this study, analysis does not control for cause or age of onset of blindness.

To increase the pool of potential participants, previous studies have asked sighted blindfolded participants to evaluate prototypes [36, 12]. In combination with feedback from a collaborator who is blind during the design of this prototype and blind participants making up half of the participant population, this experiment also turns to sighted blindfolded participants to help evaluate the audio symbology. Within the experimental results, there were no planned comparison between the performance of the blindfolded sighted participants and that of blind participants.

There are differing beliefs about the validity of asking blindfolded sighted users to evaluate interfaces designed for users who are blind or low vision. Observations from this study do not discourage this practice and are discussed in more detail in Chapter VII.

5.2.2. Simplified Lines as Test Data

The objective of this experiment was to present understandable information through the display rather than testing working memory (see also Section 2.3.1). The more complex the layout of a map, the more sighted users benefit from their ability to perceive multiple spatial attributes in parallel through vision [74]. For simplicity, this experiment excluded multiple lines, intersections between lines, endpoints that did not coincide with the edge of the display, orientation (radial direction), and directionality. Data complexity was constrained to the number of direction changes (straight line, one corner, two corners, or curve) and the angle between concurrent line segments. Examples of the test data are show in the 'dictionary of sample lines' used for training (Figure 5.5).

Both endpoints of each line were coincident with an edge of the display. This choice postponed investigation of how to represent endpoints, which has been found to be important in haptic displays to distinguish endpoints from sharp corners [12]. This constraint on the display contents introduced the possibility of determining categories of line types (e.g. both endpoints lie on a the same edge) to aid identification of the line during the matching task. In the matching task, at least one of the distractors shared the same endpoints as the true match in an effort to minimize the efficacy of this strategy and emphasize the more subtle characteristics of the line shape.

The underlying data across the conditions were rotations of lines from a single data set. To some extent, rotating the maps may help mitigate confounding issues of complexity but did not entirely eliminate them. Rotation reduced the effects of familiarity with the data, maintained consistent complexity, and allowed direct comparison between the conditions.

Although the data was strongly simplified, it maintained a moderate level of difficulty by placing the distinctive details in the central portion of the display. Features in the center of a spatial layout are more difficult to remember than those near the edges [50]. Features along the edges (such as the endpoints of the lines in this experiment), are expected to be placed more easily within the reference frame of the external edges of the display area.

5.2.3. Input Hardware

In existing audio displays, direction along the line(s) has been explicitly controlled by the user (e.g. [14]) or passively controlled in the time domain (by playing ordered feedback at a set rate; e.g. [2]). The test instrument designed for iSonic [86] provided free movement in an additional dimension (movement of the cursor to positions that did not necessarily coincide with a point on the line). Zhao et al. evaluated both keyboard and tablet input devices and found a preference for the tablet input device [83]. The use of the stylus provides proprioceptive feedback in response to movement through space.

The study described in this thesis used a tablet (190 millimeter x 130 millimeter) and stylus as the input hardware. The small size was expected to form a manageable reference frame in which participants would move the stylus. The stylus also allowed participants to anchor with their hands and fingers while exploring.

5.2.4. Tools to Collect Responses

When evaluating spatial skills, it is challenging to measure participants' mental representations. The experimental design required an effective measure that would neither substantially influence exploration nor confound data analysis. Several tools were considered and ultimately an interactive following task was paired with a matching task.

Recreation tasks (e.g. drawing or building with tangible objects) are one approach to recording a mental representation. These tasks may take the form of drawing or arranging tangible objects to reproduce the layout from memory. Some individuals who are blind or low vision are very successful artists [42], however, for others drawing is unfamiliar [75]. If drawing was unfamiliar to a participant, it may be difficult to distinguish whether they had struggled with drawing or struggled to understand the information itself. An unfamiliar task may also be detrimental to participants' confidence as they completed the tasks. The possibility of sketching the map features back on the tablet without audio feedback was considered, but this approach introduced the possibility that participants could rely on 'muscle memory' (i.e. a kinesthetic representation [45]). The line shape could be successfully reproduced by repeating movements; performance would not reflect the participant's true grasp on the overall shape of the line.

Previous shape recognition tasks have focused on recognition of letters from the English. An orientation and mobility training curriculum proffers letters of the alphabet as potential references for teaching spatial concepts (e.g. describe a river as having the shape of an 'S' or an intersection having the shape of a 'T') [27]. Letters have also been used as stimuli in working memory tasks with participants who are blind or low vision [8]. Familiarity with the written letters, however, may vary among participants who are blind or low vision and for that reason letters have been excluded from the descriptors for lines in this experiment

The ability to perceive and follow audio feedback do not necessarily imply that a mental representation of the overall shape was formed. Similar to the results found by Alty and Rigas [2], users' performance suggested that they were able to interpret individual components of the display but had difficulty synthesizing those components into a single complex whole. The matching task designed for this study included distractors that challenged the recollection of multiple attributes of each line. For example, the positions of the endpoints and overall concavity of the line matched those in the display, but the number of corners differed between the true match and a distractor (Figure 5.6). The matching task did require users to scale the image. Scaling was not expected to have a substantial impact on shape recognition due to the poor resolution of euclidean distance when perceived through kinesthetic feedback [49].

Traces from following the lines with the stylus based on real time feedback represented the accuracy with which participants received information from the display. This approach lessened the demand on memory and eliminated reliance on a potentially unfamiliar set of shapes. A matching task captured participants' ability to judge the overall shapes of the lines.

5.3. Observations from Pilot Studies

In addition to published related work from other labs, several pilot tests highlighted difficulties in using an audio display and refined the possible combinations of parameters for an audio display. Observations from these pilot tests were informal but still informative to development of the techniques used in the experiment.

5.3.1. Learning Effect

Over several rounds of pilot testing, substantial increases in performance over time were observed as participants became familiar with the audio display and input hardware (i.e. tablet and stylus). In these tests, any differences between the symbols that were being evaluated were overwhelmed by a strong learning effect. This finding reinforced the need to counterbalance the order in which conditions are presented to participants and demonstrated the need for some level of training prior to testing.

Practice improved facility with the display, but there was a point of diminishing returns. While familiarity with the system and test format increased over time, it was also observed that the more time spent on any one map display the more likely a participant was to encounter feedback that did not match the mental map they had formed up to that point. For example, the participant may have initially understood the map to include a border between an area on the left and another on the right (Figure 5.1: a). On further exploration with a movement that drifted upward, the feedback

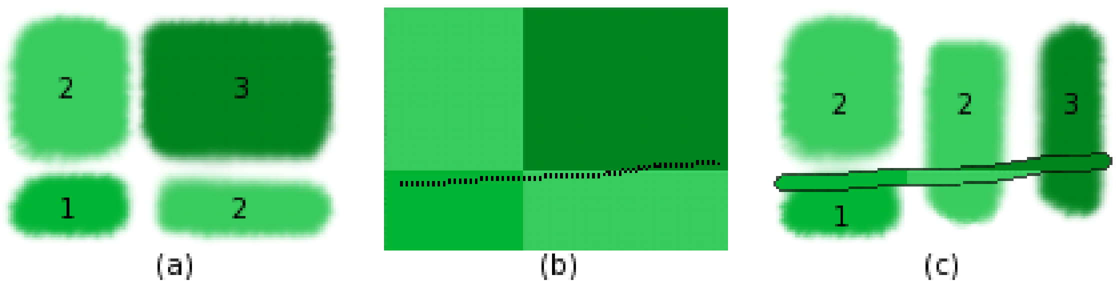

FIGURE 5.1. Extended exploration times introduced the potential for participants to encounter a conflict between the display contents and the mental representations that they had developed up to that point. A scenario where such confusion might (and did) arise is when one version of a mental map (a) conflicts with feedback when the stylus movement drifted upward (b). The participant must then try to reconcile the conflict (c) and may lose confidence in their understanding of the layout.

could indicate that there was a third area on the far right (Figure 5.1:b , c). This type of discovery caused confusion and reduced participants' confidence in their mental maps.

5.3.2. Fatigue

Participants expressed fatigue as they approached the end of the one hour sessions in which they explored and reproduced nine maps. The maps contained up to five sections that each belonged to one of three classes (high, medium, or low; represented as pitch). The experiment conducted to evaluate audio symbology for this thesis limited sessions to one hour in an effort to minimize fatigue.

5.3.3. Grid of Clicks

Orientation is important in map based exploration and a user who is blind or low vision "must navigate from memory due to lack of access to 'updating cues' [32]." To reduce the demand on memory, an audio grid of clicks (short duration audio events) was incorporated into a software prototype that presented two-dimensional data (i.e. polygons). The clicks were spaced at even distance intervals creating horizontal and vertical lines. The time interval between grid ticks was dependent on cursor speed.

Traces of the stylus movements and sketched recreations of the layout after exploration did not indicate changes in exploration style or improved accuracy that could be attributed to the grid lines. The task neither placed attention on the grid (i.e. no use of the grid was explicitly stated in the instructions) nor asked questions that required use of the grid (e.g. absolute location or distance

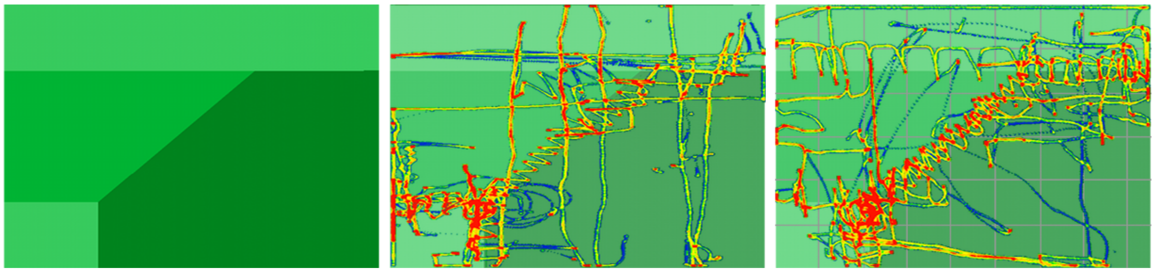

FIGURE 5.2. These heatmaps of cursor positions that were observed in a pilot study of an audio grid reflect patterns in exploration. The map data consisted of four regions (left). The high concentration of points that are shown as red traces near the region boundaries are similar between the condition that displayed no grid (center) and the condition with an audio grid (right).

judgements). The log files from exploration both with and without an audio grid revealed that users spent more time around the borders between polygons than in the middle of those polygons (Figure 5.2).

In the absence of point features (e.g. cities or landmarks), participants seemed to select the borders as the primary features of the map that supported the sketch map recreation task. Without instruction or coaching, participants developed several exploration strategies that emphasized the importance of borders.

Grid lines did not attract the same attention. Several users commented that they heard the ticks but were able to ignore them in favor of the other feedback. One user expressed confusion over how to interpret the pattern created by the ticks; rather than hearing the click at regular intervals on an evenly spaced grid, the user moved the stylus at different speeds over different areas of the display and interpreted the ticks as a texture. This demonstrated that audio events can be attended (and have more than one interpretation) or ignored. Observations from this pilot test suggested that the grid was not used in the given task and that users can selectively attend to different sounds in the display. This outcome is consistent with previous studies [54, 29, 50] and highlights a challenge in display design: increasing the amount of information presented is only helpful if the additional information directly aids task completion.

5.3.4. Symbology of Lines and Boundaries

The boundary of a two dimensional object does not inherently have width. For presentation in a graphic display, the boundary must either be given width or be represented by a change in

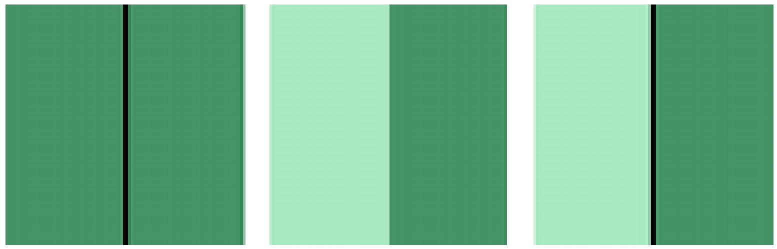

FIGURE 5.3. Representation of boundaries in visual displays involves giving a boundary width (left), representing the boundary by transition between two different fill colors (center), or both (right).

background fill of adjacent areas (Figure 5.3). If two adjacent areas shared the same fill, only the former would be applicable.

In audio, similar options exist to either play a distinct sound to represent the boundary (either playing it concurrently with the continuous background symbol or interrupting the background symbol) or to represent transition with a change in the background feedback (e.g. changing pitch of the continuous note).

The boundary sound could be a continuous sound that plays while the cursor lies on (or near) the boundary or it could be a short duration event when the cursor crosses the boundary. If the feedback plays only on transition across the line (indicated by either a short duration note or a change in the background symbol), participants have no way to receive feedback when the cursor is not moving. As will be seen in the experimental results, this difference has a substantial influence on the way that participants interacted with the display.

Two conditions in the experiment involve buffering the line (i.e. giving the line width) and one condition represents the line with a time dependent short duration audio event (click) that fades shortly after it is triggered. Because choropleth maps are likely to include data with homogeneous boundaries (i.e. boundaries shared by areas that have the same attribute value), the representation of a boundary based solely on a change in the background (center image of Figure 5.3) was not evaluated.

FIGURE 5.4. A Wacom Bamboo tablet and stylus served as the input device in this experiment. Tactilely distinct markers were placed at the midpoint of each side of the tablet as reference points.

5.4. Experimental Design

5.4.1. Test Instrument

Hardware: A Wacom Bamboo Fun Touch Tablet (190 millimeter x 130 millimeter) accepted input from the user during exploration. The tactilely distinct edges of the tablet surface determined the extent of the exploration space. In addition to the edge of the tablet surface, tactile markers indicated the midpoint of each side of the tablet (Figure 5.4). Stylus position on the tablet was sampled at up to 60 Hz (approximately every 16 ms; influenced by processor load). While using the stylus, participants were free to anchor with their fingers or opposite hand without confounding the detection of cursor input.

Software: The interface was built in Java using Swing graphical components and the GeoTools library for processing spatial data. Audio was synthesized through the sampled and MIDI sound libraries within the Java Sound API and controlled using the stylus and touch tablet. Sounds were produced only when the participant provided input by touching the stylus to the tablet.

The application ran on a MacBook Pro laptop (2.53 Ghz processor, 3MB L2 cache). Audio feedback was delivered on the built in speakers (Intel High Definition Audio).

Test Data: A set of fourteen lines were designed to evaluate the line symbology (see also Section 5.2.2). The test presented lines in a randomized order to each participant. Participants saw each line at most once in each condition and each rotation of a line was seen at most once across the entire experimental session. The lines varied one of the characteristics tested during the matching task (see also Section 5.4.4). Each subset of lines presented in the tests were representative of the overall set. The size of subsets greater than one ensured that participants experienced a variety of line characteristics (e.g. acute angles and obtuse angles) in each condition; the repeated measures were combined into a single composite score during analysis.

5.4.2. Symbology

Three distinct audio symbols represented lines in this experiment. Line symbol “on-off” (distance dependent) produced continuous sound at a constant volume any time the cursor was within a specified distance (one buffer width = 10 pixels) from the line. The interface did not produce the audio symbol when the cursor was more than one buffer width from the line. The width of the buffer used in producing the audio symbology was constant across participants and conditions.

Filtering low frequency harmonics of a continuous sound created line symbol “shape” (distance dependent). A side effect of suppressing the low frequency harmonics is decreased amplitude of the sound wave (i.e. amplitudes were not adjusted to maintain loudness levels). Similar to the “on-off” condition, the “shape” symbol was only produced when the stylus was within one buffer width (10 pixels) of the line.

The spectral envelope of both continuous line symbols simulated a sustained bassoon timbre (see also Sections 2.3.3.3 and 4.1.2). The proportions of the harmonics were based on those specified by Nave [58]. The continuous line symbols were played at 123.47 Hz (B2).

The third symbol, “click”, presented a short duration sound when the cursor moved from one side of the line to the other (position and time dependent). The short duration symbol was the MIDI ‘woodblock’ note from the Java Sound API Soundbank (see also Section 4.1.1).

In addition to the lines symbols, background symbology was also present in the display. The presence of the background simulates the audio display of data within a choropleth map. With the exception of the line symbology training, all displays were fully tiled with audio feedback (i.e. no area of the tablet lacked an associated sound). A background sound(s) covered the entire display and the line symbology was additive (i.e played concurrently) with the background.

The background symbols differed from those of the lines in spectral envelope, frequency, and volume. The background required continuous feedback and needed no specific characteristic beyond being distinct from the line symbology when presented concurrently. Ten evenly spaced harmonics of equal proportion were selected to represent the background (see also Section 4.1.2). The ability of participants to discriminate between the two timbres was not explicitly tested in the experiment.

The two background symbols (A and B) were played at 146.83 Hz (D3) and 164.81 Hz (E3). The background conditions were either a pair of the same frequency ("A:A") or different frequencies ("A:B"). Frequencies of the background symbols were separated from one another and from the line symbols by at least a minor 3rd (200 cents; see also Section 2.3.3.3). Two frequencies differentiate between the different backgrounds (fills) but the number of frequencies could scale to represent additional levels of a continuous variable.

The volume of the background (as determined by the amplitude of the sampled audio) was lower than that of the lines (approximately one third the maximum amplitude in the sampled waveform). As a side effect of the suppression of low frequency harmonics, volume of the "shape" symbol decreased the volume of that symbol and reduces the difference in volume between the line symbol and the background symbol.

5.4.3. Training

All participants first received a basic introduction to the test instrument that was limited to four minutes. A set of example lines illustrated a collection of lines and provided generic context for the following and matching tasks. Sighted participants viewed a graphic and the participants who are blind read a tactile version of the same graphic (Figure 5.5). The researcher stated that the lines could be characterized by observing whether they were curved or straight, counting how many segments they contained, and determining the direction of concavity (if any). None of the images on the training display appeared in the test data.

The hardware was described verbally while the participant explored the tactile characteristics of the tablet and stylus. The researcher informed the participant that a tap produces a short duration sound and that a touch and hold or a drag produces a continuous sound. Further, the researcher stated that the parameters of the audio feedback would vary based on the cursor position relative to features on the map. The participant then practiced using the stylus to explore a display that provided a background-only display (single continuous feedback that was constant across the entire exploration area). The participant was told that the basic training display did not contain a line feature.

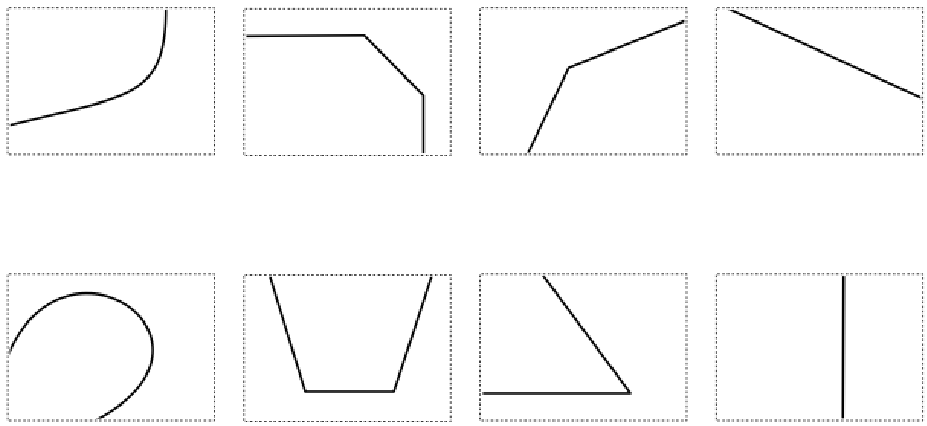

FIGURE 5.5. Participants were shown a dictionary of sample lines during the training period. The solid lines illustrated the types of lines that would be displayed during testing. Dotted lines represented the extent of the active area of the tablet.

Prior to each test condition, participants were given a brief introduction to the line symbology for that condition. A single segment vertical line was presented in the display. Similar to the test environment, feedback was produced in response to input from the stylus. During the introduction, the researcher directed participants to the location of the line and participants experienced the sound of the line symbology in the absence of any background audio feedback.

Participants were given 1 minute to explore each line. The specified time was a maximum limit; at their own discretion, participants could indicate that they were finished exploring and respond to the line following and matching tasks before the time expired.

The researcher did not suggest or encourage any particular exploration strategy or mental representation of the information presented in the audio display.

5.4.4. Data Collection

Log files: Log files were collected by the software while the participant explored the display. This information included trial configuration, type of input (e.g. drag stylus), cursor position, and timestamp.

Participant Responses: Following exploration, participants were asked to find and follow the line with the stylus. Lastly, participants pointed and provided a verbal response to a visual (sighted participants) or tactile (blind participants) matching task. The options presented in the

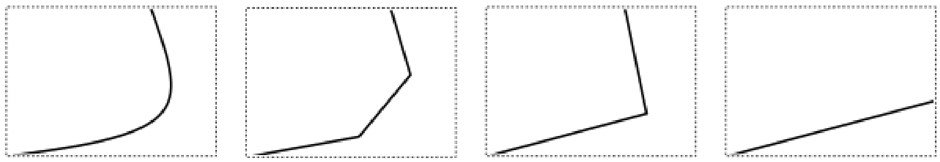

FIGURE 5.6. This example graphic from the matching task shows one correct match and three distractors. After following the line guided by audio feedback, participants were asked to select the line they had just explored from among the four choices.

matching task included one perfect match (Figure 5.6). Orientation of the perfect match aligned with that seen during exploration; participants were not asked to perform a mental rotation.

5.4.5. Research in Human Subjects

All participants gave informed consent to participate in the study. The protocol for interaction with human subjects was reviewed and approved by the Institutional Review Board in the Office for Protection of Human Subjects at the University of Oregon.

5.4.6. Round I

Experimental Design: The first round used a 3 (line symbol) × 2 (background symbol) factorial within subjects experimental design to evaluate line symbology in the context of background sounds. The three levels were discrete change in volume ("on-off"), gradual changes in the strength of low frequency harmonics ("shape"), and short duration, movement-based audio ("click"). The background configuration was a pair of continuous sounds that had either the same frequency ("A:A") or different frequencies ("A:B"). Although it also introduced learning effects, the within subjects design allowed inspection of the differences between the conditions while controlling for individual differences.

Each condition was presented to every participant. The order of presentation was counterbalanced to minimize the influence of order effects on the results. A selection of six lines (a subset of the fourteen possible lines; see also Section 5.4.1) was presented in each line symbology condition; three lines were presented in each background condition. Each line that a participant explored differed from all lines seen previously in either shape or orientation.

Experimental tests lasted less than one hour (Table 5.1). First, the researcher described the format of the test and introduced the test instrument. Line symbols were introduced prior to

<table>
<tr><th colspan="3">Activity</th><th>Time</th></tr>
<tr><td colspan="3">Introduction</td><td>4 minutes</td></tr>
<tr><td rowspan="6">3 ×</td><td colspan="2">Line Symbol</td><td>1 minute</td></tr>
<tr><td rowspan="2">3 ×</td><td>Explore (Background Level 1)</td><td>1 minute</td></tr>
<tr><td>Trace and Match</td><td>1-2 minutes</td></tr>
<tr><td rowspan="2">3 ×</td><td>Explore (Background Level 2)</td><td>1 minute</td></tr>
<tr><td>Trace and Match</td><td>1-2 minutes</td></tr>
<tr><td colspan="2">Rest</td><td>1 minute</td></tr>
<tr><td colspan="3">Total Time</td><td>45-60 minutes</td></tr>
</table>

TABLE 5.1. Allocation of time in Round I divided the one hour experimental sessions into an introduction, training each line symbol, and then completing trials for both background conditions.

exploration using each symbol. A straight line running from the top to the bottom of the tablet served as an example line and demonstrated the line symbol in the absence of a background symbol. Participants were given up to one minute to familiarize themselves with the line symbol. For each line in the test data, participants spent up to one minute freely exploring the display. After exploration, participants were asked to find and follow the line and select a matching line among three distractors.

Participants: Participants were recruited from the University or Oregon student population and local community. Testing was conducted with six blindfolded sighted participants (2 female, 4 male; mean age 30.66, range=(23,47)) and one participant who is blind (male; age=40).

Level of experience using a tablet and stylus varied. Two participants had never used a stylus, four participants (including the blind participant) had minimal previous experience using a stylus, and one participant reported extensive use of a stylus. Six of the seven participants reported four or more years of music experience; the participant who is blind also reported having perfect pitch.

5.4.7. Round II

Experimental Design: The second round evaluated two line symbols (with only one background configuration) using a within subjects experimental design. The two levels were gradual changes in the strength of low frequency harmonics ("shape"), and short duration, movement-based audio ("click"). The background was homogeneous across the entire display (equivalent to the "A:A" background in Round I).

Each condition was presented to every participant. The order of presentation was counterbalanced to minimize the influence of order effects on the results. A selection of seven lines was presented in each line symbology condition. Each line that a participant explored differed from all lines seen previously in either shape or orientation.

Experimental tests lasted less than one hour (Table 5.2; deviation from this planned timeline is discussed in Section 7.3.1). Similar to round one, the researcher described the test, line symbols were introduced with a sample line running from the top to the bottom of the tablet, and trials consisted of exploration (one minute), line following, and a matching task. Based on the observations taken during the first round, changes were made to the scope (number of conditions) and training for the second round.

The time spent completing the trials was substantially different between the blindfolded sighted participants and the blind participant in Round I. To stay within the one hour session time, the total number of trials needed to be more flexible. Two risks were anticipated. First, some participants would not complete all conditions resulting in an incomplete dataset. Second, participants might complete only one line in a condition increasing reliance on a single line to represent the entire set of test data (results may be dominated by effects of a single characteristic of the line). To address these risks, the number of conditions was reduced from six to two (see also Section 7.3.4). The number of trails in each condition increased from six to seven.

In addition to the training display that contained only a simple sample of the line symbol (no background) for each condition, Round II also presented a multimodal introduction to the condition. The multimodal training consisted of a sample line that was presented in both the audio display and as a tactile graphic. Participants had up to two minutes to go back and forth between the two displays of the same line. This training was repeated twice (one straight line and one curved line) for each condition.

Participants: Participants were recruited from the local community in Eugene, Oregon. Testing was conducted with four blind participants (4 females; mean age = 59, range=(58,62)). None of the participants in Round II had experience using a tablet and stylus; none of the participants reported having perfect pitch.

<table>
<tr><th colspan="3">Activity</th><th>Time</th></tr>
<tr><td colspan="3">Introduction</td><td>4 minutes</td></tr>
<tr><td rowspan="5">2 ×</td><td colspan="2">Line Symbol</td><td>1 minute</td></tr>
<tr><td>2 ×</td><td>Multimodal Training</td><td>1-2 minutes</td></tr>
<tr><td rowspan="2">7 ×</td><td>Explore</td><td>1 minute</td></tr>
<tr><td>Trace and Match</td><td>1-2 minutes</td></tr>
<tr><td colspan="2">Rest</td><td>1 minute</td></tr>
<tr><td colspan="3">Total Time</td><td>40-60 minutes</td></tr>
</table>

TABLE 5.2. Allocation of time in Round II divided the one hour experimental sessions into an introduction, multimodal training and completing trials for each line symbol. This table reflects the anticipated duration of each step. In contrast to Round I, there are fewer conditions but an additional training step.

CHAPTER VI

RESULTS

One post-hoc comparison was made between groups of participants to explore differences in time spent completing each trial. This analysis included only those conditions that were the same in both rounds. All other analysis was specific to one of the two rounds of testing. In all of the following analyses, the repeated measures have been controlled by computing a single score for each participant in each condition of the analysis. Details of the computations of these composite scores along with the models used to describe and evaluate the observations are given in Appendix A.

Analysis of the speed and accuracy of line following are computed using Model A/Model C comparisons (for a description of this approach, see [39]). Results are given (either inline or in tables) as triples of F^* scores, precent reduction in error (*PRE*), and probability (p). The F^* score is a measure of the improvement of one model over the other that accounts for the difference in the size of the models (number of additional parameters) and the number of remaining parameters that have not been added to the model. *PRE* reflects the improvement of an augmented model (Model A) that controls for a specified factor over a compact model (Model C) that assumes the specified factor had no effect. The F^* scores and *PRE* values are redundant representations of the same value.

The reported probability (p) reflects the likelihood of obtaining by chance an improvement that is as extreme as that observed in the data. The overall level of confidence is $\alpha = 0.05$. In Round I, there were six planned comparisons (two tests to compare the three line symbols on each of three variables) so the significance level is reduced to $\alpha' = 0.008$ after adjustment for the Bonferroni inequality. The significance level in Round II is reduced to $\alpha' = 0.017$ (there is a single planned comparison for each of three measures).

6.1. Power Analysis

Quantitative tests were designed to evaluate the three hypotheses. Based on a significance level of $\alpha = 0.05$ and the small number of participants (Round I: $n = 6$, Round II: $n = 4$), the power of the statistical analysis is very small. Even if the true difference between the conditions were large ($\eta^2 \approx 0.3$), the likelihood of finding a significant result would be around 0.26 and 0.15 respectively.

Although the number of participants in the study was too small to expect statistically significant results, the analysis was still conducted to prepare for later studies that include larger numbers of observations.

6.2. Preparation of Data

In preparation for analysis, the raw traces were processed to remove extraneous movements and account for repeated traversals performed by some of the participants.

6.2.1. Trimming

After the initial period of exploration, participants were asked to find and follow (or trace) the line. Although they were expected to have a general idea where the line was in the display, the initial part of the trace reflected a brief search between the time the stylus first touched the tablet to the first occurrence of line symbology in the feedback. Similarly, after the last occurrence of feedback from the line, some participants allowed the cursor to trail off away from the line (before picking the stylus up off the tablet) while they stated that they had completed the line following task. Both the initial 'find' and final 'trailing off' were trimmed from the traces prior to analysis.

6.2.2. Multiple Traces

When participants were asked to trace the line, the instruction did not specify whether the line should be traced only once or could be traced multiple times. This led to different approaches by different participants. Rather than taking a total sum of squared displacements, the sum was averaged over the length of the line that the stylus traversed. For each position of the cursor in the trace, the point on the line that was nearest the cursor was computed. The distance traveled along the line was then approximated by the consecutive distances between these points on the line. This accumulated a total distance (either forward or backward) along the line. Accuracy for each following task as determined from the trimmed traces (see also Section 6.2.1) was summarized as the sum of squared displacements (d) divided by the line distance ($distance$) traveled.

$$accuracy = \frac{\sum_i d_i^2}{distance}$$

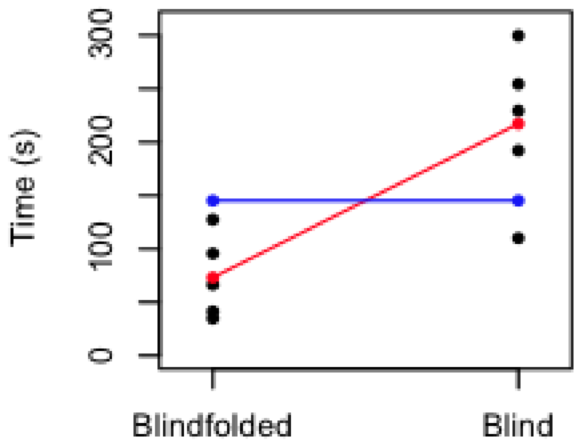

FIGURE 6.1. The time spent following lines differed between the two groups of participants (blindfolded sighted participants and blind participants). The plot shows both the overall mean (blue) and the differing group means (red).

Computation of the time spent following the line also accounted for the multiple passes. The total distance (*distance*) traveled along the line was divided by the total time (*time*) elapsed from the start to the end of the trimmed traces.

$$speed = \frac{distance}{time}$$

6.3. Line Following Speed (Between Groups)

One direct comparison between the performance of participants who are sighted and those who are blind evaluated the difference in the speed (time spent per trial) with which participants completed the line following task. Comparisons between groups were not the primary purpose of this study, but were computed post-hoc to illustrate how strongly the times differed.

Considering only those conditions completed by both groups, and controlling for repeated measures, the time spent completing the line following task was substantially greater for blind participants that for blindfolded sighted participants ($F^*_{(1,8)} = 15.709, PRE = 0.663$) (Figure 6.1). As a post hoc test, the F^* value exceeds the Scheffé adjusted critical value of $3 \cdot F^*_{3,8;0.05} = 12.21$. The time spent by blind participants was significantly greater than the time spent by blindfolded sighted participants.

Code	Comparison between conditions		$F^*_{(1,5)}$	PRE	p
X_1	"A:A" vs "A:B"	W_1	0.365	0.068	0.572
X_2	"click" vs continuous	W_2	0.155	0.030	0.710
X_3	"on-off" vs "shape"	W_3	1.527	0.234	0.271
X_4	background×"click" vs continuous	W_4	4.345	0.465	0.092
X_5	background×"one-off" vs "shape"	W_5	1.057	0.174	0.351

TABLE 6.1. Analysis of following accuracy between the conditions in Round I did not reveal significant differences between the conditions. The observed interaction X_4 did not reach statistical significance.

6.4. Round I: Background × Line Symbology

Participants completed multiple trails in each of the six conditions (three levels of line symbol ("click", "on-off", and "shape") and two levels of background ("A:A" and "A:B"). Data from one of the six blindfolded sighted participants were deemed outliers (see also Section 7.3.2) and were excluded from the analysis.

6.4.1. Hypothesis 1: Speed and Accuracy in Line Following

Following Accuracy: Differences observed in this data did not reach statistical significance (Table 6.1). The data did, however, suggest that the difference between the "click" symbol and the two continuous line symbols may be influenced by the background condition ($F^*_{(1,5)} = 4.252, PRE = 0.460, p = 0.094$).

Following Speed: In tandem with accuracy, speed is also considered as a measure of audio symbol effectiveness. The same within-subject contrast codes used to analyze line following accuracy represented the conditions and interactions. Without controlling for vision, no significant results were observed (Table 6.2). The effect is expected to be stronger when controlling for level of vision, however with only one participant representing the blind population, statistical significance would not be conclusive (and thus has not been computed). The planned model comparison that would control for vision is described in Appendix A.2.1.

6.4.2. Hypothesis 2: Accuracy of Judgements about Shape

Participant responses to the matching task were scored as either "Correct" or "Incorrect". A summary of the responses is given in Table 6.3. With the small number of independent observations, further research is required to draw conclusions about the results of the matching task.

	$F^*_{(1,5)}$	PRE	p
W_1	0.193	0.037	0.680
W_2	0.443	0.081	0.535
W_3	1.023	0.170	0.358
W_4	0.148	0.029	0.717
W_5	0.002	<0.001	0.970

TABLE 6.2. Analysis of the time spent following lines in Round I revealed no statistically significant difference between the conditions.

		Responses			
Condition		**Correct**	**Incorrect**	**Total**	
				Correct	**Incorrect**
Click	**A A**	8	7	22	8
	A B	14	1		
Shape	**A A**	13	2	23	7
	A B	10	5		
On Off	**A A**	13	2	26	4
	A B	13	2		
Total	**A A**	34	11	70	20
	A B	37	8		

TABLE 6.3. The matching accuracy in Round I is reflected by the number of correct and incorrect responses from 15 trials (three attempts for each of five sighted participants). Each line symbol is presented in a row that is divided between the homogeneous ("A:A") and heterogeneous ("A:B") background configurations.

6.5. Round II: Line Symbols

Participants completed multiple trials in each of the two conditions ("click" and "shape"). The repeated measures were combined into a single composite score, W_{0i}, for each condition to account for the nested non-independence. The time taken to complete the trials varied between participants. The total time spent by any participant was limited to one hour, so the number of observations (h) varies from one cell to the next (minimum number of observations = 3, maximum = 7). Again, the number of observations is too small to give sufficient power to the statistical analysis (df = 3). Computed values are intended only to reflect the planned calculations for a larger study.

6.5.1. Hypothesis 1: Speed and Accuracy in Line Following

With the very low number of participants, it was not surprising that the analysis did not reveal statistically significant results. No significant difference in was observed in accuracy between the two line symbols ($F^*_{(1,3)} = 0.450, PRE = 0.130, p = 0.551$). Nor was there a significant different in time spent per trial ($F^*_{(1,3)} = 0.028, PRE = 0.009, p = 0.877$).

6.5.2. Hypothesis 2: Accuracy of Judgements about Shape

The number of independent observations for the matching task were even smaller for Round II than for Round I. Due to issues with aggregation, these results were not analyzed.

6.6. Qualitative Observations

Throughout pilot testing and the final experiment, participants were able to perform tasks with high accuracy. This strong performance on objective accuracy measurements made it difficult to observe differences between performance on the different conditions. In future work it will be important to also measure subjective or perceived difficulty.

Four of the five blind participants expressed a preference for the continuous sounds over the "click" symbol. When first introduced to the "click" symbol, one participant expressed an immediate dislike of the symbol:

> "That's it? Oh, that sucks. It's insufficient feedback"

Another blind participant who had seen the "click" symbol first, expressed a preference for the continuous sound:

> "That will be a [lot] easier."

There were no formal metrics to collect participant preferences, but inclusion of such metrics may benefit further development.

CHAPTER VII

DISCUSSION

While conducting the experiment, a number of trends were observed regarding how the participants used the display. Even through statistical power was low, processing the data shed light on subtleties of presenting spatial data and helped reveal new questions about audio symbology. As articulated by Kuipers [47], several pieces of 'partial knowledge' may be required to form a complete understanding of spatial data. Individually, these measures do not represent participants' mental maps, but as a group it is hoped that patterns in performance may highlight aspects of the display and of the selected audio symbology that contribute to or detract from usability of the interface. This chapter presents a review of the trends observed (both formally and informally) through the course of the experiment.

7.1. Hypothesis 1: Speed and Accuracy of Line Tracing

The quantitative analysis did not reveal significant differences between the various conditions in either Round I or Round II. Participant feedback and anecdotal observations did, however, hint at trends in the data and suggest changes that could improve the prototype and guide future study. This section describes feedback and observations related to speed and accuracy in the line following task.

Future work could revisit the measure used to determine accuracy in line following. An updated model of error (e.g. a more accurate account of deviation from the line through a windowed average; see Section 7.3.3) could provide better insight into what made line following difficult and what the observed measurements reflect about participants' understanding of the line shape. Further, an explicit instruction to trace the line exactly once would help reduce the confounding influence of backtracking or repeated tracing.

Analysis may also benefit from controlling for different attributes of the line. In addition to an overall accuracy score for the line following task, local performance near various components of the line features (e.g. endpoints and corners) may help identify what makes a line difficult to follow in the audio display. This could be a count of the number of times the stylus departs from the line (somehow controlling for the zig-zag pattern in exploration) and a time spent reacquiring a position over the line after a departure. Similar to the pilot study in which more time was spent near borders than in open areas, more time is spent near corners. The degree of the angle (whether it was acute or obtuse) also seemed to influence exploration patterns. Consistent with Brewster's [12] observations about "sharp corners" and endpoints, this trend suggests that corners may require special treatment in the audio symbology. This analysis is left to future work.

The different line symbols were expected to influence line following by modifying how quickly participants could realize that the stylus had departed from the line. The "click" symbol would give quick feedback when the stylus was constantly on the move (e.g. rapid zig-zag) but the line could never be acquired; once the stylus stopped moving, feedback from the line symbol would cease. The continuous symbols provided feedback over time. The "on-off" symbol gave the line width and had a discrete maximum distance from the line at which feedback stopped. Feedback continued when the stylus stopped moving but there was no warning that the stylus was approaching the maximum distance for the line symbol.

These three line symbols varied in the granularity of feedback that conveyed to users that the stylus had departed from the line. Both the zig-zag approach and the "on-off" line symbol covered a ribbon of area on the display. The alignment between the line feature and the 'ribbon' of feedback could drift without being noticed by the user. The imprecision stood out when the participants were tracing lines with corners that formed acute angles. At corners, it was difficult to determine whether the corner was sharp or rounded. There seemed to be an advantage to the "click" symbol for acute angles; when the stylus moved across a sharp angle, the feedback was two rapid clicks. The "shape" symbol was intended to provide richer feedback that would convey information about the alignment. The variation of the feedback within the "shape" symbol was too subtle to be noticed when the stylus was moving quickly. It was only with slower stylus movements that the "shape" seemed to aid line following in the way that was anticipated.

7.2. Hypothesis 2: Matching Accuracy

It cannot be inferred that success in matching was the result of any specific mental map. In fact, the verbal commentary of some participant's suggested that their success may in fact have been the result of procedural knowledge that reconstructed the configuration.

For example, the endpoints of the line features aligned with the edges of the tablet and some participants expressed using that fact to help deduce the shape of the line (and determine potential matches). By increasing the complexity of the data (e.g. allowing lines to start/end in the middle of the display), this shortcut to determining matches could be mitigated. Participants would have to attend to more details of the line shape.

In future studies, it may be beneficial to code incorrect responses to a matching task in a way that revealed trends in misinterpretations. For example, if obtuse angles were consistently judged to be curves, this trend could help illuminate the way that participants are using the feedback.

7.3. General Discussion

7.3.1. Exploration Time

Consistent with previous research, speed (in this case measured as time per trial) differed between the blindfolded sighted participants and the blind participants (the comparison considered only those observations for conditions that were seen by all participants: “A:A” background and “click” and “shape” line symbols). The observed difference in exploration times could be attributed to differences in exploration or memory strategies [75, 40, 1]. Regardless of the cause, additional time may effect performance in several ways.

Over time, items in working memory decay unless they are refreshed or rehearsed [3]. If the user is able to synthesize new input into a composite piece of information or augment an existing piece of information (potentially accessing long term memory structures), the long stream of small data points received through the audio display may be comprehensible. If, however, the participant is reliant on remembering many small pieces, the more time spent exploring, the more previous information is forgotten. This highlights two important facets of usability: synthesis of information (limiting the demand on memory resources) and efficient presentation of information (limiting the time spent exploring the display for the salient details).

Although it was less noticeable in the evaluation of line symbology than in previous pilot studies looking at map layout (see also Section 5.3.1), the longer participants spent exploring the map, the more frequently they encountered feedback that did not agree with the mental map they had already developed (see also Section 5.3.1). The contradiction itself is not a negative event (resolution of the conflict would improve the mental map), but the impact on confidence may be a substantial setback to participant's willingness to explore or acceptance of the application.

Participants tended to move the cursor more quickly for the "click" condition than either of the continuous line symbol conditions. The continuous movement of the zig-zag pattern (see also Section 7.3.3) was performed faster than the precision movements within the continuous line symbols. Informally, the result was faster exploration times for the "click" condition. This difference may have had a positive effect on perception of line shapes (more time was spent with precision movements and the parts experienced early in the exploration are forgotten; see also Section 2.3.1).

7.3.2. Line Symbols vs. Background Symbols

Five of the six sighted blindfolded participant and all four participants who are blind attained a good level of accuracy in the line following task across all six (Round I) or both (Round II) conditions (Figure 7.1). This consistent level of performance suggests that the three line symbols designed for this experiment were distinguishable from the background symbols when produced concurrently. Although it was not explicitly tested, observations support the hypothesis that a 'bassoon' timbre is sufficiently distinct from a collection of evenly spaced and equally proportioned harmonics.

One participant, however, struggled to follow the line shapes. Within one condition, some traces vaguely reflected the general shape of the line while others did not (Figure 7.2). It was unclear if the difficulty stemmed from inaccurate discrimination between line and background symbols, or unfamiliarity with the task, input device, or display mode. When asked if they could hear the sound of the line while the background was playing, the participant responded in the affirmative. Due to the stark contrast between the performance of the one participant (both in following accuracy and matching accuracy) in comparison to the other ten, this participant's data was excluded from the analysis. In future studies evaluation of participants' ability to discriminate between the symbols and correctly understand them before proceeding to the trials could help identify any misinterpretations of the audio feedback.

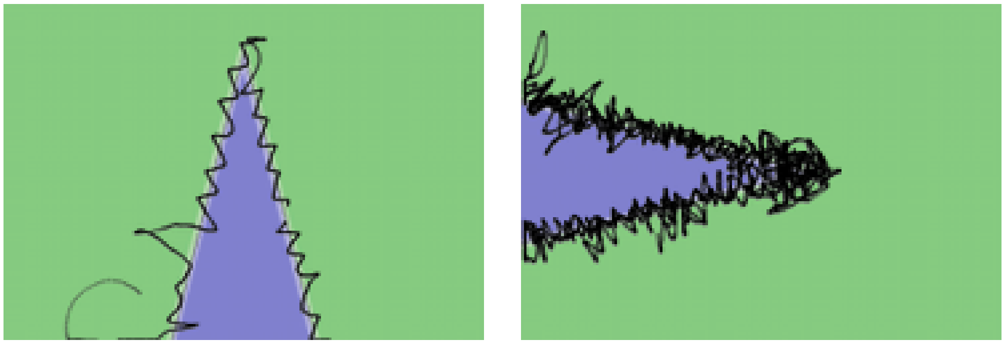

FIGURE 7.1. These visualizations of two cursor traces illustrate successful completion of the line following task. The trace on the left was performed by a blindfolded sighted participant ("A:B" × "on-off") and the one on the right by a blind participant ("A:A" × "shape").

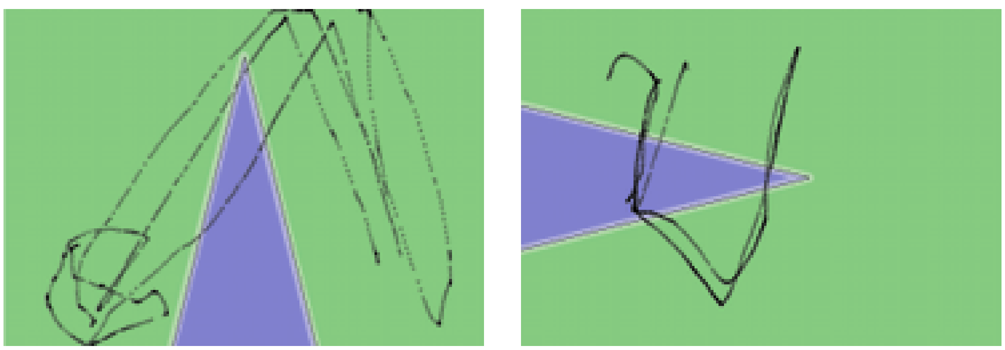

FIGURE 7.2. These visualizations of two cursor traces performed by one of the sighted blindfolded participants that were deemed outliers illustrate difficulty in the line following task. Although some traces resembled the shape of the line that had been displayed (left; "A:A" × "on-off" condition), it was unclear if the participant understood the audio feedback (right; "A:B" × "click").

Strong performance on the following and matching tasks by a majority of participants suggested that participants could discriminate between the two timbres, but discrimination between sounds was not explicitly evaluated. Without formal evaluation, it was not clear whether it was the spectral envelope or some other characteristic (e.g. harmonic complexity resulting from the difference between a single note and a chord) of the combined line and background symbols that allowed participants to segregate the different symbols into separate audio streams.

One participant suggested that the difference between the frequency of the line symbol and the frequency of the background symbol be greater. Larger separation between the frequencies may make it easier for participants to force the sounds into separate audio streams (see also Section 2.3.3).

7.3.3. Strategy

Aside from introducing the stylus and tablet, which suggest a movement heuristic [49], the researcher did not encourage any particular encoding strategy (see also Section 2.3.2) or exploration technique. To better understand differences in performance (both speed and accuracy), analysis could control for the encoding strategy. Although more prevalent in the pilot studies than in the line symbology experiment, participants were observed moving the stylus (or a pen) in the air or singing back the different tones used in the display to assist with recalling positions and attributes of display components. Using those actions, participants seemed to be recalling and translating from the movements they had completed or sounds they had heard during exploration. In response to the same task, different participants chose different strategies resulting in differing performance as translation from one to the other takes time [1].

The type of symbology encouraged exploration strategy. When the line (i.e. boundary) was presented in response to transition from one side to the other as either a click or a discrete change in background frequency, participants had to traverse the line to receive feedback. They did not receive feedback unless they were moving. Across all of the participants, the transition based feedback encouraged a 'zig-zag' pattern of stylus movements to follow the line.

The zig-zag traces modified the requirements for measuring accuracy in the line following task. In the current analysis, the sum of squared displacement from the line was used to score accuracy. This approach to accumulating error penalizes large displacements more than small displacements. In the case of zig-zag traces, this may not appropriately reflect the locations in the display that the user was interpreting as the position of interest.

The zig-zag strategy introduced frequent, intentional, regular departure from and return to the line as the user moved the stylus back and forth across a line. Because of this influence, raw distance measurements were not a strong measure of accuracy. The effective position of the cursor could be anywhere within the width of the zig-zag. Instead, a windowed average of the distances could be computed. Based on the expectation that the zig-zag was an intentional regular pattern, the resulting average trace (i.e. midline of the zig-zag path) could approximate the path of the users focus.

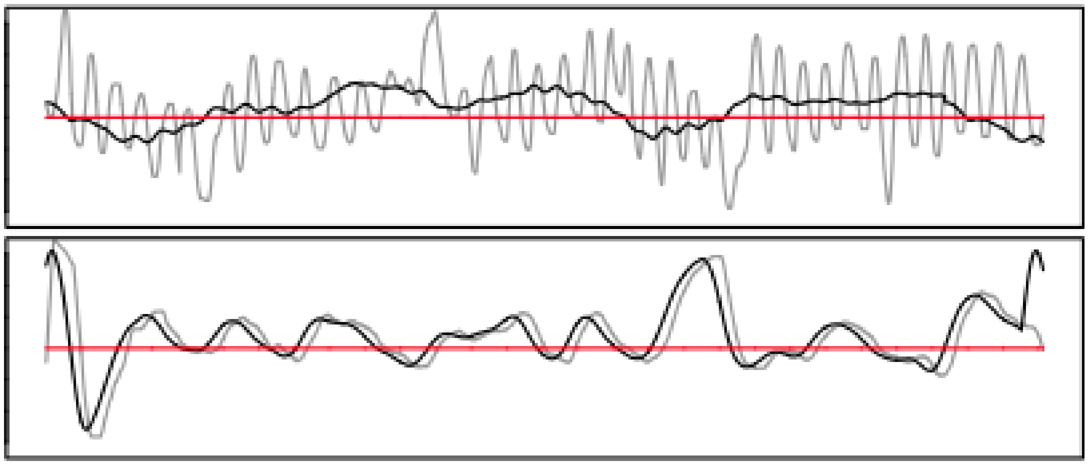

FIGURE 7.3. A windowed average of the cursor traces may aid quantification of following accuracy. Two example traces illustrate a zig-zag strategy (top) and a smooth trace (bottom). The averaged cursor position (black) is shown over the raw cursor positions (gray). The red line represents points that lie on the line feature that the user was following.

Although use of the zig-zag strategy was the most common for the "click" condition, it was also observed in both of the continuous feedback conditions. Once a participant had developed a zig-zag strategy, they tended to use that strategy during subsequent trials regardless of the line or background symbols. The traces would need to be reliably coded as a zig-zag technique or the average would need to be applied to displacement data across all conditions.

A windowed average could help moderate the effects of cursor movement patterns (i.e. whether or not the participant used a zig-zag strategy). The different widths of the zig-zag pattern can be accommodated by setting a window size based on the number of samples between reversals (changes in direction of the cursor relative to the line). The windowed average would produce a smoothed trace that would approximate the center of the zig-zag. When an alternative strategy was in use, the windowed average produced a slightly simplified the trace. The windowed average would ideally have little effect on the traces in which the user did not employ a zig-zag strategy. Examples of the windowed average that were computed to illustrate the approach are shown in Figure 7.3.

7.3.4. Experimental Conditions

The initial experiment presented six conditions (two background levels × three line symbols). Participants tended to develop different strategies for each of the two background symbols. With a focus on line symbology, Round II reduced the design to a single background level. Among the line symbols, participants interacted differently with the continuous symbols ("shape" and "on-off") than they did with the short duration symbol ("click"). Rather than looking at subtle differences between the two continuous symbols, the difference between the two classes of line symbols was explored.

Background Levels: Informal observations revealed that users' experience of the line that separated the two background symbols in level "A:B" was dominated by the frequency change between the background symbols (in contrast to the line symbol itself). This bodes well for maps that display heterogeneous boundaries (boundaries between enumeration units whose attribute values differ), but an approach to symbolizing lines that depends on changes in the attribute values of the surrounding areas may not be sufficient to depict choropleth maps. Displays must also be able to present homogeneous boundaries (boundaries between enumerations that both have the same attribute value). Between the two, the homogeneous boundaries seem more difficult to display through audio.

The heterogeneous border may be conveyed by any of several audio events (e.g. change in frequency or addition of new sounds). There are fewer possible audio events to represent homogeneous borders (i.e. we cannot rely on a secondary effect of symbols designed for other purposes). In applications it will be less important which symbol (line symbol or the change in background symbol) effects the perception of a boundary; more important is that the user can determine that there is a boundary. Round II focused on lines that could represent the more challenging homogenous boundaries and limited the background to a single level ("A:A").

The presence of a background sound was maintained even when there was only one level of background. The "A:A" level of the background served as a distractor as participants attended to the various line symbols. Perceiving and understanding the line symbols in isolation was seen as an oversimplification. Users ability to attend to the line symbol despite the presence of a concurrent distractor will be necessary in a minimal geographic information system where multiple display elements overlap.

Line Symbology: Three line symbols were used in Round I, while Round II presented only two. The difference between the two continuous line symbols ("shape" and "on-off") was expected to be smaller than the difference between the continuous line symbols and the "click" symbol. In favor of a simpler design, the "on-off" symbol was dropped from the design for Round II.

Number of Trials Completed

	Minimum	Maximum
per condition	3	7
total	9	14

TABLE 7.1. The number of trials completed within the one hour session varied across the four participants in Round II.

Number of Conditions: The total number of conditions (six and two for Round I and II, respectively) emphasized different aspects of the evaluation. A larger number of conditions allows comparison of interaction effects but also necessitates a larger number of observations from each participant for within subjects analysis. Limiting time to one hour then means that the trials must be completed quickly. Two conditions provides much more focused analysis on one aspect of the display and relaxes the time pressure (created by the one hour time limit).

The time each participant spent completing the trials varied between participant groups and between individual participants. All blindfolded sighted participant completed all of the planned trials (18 trials) within the one hour session. The blind participant in Round I completed 13 of the 18 trials. The number of trials that blind participants completed in Round II varied (Table 7.1). The greater repetition of fewer conditions ensured that each participant experienced each condition more than once (i.e. for more than one trial). It was important to collect multiple observations for a variety of lines to minimize a bias due to the complexity of a single line.

7.3.5. Input Device

The interface described in this thesis and used in the experiment is a stylus based interface that evolved from a visual design. There may be a more fundamental model of the data that would lead to a very different display [51, 67]. Future work will explore additional ways to isolate critical elements of a spatial display, optimize the symbology, and develop a educational tool to teach spatial thinking skills.

7.3.6. Correlation between Following and Matching Accuracy

In addition to the hypotheses presented in Chapter V, the following accuracy and matching accuracy could be informative. Correlation between performance scores for these two tasks would help validate the metrics used to measure participants' understanding of the display layout. For future studies, a third hypothesis is proposed:

> **Proposed Hypothesis 3:** Accuracy following a line has a positive correlation with accuracy in shape recognition.

7.3.7. Inclusion of Blindfolded Sighted Participants

Systematic differences in performance between participants who are blind and those who are sighted challenge validity of using sighted participants during pilot testing of applications designed for accessibility. On one hand, a map user who is sighted gathers and understands spatial information differently than a blind or low vision map user and may not consider the same details of the audio display to be salient, necessary, or sufficient. On the other hand, characteristics of perception and ability to comprehend spatial information is comparable across both groups. During experiment design and analysis, statistical models may benefit from controlling for some of the systematic differences (e.g. differences in strategy; see also Section 7.3.3). Testing interfaces with sighted participants cannot replace testing with the target population. Blind participants should be involved from the early phases of design to moderate the reliance on vision-based metaphors or other vision-specific design decisions (it is incredibly difficult as a sighted person to relinquish a tight grasp on vision based approaches), but with carefully designed experiments it may be reasonable to include sighted participants in iterative testing.

CHAPTER VIII

CONCLUSIONS

The presentation of spatial information through an audio display is a difficult task. Multiple interdependent aspects of the design and implementation play critical roles in the success of the communication between the information designer (e.g. map designer), the computer interface, and the user. This thesis has reviewed related works and designed and tested a simple audio symbology to represent lines in a two dimensional display.

The experiment described in this thesis focused on comparing three line symbols through a quantitative analysis. The observed differences between the symbols were subtle, but the feedback from users revealed a preference for the continuous sounds. In addition to the quantitative data, more formal qualitative analysis may help capture users' impressions and help us understand the issues of usability and aesthetics in the design of audio symbols.

Designing an audio symbology is more than just selecting audio parameters. It also entails understanding how people understand space and what elements mean within the context of a display. Awareness of cognitive processing and spatial thinking are a broad fields that will inform this research.

This thesis is only the first step in a much larger project. The symbols designed for evaluation in this thesis will be incorporated into a minimal geographic information system. The insight provided by participants regarding the effectiveness of the symbology and experience gained working with real time audio feedback will augment and refine an existing prototype. The next steps will include continued evaluation of symbology, documentation of lessons learned, and integration of an audio symbology into existing libraries for geographic information systems.

APPENDIX

ANALYSIS DETAILS

This appendix presents details of the analysis completed to evaluate the relative effectiveness of the audio symbols. The approach follows the Data Analysis textbook by Judd, McClelland, and Ryan [39].

A.1. Line Following Speed (Between Groups)

The comparison between groups of time spent line following used observations from only those conditions completed by both groups ("A:A" × "click" and "A:A" × "shape"). The repeated measures (nested non-independence) were combined into a single composite observation for each subject (h = number of observations, i = subject for which observation was taken).

$$W_{0i} = \frac{\sum_h \text{observation}_{hi}}{\sqrt{h}}$$

Participants level of vision was contrast coded and stored the variable *visionCC*. The time spent following the line was expected to be larger for participants who were blind than for those who were blindfolded; observations from blind participants were assigned the positive value contrast code.

$$blind = 1, blindfolded = -1$$

A Model A/Model C comparison looked for evidence that there was systematic deviation from the overall mean that could be attributed to whether or not the participant was blind.

$$\textbf{Model A} : W_{0i} = \beta_0 + \beta_1 * visionCC_i + \epsilon_i$$

$$\textbf{Model C} : W_{0i} = \beta_0 + \epsilon_i$$

$$\mathbf{H_0} : \beta_1 = 0$$

The two β values were approximated from the data. The use of a code centered at zero for *visionCC* meant that β_0 in Model A was the same as β_0 in Model C.

$$b_0 = 145146, b_1 = 71990$$

Subject	"A:A" "click"	"A:A" "on-off"	"A:A" "shape"	"A:B" "click"	"A:B" "on-off"	"A:B" "shape"
i	$C_{A:A,1i}$, $C_{A:A,2i}$, $C_{A:A,3i}$	$O_{A:A,1i}$, $O_{A:A,2i}$, $O_{A:A,3i}$	$S_{A:A,1i}$, $S_{A:A,2i}$, $S_{A:A,3i}$	$C_{A:B,1i}$, $C_{A:B,2i}$, $C_{A:B,3i}$	$O_{A:B,1i}$, $O_{A:B,2i}$, $O_{A:B,3i}$	$S_{A:B,1i}$, $S_{A:B,2i}$, $S_{A:B,3i}$
	$W_{0iC_{A:A}}$	$W_{0iO_{A:A}}$	$W_{0iS_{A:A}}$	$W_{0iC_{A:B}}$	$W_{0iO_{A:B}}$	$W_{0iS_{A:B}}$

TABLE A.1. The multiple scores (nested) for each subject were combined to create composite scores for each subject in each condition in Round I .

The time spent by participants who are blind was greater than that spent by blindfolded participants ($F^*_{(1,8)} = 15.709, PRE = 0.663$). As a post hoc test for a design with four levels (Round I: $m = 4$) for which there were ten observations ($n = 10$) in a model with two parameters ($PA = 2$) the critical value of F^* is larger than that of the planned comparisons. The result ($F^* = 15.709$) exceeds the Scheffé adjusted critical value of

$$(m-1) \cdot F^*_{m-1;n-PA;\alpha} = (4-1) \cdot F^*_{(4-1),(10-2);0.05} = 3 \cdot F^*_{3,8;0.05} = 3 \cdot 4.07 = 12.21$$

A.2. Line Following Speed and Accuracy

A.2.1. Round I: Background × Line Symbology

Similar to the composite scores computed in the analysis of overall time spent (Section A.1), composite scores (W_{0i}) were computed for each of the six participants in Round 1 (Table A.1; each observation was of the $C =$ "click", $O =$ "on-off", or $S =$ "shape" condition and belonged to one of the two background conditions).

$$W_{0i} = \frac{\sum_h \text{observation}_{hi}}{\sqrt{h}}$$

Each of the levels and interactions were contrast coded (Table A.2). Those contrast codes were used to compute a total of six W_{hi} scores (one for each participant; excluding data from the participant whose results were outliers in this sample) for a within subjects analysis. With one score for each participant, there are five degrees of freedom ($df = 5$).

		"A:A"			"A:B"		
		"click"	"on-off"	"shape"	"click"	"on-off"	"shape"
X_1	"A:A" vs "A:B"	-1	-1	-1	1	1	1
X_2	"click" vs continuous	-2	1	1	-2	1	1
X_3	"on-off" vs "shape"	0	-1	1	0	-1	1
X_4	background×"click" vs continuous	2	-1	-1	-2	1	1
X_5	background×"one-off" vs "shape"	0	1	1	0	-1	-1

TABLE A.2. The conditions of the factorial design and their interactions were given contrast codes for analysis of data for Round I.

Following Accuracy: The composite scores were evaluated in a Model A/Model C comparison looking for a significant difference between the means for the contrast coded variables controlling for vision.

$$\textbf{Model A} : W_{hi} = \beta_0 + \epsilon_i$$

$$\textbf{Model C} : W_{hi} = 0 + \epsilon_i$$

$$\mathbf{H_0} : \beta_0 = 0$$

The within subjects analysis does not find a significant between the conditions or their interactions (see Table 6.1 in the main text). Although not significant, the interaction between background level and "click" vs. continuous line symbol was strong ($F^*_{(1,5)} = 4.252, PRE = 0.460, p = 0.094$).

Speed (Time per Trial): Controlling for level of vision, the time spent following lines for each condition could be compared by including a term in the models to represent a contrast code for level of vision (*visionCC*). For example, the following models and null hypothesis would test whether within subjects differences in score differed from zero.

The differences between the "click" symbol and the continuous symbols look for a difference between the contrasts for "click" (for which a mean value of zero would imply no difference) and the average of the mean scores for the two continuous symbols. The other W_{hi} scores were similarly analyzed producing the results in Table 6.2.

$$\textbf{Model A} : W_{2i} = \beta_0 + \beta_1 * visionCC_i + \epsilon_i$$

$$\textbf{Model C} : W_{2i} = 0 + \beta_1 * visionCC_i + \epsilon_i$$

$$\mathbf{H_0} : \beta_0 = 0, \mu_{\text{“}click\text{”}} = \frac{\mu_{\text{“}shape\text{”}} + \mu_{\text{“}on-off\text{”}}}{2}$$

A.2.2. Round II: Line Symbols

In the same manner as presented for Round I, the repeated measures were combined to form a composite score. With two levels of a single condition, the analysis was much simpler. The difference between the two conditions were computed based on contrast coded conditions ("shape" = 1, "click" = -1) to compute a composite score for each participant.

$$W_{1i} = \frac{(-1)C_i + (1)S_i}{\sqrt{(-1)^2 + (1)^2}}$$

Using the W_{1i} scores, a single parameter model (Model A) was compared to a zero parameter model (Model C) to look for a statistically significant deviation of the mean difference from zero, which would reflect a statistically significant difference between the two conditions. With four participants, there were three degrees of freedom (df = 3).

$$\textbf{Model A}: W_{1i} = \beta_0 + \epsilon_i$$

$$\textbf{Model C}: W_{1i} = 0 + \epsilon_i$$

Similar analyses were conduced for both speed and accuracy data.

REFERENCES CITED

[1] Amandine Afonso, Florence Gaunet, and Michel Denis. The mental comparison of distances in a verbally described spatial layout: effects of visual deprivation. *Imagination, Cognition and Personality*, 23(2-3):173–182, 2003-2004.

[2] James L. Alty and Dimitrios Rigas. Exploring the use of structured musical stimuli to communicate simple diagrams: the role of context. *International Journal of Human-Computer Studies*, 62(1):21–40, 2005.

[3] Alan Baddeley. Working memory. *Current Biology*, 20(4):R136 – R140, 2010.

[4] Jeffrey J. Berkley. Haptic devices. White paper, Mimic Technologies, Inc., Seattle, WA, 2003.

[5] Jennifer K. Bizley, Kerry M. M. Walker, Bernard W. Silverman, Andrew J. King, and Jan W. H. Schnupp. Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *Journal of Neuroscience*, 29(7):2064–2075, February 2009.

[6] Mark Blades, Simon Ungar, and Christopher Spencer. Map use by adults with visual impairments. *The Professional Geographer*, 51(4):539–553, 1999.

[7] Meera M. Blattner, Denise A. Sumikawa, and Robert M. Greenberg. Earcons and icons: Their structure and common design principles. *Hum.-Comput. Interact.*, 4:11–44, March 1989.

[8] Irina Bliss, Teija Kujala, and Keikki Hamalainen. Comparison of blind and sighted participants' performance in a letter recognition working memory task. *Cognitive Brain Research*, 18:273–277, 2004.

[9] Eoin Brazil, Mikael Fernstrom, and John Bowers. Exploring concurrent auditory icon recognition. In *Proceedings of the 15th International Conference on Auditory Display*. Re:New - Digital Arts Forum, May 2009.

[10] Albert S. Bregman. Auditory scene analysis. In Neil J Smelser and Paul B Baltes, editors, *International Encyclopedia of the Social and Behavioral Sciences*, volume 2, pages 940–942. Elsevier, Amsterdam / New York, 2001.

[11] Bill Brewer and Julian Pears. Frames of reference. In Naomi Eilan, Rosaleen A. McCarthy, and Bill Brewer, editors, *Spatial Representations: Problems in Philosophy and Psychology*, pages 25–30. Blackwell, Cambridge, Massachusetts, 1993.

[12] Stephen A. Brewster. Visualization tools for blind people using multiple modalities. *Disability and Rehabilitation*, 24(11-12):613–621, 2002. http://dx.doi.org/10.1080/09638280110111388.

[13] Stephen A. Brewster, Peter C. Wright, and Alistair D.N. Edwards. Experimentally derived guidelines for the creation of earcons. In *Proc. of HCI'95*, 1995.

[14] Lorna Brown, Stephen Brewster, Ramesh Ramloll, Mike Burton, and Beate Riedel. Design guidelines for audio presentation of graphs and tables. In *Proceedings of the 2003 International Conference on Auditory Display*, pages 6–9, Boston, MA, USA, July 2003.

[15] Lorna Brown, Stephen Brewster, Ramesh Ramloll, Wai Yu, and Beate Riedel. Browsing modes for exploring sonified line graphs. In *Proceedings of the 16th British HCI Conference*, pages 6–9, London, England, 2002.

[16] David S. Burch and Dianne T.V. Pawluk. A cheap, portable haptic device for a method to relay 2-d texture-enriched graphical information to individuals who are visually impaired. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, Assets '09, pages 215–216, Pittsburgh, Pennsylvania, USA, 2009. ACM.

[17] Matt Calder, Rober F. Cohen, Jessica Lanzoni, Neal Landry, and Joelle Skaff. Teaching data structures to students who are blind. In *Proceedings of the 12th annual SIGCSE conference on Innovation and technology in computer science education (ITiCSE '07)*, pages 87–90, Dundee, Scotland, 2007. ACM.

[18] Murray Campbell and Clive Greated. *The Musician's Guide to Acoustics*. Schirmer Books, New York New York New York New York, 1987.

[19] E. Collin Cherry. Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acourstical Society of America*, 25(5):975–979, 1953.

[20] Douglas H. Clements and Julie Sarama, editors. *Engaging Young Children in Mathematics*. Lawrence Erlbaum Associates, Mahwah, New Jersey, 2004.

[21] Regina Araujo de Almeida (Vasconcellos) and Bruce Tsuji. Interactive mapping for people who are blind or visually impaired. In D.R.F. Taylor, editor, *Cybercartography: Theory and Practice*, chapter 18, pages 411–431. Elsevier, Amsterdam, 2005.

[22] Franco Delogu, Massimiliano Palmiero, Stefano Federici, Catherine Plaisant, Haixia Zhao, and Olivetti Belardinelli. Non-visual exploration of geographic maps: Does sonification help? *Disability and Rehabilitation: Assistive Technology*, 5(3):164–174, May 2010.

[23] Borden D. Dent. Visual organization and thematic map communication. *Annals of the Association of American Geographers*, 62(1):pp. 79–93, 1972.

[24] Borden D. Dent. *Cartography: Thematic map Design*. Wm. C. Brown Publishers, Debuque, Iowa, 3rd edition, 1993.

[25] Tilman Dingler, Jeffrey Lindsay, and Bruce N. Walker. Learnabiltiy of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In *Proceedings of the 14th International Conference on Auditory Display*, Paris, France, June 2008.

[26] Bernhard Dürnegger, Christina Feilmayr, and Wolfram Wöß. Guided generation and evaluation of accessible scalable vector graphics. In Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer, editors, *Computers Helping People with Special Needs*, volume 6179 of *Lecture Notes in Computer Science*, pages 27–34. Springer, Berlin / Heidelberg, 2010.

[27] Diane L. Fazzi and Barbara A Petersmeyer. *Imagining the Possibilities: Creative approaches to orientation and mobility instruction for persons who are visually impaired.* AFB Press, New York, NY, 2001.

[28] F. De Felice, F. Renna, G. Attolico, and A. Distante. A haptic/acoustic application to allow blind the access to spatial information. In *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)*, Tsukuba, Japan, March 2007. IEEE.

[29] Ludovic Ferrand, Kevin J. Riggs, and Julie Castronovo. Subitizing in congenitally blind adults. *Psychonomic Bulletin and Review*, 17(6):840–845, 2010.

[30] Scott M. Freundschuh and Max J. Egenhofer. Human conceptions of spaces: Implications for geographic information systems. *Transactions in GIS*, 2(4):361–375, 1997.

[31] Miguel Angel Garcia-Ruiz and Jorge Rafael Gutierrez-Pulido. An overview of auditory display to assist comprehension of molecular information. *Interact. Comput.*, 18:853–868, July 2006.

[32] Reginald G. Golledge. Geography and the disabled: A survey with special reference to vision impaired and blind populations. *Transactions of the Institute of British Geographers, New Series*, 18(1):63–58, 1993.

[33] Reginald G. Golledge. Spatial cognition and converging technologies. In Mihail C. Roco and Editors William Sims Bainbridge, editors, *Converging Technologies for Improving Human Performance*, pages 122–140. Dordrecht, Boston, MA, August 2003.

[34] Reginald G. Golledge, Matthew Rice, and R. Daniel Jacobson. A commentary on the use of touch for accessing on-screen spatial representations: The process of experiencing haptic maps and graphics. *The Professional Geographer*, 57(3):339–349, August 2005.

[35] Donald E. Hall. *Musical Acoustics*. Physics Series. Brooks/Cole Publishing Company, 2002.

[36] Susumu Harada, Hironobu Takagi, and Chieko Asakawa. On the audio representation of radial direction. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 2779–2788, Vancouver, BC, Canada, 2011. ACM.

[37] R. Dan Jacobson. Navigating maps with little or no sight: An audio-tactile approach. In *Proceedings of Content Visualization and Intermedia Representations*, Montreal, Quebec, Canada, 1998.

[38] Yvonne Jansen, Thorsten Karrer, and Jan Borchers. Mudpad: localized tactile feedback on touch surfaces. In *Adjunct proceedings of the 23nd annual ACM symposium on User interface software and technology*, UIST '10, pages 385–386, New York, New York, USA, 2010. ACM.

[39] Charles M. Judd, Gary H. McClelland, and Carey S. Ryan. *Data Analysis: A Model Comparison Approach*. Routledge, New York, NY, 2nd edition, 2009.

[40] Authur Karshmer and Ken Paap. Automathic blocks: Supporting learning games for young blind students. In Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer, editors, *Computers Helping People with Special Needs*, Lecture Notes in Computer Science, pages 459–465. Springer Berlin / Heidelberg, 2010.

[41] Brian F. G. Katz, Emmanuel Rio, Lorenzo Picinali, and Olivier Warusfel. The effect of spatialization in a data sonification exploration task. In *Proceedings of the 14th International Conference on Auditory Display*, page 7, Paris, France, 2008.

[42] John M. Kennedy. *Drawing & the Blind: Pictures to Touch*. Yale Univeristy Press, New Haven, 1993.

[43] Rob Kitchin and Scott Freundschuh. Cognitive mapping. In Rob Kitchin and Scott Freundschuh, editors, *Cognitive Mapping: Past, Present, Future*, chapter 1, pages 1–8. Routledge, New York, NY, 2000.

[44] Robert M. Kitchin, Mark Blades, and Reginald G. Golledge. Understanding spatial concepts at the geographic scale without the use of vision. *Progress in Human Geography*, 21(2):225–242, 1997.

[45] Roberta L. Klatzky and Susan J. Lederman. Representing spatial location and layout from sparse kinesthetic contacts. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2):310–325, April 2003.

[46] John B. Krygier. Sound and geographic visualization. In Alan M. MacEachren and D. R. Fraser Taylor, editors, *Visualization in Modern Cartography*, chapter 8, pages 149–166. Pergamon, Oxford, UK, 1994.

[47] Benjamin Kuipers. The cognitive map: Could it have been any other way? In Herbert L. Pick and Linda P Acredolo, editors, *Spatial Orientation: Theory, Research, and Application*, pages 345–359. Plenum Press, 1983.

[48] Sukhbinder Kumar, Klass E. Stephan, Jason D. Warren, Karl J. Friston, and Timothy D. Griffiths. Hierarchical processing of auditory objects. *PLoS Computational Biology*, 3(6):977–985, 2007.

[49] Susan J. Lederman, Roberta L. Klatzky, and Paul O. Barber. Spatial movement-based heuristics for encoding pattern information through touch. *Journal of Experimental Psychology: General*, 114(1):33–49, March 1985.

[50] Svenja Leifert. The influence of grids on spatial and content memory. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, pages 941–946, Vancouver, BC, Canada, 2011. ACM.

[51] Stefan Leuthold, Javier A. Bargas-Avila, and Klaus Opwis. Beyond web content accessibility guidelines: Design of enhanced text user interfaces for blind internet users. *International Journal of Human-Computer Studies*, 66:257–270, April 2008.

[52] Daniel J. Levitin. *This is your Brain on Music: The Science of a Human Obsession*. Dutton, Penguin Group, New York, NY, 2006.

[53] Gareth Loy. Musicians make a standard: The midi phenomenon. *Computer Music Journal*, 9(4):pp. 8–26, 1985.

[54] Murray T. Maybery, Peter J. Clissa, Fabrice B.R. Parmentier, Doris Leung, Grefin Harsa, Allison M. Fox, and Dylan M. Jones. Binding of verbal and spatial features in auditory working memory. *Journal of Memory and Language*, 61:112–133, 2009.

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

[56] D. J. Medyckyj-Scott and M. Blades. Cognitive representations of space in the design and use of geographical information systems. In Dan Diaper and Nick Hammond, editors, *People and Computers VI: Proceedings of HCI '91*, pages 421–433. Cambridge University Press on behalf of the British Computer Society, 1991.

[57] Elizabeth D. Mynatt and W. Keith Edwards. Mapping GUIs to auditory interfaces. In *Proceedings of the 5th Annual ACM Symposium on User Interface Software and Technology (UIST '92)*, pages 61–70, Monteray, California, United States, 1992. ACM.

[58] Carl R. Nave. Hyperphysics: The bassoon, 2010. Available http://hyperphysics.phy-astr.gsu.edu/hbase/music/basoon.html.

[59] Michael A. Nees and Bruce N. Walker. Relative intensity of auditory context for auditory graph design. In Tony Stockman, Louise Valgerdur Nickerson, Christopher Frauenberger, Alistair D. N. Edwards, and Derek Brock, editors, *Proceedings of the 12th International Conference on Auditory Display*, pages 95–98, London, UK, June 2006.

[60] Michael A. Nees and Bruce N. Walker. Mental scanning of sonifications reveals flexible encoding of nonspeech sounds and a universal per-item scanning cost. *Acta Psychologica*, 137(3):309–317, July 2011.

[61] Donald A. Norman. Knowledge in the head and in the world. In *Psychology of Everyday Things*, chapter 3, pages 54–80. Basic Books, Inc., New York, NY, 1988.

[62] Patrick Péruch, Vanessa Chabanne, Marie-Pascale Nesa, Catherine Thinus-Blanc, and Michel Denis. Comparing distances in mental images constructed from visual experience or verbal descriptions: The impact of survey versus route perspective. *The Quarterly Journal of Experimental Psychology*, 59(11):1950–1967, 2006.

[63] Herbert L. Pick Jr. Organization of spatial knowledge in children. In Eilan et al. [11], chapter 1, pages 31–42.

[64] Ian J. Pitt and Alistair D. N. Edwards. Navigating the interface by sound for blind users. In D. Diaper and N. Hammond, editors, *People and Computers VI: Proceedings of HCI '91*, pages 373–383. Cambridge: Cambridge University Press, 1991.

[65] Matt Rice, R. Daniel Jacobson, Reginald G. Golledge, and David Jones. Design considerations for haptic and auditory map interfaces. *Cartography and Geographic Information Science*, 32(4):381–391, 2005.

[66] Dimitrios Rigas and James L. Alty. The rising pitch metaphor. *International Journal of Human-Computer Studies*, 62(1):1–20, 2005.

[67] Anthony Savidis and Constantine Stephanidis. Developing dual user interfaces for integrating blind and sighted users: The HOMER UIMS. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 106–113, Denver, Colorado, United States, 1995. ACM Press/Addison-Wesley Publishing Co.

[68] Ben Schneiderman. *Software Psychology: Human Factors in Computer and Information Systems*. Winthrop Publishers, Inc, Cambridge, Massachusetts, 1980.

[69] Ben Schneiderman and Catherine Plaisant. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley, 5th edition, 2010.

[70] Roger Shepard. Stream segregation and ambiguity in audition. In P. R. Cook, editor, *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics*. MIT Press, Cambridge, Massachusetts, 1999.

[71] Eva Siekierska, Richard Labelle, Louis Brunet, Bill Mccurdy, Peter Pulsifer, Monika K. Rieger, and Linda O'Neil. Enhancing spatial learning and mobility training of visually impaired people - a technical paper on the internet-based tactile and audio-tactile mapping. *Canadian Geographer*, 47(4):480–493, 2003.

[72] Tony Stockman, Louise Valgerdur Nickerson, and Greg Hind. Auditory graphs: A summary of current experience and towards a research agenda. In *Proceedings of the 11th International Conference on Auditory Display*, Limerick, Ireland, July 2005.

[73] Martin Talbot and William Cowan. On the audio representation of distance for blind users. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI '09)*, pages 1839–1848, Boston, MA, USA, 2009. ACM.

[74] Carla Tinti, Mauro Adenzato, Marco Tamiettob, and Cesare Cornoldi. Visual experience is not necessary for efficient survey spatial cognition: Evidence from blindness. *The Quarterly Journal of Experimental Psychology*, 59(7):1306–1328, 2005.

[75] Simon Ungar. Cognitive mapping without visual experience. In Rob Kitchin and Scott Freundschuh, editors, *Cognitive Mapping: Past, Present, Future*, chapter 13, pages 221–248. Routledge, New York, NY, 2000.

[76] Andrea Valle, Vincenzo Lombardo, and Mattia Schirosa. A graph-based system for the dynamic generation of soundscapes. In Mitsuko Aramaki, Richard Kronland-Martinet, Sølvi Ystad, and Kristoffer Jensen, editors, *Proceedings of the 15th International Conference on Auditory Display*, Copenhagen, Denmark, 18—21 May 2009.

[77] W3C. Web content accessibility guidelines (WCAG) 2.0. W3C Recommendation, The World Wide Web Consortium (W3C), 2008. Available http://www.w3.org/TR/WCAG/.

[78] Bruce N. Walker and Lisa M. Mauney. Universal design of auditory graphs: A comparison of sonification mappings for visually impaired and sighted listeners. *ACM Transactions on Accessible Computing*, 2(3):12:1–12:16, March 2010.

[79] Bruce N. Walker, Amanda Nance, and Jeffrey Lindsay. Spearcons: Speech-based earcons improve navigation performance in auditory menus. *Proceedings of the 12th International Conference on Auditory Display (ICAD2006)*, 2006.

[80] Fangju Wang. A distributed geographic information system on the common object request broker architecture (corba). *GeoInformatica*, 14(1):89–115, 2000.

[81] J.D. Warren, A.R. Jennings, and T.D. Griffiths. Analysis of the spectral envelope of sounds by the human brain. *NeuroImage*, 24(4):1052 – 1057, 2005.

[82] István Winkler, Titia L. van Zuijen, Elyse Sussman, János Horvath, and Risto Näätänen. Object representation in the human auditory system. *European Journal of Neuroscience*, 24:625–634, 2006.

[83] Haixia Zhao, Catherine Plaisant, and Ben Shneiderman. I hear the pattern: Interactive sonification of geographical data patterns. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*, pages 1905–1908, Portland, OR, USA, 2005. ACM.

[84] Haixia Zhao, Catherine Plaisant, and Ben Shneiderman. iSonic: interactive sonification for non-visual data exploration. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, Assets '05, pages 194–195, Baltimore, MD, USA, 2005. ACM.

[85] Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and Jonathan Lazar. Data sonification for users with visual impairment: A case study with georeferenced data. *ACM Trans. Comput.-Hum. Interact.*, 15:4:1–4:28, May 2008.

[86] Haixia Zhao, B.K. Smith, K. Norman, C. Plaisant, and B. Shneiderman. Interactive sonification of choropleth maps. *IEEE Multimedia*, 12(2):26 – 35, April-June 2005.