12-2013

Making Graphical Information Accessible Without Vision Using Touch-based Devices

Hari Prasath Palani

MAKING GRAPHICAL INFORMATION ACCESSIBLE WITHOUT VISION USING TOUCH-BASED DEVICES

By

Hari Prasath Palani

B.E. Anna University, India, May 2009

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

(in Spatial Information Science and Engineering)

The Graduate School

The University of Maine

December 2013

Advisory Committee:

Nicholas A. Giudice, Associate Professor of School of Computing and Information Science, Advisor

Kate Beard-Tisdale, Professor of School of Computing and Information Science

Reinhard Moratz, Associate Professor of School of Computing and Information Science

THESIS ACCEPTANCE STATEMENT

On behalf of the Graduate Committee for Hari Prasath Palani, I affirm that this manuscript is the final and accepted thesis. Signatures of all committee members are on file with the Graduate School at the University of Maine, 42 Stodder Hall, Orono, Maine.

Dr. Nicholas A. Giudice, Date

Associate Professor of School of Computing and Information Science

LIBRARY RIGHTS STATEMENT

MAKING GRAPHICAL INFORMATION ACCESSIBLE WITHOUT VISION

USING TOUCH-BASED DEVICES

By Hari Prasath Palani

Thesis Advisor: Dr. Nicholas A. Giudice

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Master of Science
(in Spatial Information Science and Engineering)
December 2013

Accessing graphical material such as graphs, figures, maps, and images is a major challenge for blind and visually impaired people. The traditional approaches that have addressed this issue have been plagued with various shortcomings (such as use of unintuitive sensory translation rules, prohibitive costs and limited portability), all hindering progress in reaching the blind and visually-impaired users. This thesis addresses aspects of these shortcomings, by designing and experimentally evaluating an intuitive approach —called a vibro-audio interface— for non-visual access to graphical material. The approach is based on commercially available touch-based devices (such as smartphones and tablets) where hand and finger movements over the display provide position and orientation cues by synchronously triggering vibration patterns, speech output and auditory cues, whenever an on-screen visual element is touched. Three human behavioral studies (Exp 1, 2, and 3) assessed usability of the vibro-audio interface by investigating whether its *use leads to development of an accurate spatial*

representation of the graphical information being conveyed. Results demonstrated efficacy of the interface and importantly, showed that performance was functionally equivalent with that found using traditional hardcopy tactile graphics, which are the gold standard of non-visual graphical learning.

One limitation of this approach is the limited screen real estate of commercial touch-screen devices. This means large and deep format graphics (e.g., maps) will not fit within the screen. Panning and zooming operations are traditional techniques to deal with this challenge but, performing these operations without vision (i.e., using touch) represents several computational challenges relating both to cognitive constraints of the user and technological constraints of the interface. To address these issues, two human behavioral experiments were conducted, that assessed the influence of panning (Exp 4) and zooming (Exp 5) operations in non-visual learning of graphical material and its related human factors. Results from experiments 4 and 5 indicated that the incorporation of panning and zooming operations enhances the non-visual learning process and leads to development of more accurate spatial representation. Together, this thesis demonstrates that the proposed approach —using a vibro-audio interface— is a viable multimodal solution for presenting dynamic graphical information to blind and visually-impaired persons and supporting development of accurate spatial representations of otherwise inaccessible graphical materials.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my mentor and advisor Dr. Nicholas A. Giudice for his support and patience throughout this work. Without his timely advice, it would have been impossible to perform this research and write this thesis. I have enjoyed my discussions and arguments with him and they have been influential in helping me grow as a researcher. I would also like to thank the advising committee members: Dr. Kate Beard- Tisdale and Dr. Reinhard Moratz for their support throughout my graduate career and in this work.

I also take this opportunity to thank all the professors in the Department of Spatial Information Science and Engineering for their support in the success of my work. Special thanks to my colleagues and friends at VEMI Lab, department, and the University: Matt Dube, Christopher Bennett, Hengshan Li, Balaji Venkatesan, Bill Whalen, Shreyans Jain, Jonathon Cole, Joshua Leger, Meghan White, Monoj Raja, Rick Corey, Riju Shreshta, RJ Perry, Tim McGrath, and Upasana Pandey.

I thank Advanced Medical Electronics Corporation for providing the required software and Tablet devices. I would also like to acknowledge the support from NIDRR grant H133S100049 and NSF grant CDI-0835689 on this project.

Finally, I would like to thank Saranya Kesavan and my parents for their unwavering support in all of my endeavors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Graphics (or Infographics) are a visual representation of information, data or knowledge that intends to present complex information quickly and clearly ("Infographics," 2013). Although graphics are usually rendered for visual use, they are not inherently visual. In most cases however, graphics are visual representations that allow people to conceptualize and learn from quantitative data. In our technical world, graphics have ascended to dominant importance as an essential way to communicate information. Graphics can adopt many forms, ranging from simple line drawings to complex maps and are used in almost all fields for effective communication. This evolution towards graphic communication is bringing forward interesting research challenges, especially in the field of accessibility. For instance, the visual nature of graphical elements makes them inaccessible to numerous blind and visually-impaired (i.e., those with limited functional vision) persons.

By contrast, access to printed material has largely been solved with the advancement of electronic text via screen readers —such as JAWS for Windows

(“JAWS,” 2013) or VoiceOver for the Mac and iOS-based portable devices (“VoiceOver,” 2013)— and/or electronic refreshable Braille displays. But these programs do not have the ability to convey meaningful information about graphics or non-text-based material. Currently, the most common method to substitute for visual graphics is by producing tactile representations of the graphics (Edman, 1992). However, compared to visual graphics, interpreting tactile graphics is a difficult process (Loomis, Klatzky, & Lederman, 1991) and also making tactile equivalence of visual representation is a cumbersome process involving removal of crucial details (see Chapter 2 for discussion). In addition, both paper-based and swell-based tactile graphics are non-refreshable, meaning that they are static renderings that are both cumbersome and expensive to produce. The advent of electronic refreshable displays presented an opportunity to overcome the drawbacks of tactile graphics by their design to work in dynamically changing environments (Rastogi, Pawluk, & Ketchum, 2013). Although many research groups have focused on developing virtual tactile graphics based on electronic haptic displays (G. Jansson, I. Juhasz, & A. Cammilton, 2006; Owen, Petro, Souza, Rastogi, & Pawluk, 2009; Petit, Dufresne, Levesque, Hayward, & Trudeau, 2008; S. Walker & Salisbury, 2003), these approaches still suffer from various shortcomings (such as lack of intuitiveness, limited portability and prohibitive cost) that has significantly hindered progress in reaching blind users.

This thesis addresses the challenges in non-visual access to graphical materials in the context of fundamental perceptual and cognitive capabilities of human users. To overcome existing challenges, this thesis proposes a novel touch-based vibro-audio interface, developed with consideration of basic human information processing and user-centered design principles in mind.

1.1 Motivation

An increasing amount of information content used in the workplace, educational settings, and for everyday living is presented in graphical form (Hasty, 2009). For instance, it is estimated that 70% of the content of current textbooks is presented solely in graphical form (Hasty, 2009). Unlike text, graphics enhances the human ability to detect patterns and trends. Research has revealed that humans process graphics 60,000 times faster than text (Parkinson, 2013). Furthermore, it is estimated that 65% of the population are visual learners (as opposed to auditory or kinesthetic), so the visual nature of graphics caters to a large portion of the population (Smiciklas, 2012). With the advantages of graphics being substantial, even existing text-based information is being converted to graphical representation (such as graphs, figures and charts) and the use of graphics in all fields is only going to continue to increase. On the other hand, approaches for providing non-visual access to graphical material have not made much progress in reaching blind and visually-impaired people. As this demographic is estimated to number around 285 million people worldwide (WORLD HEALTH ORGANIZATION, 2011) the need for developing devices that are both accessible and usable for non-visual graphical rendering is critical for educational, social, and vocational purposes. Being in such an information-driven culture, blind and visually-impaired users will continue to miss out on this major component of information unless new non-visual solutions providing access to such graphical information are developed.

Much of the previous research and development projects have focused on designing new hardware/software systems that allow blind users to explore graphical elements using auditory (verbal or non-verbal), haptic, or multi-modal cues accessed via keyboard, mouse and/or force-feedback devices (Nees & Walker, 2005; Owen et al., 2009; Rastogi et al., 2013; S. Walker & Salisbury, 2003; Wall & Brewster, 2006; Wilson, Brewster, Halvey, Crossan, & Stewart, 2011). These systems have a steep learning curve because of unintuitive sensory translations (see Chapter 2 for details) and are often non-portable. Also, the approaches that address the development of accessible graphics have various shortcomings that hinder progress in reaching end-users (Hoggan, Crossan, Brewster, & Kaaresoja, 2009; Nees & Walker, 2008; Williamson, Crossan, & Brewster, 2011). For instance, many of these approaches involve purchase of expensive single-purpose hardware whose design and development was primarily driven by engineering principles rather than theoretical knowledge of human information processing and awareness about the needs and behaviors of end-user's (Giudice & Legge, 2008). In addition, most of the previous research work has emphasized technical design features and algorithms rather than conducting empirical experiments and behavioral evaluations. This has led to a huge information gap in accessing graphical information for visually-impaired persons (Raja, 2011).

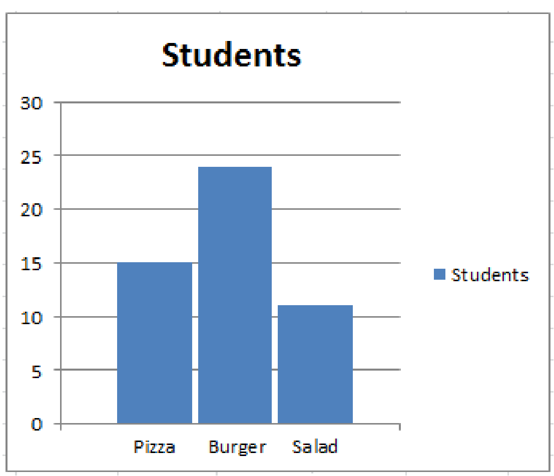

Figure 1.1 Bar graph summarizing data: students vote for their favorite food

The information gap is mainly attributed to lack of basic research on theoretical knowledge of human information processing and improper/insufficient sensory substitution. For instance, consider a bar graph (see figure 1.1) summarizing data collected in a class where students voted for their favorite food. Visual representation of such bar graphs can have two bars separated by as low as ~ 0.116 mm —the smallest object resolution at a viewing distance of ~400 mm— ("Naked Eye," 2013), but when this visual representation is directly translated into a tactile representation it becomes inaccessible due to the coarser nature of tactile resolution —1-2 mm— (Craig, 1999; Loomis, 1981). Also, a "two-point touch" test, from the encyclopedia of human biology, revealed that the smallest two-point separation that can be detected on the human fingertip —one of the most sensitive touch sensors known— is approximately 2-3 mm (S. J. Lederman, 1991). This limitation of lower spatial resolution of touch, as compared to vision (Jones & Lederman, 2006; Loomis, Klatzky, & Giudice, 2012), leads to restrictions in translating visual representations into tactile representations as there are no clear rules governing the down sampling of information for visual to tactile sensory substitution. In addition, most tactile representations are processed serially by contour following as opposed to the parallel processing used with vision. Because of serial processing, gaining information through touch is memory intensive, prone to error and often slow when compared with vision (Jones &

Lederman, 2006). Understanding such theoretical knowledge of sensory psychophysics, human behaviors and human information processing is critical for developing accessible graphics that are truly usable.

1.2 Approach

This thesis aims to bridge the information gap in accessing graphical information between blind / visually impaired users and their sighted peers by providing non-visual access to graphical material using an intuitive interface that: (1) provides dynamically updatable information on a device which is inexpensive (i.e. is based on off-the-shelf commercial hardware vs. highly specialized adaptive equipment), (2) is portable enough to be used in many contexts and environments, (3) is dual-purposed (meaning that the underlying hardware can be used for other applications), and (4) supports universal design principles (i.e., is highly customizable and includes many accessibility features in the native interface).

This thesis proposes a touch-based vibro-audio interface for presenting dynamic graphical information via commercially available touch-screen devices (such as smartphones and tablets) which satisfies the design criteria mentioned above. The approach focuses on spatial properties of the graphical material being conveyed through touch. Unlike other approaches that have focused on perception of the stimuli, the focus here is on the mental representation of the stimuli (graphical material) and how it can be used to support human spatial behaviors. The logic is that for an approach to be truly useful, learning must lead to an accurate representation in memory, similar to that derived from visual access, which supports subsequent mental transformations, computations, and behaviors (Giudice, Palani, Brenner, & Kramer, 2012). The current work involves empirical investigation of this logic by conducting a series of human behavioral experiments. Refining the perceptibility, usability and acceptability of the interface based on empirical evaluations, along with consideration of the design factors will lead to a better solution for filling the graphical information gap between blind persons and their sighted peers.

1.3 Goals and Hypotheses

The goal of this thesis is to address the problems stated in section 1.1 by designing a vibro-audio interface with the consideration of human information processing and user-centered design principles. The work involves experimentally evaluating whether use of the interface leads to development of accurate spatial representation of the graphical information in user's memory. The main focus of this thesis is on evaluating the fundamental perceptual and cognitive capabilities of human users via human behavioral experiments. These experiments assess human spatial behaviors that involve accessing the spatial representations developed from learning using the interface.

This thesis hypothesizes that *use of the interface leads to development of an accurate mental representation of the graphical information being conveyed*. That is the spatial representation developed from learning using the interface is functionally equivalent to that of developed from traditional hardcopy tactile graphics. This work also hypothesizes that incorporation of panning and zooming methods enhances learnability of large and deep format graphical material and produces accurate spatial representation in the user's memory. That is, using panning and zooming operations will yield accurate learning in non-visual settings than when not using these operations. This thesis also documents the haptic illusions that arise from the pattern of finger movement on the smooth touch-screen displays and analyses the underlying constraints of haptic perception.

1.4 Scope of thesis

Broadly defined, Information is represented as two types; namely, 1) sentential, 2) graphical. Sentential representations are sequential, such as the propositions in a text whereas graphical representations are indexed by location in a plane. The fundamental difference between these two types is that graphical representations preserve the geometric and topological relations among the components of the information being conveyed, while the sentential representation does not (Larkin & Simon, 1987). Graphical representations include diagrams, maps, plans, animations and virtual reality (Scaife & Rogers, 1996). Although each of these terms (such as image, diagram, picture, etc.) signify something on their own, it is important to realize that these terms are broad and often overlap both in common usage and meaning. For instance, a diagram represents appearance and structure or explains how something works while an image represents the external form of a person, scene or object (Wordweb dictionaries, 2013). Despite the difference in definition, both types fall under graphical representations as they are rendered in graphical form as opposed to textual form. All these graphical representations are composed of points, lines, and regions which involve spatial aspects. This thesis concentrates only on these spatial aspects of the graphical formation, as this is most conducive to haptic rendering and perceptual comprehension. As discussed in section 1.2, the focus of this thesis work is to evaluate the fundamental perceptual and cognitive capabilities of human users in accessing and learning the spatial information conveyed by graphical representations. To perform this evaluation, this thesis concentrates only on graphical formations such as graphs, shapes and maps and does not include other graphical formations such as diagrams, images, pictures and animations.

1.5 Intended audience

This thesis primarily addresses an audience that is related to the domain of spatial information science and engineering, especially researchers who are involved in studying non-visual spatial information processing. This thesis is also intended for researchers and scientists involved in the field of accessibility. Touch-screen-based Industries may find the design principles and learning strategies derived from the exhibited human behaviors as useful in developing hardware/software for touch-based devices. Such an audience can include, but is not limited to, blind and visually-impaired persons, researchers and industries working on eyes-free notification, multimodal gaming, and many others connected with non-visual interfaces.

1.6 Organization of remaining chapters

Chapter 2 provides a discussion of earlier research on addressing the issue of non-visual access to graphical material and how the current design of the vibro-audio interface has evolved from this literature. Chapter 3 describes the initial investigation on usability of the interface and describes the methods and results for the first three behavioral experiments (Exp-1, 2, and 3). Chapter 4 elaborates on the limitations of the interface and proposes potential solutions for these limitations. It then describe the behavioral experiments (Exp-4) conducted to determine the efficacy of one of the solutions proposed in Chapter 4. Chapter 6 elaborates the other solution proposed in Chapter 4 and then describes the behavioral experiments (Exp-5) conducted to determine the efficacy of the solution. Chapter 7 summarizes the major findings of the thesis and discusses possible future directions that could be extended based on the research related to non-visual graphical access.

CHAPTER 2

LITERATURE REVIEW

This chapter reviews some of the previous approaches to accessible graphics and highlights their pros and cons with respect to the focus of this thesis.

2.1 Non-visual graphical interfaces

Compared to advancements in access to text-based material, there has been far less development in access to graphical material. This is mainly because of prohibitive costs of the technology, which hinders the interfaces (or devices) from reaching the blind user. In addition to cost constraints, many approaches have not emphasized critical human perceptual factors in their design and have ignored end-user needs. Understanding the challenges to non-visual graphical accessibility requires appreciation of the amount of spatial information available from vision. Despite having five major external sensory subsystems (Visual, Auditory, Somatosensory, Gustatory and Olfactory), humans primarily use their visual, haptic and auditory subsystems for gaining spatial information about the surrounding world (Coren, Ward, & Enns., 2004; Hatwell, 1993). Of the three "spatial senses," vision is generally accepted as the primary source for acquiring spatial information as it allows simultaneous perception (parallel information processing) of multiple details over a large field of view (Thinus-Blanc & Gaunet, 1997). For instance, consider looking at a “You are here” map of a new building that you are visiting. With vision, it is trivial to see and immediately grasp the spatial configuration of graphical elements and their relations within the map. In addition to vision, one can also use haptic or auditory cues for gaining spatial information about the environment. But sighted individuals likely do not pay much attention to these cues as they convey very little information compared to vision and are less accurate. However, in the absence of vision, one is forced to rely on haptic or auditory cues.

During visual learning of graphics, vision performs two activities synchronously; (1) it allows identification of the graphical elements based on their visual parameters (such as color and pattern) and (2) relates the graphical elements based on their position, structure and orientation subtended with respect to the visual axis. In conjunction, these two activities allow for the building up of global spatial images in the perceivers memory. In contrast, during non-visual learning, these two activities must be performed by at least two different sensory subsystems (e.g., haptic and auditory). For instance, while accessing tactile graphics using one or more fingers, the mechanoreceptors of the fingers (cutaneous sense) identifies the graphical elements, and the kinesthetic sensory system, which detects limb and joint movements, relates these graphical elements to each other based on their position, structure and orientation. Although, studies have demonstrated that haptic input can lead to internal spatial representations that are functionally equivalent to those obtained from visual input (Cattaneo et al., 2008; Giudice, Betty, & Loomis, 2011), haptic input coupled with audio cues are considered better than either haptic or audio input in isolation. Also, studies have demonstrated that presenting information through multiple senses (Multimodal interfaces) increases the readability of the graphical material (Zeng & Weber, 2010). This understanding of multimodal non-visual information processing in humans forms the foundation for designing the vibro-audio interface at the heart of this thesis.

Many non-visual interfaces have been developed for providing access to graphical material, but only a few are still in existence. Part of the reason may be due to a disconnect between engineering factors and a device's perceptual and functional utility. This means more basic research should be conducted investigating whether these interfaces are providing access to the graphical material by considering the needs of the intended users, rather than simply implementing an elegant algorithm. In the following discussion, the most common or most promising approaches are categorized based on their sensory characteristics. For each, pros and cons are highlighted with respect to the sensory translation rules, cost, usability, and device functionality.

2.2 Haptic-based approaches

Tactile graphics are considered the most frequently used approach to accessible graphics and are commonly used in the education sector ("Perkins Museum," 2013). They allow the user to feel the graphic rendering and have been in use for over 200 years (Eriksson, 1998). A typical example is a paper based tactile map that is used to teach spatial concepts (Golledge, 1991). Tactile graphics are usually displayed on embossed tactile paper in which embossers punch the paper with varying height dots to create raised shapes or thermo-form (swell) paper which contains thermo capsules that rise when heat is applied (Goncu & Marriott, 2011). The major drawback of these approaches is that they are based on non-refreshable media which do not support interactive use of graphics, i.e. once authored, they are static and cannot be updated unless completely reproduced. The graphical material also requires being authored by specialists in order to be embossed on paper or swell media, which is an expensive and extremely time consuming process.

2.2.1. Force-feedback devices

Most of the research addressing haptic graphic rendering beyond traditional hardcopy tactile graphics has used force-feedback devices. These devices provide a fixed or controllable frame of reference. The PHANTOM from Sensable Technologies ("Phantom Omni," 2013), or the Logitech WingMan force-feedback mouse (Yu & Brewster, 2002) represent some examples of these force-feedback technologies. The BATS (Parente & Bishop, 2003) project used force-feedback joysticks coupled to a pointer for providing tactile bumps and feedback over an interface as the cursor crossed environmental boundaries or feature changes. Another study utilized a force-feedback 3-dimensional pen to guide the user's hand in a trajectory, outlining geometry of simple shapes (Crossan & Brewster, 2008). These devices suffer from the technological limitation of the hardware in that they require expensive or non-portable add-on equipment that is generally bulky. The price for the desktop version of PHANToM, which is the cheapest one in the range, is over $10,000 US. In addition, they use an indirect interaction between the user and the interface, which is less intuitive and potentially confusing than a direct interface, where the user interacts directly with the interface (e.g., as with a touch-screen). These devices also have a constrained extent (i.e. a small workspace) and require frequent panning or scrolling operations to explore larger graphics. In addition, authoring the stimuli is expensive, time consuming and are not practical for accessing many graphics in real-time.

2.2.2. Refreshable displays

The advent of refreshable tactile displays presented an opportunity to overcome many of the limitations of paper based tactile graphics. Refreshable tactile displays are mainly composed of units called taxels —touch stimulation units, which replace the screen pixels— (Vidal-Verdú & Hafez, 2007). The taxels are either based on electromagnetic, piezoelectric actuators or electrostatic (Raja, 2011; Klatzky, Giudice, Bennett, & Loomis, In press). The display contains multiple actuators that dynamically change in time. When the display is activated, the user traces the area to feel what is on the display. Larger displays suitable for presenting tactile graphics are expensive (e.g. A4 size displays are around US $50,000) and have quite low resolution (Goncu & Marriott, 2011). Refreshable tactile displays are further classified into two: static and dynamic (virtual screens).

The static-refreshable displays have an array of taxels that completely cover the entire width and length of the large flat surface display, such that the entire graphical material is displayed at once. This means the display will be activated only once for a given graphic and subsequently refreshes for different graphics.

This is like fixing the display to render a digital image, but once fixed (e.g., the pins are raised), it is not able to be changed unless the pins go down and the graphic is erased. Some examples of static-refreshable displays are HyperBraille's BrailleDis 9000 (Völkel, Weber, & Baumann, 2008), METEC's DMD 12060 (Schweikhardt & Klöper, 1984) and NIST ("NIST 'Pins' Down Imaging System for the Blind," 2002). In addition to the tactile actuator arrays, the BrailleDis 9000 unit can take multi-touch gestural inputs based on finger gestures over the surface. In contrast to static-refreshable displays, the dynamic-refreshable display uses a small array of taxels (finger sized) coupled with a pointer device (i.e., a mouse) which points over a virtual tactile screen (Raja, 2011). The tactile pins actuate up and down dynamically based on position of the mouse on the virtual tactile screen. Examples of dynamic-refreshable displays include, HAPTAC (Hasser, 1995), TACTACT (Kammermeier, Buss, & Schmidt, 2000), Virtouch mouse (Kammermeier & Schmidt, 2002) and VITAL (Benali-Khoudja, Hafez, Alexandre, Kheddar, & Moreau, 2004). The major drawback of most static-refreshable displays is the cost and the resolution capabilities. These devices are very expensive due to the high cost of taxel actuator units and the density of these units required to cover the entire extent of the graphic with a reasonable tactile resolution (Raja, 2011). Also, haptic rendering on such displays is a demanding process, as the tactile resolution is lower than vision, and significant filtering and simplification is required before presenting a graphic using a tactile display (Zeng & Weber, 2010). Also, these devices are not commercially available, and are often not portable. The main problem with dynamic-refreshable displays is that the mouse pointer only registers user's relative motion, and the user can become "lost" in the nonvisual virtual scene after a while, as there is no fixed frame of reference. Most of these displays are prototypes still in the research phase and are not commercially available.

2.3 Audio-based approaches

Many research efforts have examined audio techniques for conveying eyes-free notification, spatial information, and context-specific information. The greatest amount of work has been done with auditory graph displays utilizing different sonification techniques where changes in the visual data are mapped onto auditory parameters such as pitch, loudness, timbre, or tempo (Dinger, Lindsay, & Walker, 2008a; B. N. Walker & Mauney, 2010a; B. N. Walker, 2002). The motivation for these approaches is to create the audio equivalents of visual rendering. Audio icons, or earcons, were evaluated to explore their effectiveness in conveying metaphoric meanings, for example an ascending tri-tone conveys "up". The AUDIOGRAPH system (Alty & Dimitrios I. Rigas, 1998) explored enhancements to the earcon concept, whereby musical sequences or relationships between musical sequences convey semantic information. Virtual Audio Reality (Frauenberger & Noisternig, 2003) and Multi-way Visual Analysis (McGookin & Brewster, 2006) are a type of force-feedback device that also used audio cues for presenting graphical information. These devices use non-speech audio to construct and provide quick overviews of graphical elements. Although results from these projects indicated that sophisticated audio sequences can be used to convey complex graphical information, the main problem with the audio-based approaches is that they suffer from a steep learning curve as users need to have a good understanding of musical concepts to interpret the auditory output and also be trained in the interface along with these musical concepts. Also, this approach is not a direct mapping and the translation of spatial information in the graph being mapped onto these concepts is not necessarily intuitive.

2.3.1. Language-based displays

In addition to use of audio, many research projects have explored the use of Natural language to convey information traditionally presented visually (Ferres, Lindgaard, & Sumegi, 2010; Giudice, 2004). Virtual Verbal Displays (VVD) used verbal descriptions of indoor geometry which are updated based on the location and orientation of the user in a virtual indoor layout (Giudice, 2004). The user can move his or her position or turn in the large indoor virtual environment by the use of keyboard arrow keys. This means the user can visualize the orientation, position and structure of the environmental elements based solely on the verbal descriptions. Experiments with this VVD display demonstrated that users could obtain 76% accuracy in localizing targets in physical buildings after exploring an indoor space with the VVD using dynamically-updated verbal descriptions based on spatialized audio (Giudice & Tietz, 2008). Notably, performance with the VVD did not significantly differ on an identical localization task after learning with a visual display, demonstrating that use of dynamic non-visual displays can lead to similar learning and navigation performance as is obtained from the same tasks with visual displays (Giudice, 2004). Some of the force-feedback devices also utilized natural language to describe graphical elements. An example is TeDub (Technical Drawings Understanding for the Blind), which is a type of force-feedback device that presents node-link diagrams such as UML diagrams where speech is used to describe the node's attributes (Petrie et al., 2002). In addition, Spearcons (which are highly compressed short sequences of speech) were found to be highly effective in conveying the spoken meaning of graphical objects to the user while not imposing the cognitive load that standard speech incurs on the human listener (Dinger, Lindsay, & Walker, 2008). These studies demonstrated that language-based displays are efficient in conveying orientation and position information about one's surrounds that are traditionally conveyed through visual access. These findings are important as the vibro-audio interface evaluated in this thesis also conveys some information via speech and audio.

2.4 Touch-Screen-based Interfaces

The advent and proliferation of smooth (e.g., smartphones and tablets) touch-screen-based devices has opened the door to a new era of multimodal interfaces incorporating combinations of auditory, vibro-tactile, and kinesthetic cues. With these devices, hand and finger movements over the display provide position and orientation cues through kinesthesis and the presence of visual elements are delivered by an external synchronized cue (such as audio or vibration) when the user touches that element on the touch-screen. These interfaces differ from the haptic devices described in section 2.1 as no meaningful cutaneous information is being conveyed beyond that the finger is contacting the device surface (Raja, 2011). Also these are direct perceptual interfaces that do not need a confusing sensory mapping but directly convey the information being rendered These devices are differentiated into two categories based on the perceptual cues provided: (1) audio-kinesthetic interfaces, which couple text and sound cues with hand movement and (2) haptic-audio interfaces, which add vibro-tactile feedback (Giudice et al., 2012).

2.4.1. Audio-Kinesthetic Interfaces

These devices employ audio (sound and speech) for presenting visual elements on the touch-screen. Examples of audio-kinesthetic interfaces include Timbremap, which uses sonification for representing complex indoor layouts on a touch-screen equipped smartphone (Su, Rosenzweig, Goel, de Lara, & Truong, 2010) and the PLUMB project, which uses sonification to describe auditory graphs on a touch tablet (Cohen, Rui, Meacham, & Skaff, 2005). An experiment using Timbremap showed that 81% accuracy was achieved in shape identification, demonstrating the efficiency of touch-screen devices in conveying graphical information (Su, 2010). Similarly, another project utilized a touch-pad to convey relative positioning of points of interest on a map using sonification (Jacobson, 1998). The importance of these earlier projects is that they provide clear evidence for efficacy of touch-screen devices to support users in learning graphical material, as is the goal in the current thesis.

2.4.2. Haptic-Audio Interfaces

Touch-screen-based Haptic-Audio interfaces differ from traditional hardcopy tactile stimuli and other electronic haptic devices as the cutaneous information being conveyed is purely through vibration on a smooth display surface, rather than the traditional method of feeling embossed lines or moving or vibrating pin arrays (Giudice et al., 2012). Examples of haptic-audio interfaces include TouchOver map, which showed that blindfolded-sighted participants could understand a road network through vibration and auditory labels when feeling a smartphone touch-screen, and then were able to accurately reproduce the map using vision while simultaneously exploring the now occluded display (Poppinga, Pielot, Magnusson, & K. Rassmus-Grohn, 2011). Here, the vibration was generated by rotating electro-magneto vibration actuators which were fixed internally in the device. In other approaches, vibration was generated by rotating electro-magneto vibration actuators that were either fixed to the fingers of the users or to the back of the device. An example of the former approach is the GraVVITAS project which used external vibration motors and multiple fingers during exploration (Goncu & Marriott, 2011). This research showed that graphs, shapes, and maps could be understood by blind users when learned from a touch tablet with external vibrators affixed to the user's fingers. Similarly, the SemFeel project showed that touch-screen devices with external vibration actuators are beneficial in supporting recognition of shapes and patterns (Yatani & Truong, 2009).

TouchOver map and GraVVITAS shares similarities to this thesis work. However, the focus of these studies significantly differ from the primary goal of this thesis.

For instance, none of the previous studies required development of an accurate spatial representation to perform the tasks and did not use formal statistical procedures to analyze user data. Unlike the vibro-audio interface being evaluated here, the development of these interfaces did not involve consideration of basic human information processing and sensory psychophysics. Also, their evaluations were not focused on constraints of haptic perception using smooth touch-screen displays as is our approach here.

CHAPTER 3

LEARNING GRAPHICAL INFORMATION USING A VIBRO-AUDIO INTERFACE

This chapter details the design requirements of the vibro-audio interface that is at the heart of this thesis. It then presents the functional and implementation details of the interface and then describes its initial usability evaluation. Three human behavioral studies are described and the methods and results of the experiments (Exp-1, 2 and 3) are elaborated.

3.1. Design Requirements

One of the most basic design requirements for any non-visual graphical interface is that it should allow blind users to apprehend an accessible version of the visual graphics being rendered. This means that the accessible graphic presented via a non-visual interface should contain functionally similar information as the original visual representation. However, conveying the information alone is not sufficient. For instance, consider the bar graph example mentioned in 1.1, the same information can be made accessible using natural language (such as "Pizza has 15 votes, Burger has 24 votes and Salad has 11 votes"), which conveys the key information and can lead to functionally equivalent behavior due to development of a common spatial representation. However, this linguistic data presentation does not provide the benefits of graphical representation such as geometric and topological congruence and computational off-loading (Scaife & Rogers, 1996). It has been often suggested that graphics resemble what they represent and provide some correspondence between the structures of the representation and its target (Shimojima, 2001). In addition, many researchers have investigated the differences between graphics and text and the benefits that can make graphics more effective than text. Such benefits include indexing, mental animation, macro/micro viewing, and graphical constraining (Goncu & Marriott, 2011). To obtain such benefits, the accessible graphic should have functional equivalence with the visual graphic by maintaining the spatial and geometric nature of the original rendering. This means that the blind users should gain at least some — though not all— of the benefits (as discussed above and in section 1.1) that sighted users obtain using graphics. These benefits can be obtained by conveying functionally similar graphics (as opposed to actual equivalence of information content) that is necessary to support a spatial task. This functional equivalence is important to bridge the information gap between blind users and their sighted peers. For instance, consider a classroom setting with a mixture of sighted and blind persons, where lectures are often explained with reference to graphics. It is critical for blind individuals to have access to functionally equivalent graphics to be competitive in such a collaborative setting.

Much empirical research has shown that haptic input can lead to internal spatial representations that are functionally equivalent to those obtained from visual input (Cattaneo et al., 2008; Giudice et al., 2011). Studies have also shown that blind users generally prefer tactile presentations to audio (Goncu, Marriott, & Hurst, 2010) with audio only being preferred in some exploration and navigation tasks (Goncu & Marriott, 2011). These findings shaped the initial design requirements for the vibro-audio interface discussed in this thesis, which is to provide access to graphics using combined haptic and audio. However, as mentioned in section 1.1, 2.1 and 2.2, approaches for tactile graphics accessed via hardcopy, keyboard, mouse and/or force-feedback devices have various shortcomings. This led to the next design requirement, which is to present the graphics in a low-cost, commercially available dynamic refreshable display that is portable, customizable and multi-purpose. This requirement was fulfilled by the recent advancements in touch-screen devices, which are inexpensive, have a dynamic refreshable display and can provide simultaneous haptic and audio feedback without the need for any additional equipment. However, despite satisfying the two design requirements mentioned above, many approaches (based on touch-screen technology) have still not reached the end-users because they incorporate unintuitive sensory translation rules and focus on engineering principles rather than user-centered design, which is termed as the “Engineering trap” (Giudice & Legge, 2008; Loomis et al., 2012). This led to the final and central design requirement of this thesis, namely, understanding the basic perceptual and cognitive capacities of human end-users and designing the interface with consideration of these human information processing parameters in mind.

As stated earlier, the advent and proliferation of smooth surfaced touch-screen-based devices (e.g., smartphones and tablets) has opened the door to a new era of multimodal interfaces incorporating combinations of auditory, vibro-tactile, and kinesthetic cues. Also, these devices satisfy the basic design criteria discussed above. These devices are also capable of indicating the presence of visual elements by an external synchronized cue (such as audio or vibration) when the user touches that element on the touch-screen, which makes them an ideal platform for native multimodal interface implementation.

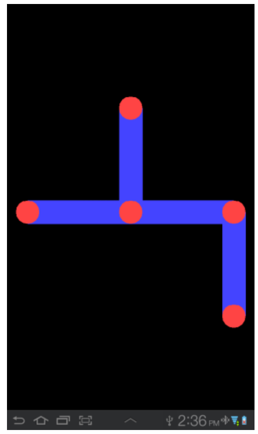

Figure 3.1 Vibro-audio interface displaying sample graphic

To begin with, a commercially available tablet was chosen as the platform which can track hand and finger movements over the display and provide position and orientation cues through kinesthesis. The prototype —vibro-audio interface— was based on a Samsung galaxy tablet with a 7.0 inch touch-screen running Android OS version 3.2, Target version 13. Vibro-tactile information was generated from the tablet's embedded electromagnetic actuator, i.e., an off-balance motor. Auditory information was produced and delivered from the device's onboard speakers. Users also received kinesthetic feedback as they moved their hand over the tablet's touch-screen, which also acted as a reference frame for positioning and orienting the graphic elements within the bounding frame of the touch-screen. Any object, visual or non-visual, that was displayed on the tablet's screen was referenced to the screen coordinate system (e.g., 1024x600 pixels in the case of the Samsung galaxy 7.0) and whenever an on-screen visual element was touched, pre-defined vibration patterns and auditory information was synchronously triggered at that coordinate (see(Raja, 2011) for details). The vibration patterns effects for the interface were based on the Universal Haptic Layer (UHL) developed by Immersion Corporation (“Immersion,” 2013). The UHL provides a set of pre-defined vibration effects that can be incorporated into the application by installing the UHL as a plugin for JAVA source code in Eclipse IDE. Since the device has only one embedded vibration motor, the vibration pattern will be felt evenly across the entire device screen. Hence, use of multiple fingers (either from the same hand or a different hand) was restricted as the haptic feedback cannot be differentiated between the different fingers. On the bright side, the constrained use of one finger provided a strong focal stimulus to the finger digit touching the screen, which is perceived as a tactile point or line as the finger moved over the stimuli. Although many stimulus variables can be manipulated by altering the haptic and audio cues, only a fixed set of parameters were considered for this prototype interface. The parameters were established in earlier psychophysical studies (Raja, 2011) that identified the vibro-tactile line width that is most conducive to line tracing and contour following and the cues (vibratory or audio) that best differentiate different visual elements. Based on these previously established parameters, all lines in the current prototype interface were rendered with a width of 8.9 mm (0.35 inch), which corresponded to 60 pixels on the tablet's screen. This was also used as the minimum inter-line distance for all stimuli. Unlike many other non-visual interfaces such as force-feedback devices and mouse-based haptic refreshable displays, the vibro-audio interface provides a natural mapping of stimulus information to what is being perceived, while also employing a relatively large (7.0 inch) haptic workspace which can be quickly and easily updated in real-time.

3.2. Usability evaluation of the vibro-audio interface

As discussed in Chapter 2, previous research on accessible graphics using auditory (verbal or non-verbal), haptic, or multi-modal cues has focused on design guidelines and user preferences of the interface (Maclean, 2008; Nees & Walker, 2005), psychophysical factors characterizing optimal display properties to be implemented (Raja, 2011) or the nature of the perceptual mapping employed (B. N. Walker, 2002), or interpretation and legibility of specific information being displayed (Hoggan, Brewster, & Johnston, 2008). However, these studies did not address the constraints of human information processing when learning with such an interface. To my knowledge, none of these studies focused on how accurately graphical information can be learned from the interface and represented in memory as a global spatial image. Accordingly, three human behavioral studies were conducted to investigate the human spatial behaviors that are involved with accessing the spatial representations developed from learning using the vibro-audio interface. The first experiment assessed the users' ability to comprehend the relative relations and global structure between elements on a bar graph (Experiment 1), the second experiment assessed the users' ability to recognize patterns via a letter identification task (Experiment 2), and the third experiment evaluated the users' ability to recognize orientations of complex geometric shapes on a shape discrimination task (Experiment 3). Each of these experiments represent a different set of human behavior that encourages users to access mental spatial representation built up from learning using the new vibro-audio interface. The performance with the interface was then compared to the performance with the traditional technique of tactile graphic rendered using information-matched hardcopy embossed material.

3.3. Experiment 1: Learning Bar Graph

Graphs and charts are the primary techniques for representing numeric data as they convey the information in the simplest possible way. Accessing such numeric data is critical in many educational and vocational contexts. Although there are many types of graphs, the bar graph was chosen for this experiment because it displays discrete and categorical data. To understand such categorical data and visualize it as a bar graph, one must be able to access and learn the individual bar's position, height, and global relations with respect to the other bars in the graph. Although one can readily understand a bar graph with vision, questions remain about the best method to present this information to a blind individual for accurate learning and representation as a global spatial image in memory. Hence, this experiment assessed whether the use of the vibro-audio interface supports accurate learning of relative relations and global structure of various bar graphs. The performance with the interface was expected to be on par when compared to the same tasks performed using hardcopy tactile stimuli.

3.3.1. Method

Twelve sighted participants (six males and six females, ages 18-35) and six additional congenitally blind participants (3 males and 3 females, ages 22-43) were recruited for the experiment. All gave informed consent and were paid for their participation. This experiment (and all experiments in this thesis) was approved by the Institutional Review Board (IRB) of the University of Maine and took between 1.5 and 2 hours. Of note, it is important to carefully consider whether blindfolded-sighted participants are a reasonable sample when generalizing to blind participants. Inclusion of sighted participants is justified here as the work focuses on testing the ability to learn and represent non-visual material which is equally accessible to both groups. In support, previous studies with auditory graphs (B. N. Walker & Mauney, 2010), and tactile maps (Giudice et al., 2011) found no differences between blind and blindfolded-sighted groups. If anything, the performance of the blindfolded-sighted participants in the current experiments represents a conservative estimate of interface efficacy, as this group is likely to be less accustomed to using haptic cues as a primary mode of information gathering (Giudice et al., 2012). Although the sample is too small to make valid statistical comparisons between groups, the similarity of performance observed between blindfolded-sighted and blind participants provides support for the validity of our subject sampling decision.

3.3.2. Conditions, stimuli and apparatus

Two display mode conditions were evaluated in this experiment, one that employs the vibro-audio interface at learning and another that employs hardcopy tactile stimuli produced by a graphics embosser (the gold standard for tactile output). In the vibro-audio condition, a Samsung Galaxy Tab 7.0 Plus tablet, with a 17.78 cm (7.0 inch) touch-screen was used as the information display. Vibro-tactile feedback was generated when the user's finger touched the stimulus on the screen and auditory information was provided by tapping the vibrating region. Lines rendered in the vibro-audio mode were given a constant vibration, based on the UHL effect "Engine_100," which uses an infinite repeating loop at

250Hz with 100 percent power. The tops of the bars in the bar graphs were indicated by a pulsing vibration, based on the UHL effect "Weapon_1," which uses a strong infinitely repeating wide pulse at a frequency of 10-20 milliseconds. Pulses were given in a 60 x 60 pixel (0.35 x 0.35 inch) region encompassing the node at the edge of each bar. In the hardcopy conditions, tactile analogs of the same stimuli were produced on paper by a graphics embosser (ViewPlus Technologies, Emprint SpotDot). The paper was then mounted on a second Galaxy tablet such that auditory information could be given in real-time matching the available information content with the vibro-audio interface. Exploration with both displays was performed using only one finger (dominant) and the user's movement behavior was tracked via the device's touch-screen as they felt the stimuli. During the experiment, participants sat on an adjustable chair and adjusted the seat height such that they could comfortably interact with the experimental devices which rested on a 76.2 cm (30 inch) height table in front of them. During the learning phase of each experimental trial, participants wore a blindfold (Mindfold Inc., Tucson AZ).

Figure 2.2 Example stimuli displayed on the Touch-based device with the vibro-audio mode for the three experiments. Analog hardcopy tactile stimuli (not depicted) were used as a comparison in each experiment.

3.3.3. Procedure

A within subjects design was used in the experiment. In each display mode condition (hardcopy and vibro-audio), participants learned bar graphs and performed subsequent testing tasks (graph trials were randomized within display mode block, with block order counterbalanced between participants). A sample bar graph displayed in the vibro-audio mode is shown in Figure 3.2. Each display mode condition had three bar graphs that included a graph with 3, 4, and 5 bars (presentation order was randomized within graph set, with set order alternating between participants). Each bar was assigned a name, with set 1 based on food: pizza, burger, salad, chocolate and ice cream; and set 2 on disciplines of study: biology, physics, chemistry, mathematics, and computer science. Whenever the user tapped on the bar, the name of the bar was spoken as an audio message.

The study consisted of a practice, learning, and testing phase for each display mode condition, for a total of 10 trials. The first practice trial in each display mode was a demo trial where the experimenter explained the task, goal, and strategies and the participant explored the stimuli with corrective feedback provided. Participants were told that the height of each bar represented how many people liked the specific food category (Set 1) or how many people were enrolled in the class (Set 2). In the second practice trial, blindfolded participants were asked to perform the complete graph learning procedure, followed by the complete test sequence performed without blindfold. The experimenter evaluated their answers immediately to ensure that they understood the task correctly before moving on to the experimental trials. During the learning phase, participants were blindfolded and asked to learn the graph. Participants were asked to indicate to the experimenter when they believed that they had learned all of the material represented. Once indicated, the experimenter removed the device and the participant was then allowed to lift their blindfold to continue with the testing phase.

The testing phase consisted of two tasks: (1) a spatial relation task and (2) a graph reconstruction task. In the spatial relation task, participants answered four questions about the bar graph they just learned. Two of the questions assessed spatial relations between bars. For instance, “What is the relation between apple and orange?” The answer required a directional response (e.g., apple is left/right of orange), and a height judgment (e.g., apple is taller/shorter than orange). The other two questions assessed the ability to think of the individual bar position in a global context. For instance, “Which is the second highest bar?” “What is the middle bar?” To reduce recall errors, the names of the bars were given in a list.

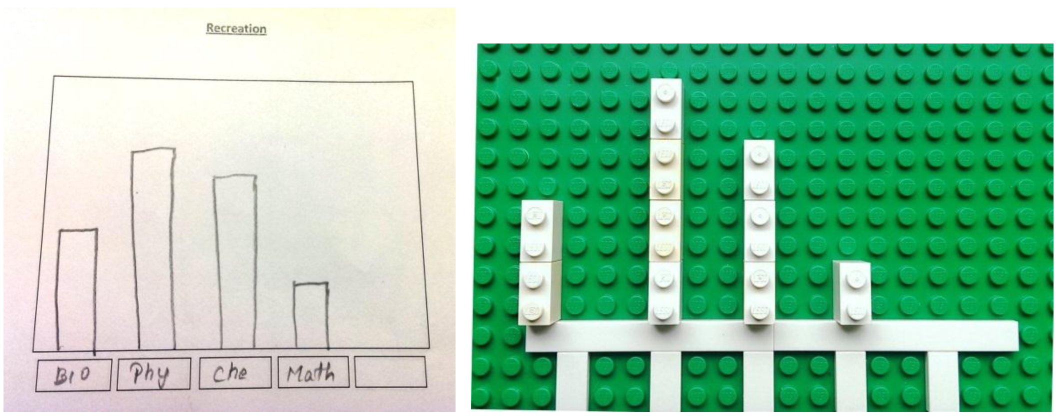

Figure 3.3 Practice reconstructed graph by sighted participants (left) and blind participants (right)

In the reconstruction task, participants were asked to draw the graph on a template canvas of the same size as the display and to label each bar. Five equidistant textbox place holders were provided to indicate the possible bar positions (see Figure 3.3). The only procedural differences for blind participants were that the questions were read aloud by the experimenter and the reproduction task was done with Lego™ pieces on a board with affixed position indicators (see Figure 3.3). They labeled each bar by verbally indicating its name. All reconstructed graphs were analyzed in terms of whether individual bars had the correct label, position, and relative height in relation to the graph's global structure.

3.3.4. Experimental measures and analyses

From this design, the following measures were evaluated as a function of display mode condition.

1. Learning time: The learning time can be interpreted as an indication of relation between cognitive effort and time taken for learning. That is, the greater the learning time, the higher the cognitive load for the condition. The learning time is the time taken from the moment they touch the screen until they confirmed that they have completed learning of the graph. The time was measured from log files of each trial that was created and stored within the device.

2. Relative height accuracy: This is the spatial height relation between any two individual bars (e.g., physics was taller than chemistry). This was measured from the reconstructed graph.

3. Relative directional accuracy: This is the spatial direction relation between any two individual bars (e.g., physics was left of chemistry). This was measured from the reconstructed graph.

4. Relative position accuracy: This is the spatial position of an individual bar with respect to its global spatial context (e.g., physics is the middle bar). This was also measured from the reconstructed graph.

5. Reconstruction accuracy: The reconstruction accuracy measures the accuracy of the global spatial representation from the reconstructed graph. This was measured by comparing the spatial pattern of the reconstructed graph with the actual graph. A discrete scoring was applied based on the correctness of the reconstruction (i.e., 1 for each correct bar in the graph).

6. Bar labeling accuracy: This is the relative quantitative information of an individual bar with respect to the global spatial context. Labels are crucial in such categorical data as changing labels will eventually change the data represented. The accuracy in labeling was measured from the reconstructed graphs.

3.3.5. Results

Performance data for each of the measures described above were analyzed and compared between the two display modes. The most important finding, as shown in Figure 3.4, is the similarity of performance across all measures for the two display modes (hardcopy mode or vibro-audio mode) and the two participant groups (blindfolded-sighted and blind).

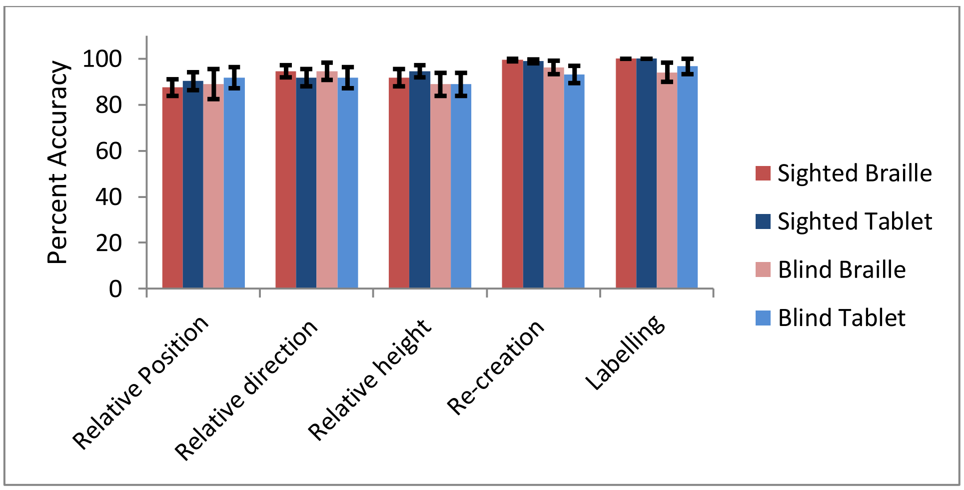

Figure 3.4. Accuracies on test measures as a function of display mode and subject group.

The results of paired-sample t-Tests between the two display modes (hardcopy and vibro-audio) were highly in-significant for all measures except learning time. Below are the t and p values for each of the measures.

Measures	Sighted			Blind		
	t	df	Sig. (2-tailed)	t	df	Sig. (2-tailed)
Learning time	-4.924	35	0.000	-4.423	17	0.000
Relative Height accuracy	-0.329	35	0.744	0.000	17	1.000
Relative directional accuracy	0.329	35	0.744	0.437	17	0.668
Relative position accuracy	-0.828	35	0.413	-0.325	17	0.749
Reconstruction accuracy	1.000	35	0.324	1.409	17	0.177
Labeling accuracy	0.000	35	1.000	-0.660	17	0.518

Table 3.1. Results of the paired sample t-Tests between display modes

Repeated measures ANOVAs were also conducted on the measures of interest to assess if there were effects of the number of bars (e.g., 3, 4, or 5) between the two display modes, but no statistically significant differences were found, all p-values >0.05.

From these results, it is evident that use of a vibro-audio interface on a touch-enabled device supports accurate learning of relative relations and global structure of a bar graph. It can be seen that in general, both blindfolded-sighted and blind subjects yielded higher accuracy values with the reconstruction task than with the spatial relations task. This result may be due to reconstruction being done sequentially, whereas performance on the spatial relation questions required making judgments about bars that often required non-contiguous and non-sequential judgments. Also, it can be inferred from figure 3.4 that subject's average accuracy with the vibro-audio mode for measures of positional accuracy, relative direction, relative height, and labeling were numerically higher than in the hardcopy tactile mode. For both groups (blind and blindfolded sighted participants), superior performance for the hardcopy mode was observed in learning time ($p < 0.001$). This outcome is not surprising, as it is easier to find and track the line using the embossed hardcopy stimuli. Despite differences in learning time, the similarity in output performance provides evidence that once learned, representations for both displays were able to support the same level of spatial behaviors. Together, it can be inferred that accuracy with the vibro-audio mode was numerically higher than with the hardcopy mode. Although the difference is not significant, the higher accuracy, and null results for any statistical differences with the vibro-audio mode provides strong support for the efficacy of the interface in supporting development of accurate spatial representations.

3.4. Experiment 2: Letter recognition

Pattern recognition is a key component of extracting data from graphics and learning about their content. To investigate this process, Experiment 2 assessed whether the use of the vibro-audio interface assisted in recognition of familiar patterns via a letter identification task. This experiment used the same vibro-audio interface and hardcopy tactile stimuli as in Experiment 1 but for recognizing patterns based on capital letters from the English alphabet. Letters represent complex but well known shapes and require participants to trace the contour of the stimuli and build up a global representation of its shape in order to correctly name the letter. Early research with the Optacon, a device that used an array of 144 electro-tactile stimulators felt by the finger, proved useful for real-time letter recognition and even limited reading (Linvill & Bliss, 1966). However, to my knowledge there have not been any studies addressing non-visual letter recognition with modern vibro-tactile touch-screen devices. Although letters are used as stimuli, the focus of this experiment is on the more general task of comparing the pattern recognition performance between the two display modes and not on reading tactile letters.

3.4.1. Method

The participants here were the same as those in Experiment 1.

3.4.2. Conditions, stimuli and apparatus

Similar to Experiment 1, two display mode conditions were evaluated (vibro-audio and hardcopy stimuli). The apparatus used here was the same as in Experiment 1. Six letters were used during the experimental trials that included: D, F, M, P, T, and W (with N and C used in the practice conditions). The letters were selected such that each display mode condition included three unique patterns including a letter with just straight lines (F or T), a letter with curves (D or P), and a letter with oriented lines (W or M). This is because oriented lines and curves are common in many graphics and the ability to trace and recognize such lines are crucial for understanding the graphics content. Similar to Experiment 1, the lines rendered in the vibro-audio mode were given a constant vibration, based on the UHL effect "Engine_100" and at each vertex a pulsing vibration (based on the UHL effect "Weapon_1") was provided. As nodes at non-orthogonal vertices were not symmetric, the width of the pulsing region varied depending on the intersecting angle of the lines. No audio cues were used in this experiment. Similar to Experiment 1, exploration with both displays was done using only one finger (dominant) and the user's movement behavior was tracked via the device's touch-screen as they felt the stimuli.

3.4.3. Procedure

Similar to Experiment 1, a within subjects design was used here. The procedure of two practice trials and three experimental trials per display mode (counterbalanced) was also the same as in Experiment 1. The task here was for blindfolded participants to explore the stimuli (one of six randomly presented letters) and to name the letter as soon as it was recognized. If the letter was misidentified, a second learning period was allowed following the same procedure. Incorrect identification on the second learning phase was considered a miss and participants moved on to the next trial.

3.4.4. Experimental measures and analyses

The following measures were evaluated as a function of display mode condition.

1. Learning time: As mentioned in Experiment 1, The Learning time is the indication of cognitive load imposed on user and is the time taken from the moment they touch the screen until they confirmed that they had completed identification of the letter. The time was measured from log files of each trial that was created within the device.

2. Pattern recognition accuracy: The accuracy in pattern recognition was measured as a correct/incorrect response by the participant.

3. Number of learning Iterations: This represents the number of times participants took to recognize the letter. Since participants were given only two chances, this measure can have only two values (1 or 2). This was measured for each of the trials.

3.4.5. Results

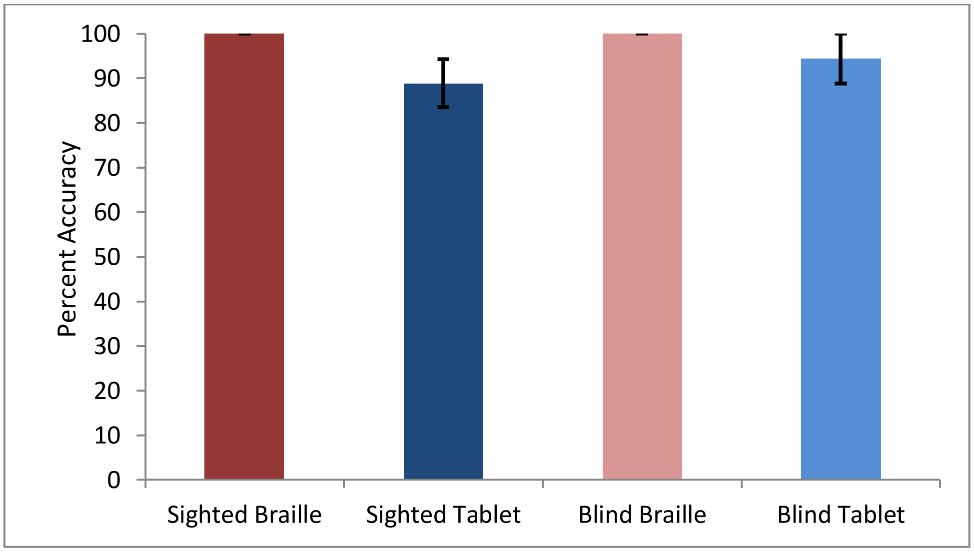

Figure 3.5. Letter recognition accuracy as a function of display mode and subject group.

Corroborating what is shown in Figure 3.5, the letter recognition accuracy performance (for both participant groups) with the vibro-tactile mode was numerically lower than the 100% accuracy observed in the hardcopy mode. The results of paired-sample t-Tests between the two display modes are as follows,

Measures	**Sighted**			**Blind**		
	t	**df**	**Sig. (2-tailed)**	**t**	**df**	**Sig. (2-tailed)**
Learning time	-6.137	35	0.000	-7.418	17	0.000
Patter recognition accuracy	2.092	35	0.044	1.000	17	0.331
Number of Iterations	0.324	35	0.096	-2.204	17	0.042

Table 3.2. Paired sample t-Tests results of Letter recognition task

The mean learning iterations (sighted vibro-audio: 1.222, sighted hardcopy: 1.083, blind vibro-audio: 1.05, blind hardcopy: 1.27) for both the modes are greater than 1 learning iteration, which suggests that even in the hardcopy modes participants made errors in their first pattern recognition attempt.

As in the previous experiment, a significant difference was observed in learning time ($p < 0.01$), again manifesting as the vibro-audio mode being slower than in the hardcopy mode.

The difference in the letter recognition accuracy performance is likely due to the impoverished orientation cues available in the vibro-audio mode, which made it harder to detect line orientation, especially if the line was slanted or curved. This can be mainly attributed to the smooth nature of the touch-screen devices (as opposed to the physical bumps in the hardcopy stimuli). Despite the differences, the performance with vibro-audio interface was nearly equivalent to hardcopy output on most measures. Indeed, although they only had a short period of practice to become accustomed to the device, the letter recognition task took only ~2-3 minutes. This was a remarkable outcome given that proficiency on shape and letter recognition with other non-visual devices using haptic sensing (e.g., the Optacon) took well over 100 hours (Bliss, Katcher, Rogers, & Shepard, 1970).

3.5. Experiment 3: Orientation discrimination

As stated in section 3.4, use of orientation information is an integral component of graphical material and visualizing orientation is crucial for gaining global spatial information. Hence, this experiment assessed whether the use of the vibro-audio interface supports learning and representing the orientation of irregular shapes. Previous research has shown that touch-screen devices with external vibration actuators are beneficial in supporting recognition of shapes and patterns (Goncu & Marriott, 2011; Yatani & Truong, 2009). Unlike these studies, the focus in the current experiment not only requires learning a oriented shape but that the representation built up from learning was sufficiently robust to identify the shape in the presence of geometrically identical alternatives.

3.5.1. Method

The participants here were the same as in the previous experiments.

3.5.2. Conditions, stimuli and apparatus

The conditions (two display modes) and apparatus were similar to that of previous experiments. Six distinct shapes were used as experimental stimuli (with two additional shapes used for practice). All the stimuli were four-sided polygons that were misaligned with the display's intrinsic frame of reference. Only the bounding contour of the shape was rendered and none of the shapes were readily namable polygons (refer to figure 3.2 and figure 3.6).

3.5.3. Procedure

Similar to the previous experiments, the within subjects design also followed the same procedure of two practice trials and three experimental trials per display mode (counterbalanced). Three distinct shapes were used in each display condition (counterbalanced). The task in this experiment was for blindfolded participants to explore the shape during a learning phase and to indicate once they felt that they were familiar with its global geometry and orientation.

Figure 3.6. Alternatives for the example shape displayed in figure 1

During learning, participants were asked to imagine the vertices, length of the sides, and the orientation of the shape on the display. Upon verbal indication that the shape was learned, the experimenter removed the device and placed an A4 size paper containing the same shape along with three geometrically identical alternatives. The shapes were numbered from 1 to 4 in a column (all stimuli were size-matched). Participants removed their blindfold and marked the shape which matched the orientation of the shape previously learned. Blind participants performed the same task but made their comparison based on a board with 3D cut-outs of the four shapes.

3.5.4. Experimental measures and analyses

Experimental measures analyzed in this experiment include:

1. Learning time: Similar to previous experiments, this is the indicator for the cognitive load imposed on the user in learning the experimental stimuli and is the time taken from the moment they touch the screen until they confirmed that they had completed learning of the shape. It was also measured from log files of each trial that was created within the device.

2. Orientation accuracy: This is the accuracy in identifying the shape with correct orientation by eliminating the alternatives

3.5.5. Results

No reliable differences were observed between the two display modes for orientation accuracy as assessed by a paired samples t-Test,

Measures	Sighted			Blind		
	t	df	Sig. (2-tailed)	t	df	Sig. (2-tailed)
Learning time	-7.170	35	0.000	-5.076	17	0.000
Orientation accuracy	0.298	35	0.768	0.000	17	1.000

Table 3.3. Paired sample t-Tests results of shape identification task

These results suggest that learning with the vibro-audio mode was functionally equivalent to learning with the hardcopy mode for apprehending shapes and for identifying the reference shape from geometrically identical alternatives.

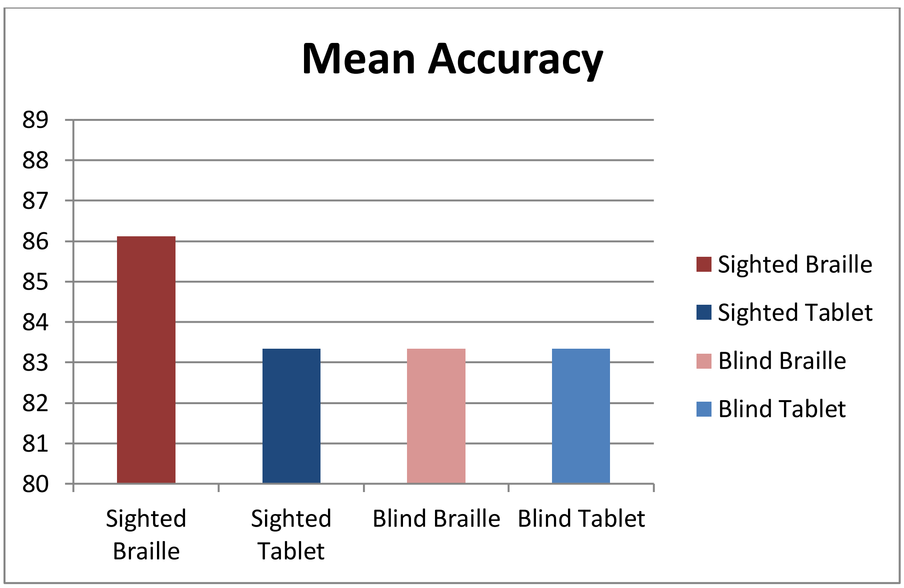

Figure 3.7. Orientation accuracy as a function of display mode and subject group.

The orientation performance of blindfolded-sighted participants yielded lower numeric means (~83% mean accuracy) with the vibro-audio mode, contrasting with ~86% accuracy for the hardcopy mode. However, as is shown in Figure 3.7, the performance of blind participants was equal (~83% mean accuracy) for both display modes, suggesting the efficacy of the interface. As with the previous experiments, learning time with the hardcopy mode was significantly ($p < 0.001$) faster than with the vibro-audio interface. Importantly, as with the previous experiments, no significant differences were found between the two display modes in this experiment, which again demonstrates the efficacy of the vibro-audio interface in supporting development of accurate spatial representations.

3.6. Discussion

Results from three experiments provided strong support for the efficacy of the vibro-audio interface for learning the experimental stimuli and in building up accurate mental representations for both blindfolded-sighted and blind participants. These findings are important as this interface provides dynamic and readily implemented information, whereas hardcopy material is static and requires expensive, highly specialized equipment to produce. In addition, as the vibro-audio interface is based on inexpensive, multi-purpose, and commercially-available hardware, it represents a viable alternative to the expense and complexity of existing auditory and haptic solutions which have various shortcomings, as described earlier. Also, it is remarkable how well this device faired compared to the tried and true hardcopy tactile output, especially given that it was a completely new mode of access for all participants

Some behavioral ambiguities which were observed during the experiments and their potential solutions are as follows:

1. Staying Oriented: Although all participants were able to use the vibro-audio interface to efficiently learn the bar graphs (Exp 1), their strategy of moving perpendicularly between the tops of the bars (i.e., to gauge their relative heights) was sometimes challenging as they had trouble moving laterally, often deviating upward during their trace because of the smooth touch-screen. This behavior was not observed in the hardcopy condition, as the physical lines provided a fixed reference on the paper. Similar challenges were observed in the vibro-audio mode for following slanted and curved lines in the letter and shape recognition tasks (Experiments 2 & 3). Although the pulsing vibration at the vertices helped in determining an intersection or end node, there were no orientation cues to assist with non-rectilinear stimuli, which is particularly challenging in the vibro-audio interface. In the hardcopy condition, the embossed lines make it easier to detect line orientation and to follow the lines when they change direction. This suggests the need for developing a secondary cue to assist with contour tracing and staying oriented when exploring non-

rectilinear stimuli. Corroborating this interpretation, multiple subjects self-reported difficulty in identifying the slanting lines as they felt that the perceptual cues from the vibro-audio interface were not as “sharp” as with the hardcopy stimuli. Implementing additional complimenting audio or haptic cues could likely resolve this issue. However, this needs to be addressed in future studies through more basic research regarding cue salience.

2. Haptic illusion: A phenomenon was observed in the data arising from the pattern of finger movement that turns slight orientations or curves in the stimuli (10 to 20 degrees) into a straight line. Such illusions are observed in both touch-screen-based graphics and paper-based tactile graphics (see (Sanders & Kappers, 2007) for details). This problem could be resolved in the future by using additional cues to indicate deviation from a given line orientation.

3. Pattern errors: Letters such as “D” and “P” were interpreted as the same since they have a line and a curve in common. Since these pattern errors were only observed in the first learning attempt, and correct recognition was very high after the second learning iteration, this problem may be more due to lack of familiarity with the vibro-audio interface than to actual problems interpreting the information conveyed. Also, the letters with symmetric patterns contributed to the wrong interpretation. For example, the W was often interpreted as V, U

or M. This occurred because subjects might trace only half (or part) of the object and then guess that it is U or V, but when traced fully, subjects tended to count the number of lines and use this as a strategy to narrow the possible letter alternatives. This suggests the need for accessing the entire image serially as incomplete exploration might lead to incorrect inference about global meaning of the graphical material.

4. Learning time: The time taken to learn was significantly different between the hardcopy and vibro-audio modes for all conditions. Although the learning time with the vibro-audio mode was approximately four times greater than the time taken in the hardcopy mode, this was not unexpected owing to differences in the way information is conveyed and extracted between modes. As discussed earlier, adding additional complementing cues and allowing greater experience with the vibro-audio interface is predicted to narrow this gap.

3.7. Summary

In sum, error performance in the three experiments did not reliably differ between display modes on any of the measures tested, demonstrating that the vibro-audio interface provides a comparable level of access to graphical material as is possible from a traditional hardcopy medium. Thus, with the addition of new auditory cues to complement the vibro-tactile information, and more training with the interface, it is likely that many of these ambiguities would be ameliorated. This demonstrates that the interface is a viable solution improving the information gap between blind and their sighted peers. Although the interface supported accurate learning and representation of simple and small format graphics, the question arises on what happens if the material being rendered extends beyond the touch-screen on the device. This issue is taken up in the next chapters.

CHAPTER 4

LEARNING LARGE FORMAT GRAPHICS USING NON-VISUAL PANNING OPERATIONS

This chapter elaborates the limitations in generalizing the interface for different kinds of large format graphics such as maps and highlights the pros and cons of existing solutions to overcome these limitations. It then investigates the human factors involved in performing non-visual panning operations through a human behavioral study (Experiment 4). The following sections introduce the panning methods designed as a part of this thesis, and describe the method, procedure, results, and discussion for Experiment 4.

4.1. Limitations of the vibro-audio interface

Results from Experiments 1, 2, and 3 provided strong evidence for the efficacy of the vibro-audio interface for learning the experimental stimuli and in building up accurate mental representations, supporting various spatial behaviors for both blindfolded-sighted and blind participants. However, most non-visual interfaces will have some inherent limitations and the vibro-audio display studied in this thesis is not an exception. The limited display size of touch-screen devices hinders the blind user from accessing graphical materials that are larger than the screen size. For instance, consider the prototype vibro-audio interface (implemented on the Samsung galaxy 7.0 tablet) where the display width of the device is ~3.5 inch (600 pixels). With the lines (and inter-line spacing) being rendered with a width of 0.35 inches (60 pixels), only a maximum of five bars can be displayed on the device's screen (refer to figure 4.1). This means bar graphs with greater than 5 bars cannot be displayed in their entirety. This necessitates zooming or panning of the image to apprehend its global structure. This restriction of hardware display size is common for almost all electronic refreshable displays, such as refreshable tactile displays, mouse-based virtual screen displays, and touch-screen displays (see chapter 2). Using panning or zooming is very common for visually-rendered material on portable devices, or even on standard computer monitors. By contrast, these operations are not used in most assistive technology, since they are usually fixed and cannot be panned or zoomed. However, to access large format graphical material in touch-screen devices both in visual and non-visual settings, it is necessary to incorporate panning or zooming operations. There is a huge difference in the sensory resolution between vision and touch. For instance, in the Samsung galaxy 7.0 tablet screen, vision can be used to perceive ~386 lines of width at 0.116 mm at a viewing distance of ~400 mm. Whereas touch can only perceive 5 lines of width at 0.35 inches. Because of this difference, many graphical materials that can be perceived with vision on a single screen cannot be perceived through touch. Thus, for a touch-based device to be truly useful, it is essential for the interface to provide access to graphical elements that extend beyond the device's screen extent.

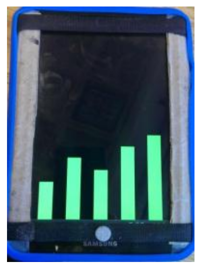

Figure 4.1. Example bar graph on 7.0 inch Samsung galaxy tablet

4.2. Panning and Zooming

Incorporating panning and zooming operations are traditional methods to deal with the limitation of touch-screen size. Visual applications (e.g., Google maps) generally implement these two operations in order for users to explore large format graphics (e.g., maps) within available screen size. Zooming is the ability to magnify or shrink the graphical material (i.e., ability to do image scaling). Zooming commonly requires a change in the image dimensions by a non-integer factor, such as a 50 % zoom where the dimensions must be 1.5 times the original image. Conversely, panning is the ability to drag the graphical material in any direction and distance without altering its scale. However, these operations are almost always performed visually. In order to better conceptualize this limitation, the reader is invited to try the following task, try to pan a map in a map application (e.g., Google maps) with your eyes closed, after panning you will likely lose control over the map as there is no reference between the graphical elements perceived before and after panning. As stated earlier, interpreting tactile graphics is a challenging process by itself (Loomis et al., 1991). Hence, adding additional operations like panning and zooming will further increase the difficulty in interpreting the graphics. Performing these operations without vision (i.e., touch or audio) represents several computational challenges relating both to cognitive constraints of the user and technological constraints of the interface.

4.3. Visual vs. non-visual panning

Results from Experiments 1, 2, and 3, advocated that the interface is efficient in supporting users to access graphical information when its entire contents are displayed within a single screen (i.e., rendered on the display without the need for any panning or zooming operations). However, it is unclear whether a blind user can access graphical material in a similarly efficient manner when it is larger than the touch-screen size and requires panning operations to access it in its entirety. To access graphical material beyond the screen extent, a user should be able to pan and bring the extended graphical material to the current screen view.

The user must also accurately integrate the graphical elements traced before and after panning to conceptualize the entire graphic.

In a visual setting, the parallel processing nature of vision makes it easier for sighted persons to perform panning operations and subsequently integrate the graphical elements across panning screens. Vision has fine spatial resolution and facilitates development of multiple references allowing the observer to integrate graphical information dynamically even while panning. Conversely, the spatial resolution of touch is coarse when compared to vision (S. L. Lederman, Klatzky, Chataway, & Summers, 1990; Rastogi & Pawluk, 2013). With one finger being the source of information (in both taxel and touch-screen-based interfaces), it is difficult to develop multiple references and integrate information dynamically. Since the finger location acts as the primary and only reference for the user at any given point of time, it is necessary for the user to always remember where they are within the given graphic. Allowing users to keep track of their finger location is a key design requirement for any non-visual interface, especially on touch-based devices. In a standard visual setting, vision is used for learning and finger gestures are used for performing panning operations. Conversely, with the vibro-audio interface the finger is primarily used for learning, thus it cannot be simultaneously used for panning. Because of this consideration, visual panning methods using finger gestures such as swipe, flip, and drag should not be incorporated into non-visual interfaces as such finger-based actions will likely lead to confusion. In addition, tracing with one finger by itself can be considered as a gesture by the interface and thus cannot be used as a method for performing panning operations.

Much research has shown that blind individuals often have difficulty to organize and integrate graphical elements of a map (Casey, 1978) and require more decision factors (landmarks) for way-finding behavior when compared to their sighted peers (Passini & Proulx, 1988). This suggests that panning methods should be designed in such a way that the user's touch location remains the same before and after a panning operation or the user should at least be notified of where the last touch location is moved after panning. The logic here is that if the user can remember the touch location before and after panning, it will act as a decision factor (reference point) allowing the user to integrate graphical elements across the panning screens. It is postulated that controlling the panning operations based on this design requirement will lead to reduced cognitive computation and development of more accurate spatial representations of the material being explored. In addition, the design should also consider the ease of use of panning operations such that it does not impose any additional cognitive effort in the learning process. That is, the panning operation is only a tool to move and manipulate the graphical material so that it is perceptually accessible on the screen. Thus, it is important to design the panning method in such a way that it works in parallel with the learning process and is not treated as a process by itself. Otherwise, the user might concentrate more on performing the panning operation and get distracted from learning the graphical material, which is of primary importance. Similarly, the panning method should be easy to remember and apply so that the user can concentrate only on the learning process rather than thinking about how to apply the panning method. The user should be able to cognitively process the information learned before and after the panning operation. If the user focuses on how to perform the panning operation then it will affect the information processing and will eventually lead to inaccurate integration of graphical elements.

Much of the existing empirical research on non-visual interfaces has focused on learning large format graphics (such as maps and floor plans), but only a few studies have addressed the issue of performing panning operations using touch. A three finger gestural input was used for map panning in BrailleDis 9000, an example of a haptic refreshable display (Schmidt & Weber, 2009). Similarly, gesture-based panning was tested in the audio-haptic browser (Zeng & Weber, 2010). However, panning methods used in these studies were not evaluated for their efficiency or usability. Much research has implemented map scrolling/panning using force-feedback devices and tested for usability (Schloerb, Lahav, Desloge, & Srinivasan, 2010). A project with force-feedback device implemented panning for navigating a 3-D topographical surface (S. Walker & Salisbury, 2003). Another study examined performance with scrolling for a model world with representations for houses, roadways and walkways, with spoken sound for details about an object (Magnuson & Rassmus-Grohn, 2003). However, panning methods used in these studies were not created with consideration of the design requirements discussed in section 4.3, and were also not evaluated statistically for its influence in the actual learning process. Hence a human behavioral study (Experiment 4) was conducted here to investigate whether incorporation of panning operations in a non-visual interface supports or hinders the learning process.

4.4. Experiment 4: Evaluation of non-visual panning

This experiment investigated non-visual panning and was motivated by the following goals:

1. To assess whether incorporation of panning operations to the vibro-audio interface strengthens or weakens the learning process. The performance

in learning graphical material using a panning operation and subsequent spatial representation will be compared to performance in learning graphical material without panning operations. If performance does not statistically differ between the panning condition and no-panning condition, then it can be concluded that the incorporation of panning operations does not interfere the learning process with vibro-audio interface.

2. To investigate how the graphical information is processed and represented as a global spatial image in memory when learned using panning operations. That is how a user will integrate the graphical elements across panning screens and represent it as a global spatial image.

3. To compare and examine the efficacy of different panning methods (discussed in section 4.4.3 to 4.4.6) in supporting user's ability to integrate and learn graphical information across multiple panning screens using the vibro-audio interface.

4.4.1. Method

Fifteen sighted participants (eight males and seven females, ages 19-29) were recruited for the study. All gave informed consent and were paid for their participation. The study took between 1.5 and 2 hours.

4.4.2. Conditions

Based on an extensive literature search, to my knowledge, only two studies have used a panning operation on touch-based interfaces; one used two fingers for panning to learn a map using auditory cues (Su, 2010) and the other used a finger and a button to pan and learn an indoor map (Raja, 2011). Although these two methods satisfy the design requirements for non-visual panning discussed in section 4.3, these studies did not address the influences of the panning method on the actual learning process. Hence, to investigate whether incorporation of a panning operation in the vibro-audio interface supports or hinders the learning process, five different panning mode conditions were designed and evaluated for this study. Four panning methods were designed based on the design requirements discussed in section 4.3 that involved multitouch (Section 4.4.3), buttons (Section 4.4.4 and 4.4.5), and gestures (Section 4.4.6). An additional no-panning condition was used as a control condition. Each method represents a different set of techniques and behaviors, and involves varying control over the direction and distance of panning.

4.4.3. Two Finger-Drag panning

As the name suggests, this method uses two fingers to perform the panning operation. This method was inspired from the Timbremap project (Su, 2010) where the placement of a second finger was restricted to one of the four corners of the screen. The authors alleged that this restriction led to confusion while learning, as participants indicated that the largest difficulty they had was with the panning operation. Hence, this restriction was replaced in the current design by allowing the second finger to be placed anywhere on the screen. As was described in section 3.2, users learn the graphical material displayed in the explore mode of the vibro-audio interface by exploring with one finger. On placement of an additional finger, the panning mode was initiated. Once in panning mode, users could pan the graphic in any direction by dragging it with two fingers synchronously (refer to figure 4.2). A clicking sound was triggered to indicate that the panning mode was activated to the user. The clicking sound stopped on removal of the additional finger indicating to the user that they were back in the explore mode. The user's primary finger was not disturbed during the panning operation which is expected to provide a reference and allow the user to continue the learning process immediately after panning. This method is similar to the conventional panning method (swipe or drag) used in visually-based touch-screen devices except that an additional finger is used here.

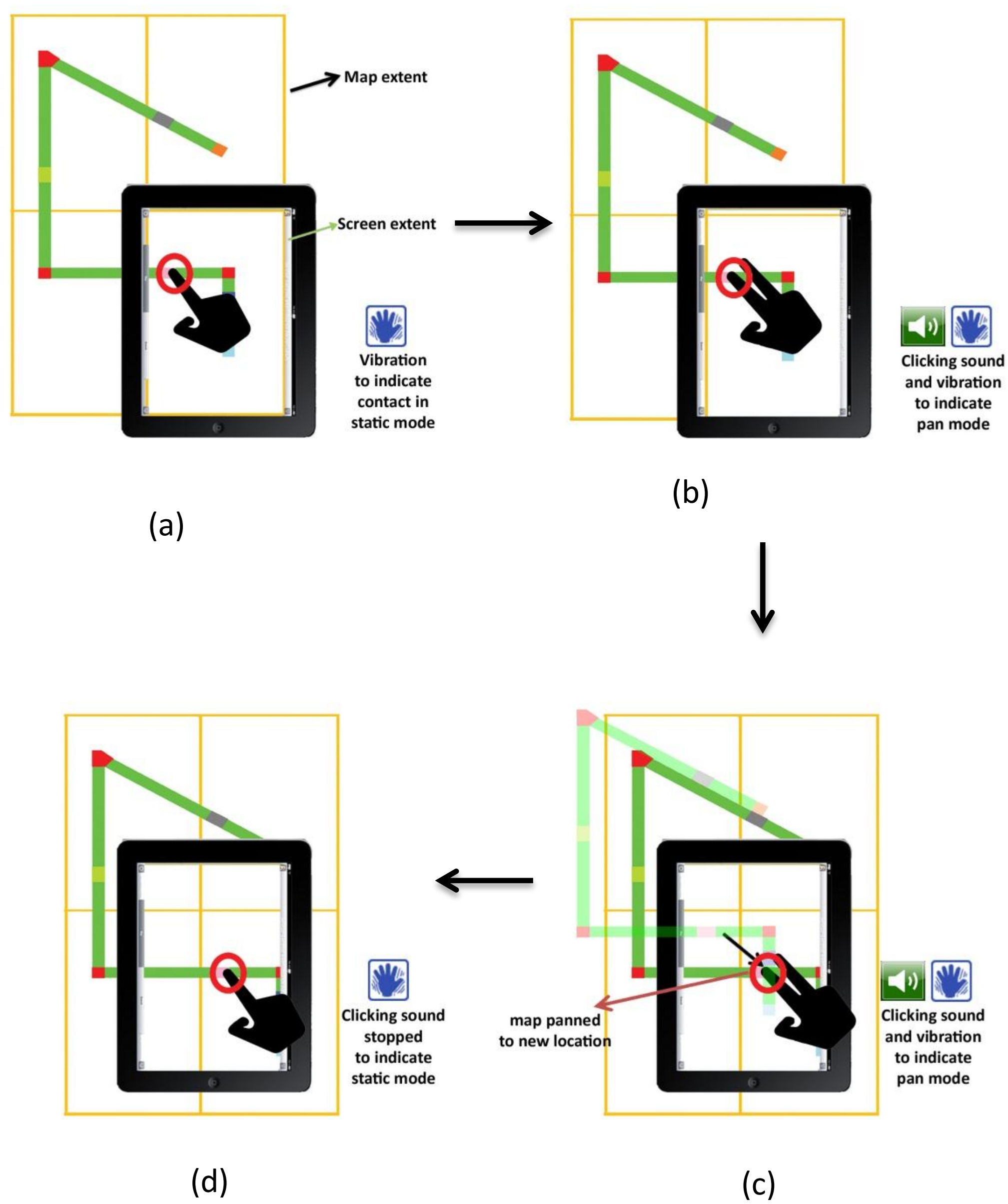

Figure 4.2 Two finger-Drag panning operation: (a) explore mode, (b) pan mode initialized with two finger, (c) map panned by dragging two finger and (d) back to explore mode on removal of second finger

4.4.4. Button-based panning

Earlier research with the vibro-audio interface, similar to the incarnation used in this thesis, demonstrated that button-based panning is an efficient method for non-visual panning (Raja, 2011), when compared with an Extended Display (a method that could compensate the need for panning operation by virtually extending the device's display size). However, the efficiency of this panning method cannot be generalized unless direct comparisons are made with other viable panning methods. Hence, the button-based panning method was included here to evaluate its efficiency in supporting the non-visual learning process using the vibro-audio interface. This panning method involves three steps; (1) remember the touch location and raise the primary finger from the touch-screen, (2) press the pan button, and (3) then place the primary finger in a different location such that the last touch-point is moved under the current touch location (refer (Raja, 2011) for detailed procedure).

4.4.5. Button-Drag

The two panning methods discussed in sections 4.4.3 and 4.4.4 represented a unique set of behaviors and previous studies have provided supporting evidence for their efficacy in non-visual panning operation. At the same time, each of these methods had some drawbacks. For instance, raising the finger in the buttonbased pan mode increases cognitive effort as the user must remember, recall and confirm their current location before and after the panning operation. Similarly, the use of an additional finger was sometimes confused with the primary finger which increased cognitive load and led to potential confusion for the user (as indicated during pilot studies in the lab with the vibro-audio interface). Hence, in this Button-Drag method, pros of the previous two methods were combined; using a button to control the panning mode and using a drag gesture to perform the panning operation. Unlike the button-based method, here users need not remove their primary finger. Pressing the pan-start button initiated the panning mode and indicated it was active to the user via a continuous clicking sound. Once in panning mode the user could pan the graphic in any direction as needed by dragging it with the primary finger. Pressing the pan-stop button simultaneously stopped the pan mode and the clicking sound, indicating to the user that they could continue learning the graphics in explore mode using the same primary finger (refer to figure 4.3). This method was expected to be faster than the previous two methods, as the user need not focus on their touch location while panning because the primary finger is always in contact with the screen. However, it is expected that user's might not achieve the same level of accuracy in mental representation like other methods since they are not focusing on the touch locations while panning.

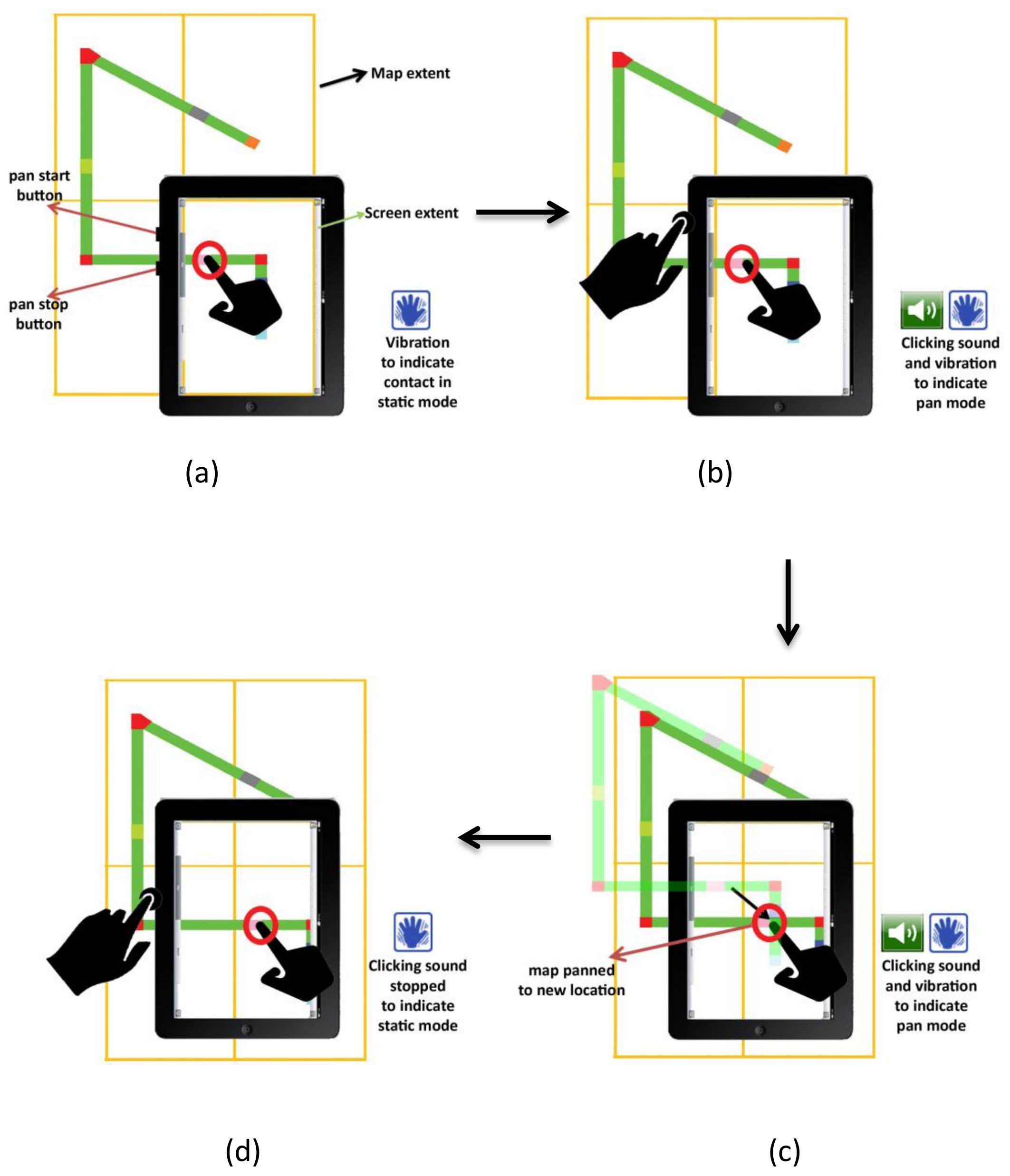

Figure 4.3. Button-Drag panning operation: (a) explore mode, (b) pan mode initialized by pressing pan start button, (c) map panned by dragging primary finger and (d) back to explore mode on pressing pan stop button

4.4.6. Grid-Tap

In the three methods discussed above, the users were allowed to pan the graphical material in any direction and to any distance they desired. However, most of the conventional non-visual panning methods in the literature have restricted these parameters. For instance, direction was restricted to either horizontal or vertical movement and the distance of panning was fixed (Magnuson & Rassmus-Grohn, 2003). This means that the user must learn grids of graphical material and integrate the grids to visualize a global spatial image. This operation is often termed as scrolling. To investigate the efficiency of such a restricted method, the grid-tap was designed to control panning distance and panning direction. The graphical material was divided into an even number of grids, where the size of each grid was matched to the device's display size such that only one grid can be displayed at a given time. The panning operation eventually moved the grids horizontally or vertically and was triggered by a double tap gesture. That is all movement is in fixed, predefined increments based on the device's screen size.

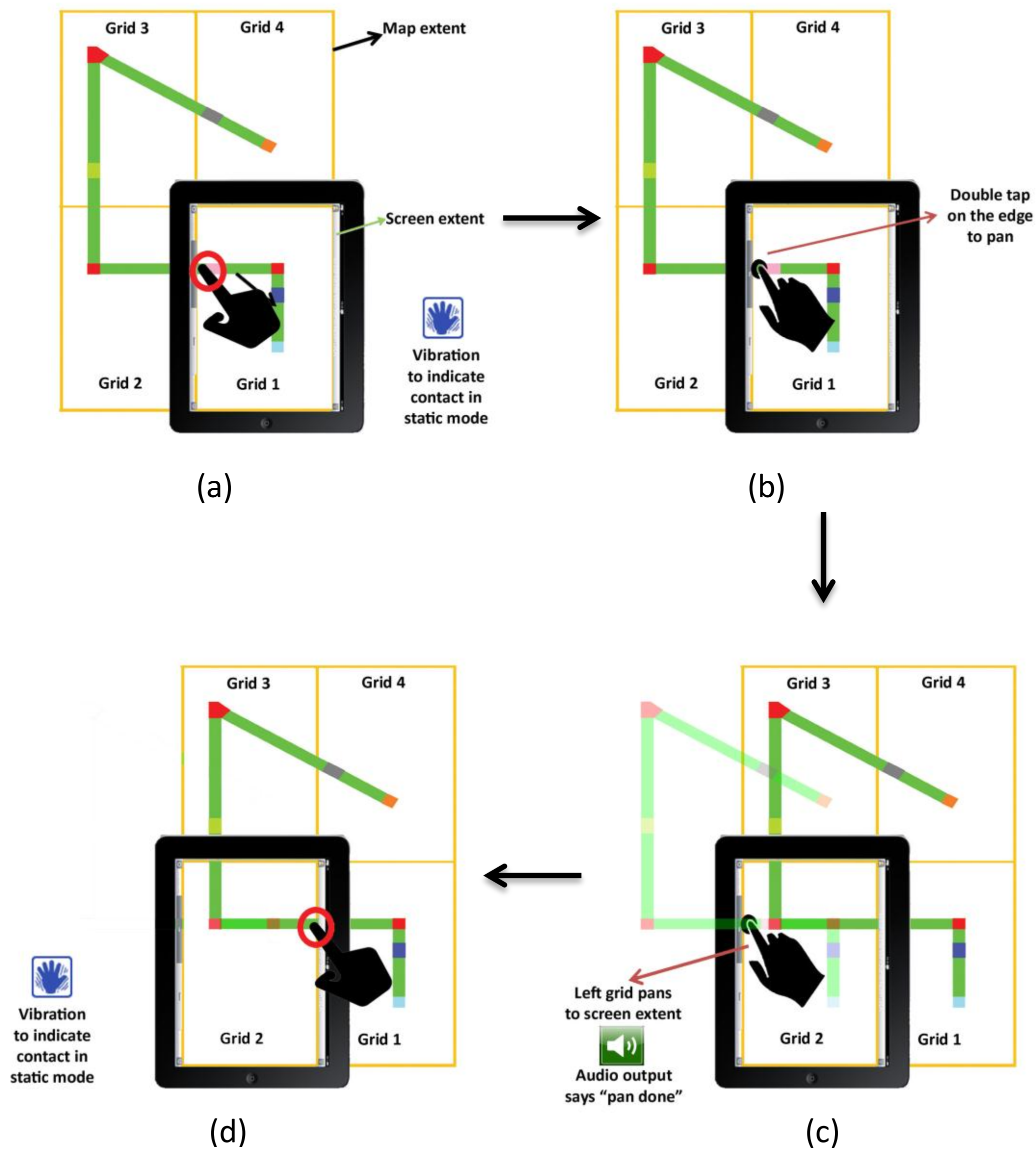

Figure 4.4. Grid-Tap panning operation: (a) explore mode, (b) panning initialized by double tap on edge, (c) map panned and indicated to user by audio and (d) back to explore mode

A Double tap gesture performed on the edge of the screen would bring the adjacent grid in that direction to the current screen focus (refer to figure 4.4). For instance, to bring the grid that is on the left of the current screen's rendered material, a double tap gesture should be performed on the left edge of the display screen. This process can be compared to flipping a page in a book. The completion of the panning operation was indicated to the user through speech output stating "pan done." This restricted panning was expected to provide better reference for image scaling, alignment, and spatial relations between graphical elements as the user is simply integrating grids of equal size to that of the display size. Also, since the grids are fixed and equally aligned it provides the user with a good reference for the alignment and direction between landmarks.

4.4.7. Stimuli and apparatus

The four panning conditions were implemented with the vibro-audio interface on a Samsung Galaxy Tab 7.0 Plus tablet, with a 17.78 cm (7.0 inch) touch-screen. A no-panning method was used as a control condition for comparing with the pan mode conditions where the entire graphical material could be accessed from one screen without panning. Hence, to present the entire graphic within the touch-screen extent, the vibro-audio interface was implemented on a bigger Samsung Galaxy Tab 10.1 tablet, with a 25.65 cm (10.1 inch) touch-screen used as the information display. The apparatus setup (table, chair and blindfold) was the same as in Experiment 1.

Five indoor corridor layout maps were used as experimental stimuli (with two additional maps for practice). Each of the maps was composed of corridors, landmarks, junctions, and dead-ends. The maps were designed by considering a frame size matching A4 paper. The five maps were carefully designed such that they were based on the same complexity but a different topology (refer to figure 4.5) and forced the user to pan in all four directions to access the entire map. The complexity was matched in terms of:

1. Number and orientation of corridor segments: Each of the maps had 3 straight corridor segments (either horizontal or vertical) and one oriented corridor segment that was misaligned with the display's intrinsic frame of reference.

2. Number of junctions: Each of the maps had 3 two-way junctions and 2 dead-ends (one start and one destination).

3. Number of landmarks: Each of the maps had 4 land marks. Each landmark was assigned a name based on a hotel theme including its corridor layout and salient landmarks: lobby, elevator, restaurant and stairwell.

4. Position of landmarks: Each of the maps had exactly 1 landmark on each of the corridor segments. Of the four landmarks, two were always on the start screen such that they can be apprehended without any panning operations. This was measured across conditions to analyze how this is represented in user's memory. The landmarks were positioned in such a way that in each map at least two landmarks were aligned (either horizontally or vertically). Again, this was measured and analyzed across conditions to investigate the efficacy of each panning method in conveying alignment information.

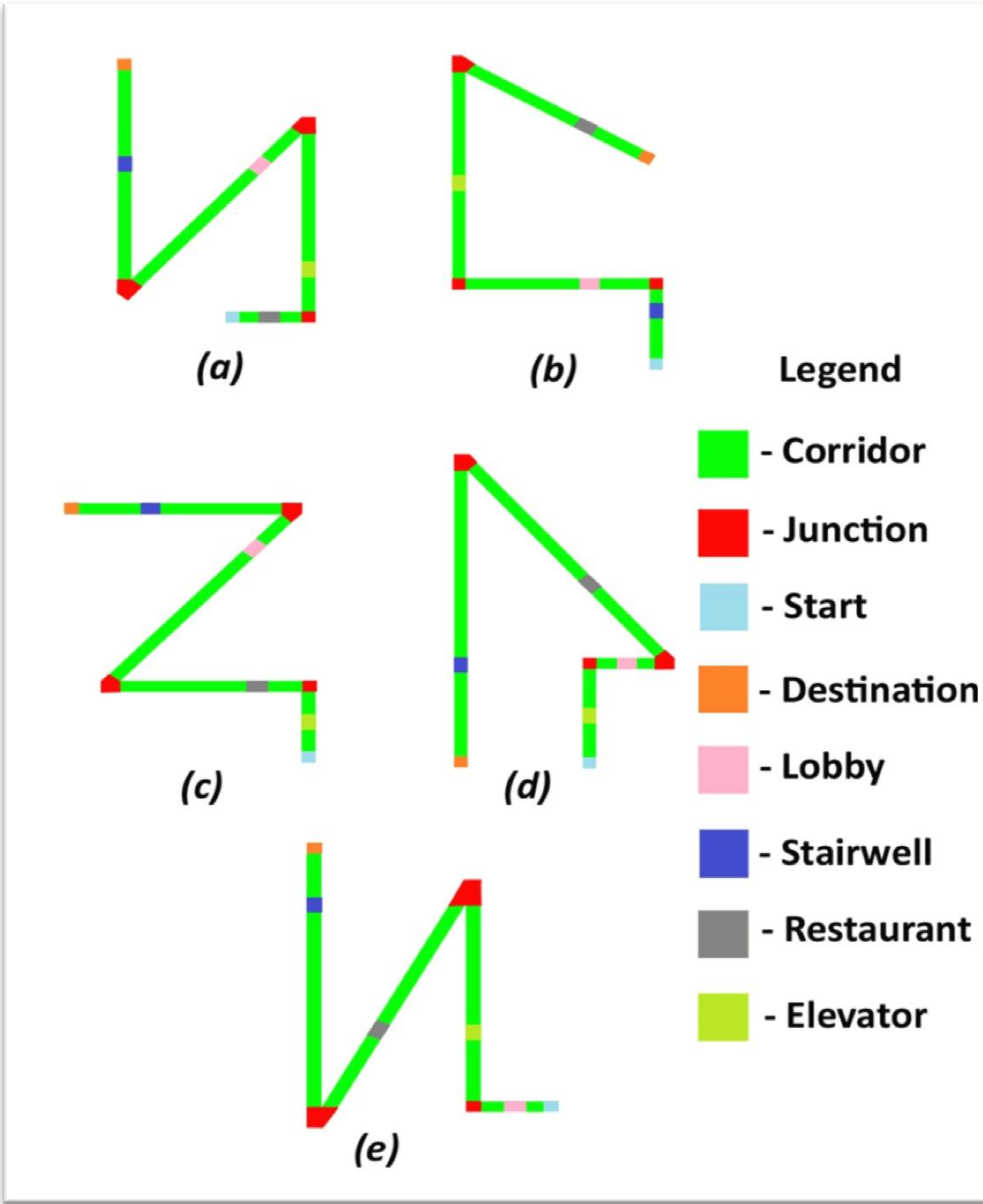

Figure 4.5. Corridor layout maps: Experimental stimuli used in Experiment 4

Similar to Experiment 1, all the maps were rendered with a line-width of 8.9 mm (0.35 inch), which corresponded to 60 pixels on the 7.0 inch touch-screen and 52 pixels on the 10.1 inch touch-screen. In both the devices, vibro-tactile feedback was generated when the user's finger touched the stimulus on the screen. Corridors were given a constant vibration, based on the UHL effect "Engine1_100". The junctions and dead-ends were indicated by a pulsing vibration, based on the UHL effect "Weapon_1". The landmarks were indicated by an auditory cue (sine tone) coupled with fast pulsing vibration, based on the UHL effect "Engine3_100". In addition, speech output (e.g., name of the landmark) was provided for the junctions, dead-ends, and landmarks by tapping the vibrating region. In both the devices, exploration was done using only one finger (dominant). The user's movement behavior was tracked via the device's touch-screen as they felt the stimuli. The system logged the learning time, finger-traces (co-ordinates), type of vibration pattern and the panning points into a text file for each of the trials.

4.4.8. Procedure

A within subjects design was used in the experiment. In each condition, participants learned a corridor layout map and performed subsequent testing tasks. The condition orders were counterbalanced between participants and the maps were randomized between conditions. The study consisted of a practice, learning, and testing phase for each condition. The first practice trial in each condition was a demo trial where the experimenter explained the task, goal, and panning strategies and the participant explored the stimuli with corrective feedback provided. The participants were instructed to visualize the corridor layout map as analogous to a hotel floor map with the four landmarks being *Lobby, Elevator, Restaurant* and *Stairwell* (order of the landmarks were randomized between maps). In the second practice trial, blindfolded participants were asked to learn the entire map, followed by the test sequence without blindfold. The experimenter evaluated the answers immediately to ensure they understood the task correctly before moving to the experimental trials.

4.4.9. Learning phase

During the learning phase, participants were first blindfolded. The experimenter then placed their primary finger at the start location of the map and instructed them to explore and learn the map. Participants were allowed to go back and forth between the start and the destination of the map without limitation. Participants were asked to indicate to the experimenter when they believed that they had learned the entire map. Once indicated, the experimenter removed the device and then the participants were allowed to lift their blindfold to continue with the testing phase.

4.4.10. Testing phase

The testing phase consisted of two tasks: (1) a pointing and (2) a map reconstruction task. In the pointing task, participants indicated the allocentric direction between landmarks using a physical pointer fixed on a wooden board (refer to figure 4.6). The pointing task consisted of a set of four pointing questions (e.g., indicate the direction from elevator to lobby) covering all four landmark pairs. The reproduced angles were analyzed for their correctness in relative position and direction between landmarks.

In the reconstruction task, participants were asked to draw the map and label landmarks on a template canvas of the same size (A4 paper) as the original map. To provide the subjects with a reference frame for the scale of the map, the start and destination points were already marked in the canvas (see Figure 4.6). The reconstructed maps were analyzed in terms of whether the maps had correct spatial pattern of corridor segments, and included the correct landmark' position and labels.

Figure 4.6. Pointing device used in the pointing task (left) and A4 Canvas for reconstruction with start and destination points(right)

4.4.11. Experimental measures and analyses

From this experimental design, the following measures were evaluated as a function of the five pan-mode conditions.

1. Learning time: The Learning time is the time taken from the moment they touch the screen until they confirmed that they had completed learning of the map. The time was measured from log files of each trial that was created within the device. The learning time ranged from ~2.5 minutes to ~15 minutes with a mean of ~7.5 minutes. The learning time can be interpreted as an indication of relation between cognitive effort and time taken for learning. That is, the greater the learning time, the higher the cognitive load for the condition.

2. Times Traversed: Participants were allowed to go back and forth between the start and the destination of the corridor map without restriction. It can be postulated that the fewer number of times they traverse the map, the more efficient was the panning method for learning. The times traversed were calculated from the log files of each trial that was created within the device.

3. Times panned: The number of times the map was panned will vary greatly between the conditions because of the nature and procedure of the panning method. For instance, the amount of pan is fixed in the grid-tap method but can be varied in other techniques. This measure can be interpreted as an indication for ease of use of the panning method. That is, the easier the panning method, the more times the participants will perform panning. The control condition is excluded from this measure as there was no panning involved. The times panned were calculated from the log files of each trial that was created within the device.

4. Relative directional accuracy: This is the spatial direction relation between any two landmarks. This was measured from the pointing tasks. The angles between landmarks reproduced by the participants were compared to the actual angles between the landmarks to measure the angular errors. These

angular errors were then analyzed in two ways: Unsigned error and Signed error (under estimating the angle representing a negative bias and over estimating representing a positive bias).

5. Reconstruction accuracy: The reconstruction accuracy is the accuracy in the global spatial representation of the map. This was measured by comparing the spatial pattern of the reconstructed map with the actual map. The reconstructed maps were analyzed in two ways; (1) Discrete scoring and (2) Bi-dimensional regression. In discrete scoring the maps were analyzed for their correctness in spatial pattern and were given a score of 1 if correct and 0 otherwise. Since binary scoring does not capture the metric accuracy or nature of the errors of the reconstructed maps, a Bi-dimensional regression analysis was used to analyze the metric accuracy as it measures the fidelity between cognitive maps and actual locations. Seven anchor points (4 landmarks and 3 junctions) were chosen on each map and the degree of correspondence of those anchor points between the actual map and the reconstructed map were calculated. The 4 junction points covered the entirety of the map and acted as a decision factor in forming the spatial pattern of corridor segments. Similarly, the landmark points were the other prominent points within the map that

assisted participants in integrating map elements across panning screens. Three metric factors were considered: 1. Scale, 2. Theta, and 3. Distortion Index. The scale factor indicates the magnitude of contraction or expansion of the reconstructed map. The theta determines how much and in which direction the reconstructed map rotates with respect to the actual map. The distortion index depicts the amount of distortion of the reconstructed map with respect to the actual map.

6. Relative positioning accuracy: As discussed in section 4.4.7, in each map at least two of the four landmarks were aligned (either horizontally or vertically). Understanding such relative position is crucial in grasping the global structure of any map. For example, the entrance and exit will be aligned in many indoor maps. Hence, the reconstructed maps were analyzed with respect to alignment between the two aligned landmarks. A discrete scoring was applied based on the correctness of the landmark alignment (i.e., 1 if aligned correctly, 0 otherwise).

7. Single screen landmark positioning: The start screen of each condition had two landmarks which can be accessed without panning. It was expected that the positioning of these two landmarks would be more accurate and consistent among all the four panning conditions as there were no

differences between conditions in the way this start screen was accessed. However, the fact that participants perform different panning operations in each condition to trace back and forth between the start and destination locations could alter the cognitive representation. To investigate this possibility, the single screen landmark positioning accuracy was measured from the reconstructed maps. It was expected that this should be more accurate with panning conditions than with the no-panning (control) condition as all four landmarks were equally accessible in the no-panning condition. Whereas the two landmarks were accessible without panning in the other four conditions and thus can be easily distinguished from the other two landmarks, which required panning to apprehend.

8. Landmark labeling accuracy: Labels are crucial as changing labels will eventually change the map represented. The accuracy in labeling was measured from the reconstructed maps. A discrete scoring was applied based on the correctness of the landmark labeling (i.e., 1 for each correct label, 4 if all four labels are correct).

9. Subjective rating for the panning methods: Participants were asked to rank the panning methods on a scale of five (with one being the best). The

ranks given by subjects were analyzed to understand the user's preference for the panning methods.

4.5. Results

Performance data for each of the measures described above were analyzed and compared between the five conditions. Univariate ANOVAs and One-way ANOVAs were conducted on each of the measures to assess the within-subjects effects between conditions. Similarly, Univariate ANOVAs were conducted on each of the measures to assess the between-subjects effects. Also, post hoc paired sample t-Tests were conducted to assess the difference in performance between each condition. The most important finding is the similarity of performance across all measures for the five conditions. The f, t and p value of the analyses is given in the tables 4.1 to 4.6 below.

Univariate ANOVA

Measures	Between Condition			
	df		f	Sig.
	Hypothesis	*Error*		
Learning Time	4	56	5.605	0.001
Relative directional accuracy	4	56	2.232	0.077
Reconstruction accuracy	4	56	1.233	0.307
Relative positioning accuracy	4	56	1.806	0.140
Single screen landmark integration	4	56	0.427	0.788
Landmark labeling	4	56	1.034	0.398
Times traversed	4	56	3.527	0.012
Times panned	4	56	3.642	0.020

Table 4.1. Univariate ANOVA between conditions for each measure

Univariate ANOVA

Measures	Between Subjects			
	df		f	Sig.
	Hypothesis	*Error*		
Learning Time	14	56	2.681	0.005
Relative directional accuracy	14	56	0.770	0.696
Reconstruction accuracy	14	56	1.803	0.061
Relative positioning accuracy	14	56	0.516	0.914
Single screen landmark integration	14	56	2.310	0.032
Landmark labeling	14	56	0.787	0.678
Times traversed	14	56	7.077	0.000
Times panned	14	56	3.566	0.001

Table 4.2. Univariate ANOVA between subjects for each measure

One-way ANOVA

Measures	Between Condition		
	df	f	Sig.
Learning Time	4	4.195	0.004
Relative directional accuracy	4	3.316	0.011
Reconstruction accuracy	4	1.063	0.382
Relative positioning accuracy	4	2.000	0.104
Single screen landmark integration	4	0.354	0.840
Landmark labeling	4	1.079	0.373
Times traversed	4	1.592	0.186
Times panned	4	2.218	0.096

Table 4.3. One-way ANOVA between conditions for each measure

t-Test - Learning time			
Pairs	df	t	Sig.
TwoFinger - ButtonBased	14	-2.601	0.021
TwoFinger - ButtonSwipe	14	-3.823	0.002
TwoFinger - Grid	14	-2.335	0.035
TwoFinger - nopan	14	1.161	0.265
ButtonBased - ButtonSwipe	14	-0.732	0.476
ButtonBased - Grid	14	0.234	0.818
ButtonBased - nopan	14	4.217	0.001
ButtonSwipe - Grid	14	1.022	0.324
ButtonSwipe - nopan	14	4.004	0.001
Grid - nopan	14	2.295	0.038

Table 4.4. Paired sample t-Tests between conditions for learning time

t-Test - Times traversed			
Pairs	df	t	Sig.
TwoFinger - ButtonBased	14	-2.870	0.012
TwoFinger - ButtonSwipe	14	-1.948	0.072
TwoFinger - Grid	14	-3.568	0.003
TwoFinger - nopan	14	-3.378	0.005
ButtonBased - ButtonSwipe	14	0.414	0.685
ButtonBased - Grid	14	0.000	1.000
ButtonBased - nopan	14	-1.309	0.212
ButtonSwipe - Grid	14	-0.459	0.653
ButtonSwipe - nopan	14	-1.586	0.135
Grid - nopan	14	-1.193	0.253

Table 4.5. Paired sample t-Tests between conditions for times traversed

t-Test - Times panned			
Pairs	**df**	**t**	**Sig.**
TwoFinger - ButtonBased	14	0.455	0.656
TwoFinger - ButtonSwipe	14	1.061	0.307
TwoFinger - Grid	14	2.648	0.019
ButtonBased - ButtonSwipe	14	0.348	0.733
ButtonBased - Grid	14	2.376	0.032
ButtonSwipe - Grid	14	3.060	0.008

Table 4.6. Paired sample t-Tests between conditions for times panned

From the results of the omnibus ANOVAs, it can be inferred that there were no significant differences between conditions for relative directional accuracy, reconstruction accuracy, relative positioning accuracy, single screen landmark integration, landmark labeling and subjective ratings. However, there was a significant difference (alpha = 0.05) between conditions in learning time, times traversed and times panned. Similarly, the results of the post hoc paired-sample t-Tests between conditions were highly in-significant (all $p>0.05$) for all measures except learning time. The mean and standard deviation for each of the measures are given in the table below as a function of pan-mode conditions.

Measures	Two finger-Drag		Button-based		Button-Drag	
	Mean	SD	Mean	SD	Mean	SD
Learning Time (in seconds)	354.67	80.20	491.27	198.61	529.20	194.13
Relative directional accuracy - Unsigned error	17.58	24.97	18.33	20.06	32.67	41.41
Relative directional accuracy - Signed error	-5.08	30.19	-3.17	27.09	-15.50	50.55
Reconstruction accuracy	0.80	0.41	1.00	0.00	0.87	0.35
Relative positioning accuracy	0.67	0.49	0.40	0.51	0.67	0.49
Single screen landmark integration	0.87	0.35	0.80	0.41	0.73	0.46
Landmark labeling	3.20	1.01	3.60	0.83	3.47	1.19
Times traversed	2.00	1.00	2.67	1.29	2.53	1.25
Times panned	20.87	12.79	19.20	12.39	18.40	7.87
Subjective rating	2.60	1.05	4.00	0.76	2.80	1.21

Measures	Grid-Tap		No-panning	
	Mean	SD	Mean	SD
Learning Time (in seconds)	472.40	224.56	324.07	99.53
Relative directional accuracy - Unsigned error	29.00	33.96	18.58	25.01
Relative directional accuracy - Signed error	-15.83	41.88	-6.25	30.61
Reconstruction accuracy	0.80	0.41	0.93	0.26
Relative positioning accuracy	0.27	0.46	0.40	0.51
Single screen landmark integration	0.87	0.35	0.87	0.35
Landmark labeling	3.60	0.83	3.87	0.52
Times traversed	2.67	1.05	3.07	1.28
Times panned	12.13	3.94	NA	NA
Subjective rating	3.80	1.36	1.60	0.99

Table 4.7. Mean and Standard deviation for each measure as a function of pan-mode condition

Corroborating what is shown in tables 4.1-4.7, the performance in learning and representing large format graphical material was similar across measures for all five conditions. Also, there were no reliable order effects based on a Univariate ANOVA that assessed the ordering effects between conditions ($F(4,70) = 0.217$, $p = 0.928$).

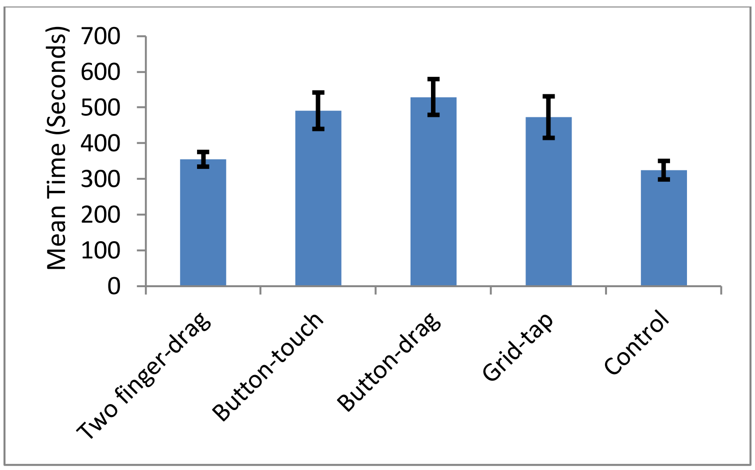

Figure 4.7. Mean learning time as a function of pan-mode, along with Standard error.

From figure 4.7, it can be inferred that no-panning and two finger-drag methods were the fastest conditions (<~400 seconds), indicating that these two methods imposed the least cognitive effort on participants. This was also evident from the results of paired sample t-Test that showed no evidence of reliable differences between the two conditions (refer to Table 4.4). The superior performance of the no-panning (control) condition in learning time can be attributed to its fixed frame of reference as users need not perform any additional operations such as use of gestures, buttons and additional finger actions. Despite performing additional panning operations, the learning time of the two finger-drag condition was similar to that of the no-panning condition, which indicates the intuitiveness of the two finger-drag method for extracting and learning information across screens. Similarly, the times traversed in the two finger-drag condition were reliably less than in other conditions (refer to Table 4.5). This means that the two finger-drag method imposed less cognitive load on the users, thereby allowing them to focus more on the learning of the map. Also, from the mean and standard deviation of times panned (refer to table 4.7) it can be inferred that the two finger-drag method was the easiest method to apply and perform panning operations.

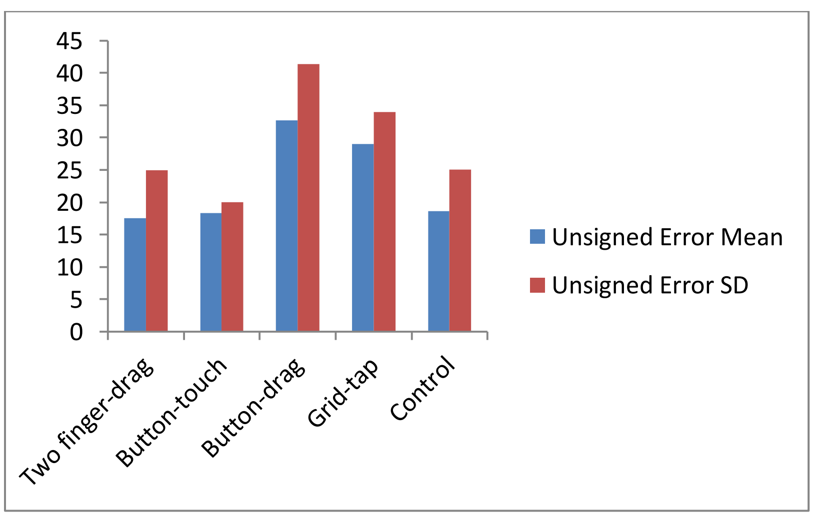

Figure 4.8. Unsigned directional error as a function of pan-mode, along with Standard deviation.

Comparing the means and standard deviations of the unsigned errors (refer to figure 4.8), it can be inferred that the participants were numerically more accurate in indicating relative directions when learning with two finger-drag, no-panning, and button-based panning conditions compared to button-drag and grid-tap methods. However, the differences were not statistically significant based on a paired sample t-tests that compared the difference between the conditions (all $p>0.05$). Also, from the signed errors (refer to table 4.7), it can be noted that participants generally under estimated the angles in all five conditions. This demonstrates the similarity in mental representation of the graphical material developed using different panning and no-panning methods but suggests a perceptual bias leading to compression of the mental representation. Remarkably, the no-panning condition was numerically less accurate than the two finger-drag and button-based conditions, indicating that incorporation of the panning operation was beneficial in identifying relative direction between landmarks and did not add any additional cognitive effort than in the control condition.

From the results of the Bi-dimensional regression it was evident that there were no significant differences between conditions in Theta ($F(4,56)= 0.876$, $p = 0.484$) and Distortion Index ($F(4,56)= 1.733$, $p = 0.156$). However, a Univariate ANOVA suggested that there was a significant difference between conditions in the Scale factor ($F(4,56) = 8.8$, $p < 0.001$). This means for each of the conditions the map was perceived as a different size. This difference in scale perception is mainly influenced by the nature of the panning operation, as the panning distance and direction differed significantly between the conditions. Comparing the mean and standard deviation of the three factors it can be inferred that participants generally contracted the map while using panning operations and in contrast expanded the map while learning without panning. This could be because the no-panning condition was carried out in a bigger device which might have created an illusion that the map was bigger than in other conditions.

Condition	**Scale**		**Theta**		**Distortion Index**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Two finger-drag	0.929	0.114	0.984	4.91	24.098	9.299
Button-touch	0.879	0.121	-2.85	5.44	30.974	10.858
Button-drag	0.885	0.09	-1.17	3.88	28.514	11.084
Grid-tap	0.806	0.088	-1.41	8.81	31.475	11.959
Control	1.02	0.112	-0.03	5.1	24.231	10.15

Table 4.8. Scale, Theta and Distortion index from bi-dimensional regression as a function of pan-mode, along with Standard deviation.

Based on the subjective ratings, it is clear that participants most preferred the no-panning (control) condition (mean = 1.6). This makes sense as this method did not require participants to perform any additional operations of the map in order to perceive its entire extent. On comparing the four panning conditions, the two finger-drag condition had the best rating (mean = 2.6), this along with the performance in all measures indicates that given a choice participants preferred panning using the two finger-drag method.

4.6. Discussion

A human behavioral experiment was conducted to address the issue of non-visual panning. The study assessed whether incorporation of a panning operation to the vibro-audio interface strengthens or weakens the learning process. Overall, the results suggest that the incorporation of panning operations in the vibro-audio interface yield positive effects in the cognitive representation of the graphical material. It is worth noting that the error performance in the panning conditions are not due to the incorporation of panning operation since in many measures the control condition performed less accurately than the panning conditions. The observed error performance could be because of inaccurate cognitive representation which is equivalent for both panning and no-panning conditions. The superior performance of panning conditions (the two finger-drag condition in particular) across measures in pointing and reconstruction tasks demonstrates that the incorporation of panning operation in vibro-audio interface strengthens the learning process. These findings are substantial given the necessity of panning operations in touch-based devices for accessing large format graphics. The overall performance and user preference suggest that the two finger-drag method was the most efficient and intuitive method for performing non-visual panning operations.

Similar to previous experiments, some behavioral ambiguities were observed in this experiment such as.

1. Human error: Some of the pointing tasks were influenced by outliers due to participants flipping the landmarks which led to a 180 degree error. However, such errors were not removed/replaced as they were consistent across all conditions and participants. This is also evident from the negative correlation between directional error and labeling accuracy.

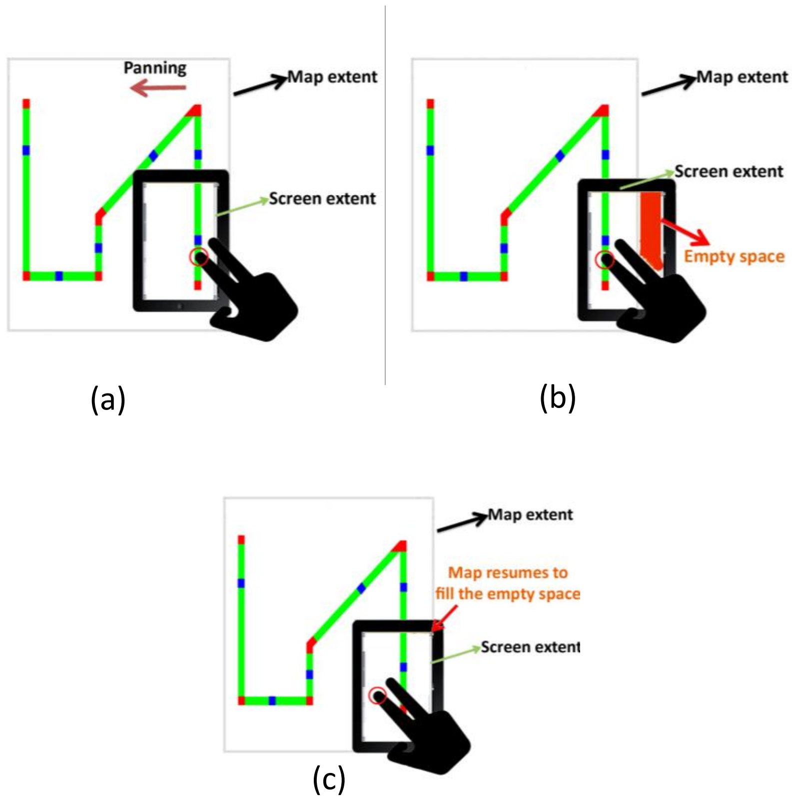

Figure 4.9. Map resume concept: (a) before panning, (b) empty space created after panning, and (c) Map resumes automatically filling the empty space.

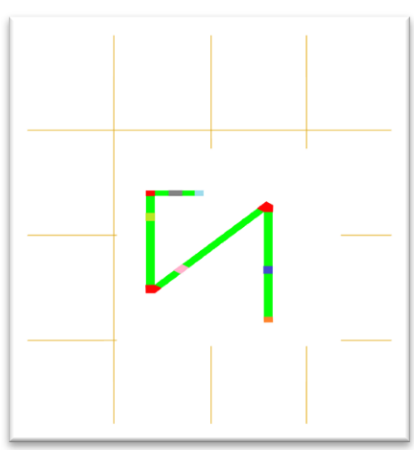

Figure 4.10. Stimuli with white space matching screen size on each side

2. Preference vs. performance: Although participants gave the highest rank for the control condition, their performance did not correlate with the ranking, indicating that the ranking was primarily influenced by users' like or dislike for a method, and the ease of access, rather than the ability to learn the map accurately. For instance, the learning time of the two finger-drag method was almost equivalent to that of the control condition demonstrating that the panning operation did not contribute much to the learning time. Also, for most of the measures, the two finger-drag and button-based conditions were better than the no-panning condition which demonstrates the efficiency of panning operations.

3. Extending the bounds: It was found from pilot studies that while panning it is possible that the map could be dragged out of its bounds. In general the map automatically resumes to fit the screen extent in such scenarios (refer to figure 4.10). But users cannot realize this change while accessing it non-visually. This could confuse the user within the screen space. To avoid this confusion, sufficient white space was included around the actual map extent such that the map will not resume even if it was pulled out of its bounds (refer to figure 4.11).

4. Re-positioning the map: It is likely that users can get lost or forget their way while tracing maps, so it is necessary for the user to get back to a known

point or to the start point to regain control on the map. Though it was not a part of the current design, by analyzing the finger traces and user's feedback, it was found that participants had difficulty in getting back to a known location when they lose control over the map. This problem can be resolved by having an additional functionality to assist the user with getting back to the start position or any other known location.

4.7. Summary

This chapter addressed the non-visual panning issue through a human behavioral study. In sum, error performance did not reliably differ between the four pan-mode conditions and no-panning condition, demonstrating that the incorporation of panning operations exhibit positive effect in the learning process. The superior performance of two finger-drag method across all measures suggests that it is the most efficient, accurate and intuitive method for performing non-visual panning. Although the interface supported accurate integration of graphical elements across panning screens, the questions arises on what happens if the material being rendered is in deep format (i.e., with multiple zoom levels). This issue is taken up in the next chapter.

CHAPTER 5

LEARNING DEEP FORMAT GRAPHICS USING NON-VISUAL ZOOMING OPERATIONS

As was discussed in chapter 4, incorporating panning and zooming operations are traditional methods to deal with the limitation of limited screen size on touch-screen devices. This chapter investigates the human information processing factors involved in non-visual zooming through a human behavioral study (Experiment 5). The Following sections detail the motivation and goals for the study, introduce the zooming techniques designed as a part of this thesis, and describe the method, procedure, results, and discussion of Experiment 5.

5.1. Motivation

As defined in Section 1.1, graphics are visual representations of data, information or knowledge. In most situations, the size of the graphical material is directly related to the data, information or knowledge represented. For example, a bar graph summarizing data from students voting on their food preferences could have 3 bars if 3 foods are compared, or could have 5 bars if 5 foods are compared, etc. The width of the bar graph depends on the amount of data being presented (number of foods being compared). This logic is true for all forms of graphical information, ranging from simple line drawings to complex maps. On the other hand, accessing graphics from both visual and non-visual scenarios involves accurate interpretation of the information represented by the graphical material. As the information becomes complex, visualization and interpretation of the information also becomes complex. For example, a map showing state boundaries will be simple, whereas the same map showing additional information such as road networks, population, street names, etc. will quickly become more complex as it has to convey all the information in a single rendering. To handle such complexities, the graphical representation of the information should consider the strengths and limitations of the human sensory systems and the perceptual factors involved in data extraction, interpretation, and representation. Also it should consider physical factors such as the display, such as screen size and the display medium. As elaborated in chapter 4, one such possibility is to make the graphical representation as large as needed and allow the users to access the information via panning operations. Another possibility is to make the graphical representation as deep as needed and allow the user to access the information via zooming operations, where information are grouped based on their spatial characteristics and the groups are accessed at different zoom levels. This means that the graphical representation should convey the information by representing it at different spatial and temporal intervals. A common way to handle the information is to group them based on spatial characteristics and represent each group at a different (or overlapping) temporal interval. That is presenting the information in the same region or display and extending it in time of presentation such that global understanding requires accurate temporal integration of the multiple spatial samples. These intervals are usually termed as zoom levels and the process of navigating between these zoom levels is termed as zoom-in (navigating deeper into the rendering) or zoom-out (navigating towards the top layer).

Figure 5.1. Google maps displaying Tokyo at zoom levels 0, 7, and 18

In a visual setting, this information grouping is usually based on the scale of the image (e.g., maps) such that each scale will represent a particular zoom level. The information represented in each of these zoom levels will vary significantly. For instance, the same location of Tokyo will have varying levels of information based on its representation at different zoom levels (refer to figure 5.1). At zoom level 0 only the overview of the globe can be represented, at level 7 only city names (around Tokyo) can be represented and at level 18 finer (deeper) details (such as street names within Tokyo) can be represented. In order to develop a global spatial image in one's memory, it is essential for the user to integrate these different levels of information represented at different zoom levels into a consolidated whole.

In addition to navigating between levels of information, zooming operations are also used for magnifying or shrinking graphical material. For instance, two lines (rendered with an inter-line distance of 0.5 mm) in a diagram can be perceived as one line if the inter-line distance is less than the threshold of human perception. But, the same can be differentiated into two distinct lines by magnifying the image, which enhances the graphical elements without affecting their topology. In general, magnifying (scale-up) the graphics is termed as a zoom-in operation, and shrinking (scale-down) is termed as a zoom-out operation. In such scenarios, there are no zoom levels (levels of information); rather a single level of information is either enhanced or reduced via gradations of magnification. An example of such a scenario could be a simple line diagram with two lines, where the information (2 lines) will remain unchanged regardless of the zooming level. Whereas in the former scenario the information will change based on zoom levels (e.g., Globe overview at level 0, and roads at level 13). In both the scenarios, to obtain the benefits of graphical representations such as geometric and topological congruence, indexing, mental animation, macro/micro view, and graphical constraining, users should be able to integrate and relate the graphical elements across different zoom levels to develop a global spatial image. Although it was evident from previous experiments (Experiment 1-4) that the vibro-audio interface is efficient in supporting users to access graphical material displayed in a single zoom level, it is unclear whether users can use the interface to navigate between different zoom levels and learn graphical elements with similar ease and accuracy as accessing it from a single zoom level. Similar to non-visual panning, non-visual zooming also presents a unique set of challenges. Unlike panning (where graphical elements remain unchanged regardless of the panning operation), zooming operations change the graphical elements completely or at least enhances some graphical elements and adds more elements to the existing graphical information. For instance, consider the Google maps example displayed in figure 5.1, where at zoom level 0 only the overview is available, and as one zooms in to level 7, road networks and labels are added. Likewise, zooming out from zoom level 7 to zoom level 0 will remove the road network and labels, and display only the overview. In both situations, the user's touch location will not remain the same after performing a zooming operation as the graphical elements being rendered on the screen completely change based on the zoom level being presented. Thus, one cannot have a fixed reference to relate the graphical elements between zoom levels (as opposed to having a reference point between panning screens). This means that the user must be able to learn graphical elements at each zoom level independently and subsequently integrate the graphical elements across zoom levels to visualize/spatialize it as a global spatial representation. With one finger being the source of information in both taxel and touch-screen-based interfaces, it is difficult to develop references and integrate information dynamically. The question remains open as to how a blind user (or anybody using non-visual zooming) can learn graphical material at each zoom level independently and then integrate it cognitively to develop a global spatial representation in memory.

Researchers have previously examined the application of visual zooming methods such as button press in electronic haptic displays (Magnuson & Rassmus-Grohn, 2003) and in virtual environment using force-feedback devices(S. Walker & Salisbury, 2003). However, these studies did not focus on the impact of the zooming operation on the learning process. Also, visual zooming methods cannot be used efficiently in non-visual settings. This is because haptic information extraction and learning requires serial processing, where one cannot take a quick glance at a particular zoom level to decide whether to explore the level further or move to the next level (Rastogi & Pawluk, 2013). Using haptics, one must at least first investigate a part of the graphic using contour following (in order to determine whether to zoom in or out), which is a slow, serial, and highly cognitively demanding process (Jones & Lederman, 2006). This could be extremely inefficient and frustrating depending on the content of the graphic and how the zoom levels are chosen. For instance, consider a scenario where different types of information are represented at different zoom levels (e.g., Structure of building at level 0, room location at level 1, floor path at level 2). Integrating information across these zoom levels is expected to be highly challenging in such scenarios. Although this integration can be achieved easily with vision, where parallel processing makes the top-down grouping of information relatively easy, it is much more difficult to perform the integration with touch, owing to its serial processing nature of information extraction and transmission. To appreciate this challenge, the reader is invited to try learning a map using zooming operations with your eyes closed. Although one can learn each zoom level separately, integrating the individual zoom levels in order to develop a single, consolidated global spatial representation is a difficult process. This means the information across levels (or adjacent levels) should have meaningful relations, and prominent features (landmarks) such that users can easily relate and integrate the levels. Therefore, for non-visual interfaces that are aimed at supporting zooming operations to be effectively used by people with visual impairments, it is necessary to maintain meaningful groupings (levels) of information. These groupings should avoid redundant zoom levels, and should also provide reference locations (or graphical elements) to assist the user in integrating and relating different zoom levels.

Much of the research on non-visual zooming has focused on usability of zooming methods (Rastogi, Street, & Pawluk, 2010) and algorithms to design meaningful groupings of information (Rastogi & Pawluk, 2013; Ziat, Gapenne, Stewart, Lenay, & Bausse, 2007). These studies addressed the efficacy of the zooming methods and algorithms with respect to computational constraints. Research projects have also focused on comparing the computational constraints of different zooming algorithms such as intuitive zooming, where zoom levels are based on functional relevance of the rendering (Rastogi et al., 2013), linear step zooming which enhances the graphical image at linear scale (Schloerb et al., 2010; Schmitz & Ertl, 2010; Ziat et al., 2007), logarithmic step zooming which enhances or shrinks the graphical image at logarithmic scale (Magnuson & Rassmus-Grohn, 2003), and smooth zooming using auditory cues (S. Walker & Salisbury, 2003). These studies demonstrated the efficiency of each of these algorithms in performing zooming operations. However, they were analyzed in an aspect relating to technological constraints of interfaces. In contrast, the focus of the current work deals with a different issue; namely, the ability of a user to accurately relate and integrate the graphical elements across different zoom levels and subsequently develop a global spatial representation in memory. The TouchOver map project investigated the complexity between two zoom levels in non-visual map learning. They found that users preferred the zoomed-in version of map over the zoomed-out version as it was easy to differentiate graphical elements in the zoomed-in version (Poppinga et al., 2011). However, similar to other studies, this work also did not investigate the human aspects related to zooming operation (i.e., how non-visual users will learn and integrate graphical elements across zoom levels). This is because, participants learned and reconstructed the graphical elements at each zoom level separately. The two zoom levels were used as different display mode conditions and the evaluation tasks did not require users to perform zooming operations or to integrate graphical elements across the zoom levels. Based on an extensive literature search, there is no research to my knowledge that addresses the cognitive constraints of the user in learning graphical information using zooming operations. For a zooming method (or algorithm) to be truly useful, it should support integration of graphical elements across different zoom levels. This means that, in addition to being intuitive and robust, the method should support the users in their learning process. To address this important issue, a human behavioral study was conducted to investigate whether incorporation of zooming operations in a non-visual interface supports or hinders the learning process.

5.2. Experiment 3: Evaluation of non-visual zooming

A human behavioral experiment was conducted to investigate the issue of non-visual zooming and was motivated by the following four goals:

1. To assess whether incorporation of zooming operation to the vibro-audio interface strengthens or weakens the learning process. The performance in learning graphical material using zooming operations and subsequent spatial representations will be compared to performance in learning the same graphical information without the need for zooming. If the performance does not differ between the zooming and no-zooming condition then it can be interpreted that the incorporation of zooming operations strengthens the learning process with the vibro-audio interface.

2. To investigate how the graphical information is processed and represented in the user's memory when learned using zooming operations. That is, how a non-visual user will integrate and relate graphical elements across

different zoom levels and subsequently develop a global spatial image of the graphical material being presented.

3. To compare and examine the efficacy of different zooming methods (discussed in sections 5.2.3 and 5.2.4) in supporting users to integrate, relate and learn graphical elements of an indoor map presented across different zoom levels using the vibro-audio interface.

4. To compare the efficiency in learning graphical material between using non-visual panning operations and non-visual zooming operations.

5.2.1. Method

Twelve sighted participants (five males and seven females, ages 19-30) were recruited for the study. All gave informed consent and were paid for their participation. The study took between 1 and 1.5 hours.

5.2.2. Conditions

To investigate whether incorporation of zooming operations in a non-visual interface supports or hinders the learning process, three different zoom-mode conditions were compared in this study; two zooming conditions and a third single zoom (control) condition. The two zooming methods were chosen from empirical research that identified the best methods as being intuitive and that were optimized for performing non-visual zooming tasks. Each of these methods represents a unique set of information redundancy and grouping.

5.2.3. Fixed zoom

The lineage of fixed zoom is rooted in visual zooming methods, where the scale of the graphical material will be stepped up (zoom-in) or stepped down (zoom-out) to enhance or reduce the level of information presented respectively. This method is commonly used with websites, image viewers, map applications, photo editors, and even for text magnifiers. This zooming method involves grouping of information based on its perceivable scale range. Zooming-in enhances the current information (graphical elements) and adds additional graphical elements based on the scale range (zoom level). Conversely, zooming-out removes some graphical elements and shrinks the other graphical elements according to the zoom level.

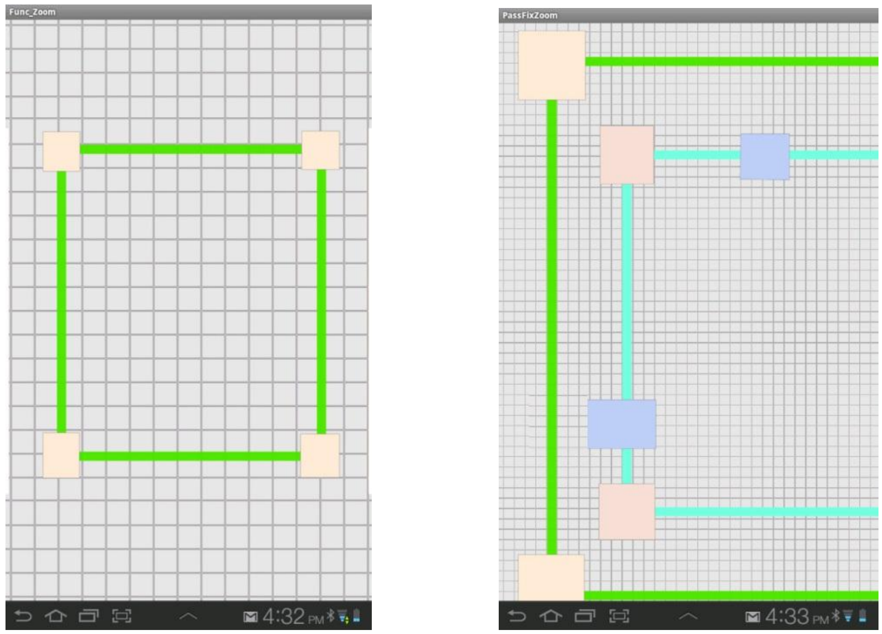

Figure 5.2. Vibro-audio interface (Fixed zoom) displaying building layout map at zoom level 0 (left) and level 1(right)

For example, if a building structure is displayed at zoom level 0, once zoomed-in to the next scale, the building structure (enhanced) along with rooms will be displayed at zoom level 1. Because of this, some of the graphical elements will expand to become larger than the screen extent (refer to figure 5.2). In such scenarios, a panning operation must also be incorporated into the interface in order to provide access to the entire graphic at each zoom level. It can be envisaged that the redundancy of graphical elements across different zoom levels will act as reference locations and support integration across those zoom levels.

5.2.4. Functional zoom

In contrast to the fixed zoom, functional zoom avoids redundancy across zoom levels and groups graphical elements based on their inter-relation and position. This method was conceptualized, developed and validated in a "mouse-like" display that senses absolute position in a virtual screen and provides feedback on an eight-pin tactile display (Owen et al., 2009; Rastogi & Pawluk, 2013). This method involves the use of what is termed as "intuitive zoom" levels, which determines the zoom levels based on an object hierarchy (see Rastogi et al., 2013 for details).

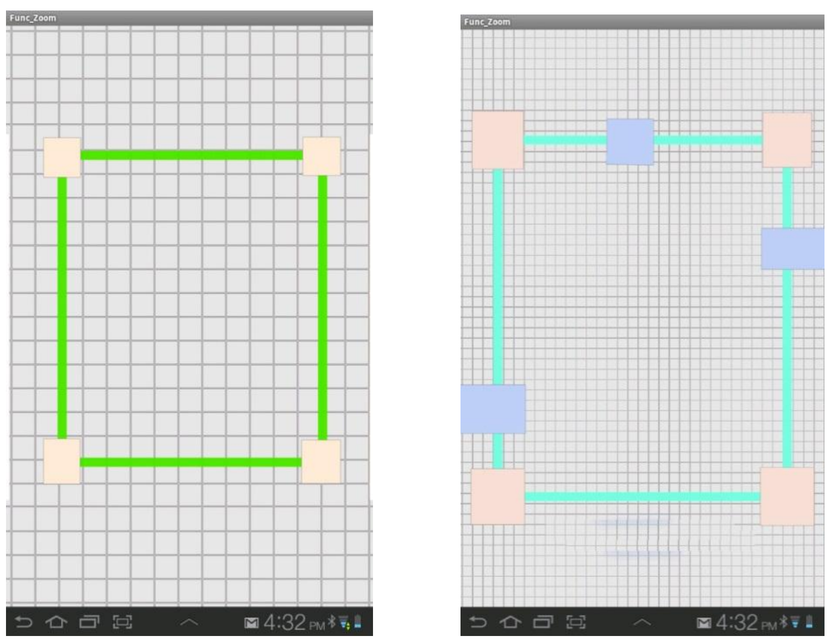

Figure 5.3. Vibro-audio interface (Functional zoom) displaying building layout map at zoom level 0 (left) and level 1(right)

This intuitive zooming algorithm involves two rules: (1) objects that are close to each other are considered as meaningful groupings and are selected as a whole to be represented in a sub-graphic; otherwise, (2) individual objects are represented in each sub-graphic. This grouping analysis is performed recursively via an algorithmic process on each sub-graphic until all graphics of the hierarchy are created. For example, as shown in figure 5.3, the building structure is grouped at zoom level 0 and corridor segments and landmarks within the corridor segments will be grouped at zoom level 1. This algorithm avoids presentation of unwanted zoom levels based on the object selection. This means the information of interest can be easily grouped and fit within the screen extent, thereby eliminating the need for panning operations. Earlier research on intuitive zooming has demonstrated this technique as an efficient method compared to fixed step zooming (Rastogi et al., 2013). However, the study evaluated the zooming method based on its usability in identifying objects within line diagrams and thus cannot be generalized to a learning process (as opposed to an identification task). Hence, this method is included here to investigate its efficiency in assisting a blind user to learn and integrate graphical elements across different zoom levels. Also, the focus here is on learning maps, which has a lot of utility in affording spatial access to one of the most common types of graphical information that is limited to blind users.

5.2.5. Stimuli and apparatus

For all three conditions, the vibro-audio interface was implemented on a Samsung Galaxy Tab 7.0 Plus tablet, with a 17.78 cm (7.0 inch) touch-screen used as the information display. The apparatus setup (table, chair and blindfold) was the same as in the previous experiments.

Three building layout maps were used as experimental stimuli (with two additional maps for practice). Each map was composed of corridors, landmarks, and junctions. Each map had three levels of information; namely (1) a layer containing the exterior wall structure of the building, (2) a layer with the corridor structure with position of important landmarks indicated, and (3) a landmark layer showing the details of each landmark (such as Restroom, Entrance and Exit).

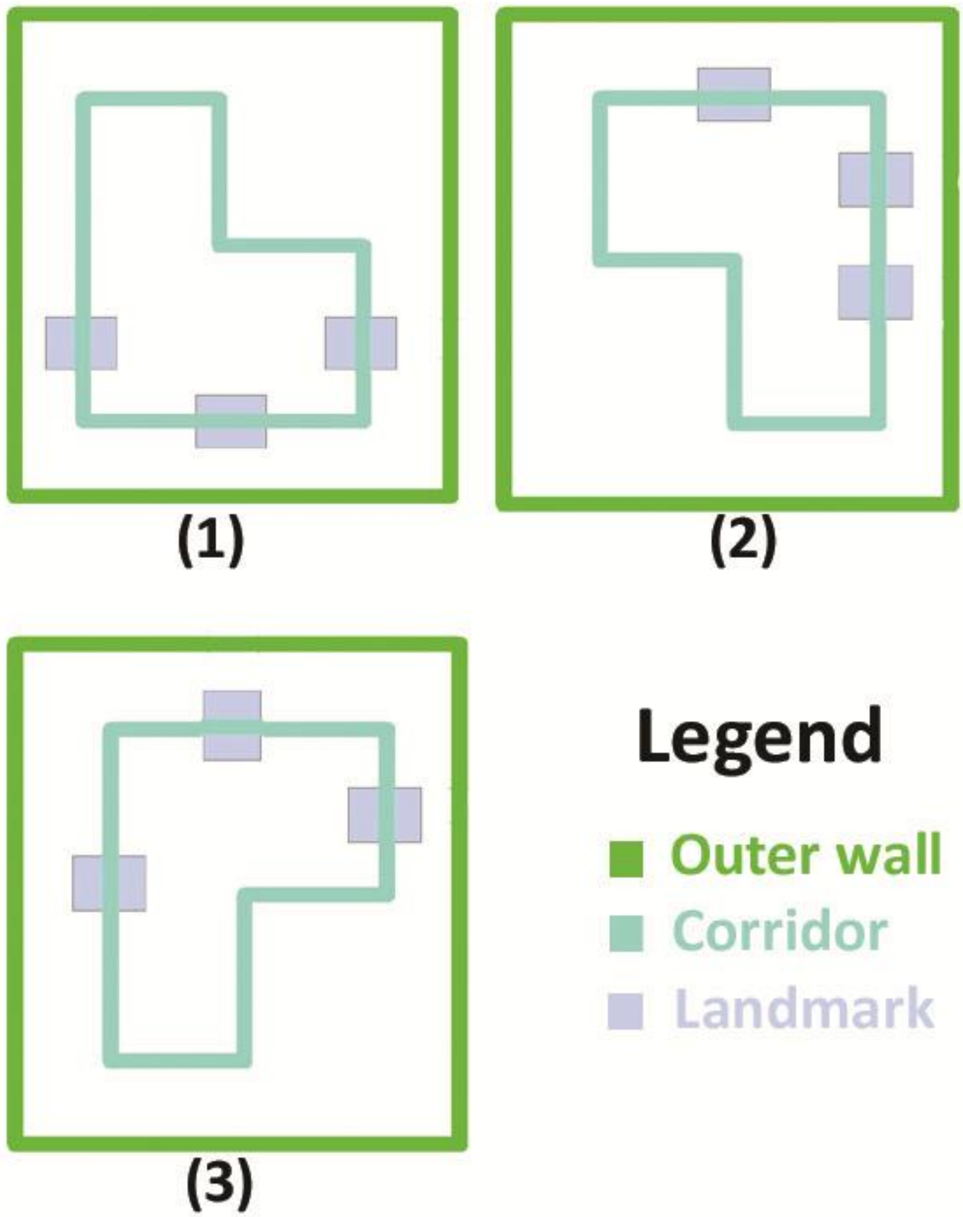

Figure 5.4. Building layout maps: Experimental stimuli used in Experiment 5

The three maps were carefully designed such that they had the same complexity but different topology (refer to figure 5.4). Each requires the user to zoom to each of the three different levels (and/or to pan in all four directions) in order to access the entire map. The complexity was matched in terms of:

1. Boundary structure: Each of the maps had a rectangular exterior wall structure (varied in aspect ratio).
2. Number and orientation of corridor segments: Each of the maps had 6 straight corridor segments (either horizontal or vertical).
3. Number of junctions: Each of the maps had 4 two-way junctions for exterior wall structure and 6 two-way junctions for corridor structure. All junctions were 90 degree right angle.
4. Number of landmarks: Each of the maps had 3 land marks. Each landmark was assigned a name based on a standard building layout theme: entrance, exit, and rest room.

Similar to previous experiments, all the maps were rendered with a line-width of 8.9 mm (0.35 inch), which corresponded to 60 pixels on the 7.0 inch touch-screen. The exterior walls were given a constant vibration, based on the UHL effect "Engine1_100". The junctions were indicated by a pulsing vibration, based on the UHL effect "Weapon_1". The corridors were indicated by a fast pulsing vibration, based on the UHL effect “Engine3_100”. The landmarks were indicated by an auditory cue (sine tone) coupled with a fast pulsing vibration, based on the UHL effect “Engine3_100”. In addition, for the junctions and landmarks, speech output (e.g., name of the landmark) was provided by tapping the vibrating region. Similarly, the zoom levels were indicated by speech output. For example, zooming-in to level 1 from level 0 was indicated by a speech output “at corridor level”. A physical sponge button affixed to the device was used as a reference (start) point. Similar to the methodology used in previous experiments, exploration was done using only one finger (dominant). The user’s movement behavior was tracked via the device’s touch-screen as they felt the stimuli, which also logged the learning time, finger-traces (co-ordinates), type of vibration pattern, zooming and the panning movements into a text file for each of the trials.

5.2.6. Procedure

A within subjects design was used in the experiment. In each condition, participants learned a building layout map and performed subsequent testing tasks. The condition orders were counterbalanced and individual maps randomized between participants. The study consisted of a practice, learning, and testing phase for each condition. The first practice trial in each condition was a demo trial where the experimenter explained the task, goal, and strategies and the participant explored the stimuli with corrective feedback provided. In the second practice trial, blindfolded participants were asked to learn the complete map, followed by a test sequence without a blindfold. The experimenter evaluated the answers immediately to ensure they understood the task correctly before moving to the experimental trials.

5.2.7. Learning phase

During the learning phase, participants were first blindfolded. The experimenter then placed their primary finger at the start location and instructed them to explore and learn the map. Participants were allowed to go back and forth between the zoom levels without restriction. Participants were asked to indicate to the experimenter when they believed that they had learned the entire map. Once indicated, the experimenter removed the device and moved on to the testing phase.

5.2.8. Learning criterion test

After learning the indoor layouts, participants performed a learning criterion test, which was done to ensure that all participants learned the map equally well based on a minimum learning level which would be required to undergo the next testing task. This test required participants to correctly indicate the allocentric direction between the reference (start) point and each of the landmarks using a physical pointer fixed on a wooden board. On passing the learning criterion test, subjects started with the next testing phase. If any of the three pointing trials were indicated incorrectly, this was considered as not passing the learning criterion test and the subject was asked to re-learn the map (an additional learning time of 5 minutes was given for re-learning).

5.2.9. Testing phase

The testing phase consisted of three tasks: a positioning, pointing, and reconstruction task.

In the positioning task, blindfolded participants answered questions and performed positioning tasks with the device. Each positioning task relied on accessing their mental spatial representation to answer questions about graphical elements in different zoom levels. The positioning task consisted of a set of three operations, each requiring zooming operation: two answering questions and one positioning question (e.g., from the landmark level, Zoom-out to the exterior wall level and mark the position of "Exit" with reference to its position on the exterior wall of the building). This task was excluded from the no-zoom condition as there was only one zoom-level. The positioning tasks were analyzed for time taken to perform a spatial task with zoom-in (or zoom-out) operations and correctly positioning the graphical element of one zoom level onto another zoom level. On completion of the positioning tasks, participants were allowed to remove the blindfold.

Similar to Experiment 4, in the pointing task participants indicated the allocentric direction between landmarks using a physical pointer fixed on a wooden board (refer to figure 4.6). The pointing task consisted of a set of three pointing questions (e.g., indicate the direction from entrance to restroom) covering all three pairs of landmarks. The reproduced angles were analyzed for their correctness in relative position and direction between landmarks.

In the reconstruction task, participants were asked to draw the map and label landmarks on a template canvas of the same size as the original map. To provide the subjects with a reference frame for map scale, the device screen size and the reference point was already marked in the canvas (see Figure 5.5). The reconstructed maps were analyzed in terms of whether the maps had the correct spatial pattern of exterior wall and corridor segments, and correct landmark position and labeling.

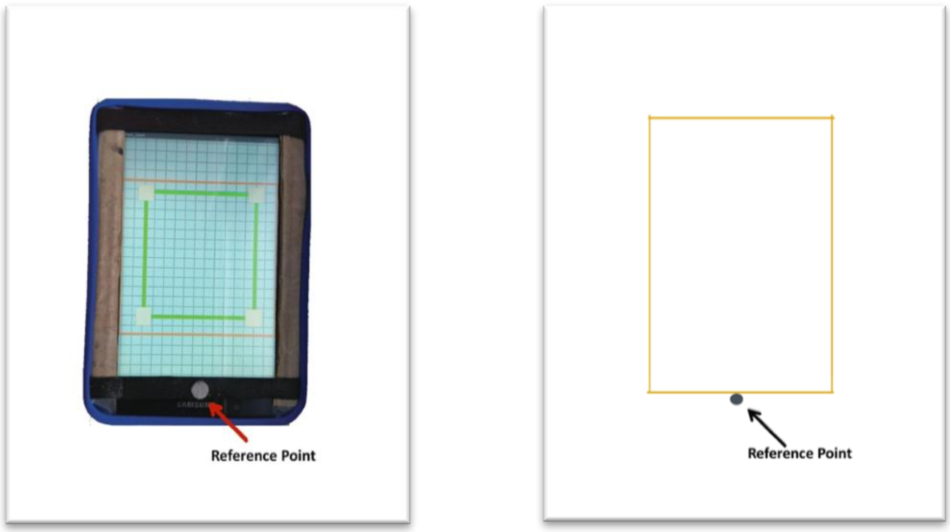

Figure 5.5. A4 canvas with frame size and reference point matching the screen size and affixed sponge button of Samsung galaxy 7.0 device

5.2.10. Experimental measures and analyses

From this design, the following measures were evaluated as a function of zoom-mode condition.

1. Learning time: The learning time represents the level of cognitive effort imposed on the user while learning the map with each zooming method. The Learning time is the time taken from the moment they touch the screen until they confirmed that they had completed learning of the map. The time was measured from log files of each trial that was created within the device. The learning time ranged from ~1.5 minutes to ~12 minutes with a mean of ~5 minutes.

2. Times panned: The number of times the map panned was compared between the fixed zoom and no-zoom conditions. As the information grouping was matched between no-zoom condition and zoom level 2 of fixed zoom condition, the number of times of performing a panning operation was expected to be the same between conditions. The times panned were calculated from the log files of each trial that was created within the device.

3. Relative positioning accuracy: As discussed in section 5.2.9, participants were asked to mark the position of landmarks from one zoom level onto another zoom level. The landmark positioning accuracy was measured by matching the marked position to its actual position. This measure was compared between the functional zoom and fixed zoom conditions. The no-zoom condition was excluded as there was no zooming operation performed. The positioning was measured from the co-ordinates recorded in the log files of each trial.

4. Positioning Time: For the three positioning tasks, the time taken to identify a landmark using zooming operations was measured and analyzed. Similar to relative positioning accuracy, this measure was also compared between

the functional zoom and fixed zoom conditions. The time was measured from the log files of each trial that was created within the device.

5. Relative directional accuracy: Similar to Experiment 4, the angles between landmarks reproduced by the participants were compared to the actual angles between the landmarks to measure the angular errors. These angular errors were then analyzed in two ways: Unsigned error and Signed error (under estimating the angle representing a negative bias and over estimating representing a positive bias).

6. Reconstruction accuracy: Similar to Experiment 4, the maps were reconstructed by participants and were analyzed in two ways; (1) Binary score and (2) Bi-dimensional regression. The measuring and analyzing procedure was similar to Experiment 4. The Only difference here is that for Bi-dimensional regression, thirteen anchor points (3 landmarks and 10 junctions) were chosen on each map (as opposed to seven anchor points in Experiment 4).

7. Landmark labeling accuracy: Similar to experiment 4, the accuracy in labeling was measured from the reconstructed maps. A discrete scoring was applied based on the correctness of the landmark labeling (i.e., 1 for each correct label, 3 if all three labels are correct).

8. Subjective rating for each condition: Participants were asked to rank the three conditions on a scale of three (with one being the best). The ranks given by subjects were analyzed to understand the user's preference.

5.3. Results

Performance data for each of the measures described above were analyzed and compared between the three conditions. Univariate ANOVAs were conducted on each of the measures to assess the within-subjects effects between conditions. Also, post hoc paired sample t-Tests were conducted to assess the difference in performance between each condition. The f, t and p values of these analyses are given in the tables below.

From the ANOVA results (see Table 5.1), it can be inferred that there were no significant differences between conditions for all measures except learning time. Similarly, the results of paired-sample t-Tests between conditions were highly in-significant (all $p>0.05$) for all measures except for learning time and times panned.

Measures	Condition			
	df		f	Sig.
	Hypothesis	*Error*		
Learning Time	2	22	8.591	0.002
Relative positioning accuracy- error in X axis	1	22	3.626	0.083
Relative positioning accuracy- error in Y axis	1	11	1.044	0.329
Positioning time	1	11	0.363	0.559
Relative directional accuracy	2	22	1.261	0.303
Reconstruction accuracy	2	22	0.186	0.831
Landmark labeling	2	22	0.000	1.000

Table 5.1. Univariate ANOVA between conditions for each measure

Measures	Fixed vs. Functional			Fixed vs. No-zoom			Functional vs. No-zoom		
	df	t	Sig.	df	t	Sig.	df	t	Sig.
Learning Time	11	4.044	0.002	11	0.787	0.448	11	3.694	0.004
Relative positioning accuracy- error in X axis	11	-1.904	0.083	*	*	*	*	*	*
Relative positioning accuracy- error in Y axis	11	-1.022	0.329	*	*	*	*	*	*
Positioning time	35	-0.457	0.650	*	*	*	*	*	*
Relative directional accuracy	11	1.989	0.072	11	-0.735	0.478	11	-1.34	0.207
Reconstruction accuracy	11	-0.432	0.674	11	0.000	1.000	11	0.561	0.586
Landmark labeling	11	0.000	1.000	11	0.000	1.000	11	0.000	1.000
Times panned	*	*	*	11	-3.802	0.003	*	*	*

*Not applicable for that condition

Table 5.2. Paired sample t-Tests between conditions for each measure

Measures	Functional Zoom		Fixed Zoom		No-Zoom	
	Mean	SD	Mean	SD	Mean	SD
Learning Time (in seconds)	222.33	99.35	335.83	145.95	368.92	175.47
Times panned	NA	NA	3.17	2.66	10.08	5.28
Relative positioning accuracy- error in X axiz	0.10	0.36	0.56	0.73	NA	NA
Relative positioning accuracy- error in Y axis	0.12	0.22	0.42	1.01	NA	NA
Positioning time	27.50	6.60	25.44	5.11	NA	NA
Relative directional accuracy - Unsigned error	3.06	5.11	5.83	6.27	9.44	29.97
Relative directional accuracy - Signed error	-0.56	5.95	-1.39	8.50	2.22	31.38
Reconstruction accuracy	0.75	0.45	0.66	0.49	0.66	0.49
Landmark labeling	3.00	0.00	3.00	0.00	3.00	0.00
Subjective rating	1.75	0.87	1.75	0.62	2.50	0.80

Table 5.3. Mean and Standard deviation for each measure as a function of zoom-mode condition

Corroborating the results from Tables 5.1-5.3 it can be inferred that participants took less time to learn using functional zoom, demonstrating the intuitiveness of the method. Also, no significant difference ($p>0.01$) was observed between the fixed and no-zoom conditions. This is notable because participants performed both zooming and panning operations in the fixed zoom condition, whereas they performed only panning in the no-zooming condition. This means that introducing a zooming operation did not impose any measurable additional cognitive load on the participants. The number of times panned was significantly less in the fixed zoom condition when compared to no-zoom condition. This makes sense as participants used panning only to integrate the graphical elements on level 2 to the elements that they already learnt from level 1. Conversely, participants used panning to learn and integrate the elements of the entire map in a single zoom level.

The results also suggested that there was no significant difference between fixed and functional zoom conditions in the relative positioning accuracy (refer to Table 5.1 and 5.2). Because of information redundancy the fixed zoom was expected to perform better than functional zoom in this measure as it provides reference points between zoom levels. But the similarity of performance between fixed and functional zooming is a remarkable finding, as it demonstrates that participants were able to integrate and relate graphical elements from one zoom level to graphical elements at another zoom level even without reference points to align position across levels. The time taken to perform a positioning task varied between the conditions for the three tasks (see figure 5.6). The first positioning task (zooming-in to a landmark from the exterior wall level) was performed fastest with the fixed zoom condition. Conversely, the last positioning task (zooming-in to a landmark from the exterior wall level) was performed faster with functional zoom. For the second positioning task (zooming-out to the exterior wall level and marking the positioning of a landmark) both functional and fixed zoom took same amount of time. However, these differences in performance time for the first and third task were not statistically significant.

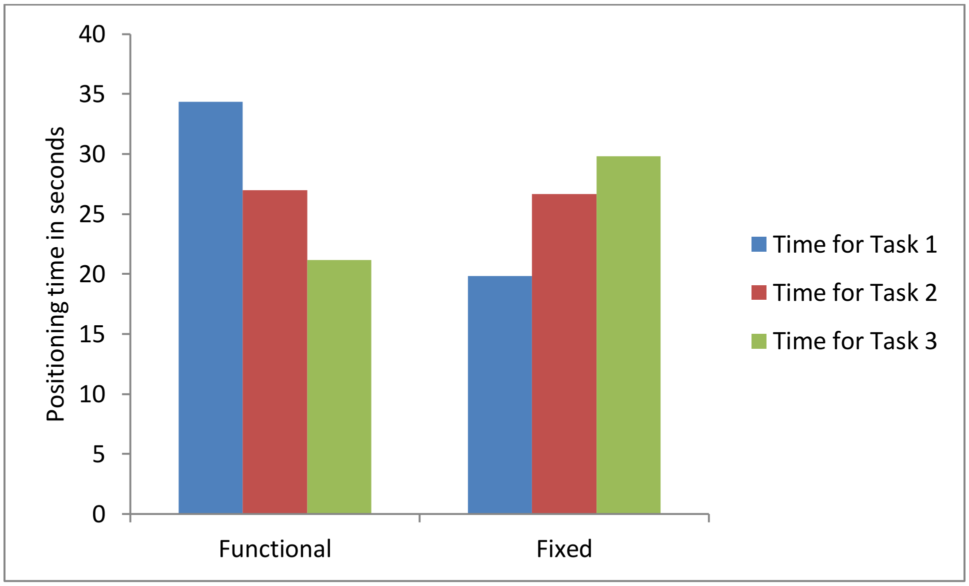

Figure 5.6. Positioning time for the three tasks as a function of zoom-mode

Comparing the means and standard deviations of the signed errors (refer to Table 5.3), it can be inferred that participants generally undershoot the direction for Fixed and Functional zoom and overshoot for the no-zoom condition. This demonstrates the differences in mental representation of the graphical material developed with and without zooming operations. That is, in the zooming conditions participants perceived the graphical elements within the screen size. Whereas in the no-zoom condition they panned many times which might have created an illusion of the graphical material as being larger than in the other two conditions. This demonstrates the difference in mental representation of the graphical material developed using different panning and zooming methods. That is a perceptual bias leading to compression of the mental representation arises while using zooming operations. From the results of the Bi-dimensional regression, it was evident that there were no significant differences between conditions for the three factors; Scale, Theta and Distortion Index. The mean and standard deviations for the three factors are given in the table below.

Condition	Scale		Theta		Distortion Index	
	Mean	SD	Mean	SD	Mean	SD
Functional	0.975442	0.11496	-0.3459	0.5594	17.29714	2.332015154
Fixed	0.983089	0.15646	-0.4215	1.221	18.6785	3.263005129
No-Zoom	1.057212	0.16818	-0.024	1.1866	17.91304	3.931660957

Table 5.4. Scale, Theta and Distortion index from bi-dimensional regression as a function of pan-mode, along with Standard deviation.

From the ranking data of user preference, participants clearly preferred zooming methods over the panning (control) method. The two zooming methods were given an equal rating suggesting participants similarly preferred both methods.

5.4. Discussion

This study evaluated the efficiency of two touch-based zooming methods in supporting non-visual map learning, navigation, and representation. The most important outcome of this experiment is the similarity of performance across testing measures for the three conditions (two zooming methods and one no-zoom method). This means, in general, both zooming and panning operations support the non-visual learning process when incorporated with a vibro-audio interface. These findings are remarkable given the necessity of panning and zooming operations in touch-based devices for accessing large and deep format graphics.

One of the main aims of this study was to investigate how graphical information is processed and represented in user's memory when learned using a zooming operation. That is how a non-visual graphical material can be accessed at each zoom level independently and then integrated to develop a global spatial representation in memory. Results from this study provide sufficient evidence that participants were able to integrate and relate the graphical elements displayed across different zoom levels and subsequently develop an accurate spatial representation of the building map displayed. On comparing the efficacy between the two zooming methods, functional zoom exhibited superior performance in all the learning time, demonstrating that functional zoom was the faster and more intuitive zooming technique for learning graphics (e.g., layout maps) using a non-visual vibro-audio interface. The superior performance of the functional zoom in learning time can be attributed to its simplicity. That is the complex information was divided into simple groups and the groups were presented to users at different temporal intervals which allowed them to conceptualize it in a better way.

In experiment 4, the two finger-drag was found to be an efficient and intuitive method to access and learn large format graphics. In contrast, results from this study suggest that participants were able to learn graphical information easily and more accurately using a zooming operation rather than learning using a two finger-drag panning operation (used as a control/no-zoom condition). This could be because in the zooming conditions the information were divided and presented as groups, making it easier for users to conceptualize and remember. This trend was demonstrated across all the testing measures. In addition to the poor performance in the no-zoom condition, participants also self-reported that having to navigate the screen and locate the graphical elements was most difficult in this condition, as there was no cue to indicate the extent of panning. Also, many participants felt that this condition required too much information on one screen.

One limitation of this study setup is that the design did not force the participants to zoom-out the map while learning. Of the 12 participants, only one participant zoomed out from the corridor level to the outer wall level in order to relate the map elements between zoom levels. Users only zoomed-out for answering one of the three positioning tasks. Although the performance between zoom-in and zoom-out operations did not significantly differ, having more zoom levels and forcing the user to zoom-out (and/or zoom-in) more than once might change the efficiency in the learning process. Because of this limitation, the results cannot be broadly generalized beyond the current stimuli and are therefore not representative of all situations. However, the current findings should be considered as an important first step for a trend to supporting the efficacy of zooming operations with the vibro-audio interface. These findings need to be evaluated in different scenarios in the future.

5.5. Summary

This chapter investigated the non-visual zooming issue through a human behavioral study. In sum, the results showed similarity of error performance across all measures for zooming conditions and panning condition, demonstrating that the incorporation of zooming operation does not weaken the learning process. The results also exhibited a trend of accurate learning of graphical information using zooming operations rather than learning using panning operations.

CHAPTER 6

DISCUSSION AND FUTURE DIRECTIONS

This chapter summarizes the major findings of the thesis and describes their importance with respect to accessible graphics. The following sections in this chapter elaborate the contributions of this work, and discuss the future research directions that could be extended based on the research related to non-visual graphical access.

6.1. Summary of the work

The Introduction of this thesis (see chapter 1) motivated the need for providing dynamic accessible graphics for blind and visually-impaired users. The traditional approaches for non-visual access to graphical information such as tactile graphics and haptic displays have had limited success in reaching the end-user because of various shortcomings (such as use of unintuitive sensory translation rules, prohibitive cost, and limited portability). Only a few approaches have made headway in overcoming these shortcomings owing to the advent of refreshable haptic displays and touch-based devices. However, these approaches also have significant limitations as their design was primarily driven by engineering principles rather than theoretical knowledge of human information processing and awareness about the needs and behaviors of end-user's as driving design decisions. Also, the previous approaches based on touch-screen devices like smart phones and tablets were designed without consideration of fundamental perceptual and cognitive capabilities of human users. To overcome these shortcomings, this thesis proposed what is called a vibro-audio Interface that was explicitly designed with considerations of human information processing and end-user' needs in mind (see section 1.3). The goal of this thesis work was to investigate the human information processing capabilities using this novel interface, with focus on non-visual graphical learning.

Three human behavioral studies were conducted that assessed comprehension of the relative relations and global structure of a bar graph (Exp 1), Pattern recognition via a letter identification task (Exp 2), and orientation of complex geometric shapes (Exp 3). Performance with the vibro-audio interface was compared to the same tasks performed using traditional hardcopy tactile graphics. Results from the three experiments showed similar error performance between the two display modes across all measures for both blindfolded-sighted and for blind users, indicating that the vibro-audio interface is a viable multimodal solution for presenting dynamic graphical information and supporting development of accurate mental spatial representations of otherwise inaccessible graphical material. These results brought forward first evidence that learning non-visual graphical material is facilitated by the vibro-audio interface. However, the implemented device has some inherent limitations, namely, the limited display size of the device screen for presenting graphical information. To overcome this limitation, panning and zooming operations have to be incorporated into the vibro-audio interface. But, since these operations are almost always performed visually, they must be modified significantly to be incorporated with the vibro-audio interface. The question remained open on how a non-visual user will perform panning and zooming operations on a touch-screen device and/or subsequently learn graphical material by integrating its elements across panning screens and zooming levels. To address this issue, two human behavioral studies were conducted that assessed the non-visual integration of elements from large format graphical material displayed across panning screens (Exp 4) and integration of elements from deep format graphical material between multiple zoom levels (Exp 5). In Experiment 4, performance in learning large format graphics was compared between four different pan-mode conditions and a no-pan (control) condition. Similarly, in Experiment 5, performance with learning deep format graphics was compared between two zooming conditions and a no-zoom (control) condition. Results from both experiments showed similar error performance across all measures for all conditions, indicating that the incorporation of panning and zooming operations in the vibro-audio interface has potential benefits in learning large and deep format graphics and subsequent development of accurate spatial representation of otherwise inaccessible graphical material.

Taken together, results from the five experiments provide compelling evidence that a non-visual user can efficiently extract graphical information from a touch-based interface and subsequently develop accurate mental representation of the graphical information being conveyed. These findings are important as this interface provides dynamic and instant rendering of information, whereas hardcopy tactile output is static and also requires expensive, highly specialized equipment to produce. In addition, as the vibro-audio interface is based on inexpensive, multi-purpose, and commercially-available hardware, it represents a viable alternative to the expensive and highly complex auditory and haptic solutions which have various shortcomings, as was described in section 1.1.

6.2. Contributions and future directions

Although the vibro-audio interface offers great promise as a non-visual graphical display, the touch-screen nature of the interface also poses many challenges. Because the surfaces are smooth, displaying graphics is inherently different from one based on traditional hardcopy tactile graphics. To access and conceptualize the information from a touch-screen device, a user must (1) use kinesthetic sensory cues to keep track of touch locations, (2) interpret the external cue (vibration and/or audio), and (3) associate the cued content with the currently contacted coordinates (Klatzky, Giudice, Bennett, & Loomis, In press). These three processing components present various challenges. To address this challenges, this thesis investigated the human factors relating to non-visual learning and human information processing through a series of behavioral studies and brought about the following contributions.

1. This work demonstrated that the touch-based vibro-audio interface is a viable multimodal solution for the long standing accessibility issue faced by blind individuals. The work also illustrated that use of the vibro-audio interface supports building up of accurate spatial representation of the graphical information and subsequently assisting users in supporting accurate spatial behaviors based on learning the graphical information.

2. Based on the study results from Experiment 4, it was found that the incorporation of panning operations in the vibro-audio interface do not exhibit any detrimental effect in the cognitive representation of the graphical material. Not only was the performance with panning operations similar to the no-panning condition on most measures, they were actually better on some, demonstrating that the incorporation of panning operations in the vibro-audio interface strengthened the learning process. This finding is important given the need for panning operations on touch-based devices for accessing larger graphics. The study also found that the two finger-drag method was the most intuitive and efficient non-visual panning method for accessing and learning larger graphical material.

3. Results from Experiment 5, demonstrated that the incorporation of zooming operations improves the learning process and support building up of more accurate spatial representations than the one build up using two finger-drag panning. The study also demonstrated that the functional zooming technique was faster and intuitive method for performing non-visual zooming.

4. The studies used three different vibration patterns to indicate different graphical objects. Although the vibration patterns were triggered using

only one embedded vibration motor, participants were able to differentiate and link the vibro-tactile cue to its relevant object. This means that even in the simplest case using a standard device with one embedded vibration motor, a high level of performance is possible. This is important as it provides evidence for usability of different vibration patterns to represent different objects. This means complex graphical information which cannot be perceived and differentiated using hardcopy output or haptic displays can be perceived and differentiated using a vibro-audio interface, as this interface provides dynamic and readily implemented information.

5. The graphical materials studied in this thesis were of different types (e.g., bar graphs, indoor maps, and simple shapes). However, these graphics were customized based on certain parameters to acknowledge both the constraints in human perception and constraints in the interface. Based on the results of the five human behavioral studies, this work recommends that several factors related to the human end-users and to the interface design should be acknowledged for a non-visual graphical access system to be successful. Following are some of the recommended considerations for research and development of non-visual interfaces.

6.3. Cognitive considerations

1. Small is better: Results from Experiment 3, suggested that participants preferred conditions where zooming of information was required versus conditions where the information needed to be panned. This is likely due to the fact that the graphical elements displayed in the zooming conditions fit entirely within the screen extent, whereas in panning conditions the graphical elements extended beyond the screen extent. Although the same graphical material was displayed in all conditions, participants perceived that the material used with panning operations was bigger than that in zooming. This is because in zooming conditions participants perceived the graphical elements within a single screen. Whereas in the panning conditions they panned many times to apprehend the graphics as a whole, which might have created an illusion that the graphical material was larger than in the other zooming conditions. Participants also self-reported that they felt there was a lot of information in the panning condition.

2. Unique patterns: As discussed in section 6.2, using different vibration patterns to indicate different graphical objects will help users in identifying different graphical elements with less cognitive effort.

3. Additional cues: Although vibration patterns can indicate different objects, more psychophysical work needs to be done on how many patterns and their parameters can be distinguished and interpreted. Similar to earlier research on multimodal interfaces (Raja, 2011; Zeng & Weber, 2010), this thesis also suggests that haptic inputs coupled with audio cues are considered better than either haptic or audio in isolation. The advantage of using a multimodal display is that one can add semantic labels to elements or augment the vibration to provide a much richer and more robust stimulus set of cues to be used to represent the graphical information. Hence, it is necessary to use complementing cues such as speech or audio to present information or to indicate an object. Most of the participants self-reported that having the additional auditory cue to the vibro-tactile information was very helpful in identifying the landmarks and junctions.

4. Topology: Maintaining a meaningful grouping between graphical elements is mandatory for non-visual learning. In order to avoid the difficulty of not knowing how much to pan or zoom, it is necessary to maintain the topology between individual graphical elements and to its sub graphics. This will likely reduce the learning time and cognitive effort of the user.

6.4. Tactile considerations

1. Angular lines and junctions: From ad-hoc analysis of the log files, it was found that participants spent more time in tracing oriented lines and their junctions when compared to straight lines and right angled junctions. Participants also self-reported that tracing slanting lines was challenging. This adds to the evidence from (Giudice et al., 2012) suggesting the need for developing a secondary cue to assist with contour tracing and staying oriented when exploring non-rectilinear lines and junctions.

2. Multitouch: Although it was not in the current design due to only one embedded vibrator in the devices used, these touch-based devices are capable of detecting different touch points and their locations. This means that the multitouch feature of touch-based displays could be utilized more efficiently in the future to allow the user to obtain stimulation on more than one finger and to simultaneously access different objects on the screen.

3. Meaningful cues: The fact that participants can detect different tactile feedback should be utilized efficiently to provide meaningful cues. This means patterns should match the functionality of graphical objects. For instance, a railway path might have a vibration pattern matching a train

sound. However, more psychophysical work needs to be done on how many vibration patterns and their parameters can be distinguished and interpreted on a touch-based device.

6.5. Interface considerations

1. Task oriented design: Designing graphical material for the use of visually-impaired individuals will increase the information gap between blind persons and their sighted peers. At the same time, learning non-visual graphics is a cognitively demanding task for blind persons. The interface should acknowledge this constraint and utilize a task oriented design approach. This means the interface should use the same graphical material as that of the visual graphic and present only the required information to blind users based on the task. This will require down sampling of information owing to the different sensory bandwidth between visual and tactual modalities, and some way of figuring out what information is salient and what is not. It also must parse the image and then map the lower resolution output to the optimized vibro-audio elements that best provide non-visual access. This is a difficult problem but one that must be addressed if there is to be automated conversion of visual images to vibro-tactile output. The graphics used in this research were all manually

authored but to have true universal access, there would need to be a more automated conversion process. This is something that should be addressed in future research.

2. Customizable: Although many participants self-reported that they preferred zooming over panning, a few participants felt they had developed a better understanding with the panning conditions. Each individual will have their own preference in using an interface. This is common for both visual and non-visual displays. Hence, the non-visual interface should be highly customizable to support the divergent needs of this heterogeneous user demographic.

6.6. Generalization of the results

The results of the five experiments described in this thesis cannot be generalized to all situations. Many assumptions were made for the prototype vibro-audio interface, which can be modified depending on the graphical information, task, or use scenario. These modifications should be user tested and statistically evaluated before being implemented in the interface. For instance, the zooming setup in my thesis work did not require users to zoom-out during the learning phase and the panning condition did not facilitate users in re-positioning the map. Increasing the zoom levels and including increased positioning functionality for panning operations could change the results. Also, the participant sample for Experiments 4 and 5 did not include any blind people. Because of these assumptions and limitations, the results are not broadly generalizable and should only be considered as first indications for measurable effects and need to be statistically evaluated in different scenarios. Following are some of the future directions that need to be addressed based on the current assumptions and known limitations of the interface.

1. Extending the bounds: As discussed in section 4.6, the bounds of the graphical material were extended to avoid automatic resuming. However, this solution is not applicable for real time scenarios. This issue should be studied further to provide a better and universal solution.

2. Re-positioning in panning: Many participants self-reported that having additional functionality to re-position the map to its start location (or last traced landmark) would have helped them to re-orient themselves within the map. This was not added in the current design. Adding this functionality may well alter the current results as getting lost within the map was the major problem faced by users in all panning conditions. Adding these improvements may

even make the panning conditions easier than the zooming conditions, but more research would need to be conducted.

3. Slanting and curved lines: Staying oriented when tracing slanting lines was one of the hardest parts of using the current design. There is a need for future research to find ways for more accurate orientation perception.

4. Regions: The Current design evaluated only the perception on lines of graphs, polygons, and maps. The question is still open on how a user would best perceive solid regions using vibration patterns on a touch-based device.

5. Real-time scenarios: The vibro-audio interface was tested in this work as an offline learning interface. This allowed the user to place the device on a table (or any flat surface) and perceive graphics using one finger. However, using this setup in real time scenarios will not provide the same level of perception as one has to hold the device in one hand. This means users may not achieve the same level of vivid perception as they achieved in the offline mode. Extending this work to online situations would represent a

significant contribution to the visually-impaired community but would also involve a host of new factors to be tested.

This thesis set out to contribute to the development of accessible graphics because in our information driven culture, this major component of information consumption has been denied to blind and visually-impaired users. Having access to graphical material means that blind persons can be competitive with their sighted peers in educational, vocational and social settings. This thesis strongly supports the efficacy of a vibro-audio interface as a viable and immediate solution to this problem. It also demonstrates the need for further development and improvement of research with this interface to be more usable and widely generalizable.

BIBLIOGRAPHY

Alty, J. L., & Dimitrios I. Rigas. (1998). Communicating graphical information to blind users using music: the role of context. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 574–581). ACM Press/Addison-Wesley Publishing Co.

Benali-Khoudja, M., Hafez, M., Alexandre, J.-M., Kheddar, A., & Moreau, V. (2004). VITAL: a new low-cost vibro-tactile display system. In *In Robotics and Automation, 2004* (pp. 721–726). IEEE International Conference.

Bliss, J. C., Katcher, M. H., Rogers, C. H., & Shepard, R. P. (1970). Optical-to-tactile image conversion for the blind. *Man-Machine Systems, IEEE Transactions on*, *11*(1), 58–65.

Casey, S. M. (1978). Cognitive mapping by the blind. *Journal of visual impairment & blindness*, *72*(8), 297 – 301.

Cattaneo, Z., Vecchi, T., Cornoldi, C., Mammarella, I., Bonino, D., Ricciardi, E., & Pietrini, P. (2008). Imagery and spatial processes in blindness and visual impairment. *Neuroscience and biobehavioral reviews*, *32*(8), 1346–60. doi:10.1016/j.neubiorev.2008.05.002

Cohen, R. F., Rui, Y., Meacham, A., & Skaff, J. (2005). PLUMB : DISPLAYING GRAPHS TO THE BLIND USING AN ACTIVE AUDITORY INTERFACE. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility* (pp. 182–183). ACM.

Coren, S., Ward, L. M., & Enns., J. T. (2004). *Sensation and perception*. John Wiley & Sons.

Craig, J. C. (1999). Grating orientation as a measure of tactile spatial acuity. *Somatosensory & motor research*, *16*(3), 197–206.

Crossan, A., & Brewster, S. (2008). Multimodal trajectory playback for teaching shape information and trajectories to visually impaired computer users. *ACM Transactions on Accessible Computing (TACCESS)*, *1*(2), 12.

Dinger, T., Lindsay, J., & Walker, B. N. (2008). 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.

Edman, P. K. (1992). *Tactile Graphics*. American Foundation for the Blind Press, USA.

Eriksson, Y. (1998). Tactile Pictures. Pictorial representations for the blind 1784-1940.

Ferres, L., Lindgaard, G., & Sumegi, L. (2010). Evaluating a Tool for Improving Accessibility to Complex Visual Objects. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility* (pp. 83–90). ACM.

Frauenberger, C., & Noisternig, M. (2003). 3D Audio Interfaces for the Blind. In *Workshop on Nomadic Data Services and Mobility* (pp. 11–12).

G. Jansson, I. Juhasz, & A. Cammilton. (2006). Reading virtual maps with a haptic mouse: Effects of some modifications of the tactile and audio-tactile information. *British Journal of Visual Impairment*, *24*(2), 60–66. doi:10.1177/0264619606064206

Giudice, N. A. (2004). *Navigating novel environments: A comparison of verbal and visual learning*. University of Minnesota.

Giudice, N. A., Betty, M. R., & Loomis, J. M. (2011). Functional equivalence of spatial images from touch and vision: evidence from spatial updating in blind and sighted individuals. *Journal of experimental psychology. Learning, memory, and cognition*, *37*(3), 621–34. doi:10.1037/a0022331

Giudice, N. A., & Legge, G. E. (2008). Blind navigation and the role of technology. In A. HELAL, M. MOKHTARI, & B. ABDULRAZAK (Eds.), *In Engineering Handbook of Smart Technology for Aging, Disability, and Independence* (pp. 479–500). John Wiley & Sons. doi:http://dx.doi.org/10.1002/9780470379424.

Giudice, N. A., Palani, H. P., Brenner, E., & Kramer, K. M. (2012). Learning non-visual graphical information using a touch-based vibro-audio interface. In

Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility (pp. 103–110). New York, New York, USA: ACM Press. doi:10.1145/2384916.2384935

Giudice, N. A., & Tietz, J. D. (2008). Learning with virtual verbal displays: effects of interface fidelity on cognitive map development. In *Spatial Cognition VI. Learning, Reasoning, and Talking about Space* (pp. 121–137). Springer.

Golledge, R. G. (1991). Tactual strip maps as navigational aids. *Journal of visual impairment & blindness*.

Goncu, C., & Marriott, K. (2011). GraVVITAS: generic multi-touch presentation of accessible graphics. *Lecture notes in computer science*, *6946*, 30–48. doi:10.1007/978-3-642-23774-4_5

Goncu, C., Marriott, K., & Hurst, J. (2010). Usability of Accessible Bar Charts. In *Diagrammatic Representation and Inference* (pp. 167–181). Springer Berlin / Heidelberg.

Hasser, C. (1995). HAPTAC: A Haptic Tactile Display for the Presentation of Two-Dimensional Virtual or Remote Environments.

Hasty, L. (2009). Personal Communication.

Hatwell, Y. (1993). *Images and non-visual spatial representations in the blind* (p. 13). COLLOQUES-INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE COLLOQUES ET SEMINAIRES.

Hoggan, E., Brewster, S. a., & Johnston, J. (2008). Investigating the effectiveness of tactile feedback for mobile touchscreens. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, 1573. doi:10.1145/1357054.1357300

Hoggan, E., Crossan, A., Brewster, S., & Kaaresoja, T. (2009). Audio or Tactile Feedback : Which Modality When ?, 2–5.

Immersion. (2013). Retrieved from http://www.immersion.com/products/haptic-sdk/

Infographics. (2013). Retrieved from http://en.wikipedia.org/wiki/Infographic

Jacobson, R. D. (1998). Navigating maps with little or no sight: An audio-tactile approach. In *Proceedings of the workshop on Content Visualization and Intermedia Representations (CVIR)*.

JAWS. (2013). Retrieved from www.freedomscientific.com

Jones, L. A., & Lederman, S. J. (2006). *Human hand function*. Oxford University Press.

Kammermeier, P., Buss, M., & Schmidt, G. (2000). Dynamic display of distributed tactile shape information by a prototypical actuator array. In *Intelligent Robots and Systems, 2000* (Vol. 2, pp. 1119–1124). IEEE/RSJ International Conference.

Kammermeier, P., & Schmidt, G. (2002). Application-specific evaluation of tactile array displays for the human fingertip. In *Intelligent Robots and Systems, 2002* (Vol. 3, pp. 2937–2942). IEEE/RSJ International Conference.

Klatzky, R. L., Giudice, N. A., Bennett, C. R., & Loomis, J. M. (In press). The promise of touch-screen technology for the dynamic display of 2D spatial information without vision. *Multisensory Research*.

Larkin, J. H., & Simon, H. A. (1987). Why a Diagmm is (Sometimes) Worth Ten Thousand Words. *Cognitive Science*, *11*(1), 65–100.

Lederman, S. J. (1991). Skin and touch. In *Encyclopedia of human biology* (pp. 51–63).

Lederman, S. L., Klatzky, R. L., Chataway, C., & Summers, C. D. (1990). Visual Mediation and the Haptic Recognition of Two-dimensional Pictures of Common Objects. *Perception & psychophysics*, *47*(1), 54–64.

Linvill, J. G., & Bliss, J. C. (1966). A direct translation reading aid for the blind. *Proceedings of the IEEE*, *54*(1), 40–51.

Loomis, J. M. (1981). Tactile pattern perception. *Perception*, *10*(1), 5–27.

Loomis, J. M., Klatzky, R. L., & Giudice, N. A. (2012). Sensory substitution of vision: Importance of perceptual and cognitive processing. In R. Manduchi & S.

Kurniawan (Eds.), *Assistive Technology for Blindness and Low Vision* (pp. 162–191). Boca Raton, Florida, USA: CRC.

Loomis, J. M., Klatzky, R. L., & Lederman, S. J. (1991). Similarity of Tactual and Visual Picture Recognition with Limited Field of View. *Perception*, (20), 167–177.

MacLean, K. E. (2008). Haptic interaction design for everyday interfaces. Reviews of Human Factors and Ergonomics, 4(1), 149-194.

Magnuson, C., & Rassmus-Grohn, K. (2003). Non-visual Zoom and Scrolling Operations in a Virtual Haptic Environment. *Eurohaptics*, *Dublin, Ir*(July), 6–9.

McGookin, D. K., & Brewster, S. A. (2006). MultiVis: Improving Access to Visualisations for Visually Impaired People. In *CHI'06 extended abstracts on Human factors in computing systems* (pp. 267–270). ACM.

Naked Eye. (2013). Retrieved from http://en.wikipedia.org/wiki/Naked_eye

Nees, M. A., & Walker, B. N. (2005). Auditory Interfaces and Sonification,In *The universal access handbook,* 32(1), 507–522.

Nees, M. A., & Walker, B. N. (2008). Encoding of Information in Auditory Displays : Initial Research on Flexibility and Processing Codes in Dual-task Scenarios Georgia Institute of Technology, 1820–1824.

NIST "Pins" Down Imaging System for the Blind. (2002). Retrieved from http://www.nist.gov/public_affairs/factsheet/visualdisplay.cfm

Owen, J. M., Petro, J. A., Souza, S. M. D., Rastogi, R., & Pawluk, D. T. V. (2009). An Improved , Low-cost Tactile " Mouse " for Use by Individuals Who are Blind and Visually Impaired. *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility (Assets'09)*, *October*(25-28), 223–224.

Parente, P., & Bishop, G. (2003). BATS: The blind audio tactile mapping system. In *Proc. of the ACM Southeast Regional Conference*.

Parkinson, M. (2013). Do-It-Yourself Billion Dollar Business Graphics: 3 Fast and Easy Steps to Turn Your Text and Ideas Into Graphics That Sell. Retrieved from http://www.billiondollargraphics.com/

Passini, R., & Proulx, G. (1988). Wayfinding without Vision: An Experiment with Congenitally Totally Blind People. *Environment and Behavior*, *20*(2), 227–252. doi:10.1177/0013916588202006

Perkins Museum. (2013). Retrieved from http://www.perkinsmuseum.org/

Petit, G., Dufresne, A., Levesque, V., Hayward, V., & Trudeau, N. (2008). Refreshable tactile graphics applied to schoolbook illustrations for students with visual impairment. *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility (Assets ’08)*, 89 – 96. doi:10.1145/1414471.1414489

Petrie, H., Schlieder, C., Blenkhorn, P., Evans, G., King, A., O’Neill, A. M., & Alafaci, M. (2002). Tedub: A system for presenting and exploring technical drawings for blind people. In *Computers Helping People with Special Needs* (pp. 537–539).

Phantom Omni. (2013). Retrieved from http://geomagic.com/en/products-landing-pages/sensable

Poppinga, B., Pielot, M., Magnusson, C., & K. Rassmus-Grohn. (2011). TouchOver Map : Audio-Tactile Exploration of Interactive Maps. *In Proceedings of the 12th international conference on Human computer interaction with mobile devices ACM, Stockholm, Sweden*, 545–550.

Raja, M. K. (2011). *The development and validation of a new smartphone based non-visual spatial interface for learning indoor layouts*. University of Maine.

Rastogi, R., Pawluk, T. V. D., & Ketchum, J. (2013). Intuitive tactile zooming for graphics accessed by individuals who are blind and visually impaired. *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, *21*(4), 655–63. doi:10.1109/TNSRE.2013.2250520

Rastogi, R., & Pawluk, D. T. V. (2013). Toward an Improved Haptic Zooming Algorithm for Graphical Information Accessed by Individuals who are Blind and Visually Impaired. *Assistive Technology*, *25*(1), 9–15. doi:10.1080/10400435.2012.680659

Rastogi, R., Street, W. M., & Pawluk, D. T. V. (2010). Automatic , Intuitive Zooming for People Who are Blind or Visually Impaired. *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility (Assets'10)*, 239–240.

Sanders, A. F., & Kappers, A. M. (2007). Haptically straight lines. *PERCEPTION-LONDON*, *36*(11), 1682.

Scaife, M., & Rogers, Y. (1996). External cognition: how do graphical representations work? *International journal of human-computer studies*, *45*(2), 185–213.

Schloerb, D. W., Lahav, O., Desloge, J. G., & Srinivasan, M. A. (2010). BlindAid : Virtual Environment System For Self-Reliant Trip Planning And Orientation And Mobility Training. In *Haptic Symposium* (pp. 363–370). IEEE.

Schmidt, M., & Weber, G. (2009). Multitouch haptic interaction. In *Universal Access in Human-Computer Interaction* (pp. 574–582). Springer Berlin / Heidelberg.

Schmitz, B., & Ertl, T. (2010). Making Digital Maps Accessible Using Vibrations. *ICCHP 2010*, *Part I, LN*(6179), 100–107.

Schweikhardt, W., & Klöper. (1984). Rechnerunterstützte aufbereitung von bildshirmtext-grafiken in eine tastbare darstellung. *Institut für Informatik Universität Stuttgart*.

Shimojima, A. (2001). The graphic-linguistic distinction. In *Thinking with diagrams* (pp. 5–27). Springer Netherlands.

Smiciklas, M. (2012). The Power of Infographics: Using Pictures to Communicate and Connect with Your Audience.

Su, J. (2010). *Considering Mobile Devices, Context Awareness, and Mobile Users*. University of Toronto.

Su, J., Rosenzweig, A., Goel, A., de Lara, E., & Truong, K. N. (2010). Timbremap: enabling the visually-impaired to use maps on touch-enabled devices. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services* (pp. 17–26). ACM. Retrieved from http://portal.acm.org/citation.cfm?id=1851606

Thinus-Blanc, C., & Gaunet, F. (1997). Representation of space in blind persons: vision as a spatial sense? *Psychological bulletin*, *121*(1), 20–42. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9064698

Vidal-Verdú, F., & Hafez, M. (2007). Graphical tactile displays for visually-impaired people. *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, *15*(1), 119–30. doi:10.1109/TNSRE.2007.891375

VoiceOver. (2013). Retrieved from www.apple.com/accessibility/voiceover/

Völkel, T., Weber, G., & Baumann, U. (2008). Tactile graphics revised: the novel brailledis 9000 pin-matrix device with multitouch input. In *Computers Helping People with Special Needs* (pp. 835–842). Springer Berlin / Heidelberg.

Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied*, *8*(4), 211–221. doi:10.1037/1076-898X.8.4.211

Walker, B. N., & Mauney, L. M. (2010). Universal Design of Auditory Graphs : A Comparison of Sonification Mappings for Visually Impaired and Sighted Listeners. *ACM Transactions on Accessible Computing (TACCESS)*, *2*(3), 1–16. doi:10.1145/1714458.1714459.http

Walker, S., & Salisbury, J. K. (2003). Large Haptic Topographic Maps: Marsview and the Proxy Graph Algorithm. *ACM Siggraph*, 83–92.

Wall, S. a., & Brewster, S. (2006). Sensory substitution using tactile pin arrays: Human factors, technology and applications. *Signal Processing*, *86*(12), 3674–3695. doi:10.1016/j.sigpro.2006.02.048

Williamson, J. R., Crossan, A., & Brewster, S. (2011). Multimodal Mobile Interactions : Usability Studies in Real World Settings. *Proceedings of the*

13th international conference on multimodal interfaces, *ICMI '11*, 361–368. doi:10.1145/2070481.2070551

Wilson, G., Brewster, S. a., Halvey, M., Crossan, A., & Stewart, C. (2011). The effects of walking, feedback and control method on pressure-based interaction. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, *MobileHCI*, 147–156. doi:10.1145/2037373.2037397

Wordweb dictionaries. (2013). Diagram. Retrieved from http://www.wordwebonline.com/

Wordweb dictionaries. (2013). Image. Retrieved from http://www.wordwebonline.com/

WORLD HEALTH ORGANIZATION. (2011). Visual impairment and blindness Fact Sheet. *http://www.who.int/mediacentre/factsheets/fs282/en/*.

Yatani, K., & Truong, K. N. (2009). SemFeel: a user interface with semantic tactile feedback for mobile touch-screen devices. *Proceedings of the 22nd annual ACM symposium on User interface software and technology (2009)*, 111–120. doi:10.1145/1622176.1622198

Yu, W., & Brewster, S. (2002). Comparing two haptic interfaces for multimodal graph rendering. *Proceedings 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. HAPTICS 2002*, 3–9. doi:10.1109/HAPTIC.2002.998934

Zeng, L., & Weber, G. (2010). Audio-haptic browser for a geographical information system. In *Computers Helping People with Special Needs* (pp. 466–473). Springer Berlin / Heidelberg.

Ziat, M., Gapenne, O., Stewart, J., Lenay, C., & Bausse, J. (2007). Design of a Haptic Zoom: levels and steps. In *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)* (pp. 102–108). IEEE. doi:10.1109/WHC.2007.36

BIOGRAPHY OF THE AUTHOR

Hari Prasath Palani was born in Hosur, India on 21th June, 1988. He was raised in Hosur and he graduated from Vijaya Vidyalaya matriculation higher secondary school in 2005. He then attended College of Engineering, Guindy in Chennai, which is one of the four constituent colleges of Anna University, Chennai, India and graduated in 2009 with a Bachelor of Engineering in Geo Informatics. He worked as a Software Engineer in CapGemini Private Limited in India from February 2010 to July 2011. He was then enrolled as a Master of Science graduate student in the Department of Spatial Information Science and Engineering at The University of Maine in Fall 2011. Hari Prasath Palani is a candidate for the Master of Science Degree in Spatial Information Science and Engineering from The University of Maine in December 2013.