

University of Nevada, Reno

**Augmenting the Spatial Perception Capabilities of Users Who Are Blind**

A Thesis Submitted in Partial Fulfillment  
of the Requirements for the Degree of Master of Science in  
Computer Science and Engineering

by

Alexander John Fiannaca

Eelke Folmer, Ph.D., Thesis Advisor

May 2014

© 2014 Alexander John Fiannaca

ALL RIGHTS RESERVED

**UNIVERSITY  
OF NEVADA  
RENO**

**THE GRADUATE SCHOOL**

We recommend that the thesis prepared  
under our supervision by

**ALEXANDER JOHN FIANNACA**

entitled

**Augmenting the Spatial Perception Capabilities of Users Who Are Blind**

be accepted in partial fulfillment of the  
requirements for the degree of

**MASTER OF SCIENCE**

Eelke Folmer, Ph.D. – Advisor

Nancy LaTourrette, M.S. – Committee Member

Jonathan Weinstein, Ph.D. – Graduate School Representative

Marsha H. Read, Ph.D. – Dean, Graduate School

May 2014

## ABSTRACT

People who are blind face a series of challenges and limitations resulting from their lack of being able to see, forcing them to either seek the assistance of a sighted individual or work around the challenge by way of an inefficient adaptation (e.g. following the walls in a room in order to reach a door rather than walking in a straight line to the door). These challenges are directly related to blind users' lack of the spatial perception capabilities normally provided by the human vision system. In order to overcome these spatial perception related challenges, modern technologies can be used to convey spatial perception data through sensory substitution interfaces. This work is the culmination of several projects which address varying spatial perception problems for blind users. First we consider the development of non-visual natural user interfaces for interacting with large displays. This work explores the haptic interaction space in order to find useful and efficient haptic encodings for the spatial layout of items on large displays. Multiple interaction techniques are presented which build on prior research [21], and the efficiency and usability of the most efficient of these encodings is evaluated with blind children. Next we evaluate the use of wearable technology in aiding navigation of blind individuals through large open spaces lacking tactile landmarks used during traditional white cane navigation. We explore the design of a computer vision application with an unobtrusive aural interface to minimize veering of the user while crossing a large open space. Together, these projects represent an exploration into the use of modern technology in augmenting the spatial perception capabilities of blind users.

This is dedicated to those fearless graduate students everywhere who are trudging through the black abyss that is thesis and dissertation writing.

May peace be with you.

And, of course, this is dedicated to my family and friends who kept me sane during each of my common and unfortunately predictable pre-conference deadline panic attacks.

## ACKNOWLEDGEMENTS

This work was supported by NSF Grant 1118074, a Microsoft Research Software Engineering Innovation Foundation Award (SEIF), and a Google Faculty Research Award.

## TABLE OF CONTENTS

|  |           |
|--|-----------|
| Abstract . . . . .   | i         |
| Dedication . . . . .   | ii        |
| Acknowledgements . . . . .   | iii       |
| Table of Contents . . . . .  | iv        |
| List of Tables . . . . .   | v         |
| List of Figures . . . . .  | vi        |
| <br>   |           |
| <b>1 Introduction</b>  | <b>1</b>  |
| <br>   |           |
| <b>2 Facilitating Spatial Interactions with Large Displays through Natural User Interfaces</b>           | <b>4</b>  |
| 2.1 Haptic Target Acquisition to Enable Spatial Gestures in Non-Visual Displays . . . . .                | 4         |
| 2.1.1 Overview and Objectives . . . . .  | 6         |
| 2.1.2 Background . . . . .   | 8         |
| 2.1.3 Evaluation: Unimanual 3D Target Acquisition . . . . .  | 12        |
| 2.1.4 Evaluation: Bimanual 2D Target Acquisition . . . . .   | 23        |
| 2.1.5 Discussion . . . . .   | 27        |
| 2.2 Applying Haptic Target Acquisition to Create Interactions for Children who are Blind . . . . .       | 31        |
| 2.2.1 Overview and Objectives . . . . .  | 32        |
| 2.2.2 Background . . . . .   | 33        |
| 2.2.3 Evaluation . . . . .   | 34        |
| 2.2.4 Discussion . . . . .   | 39        |
| 2.3 Potential Applications . . . . .   | 41        |
| <br>   |           |
| <b>3 Facilitating Difficult Spatial Navigation Tasks with Wearable Technology</b>                        | <b>44</b> |
| 3.1 HeadLock: A Wearable Interface for Helping Blind Individuals to Traverse Large Open Spaces . . . . . | 44        |
| 3.1.1 Overview and Objectives . . . . .  | 45        |
| 3.1.2 Background . . . . .   | 47        |
| 3.1.3 Design . . . . .   | 50        |
| 3.1.4 Implementation . . . . .   | 54        |
| 3.1.5 Evaluation . . . . .   | 60        |
| 3.1.6 Discussion and Future Work . . . . .   | 66        |
| <br>   |           |
| <b>4 Conclusion</b>  | <b>71</b> |
| 4.1 Future Work . . . . .  | 71        |
| 4.2 Summary . . . . .  | 72        |
| <br>   |           |
| <b>Bibliography</b>  | <b>74</b> |

**LIST OF TABLES**

|     |  |    |
|-----|--|----|
| 2.1 | Mean Corrected Search Time for Unimanual 3D Target Acquisition . | 21 |
| 2.2 | Average Search Time of Bimanual 2D Target Acquisition . . . . .  | 27 |
| 2.3 | Time Ratio Comparison of Bimanual to Unimanual Scanning Time .   | 38 |

## LIST OF FIGURES

|     |  |    |
|-----|--|----|
| 2.1 | Example of 3D Target Selection . . . . .   | 13 |
| 2.2 | <i>Multilinear Scanning</i> for 3D Target Acquisition . . . . .                        | 14 |
| 2.3 | <i>Projected Scanning</i> for 3D Target Acquisition . . . . .                          | 15 |
| 2.4 | Haptic Encoding Scheme for 3D Target Acquisition . . . . .                             | 16 |
| 2.5 | Typical Scanning Strategies for <i>Multilinear</i> and <i>Projected Scanning</i> . . . | 22 |
| 2.6 | <i>Split Scanning</i> for 2D Target Acquisition . . . . .                              | 23 |
| 2.7 | <i>Conjunctional Scanning</i> for 2D Target Acquisition . . . . .                      | 24 |
| 2.8 | Implementation of <i>Split Scanning</i> for ViAppleGrab. . . . .                       | 32 |
|     |  |    |
| 3.1 | The HeadLock Platform . . . . .  | 47 |
| 3.2 | A Visual Interpretation of the HeadLock Interface . . . . .                            | 50 |
| 3.3 | Explanation of the Veering Metric $\Theta(i)$ . . . . .                                | 65 |
| 3.4 | Examples of Typical Navigation Paths . . . . .   | 67 |

## CHAPTER 1

### INTRODUCTION

A particularly powerful and important sensory modality of humans is our ability to perceive the spatial relationships between objects in our environment. For instance, consider the task of clicking an icon on a computer desktop, or the task of crossing a large room to reach a doorway. To the average human, these tasks would be considered simple, due to the fact that they do not require much thought or effort to complete; however, these tasks are relatively complex due to the fact that each of these tasks require the person to be able to rapidly discern complex spatial relationships (e.g. where the icon is located on the screen relative to the cursor, or the position of the door relative to oneself). This complexity is handled by the human vision system which has evolved to be adept at rapidly discerning these relationships [56] (e.g. upon viewing the doorway, one understands their position relative to the door almost immediately without having to focus on perceiving the spatial relationship at hand). This feature of the human vision system is integral in many common human tasks such as navigation (e.g. walking and driving) and fine-grained motor interactions (e.g. interaction with the myriad of computing devices which have become tightly coupled with the human experience).

Clearly, spatial perception is an essential skill for humans. While research has shown that spatial perception is not solely dependent on visual perception, visual perception is a key factor in spatial perception due to the fact that vision acts as a primary spatial reference for humans [6]. Within this context, an interesting question arises considering the subset of the human population who lack vision: how does a lack of visual perception effect the ability of blind people to perform tasks

requiring spatial perception? Currently, there are approximately 25 million people with visual impairments in the US, including 1.3 million who are legally blind [4]. The increased prevalence of diabetes, macular degeneration, and an aging baby boomer generation is expected to double the number of individuals with visual impairments in the next decade [32, 25, 12]. Vision loss severely reduces the quality of life for blind individuals and has been associated with a number of problems, such as social isolation, depression, limited access to education, and fewer employment opportunities [15, 11, 13, 54, 35]. While, in the long term, leading causes for blindness such as retinal degeneration may be treated using promising medical advances, such as stem cell treatment [60] or retinal implants [17], in the meantime, it is imperative to provide blind people with access to affordable assistive technology that can improve their quality of life today [30]. This leads to a second question, which is more germane from an engineering perspective: how can the use of modern technology be leveraged to solve challenges related to the effect a lack of vision has on the spatial perception capabilities of blind users?

To address this question, this work describes two projects which explore challenges that can be addressed through augmentation of blind users' spatial perception capabilities with modern technology. The first project (Chapter 2, originally published in *Graphics Interface 2013* [20]) deals with the development of non-visual haptic displays and target acquisition within the context of these displays (i.e. finding and interacting with display elements through the perception of haptic feedback rather than visual feedback). The applicability of these haptic interaction techniques in creating usable interfaces for blind children is evaluated (this work is to be submitted to the *ACM Symposium on Spatial User Interaction 2014*). Finally, Chapter 2 concludes by discussing the potential for future work in developing natural user interfaces which utilize haptic feedback, in addition to other poten-

tial applications of this work. The second project (Chapter 3, to be submitted to the *ACM SIGACCESS Conference on Computers and Accessibility 2014*) deals with the challenges which blind users face when attempting to navigate across large open spaces. An indoor open space navigation application designed for a lightweight wearable platform (Google Glass) is presented and the efficiency and usability of the application is discussed. The design issues considered in the development of this application are discussed and an evaluation of the system with blind users is presented. Finally, Chapter 4 provides a summary of the research presented in Chapters 2 and 3, along with a discussion of future work related to augmenting the spatial perception capabilities of blind users.

## CHAPTER 2

### FACILITATING SPATIAL INTERACTIONS WITH LARGE DISPLAYS THROUGH NATURAL USER INTERFACES

This chapter presents an evaluation of two nonvisual techniques for developing three-dimensional unimanual haptic displays, an evaluation of two nonvisual techniques for developing two-dimensional bimanual haptic displays, and finally the subsequent application of the most efficient of the two-dimensional techniques in a non-visual haptic display-based game for blind children. The first section (2.1) includes the motivation for this work, the objectives and overview of the developed techniques, the related background research, and two quantitative studies evaluating the techniques. The second section (2.2) presents a study evaluating the real-world usability of one of these techniques in the form of a video game for blind children. The objectives, related work, and an evaluation of this real-world application are discussed. Finally, potential applications of our work in the field of haptic displays are proposed (Section 2.3).

#### **2.1 Haptic Target Acquisition to Enable Spatial Gestures in Non-Visual Displays**

Traditional user interfaces such as Windows or Mac OS were originally designed under the assumption that users would perceive the interface layout and interaction elements by visually inspecting a physical display and then provide input to the interface through keyboards and pointing devices such as the mouse [45]. Clearly, this style of interface design is not suitable for users with visual impairments due to the fact that vision is required for both providing input and perceiv-

ing output from these interfaces. Over the past two decades, an effort by software engineers to work around the major limitations of standard computer interfaces has resulted in a large array of accessibility applications being developed which allow blind users to explore visual interaction elements of interfaces using text-to-speech algorithms (e.g. the JAWs Screen Reader [1]). While these applications are widely used by people with visual impairments, they are only useful in the context of personal computing (e.g. desktop computers or smart phones). With the recent explosion of non-standard types of user interfaces (e.g. the Microsoft Kinect [40] and Sony Move [2] for gaming applications, and interactive wall and table displays for large interface applications [16]) it has become clear that the set of accessibility applications useful in personal computing contexts are not suitable for application across all interfaces. Therefore this project is motivated by the need to explore new and creative methods for making accessible interfaces which re-imagine the way users with visual impairments can interact with computer systems.

One method which has been explored extensively in the creation of new and creative accessible interfaces is sensory substitution [7]. Sensory substitution is a technique whereby information which is normally perceived using one sense, is encoded to be perceived using a separate sense (e.g. using sonification to let blind users “see” pictures [10]). In this work, we propose the creation of large interfaces for blind users which use haptics (i.e. the sense of touch) to output spatial interface information (the location of interface elements) to the user. Additionally, these new interfaces are developed to receive input via spatial gestures (body position and movement) which utilize the user’s sense of proprioception (the body’s ability to sense joint position and movement within joints). Therefore, the proposed interface substitutes the visual perception of standard computer interfaces with tactile perception (haptics) and proprioception , thereby creating a tactile-proprioceptive

display.

### 2.1.1 Overview and Objectives

Large interactive displays often make use of natural user interfaces (NUI) by allowing interaction through methods such as touch [63] or body movement [23, 42, 51]. NUIs take advantage of natural interaction patterns which humans have learned throughout life in order to create an intuitive invisible interface not requiring traditional interaction hardware such as a keyboard and mouse. The class of accessible NUIs previously developed for the blind use either sound or touch as input to non-visual displays. For this work, we specifically focus on expanding the state-of-the-art NUI techniques leveraging touch (haptics) as an output modality and body position and movement (gestures) as an input modality. Previous state-of-the-art work in non-visual haptic displays by Morelli et al. [21] (see section 2.1.2 for a complete discussion) presented a comparison of several unimanual techniques for creating two-dimensional non-visual haptic displays. These techniques were posed in terms of *target acquisition*, referring to the search for specific “target” interface elements. This investigation into haptic target acquisition aims at furthering the state-of-the-art in non-visual haptic display techniques by addressing two limitations of the original Morelli et al. technique:

1. First, the technique presented by Morelli et al. was limited in that it only explored target acquisition within the bounds of a two-dimensional display while a three-dimensional display could potentially allow for more expressive interactions. Therefore, the first objective of this work was to extend the Morelli technique from target acquisition in a two-dimensional display

to target acquisition in a three-dimensional display.

2. Second, the technique presented by Morelli et al. was limited in that it only allowed for unimanual interaction. By limiting interactions to a single hand, the interface forced targets to be interacted with in a serial fashion (i.e. only one target could be interacted with at any given time). Therefore, the second objective of this work was to extend the Morelli technique into bimanual interaction (remaining in two dimensions) potentially allowing for each hand to be simultaneously interacting with different target elements in the interface.

In order to address these objectives, four haptic target acquisition techniques were developed which allow users to sense the location of interface elements using their own arms via proprioception, the human ability to sense the position and orientation of their body parts (complementing existing accessible non-visual NUIs [23, 42, 51]). The use of proprioception in these techniques has the significant advantage of not requiring users to memorize the locations of objects in the non-visual display. Each technique leverages motion sensing controllers capable of being tracked with high precision in three dimensions and equipped with vibrotactors for generating feedback. Of the four techniques presented in this work, the first two address the first objective above, and the second two address the second objective above. The first set of techniques generate vibrotactile feedback indicating the direction which the user must move their controller in order to locate targets within a three dimensional space. The second set of techniques is similar in that it leverages vibrotactile feedback in order to encode the location of targets, however, it encodes locations within a two dimensional space and allows for more than one target location to be perceived at once. Both sets of techniques rely upon the user's sense of proprioception in order to convey the spatial locations

of targets. These techniques are unique from previous work in that they integrate both target sensing and target selection into a unified interface, thereby creating an *embodied NUI*. To evaluate these techniques, two studies were performed. The first, exploring unimanual interactions in a three-dimensional interface, and the second, exploring bimanual interactions in a two-dimensional interface. Bimanual interactions in a three-dimensional interface were not explored for reasons which are discussed in section 2.1.5.

## 2.1.2 Background

Natural user interfaces (NUIs) seek to capitalize on the innate abilities that users have acquired through interactions with the real world, by removing intermediary devices, such as a keyboard and mouse, so as to facilitate an invisible interface that is more intuitive to use. NUIs predominantly define novel input modalities [63], such as touch, gestures, and speech, but recent work [23, 42, 51] has explored gesture-based interaction without using a visual display.

Research in nonvisual NUIs initially focused on exploring how touch screen devices can be made accessible to users who are blind, for example, by providing speech feedback when users browse menus [22] or through the definition of custom gestures [33]. Several nonvisual NUIs have been developed for the purpose of increasing available input space of mobile devices without having to compromise their portability. These techniques typically appropriate the device itself into an input device using: (a) its orientation [36], (b) its relative position [26], or (c) gestures made with the device [38]. These techniques only allow for non-spatial interaction, such as scrolling through and activating items from lists.

Virtual shelves [42] is an input technique where users activate shortcuts by positioning a motion sensing controller within a circular hemisphere defined in front of the user. This motion controller is equipped with an integrated vibrotactor and is tracked using external cameras. Spatial interaction is limited to activating shortcuts and although users can learn and memorize the location of a particular shortcut using a vibrotactile cue, no spatial feedback is provided. The usefulness of this technique is evaluated with users with visual impairments in a second study [43], which demonstrated that proprioception can be used to create effective nonvisual NUIs.

Gustafson presents a so-called “imaginary interface” [23] where virtual objects are defined in a plane in front of the user, and whose positions can then be manipulated using spatial gestures. As the name suggests, this interface requires users to memorize the location of virtual objects, which may be challenging to perform, especially when multiple objects are present. An audio-based coordinate system is proposed to retrieve an object’s location, but this may be difficult to facilitate in mobile contexts. Familiar interfaces can be used in imaginary interfaces to avoid learning a new interface [24], but spatial interaction is limited to activating shortcuts.

In recent years various on-body computing approaches have been proposed that appropriate arms and hands into input devices [27, 50] but these are all vision based, as they use the user’s skin as a display surface by using micro projectors.

Recently, several techniques have been explored that appropriate the arm or hand into a non-visual display using a technique called a *tactile-proprioceptive display* [21]. Haptic feedback lends itself well to achieve eye and ear free interaction [57], but haptic feedback on most mobile devices is limited [44] as these typically

only feature a single rotary mass motor capable of providing on/off vibrotactile feedback with a fixed frequency and their latency limits the use of sophisticated drive signals [9]. Significantly larger information spaces that are capable of communicating larger amounts and richer types of information to the user can be facilitated through a combination of haptic feedback with proprioceptive information. For example, a navigation tool can be created by having users scan their environment with an orientation aware mobile device where a vibrotactile cue guides the user to point their arm holding the device at a specific point of interest. Target direction is then conveyed to the user using proprioceptive information using their own arm; effectively appropriating the human body into a display. A significant benefit of tactile-proprioceptive displays is that they can be created using hardware that is already present in most mobile devices [21].

Sweep-Shake [58] is a mobile application that points out geolocated information using a tactile-proprioceptive display. The user's location and orientation are determined using a compass and GPS. Vibrotactile feedback that encodes directional information (e.g. pulse delay: directional vibrotactile feedback using varying periods of delay time between feedback pulses of equal duration) renders points of interest. A study with four users found users could locate a 1D target on a 360° horizontal circle in 16.5 sec. Similarly to Sweep-Shake, PointNav [47] points out geolocated information, but accommodates users with visual impairments.

Ahmaniemi and Lantz [5] explored target acquisition using a mobile device that consists of a high precision inertial tracker (gyroscope, compass and accelerometer). Directional and nondirectional vibrotactile feedback (frequency and amplitude) were explored for rendering targets with varying sizes on a 90° horizontal line. A user study with eight sighted users found they were able to find targets

in 1.8 sec on average. Target sizes larger than  $15^\circ$  were most effective. Directional feedback was found to be more efficient than nondirectional feedback when target distance is furthest but it negatively affects finding targets that are close.

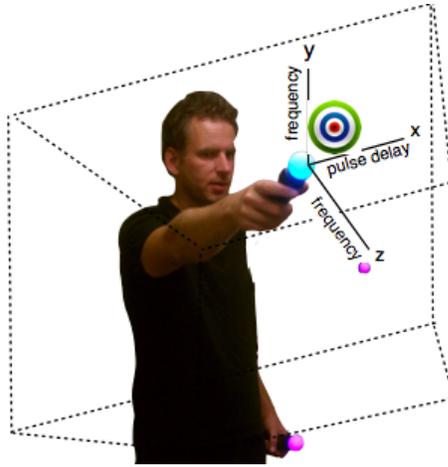
VI Bowling [51] is an exercise game for users who are blind and explores 1D target acquisition and gestured based interaction using a tactile-proprioceptive display. This game was implemented using a motion-sensing controller (Wii remote) where directional vibrotactile feedback (pulse delay) directs the player to point their controller at the location of the pins. Once the location of the pins is acquired, users hit the pins using an upper body gesture that resembles throwing a bowling ball. With a close-to-target window of  $38.6^\circ$  and a target size of  $7.2^\circ$ , a user study with six legally blind adults found that targets could be found on average in 8.8 sec and gestures were performed with an aiming error of  $9.8^\circ$ . In subsequent work [21], 2D target acquisition was explored using one arm. A tactile-proprioceptive display was implemented using a motion-sensing controller, whose position and orientation can be tracked using an external camera and inertial sensing. Its integrated vibrotactor is capable of providing directional vibrotactile feedback using pulse delay and frequency. A tactile-proprioceptive display was implemented whose size was defined by the reach of the user's arm, and defining a planar rectangular region in front of the user. Target acquisition was evaluated using an augmented reality space invader game, in which players scan to a random target defined in the display, and shoot it by pulling the controller's trigger. Two different target-scanning strategies were proposed. *Linear* scanning involves finding the target's  $X$ -coordinate using an on-target vibrotactile cue, upon which the direction to the target's  $Y$  is indicated using frequency modulation. *Multilinear* scanning uses directional vibrotactile feedback that is provided simultaneously on both axes where no pulse delay (continuous feedback) and maximum frequency

indicates the target. A between subjects study with sixteen users found multilinear scanning to be significantly faster than linear scanning. Targets were acquired on average in 7.7 sec (SD = 2.8). Additionally, a second study explored the users' ability to perform spatial gestures by having users touch a target using a thrust gesture. A user study with eight subjects using multilinear scanning found that users could perform a gesture in the direction the controller was pointing with an average aiming error of 20.74°.

### 2.1.3 Evaluation: Unimanual 3D Target Acquisition

Our first study extends prior work on 2D target selection [21] to 3D in order to investigate whether proprioceptive displays can facilitate a significantly larger interaction space. The size of the search space is therefore expanded from a plane to a frustum (the region of a pyramid remaining after removing the top section at a plane parallel to the pyramid's base), whose depth is defined by the length of the user's arm and the location of the camera (see Figure 2.1) used to track the motion sensing controller that facilitates this display. The back plane has a width that covers the entire horizontal range of the user's arm when it rotates at the shoulder joint and its height is restricted by the camera's resolution. This study is limited to rendering a single target at a time. Based on prior work [21], two different scanning strategies for 3D target acquisition were identified:

- **Multilinear** scanning uses directional vibrotactile feedback on each Cartesian axis of the frustum to indicate the target's location (see Figure 2.2). In [21], Folmer and Morelli demonstrated that users were able to scan to a target on two axes simultaneously and we naturally extend this approach to indi-

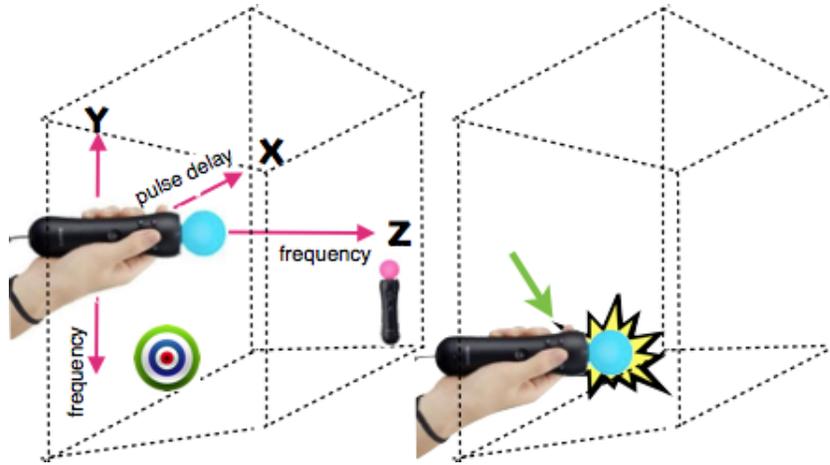


**Figure 2.1:** Example of 3D Target Selection

For 3D target selection, vibrotactile feedback guides the user to position their arm such that it touches a nonvisual object defined in a space in front of them, which then allows for manipulating this object using a spatial gesture. The dashed line indicates the frustum of the available search space.

cate a target's  $Z$ -value. Different types of haptic feedback are used on each axis to indicate the direction to the target. The user can find the direction to the target in one gesture by moving the controller in any of the 8 directions that lie between the  $X$ ,  $Y$ , and  $Z$ -axes. In theory, if the direction to the target on all axes is known, the user can scan directly to the target. This scanning type can be performed regardless of the initial start position of the controller.

- **Projected** scanning is a two-step target acquisition technique. Preliminary experiences with multilinear scanning revealed that scanning along three axes simultaneously was quite challenging to perform and required some amount of practice. To accommodate this limitation, we developed a simpler-to-perform two-step scanning technique. In previous work [21], it was found that subjects were able to perform a directed gesture in the direction their controller was pointing with reasonable accuracy, after finding a target in 2D. Projected scanning is based on these results and involves performing the fol-

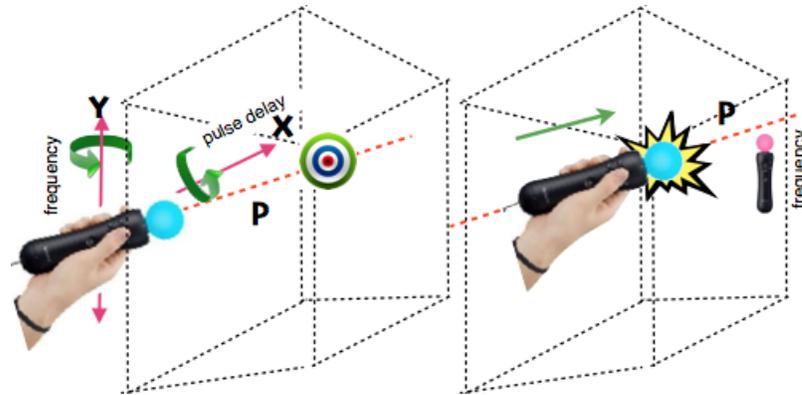


**Figure 2.2:** *Multilinear Scanning* for 3D Target Acquisition

Using multilinear scanning directional haptic feedback provided on the  $X$ ,  $Y$ , and  $Z$ -axes guides the user to select the target.

Following two steps: (1) with the controller initially outside of the frustum users rotate the controller along its own  $X$  and  $Y$ -axes as indicated using directional vibrotactile feedback until it points at the target; then (2) the user moves the controller along a projected axis ( $P$ ) that is defined by the controller's elongated shape and its current orientation (see Figure 2.3). Directional vibrotactile feedback indicates how far to move along the  $P$ -axis to select the target. Though projected scanning involves performing two consecutive steps, rotating the controller along its own  $X$  and  $Y$ -axes may be achieved faster than moving the controller along the coordinate axes of the frustum.

Each of these strategies use one controller for scanning the frustum and one controller for receiving tactile feedback corresponding to the  $Z$  or  $P$ -position of the controller scanning the frustum. Both strategies are therefore equivalent for evaluation. The goal of our user study is to evaluate which of the identified scanning techniques is faster.



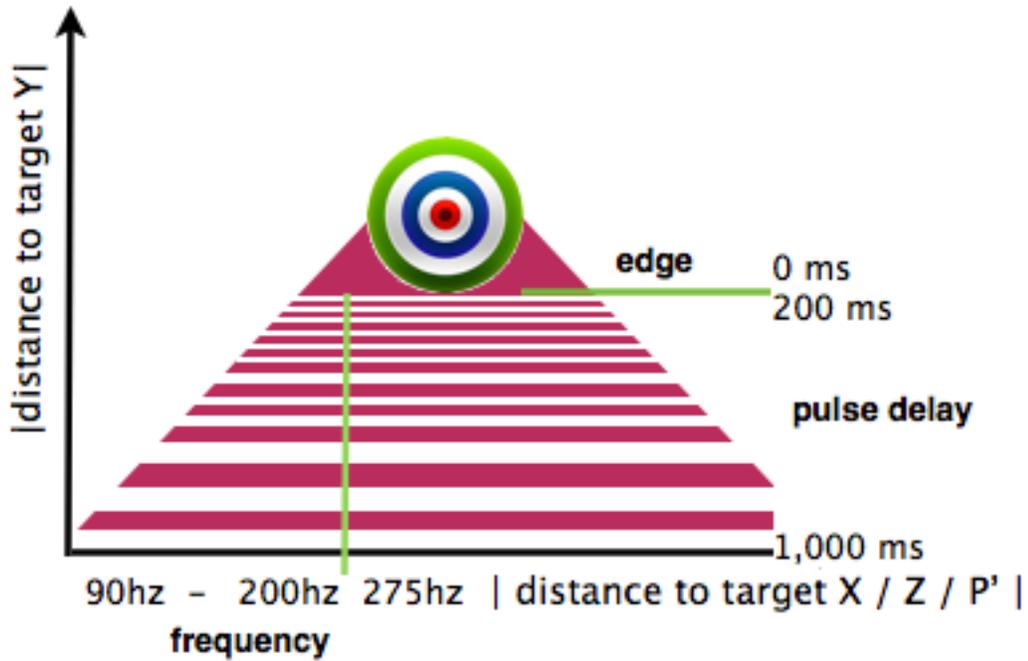
**Figure 2.3:** *Projected Scanning for 3D Target Acquisition*

Using projected scanning the user first places the controller outside the frustum and rotates the controller along its  $X$  and  $Y$ -axes until it points at the target. The user then moves the controller along the projected axis,  $P$ , as indicated using directional haptic feedback to select the target.

## Instrumentation

Our tactile-proprioceptive display is implemented using a commercially available motion sensing controller called the Sony PlayStation Move [2]. The controller's orientation is tracked using inertial sensing. It features an LED that serves as an active marker where the uniform spherical shape and known size of the LED allows for controller's position to be tracked in three dimensions with high precision ( $\pm 1.0$  mm error) using an external camera called the PlayStation Eye, which captures video at a resolution of  $640 \times 480$  at 60 fps. Directional vibrotactile feedback can be provided using pulse delay or frequency modulation with a range of 90 to 275 Hz.

The user scans the frustum with a controller held in their dominant hand, where pulse delay and frequency are used to indicate the direction of the target's  $X$  and  $Y$  coordinates. A Move controller is limited in only being able to provide two types of directional feedback, therefore both scanning techniques use a sec-



**Figure 2.4:** Haptic Encoding Scheme for 3D Target Acquisition

Haptic encoding of directional feedback on each axis, showing how haptic feedback changes in the frustum. When on target, pulse delay is zero and frequency is 275 Hz. Pulse delay increases linearly from 200 ms at the edge of the target to 1000 ms (max) at the edge of the frustum. Frequency decreases linearly from 200 Hz at the edge of the target to 90 Hz at the edge of the frustum.

ond controller in the user's non-dominant hand to indicate the target's  $Z$  position (multilinear scanning) or its  $P$  position (projected scanning) using frequency modulation (see figure 3).

$$f(d_{edge}) = \begin{cases} 90 + 110 \left(1 - \frac{d_{edge}}{x_{res}}\right) & : d_{edge} > 0 \\ 275 & : d_{edge} = 0 \end{cases}$$

$$p(d_{edge}) = \begin{cases} 200 + 3d_{edge} & : d_{edge} > 0 \\ 0 & : d_{edge} = 0 \end{cases}$$

A related study with 1D target selection using a haptic mouse [55] found that targets can be found significantly faster when the difference between the on-the-target cue and close-to-target cue is significantly increased ( $\geq 20\%$ ) at the border of a target. For target scanning on the  $Y$  and  $Z/P$ -axes frequency was modulated linearly based on the  $Y$  or  $Z/P$  distance to target with a maximum value of 200 Hz at the edge of the target, which was boosted to 275 Hz (maximum) when on target. For the  $X$ -axis, the pulse delay was 0 ms when on target and 200 ms at the edge of the target, which decreased linearly at  $3 \text{ ms/pixel}$  depending on the distance to the target's  $X$ -coordinate. The values used in our study were all informed by results from prior studies with tactile-proprioceptive displays [5, 21, 55].

Figure 2.4 illustrates the haptic encoding scheme for providing directional feedback for the various axes. For multilinear scanning when the user is on the target both controllers provide continuous (pulse delay of 0 ms) haptic feedback at 275 Hz. For projected scanning when the user points their dominant hand controller at the target, this controller provides continuous haptic feedback at 275 Hz and when the user selects the target, their other controller will provide haptic feedback at 275 Hz.

To compare both scanning types, a simple game was developed that involved destroying targets by selecting them. The faster players destroy a target the more points they score. The use of a game was motivated by the fact that games are considered powerful motivators [62], which may allow for measuring optimal performance in a user study. The game runs on a laptop and communicates with a PlayStation 3 to retrieve the current position and orientation of each Move controller and to adjust the vibrotactile feedback. As the controllers are wireless, there is a small latency in our feedback system but we found this lag to be minimal (not

noticeable by users), not having any significant effect on our study. To indicate to the player when the controller moves (multilinear) or points (projected) out of the frustum all vibrotactile feedback would be interrupted. Due to the camera's 4:3 aspect ratio, the user is more likely to move or point the controller outside of the  $Y$ -range, therefore frequency is used to render the target's  $Y$ -coordinate as this provides continuous feedback which makes being or pointing outside of the frustum more noticeable to a user than using pulse delay.

The controller's  $X$  and  $Y$  coordinates are reported in pixels and its  $Z$  coordinate in millimeters. Ahmaniemi [5] found target sizes larger than  $15^\circ$  for a 1D display of  $90^\circ$  to be most effective. For this study a target size of 100 pixels for  $X$ , 80 pixels for  $Y$  and 100 mm for the  $P/Z$ -axis (based on an average arm length of 60 cm) is used as to have a similar target size. A single target is defined at a random location within the frustum, excluding a 5% border to avoid scanning too close to the border. The use of random targets as opposed to fixed targets is motivated by the fact that it allows for assessing the user's ability to consecutively scan targets independent of the controller's initial position. Potential applications of our technique, such as an exercise game [52], typically also use random targets. If the controller is within the defined target area for 1 sec the target is destroyed, a sound effect is played, score is announced, and a new target is generated. Random background music is played to mask the sound of the vibrotactor.

## Participants

We recruited 16 participants (6 female, average age 28.5,  $SD = 3.42$ ). All subjects were right handed and none had any self-reported impairments in tactile perception or motor control. We measured players' height ( $M = 170.82$  cm,  $SD = 7.59$ )

and arm's length ( $M = 59.73$  cm,  $SD = 2.72$ ).

## **Procedure**

Participants were randomly assigned into two eight-person groups (A and B) where group A played the game using multilinear scanning and group B using projected scanning. A between-subjects study design is justified to avoid interference effects, e.g., when users have mastered one scanning technique this may disrupt their ability to learn and use another. User studies took place in a small office room. An observer was present during the study. Participants played the game using their dominant arm while standing. Due to players having different height and arm length, an application that is part of the Move SDK was used to calibrate the position of the player and to define the size of the frustum. Players were placed facing the camera at approximately 8 feet away (recommended optimal distance). Using a visual task that was displayed on the laptop's screen, players would be positioned as such to ensure the full horizontal range of their arm at the shoulder joint would match the horizontal resolution of the camera, e.g., the display ranges  $180^\circ$  by  $135^\circ$  (4:3 aspect ratio). The player would then stretch their arm and press the trigger on the controller to define the frustum's depth.

Once the position of the player was calibrated a piece of paper was placed under the player and we asked players to keep standing on it while playing the game. The laptop display was turned off to minimize distraction. Players were then instructed what the goal of the game was and how to play the game either using projected or multilinear scanning. Players familiarized themselves with the size of the frustum. For projected scanning, players were instructed to start scanning by placing the controller inside of the frustum, e.g., in front of their body, and to

then rotate the controller to be able to find the direction to the target. Finally, the users were instructed to move along the projected axis to acquire the target. For multilinear scanning players were taught how to find the direction to the target on all axes from any starting position. Players played our game briefly until they felt comfortable with scanning targets using their scanning technique. The game is then reset and users play the game until 20 targets are hit. All targets and controller states (positions and orientations) were recorded in a log file.

## Results

An analysis of collected data reveals significant variance in performance, which reduces after the eighth target. We consider this part of the learning phase and our analysis therefore focuses on the players' performance of acquiring the last twelve targets. The average search time for a target was 12.99 sec (SD = 6.90) for multilinear and 17.46 sec (SD = 4.28) for projected scanning.

Because targets were defined at random, the target distance from the initial start position could vary significantly between trials, though this variation reduces for a larger number of trials. For a more fair comparison, we compare search time corrected for distance. With the user's arm length the target distance on the  $X$  and  $Y$ -axes were converted from pixels to millimeters, which yielded corrected average search times of  $.102 \text{ mm/ms}$  (SD =  $.192$ ) for multilinear scanning and  $.075 \text{ mm/ms}$  (SD =  $.095$ ) for projected scanning. This difference was not statistically significant ( $t_{14} = -.769$ ,  $p > .05$ ) due to the large standard deviation for both the unimanual and bimanual cases. We then analyzed search performance for each axis by calculating the corrected search time based on the last time the target border was crossed in each dimension. In 14% (projected) and 10% (multilinear) of the targets

**Table 2.1:** Mean Corrected Search Time for Unimanual 3D Target Acquisition

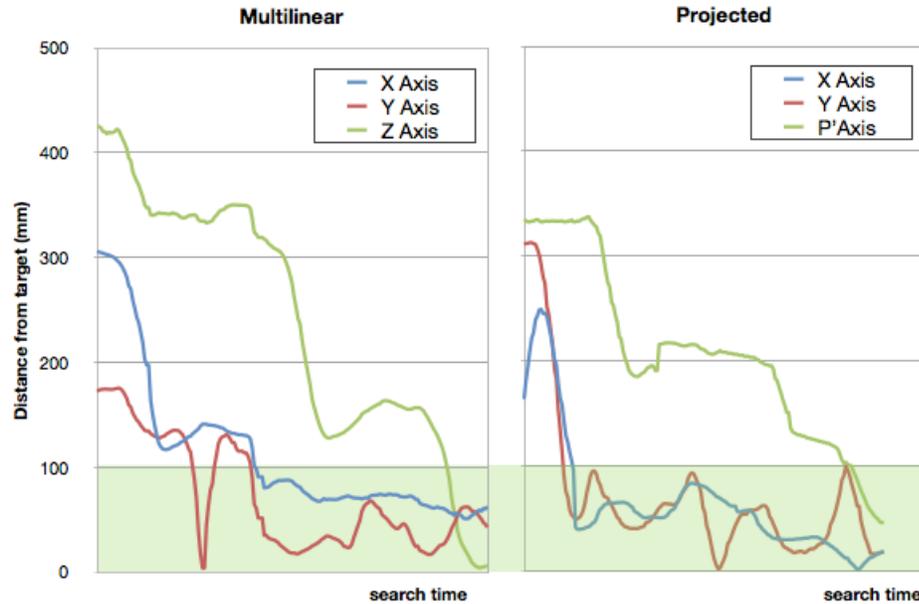
The time required for the user to locate the target positions was analyzed with respect to each axis and was corrected for by the users' arm lengths in order to calculate true distance scanning speeds (mm/ms) rather than pixel-based scanning speeds (pixels/ms)

| AXIS         | PROJECTED<br>mm/ms (SD) | MULTILINEAR<br>mm/ms (SD) |
|--------------|-------------------------|---------------------------|
| <i>X</i>     | .037 (.022)             | .066 (.072)               |
| <i>Y</i>     | .044 (.030)             | .059 (.058)               |
| <i>P / Z</i> | .034 (.014)             | .033 (.026)               |

the player was already within the target range for one specific axis, which resulted in significant outliers in corrected search time for that axis. Table 2.1 lists the results with the outliers filtered out.

A repeated measures ANOVA found no statistically significant difference between projected and multilinear scanning for corrected search times on all axes ( $F_{2,12} = .425, p > .05$ , Wilk's  $\lambda = .904$ , partial  $\varepsilon^2 = .096$ ). We then analyzed corrected search times for each axis within each scanning type, but no significant difference between axes for multilinear scanning ( $F_{2,21} = .799, p > .05$ ) or projected scanning ( $F_{2,21} = .286, p > .05$ ) was found.

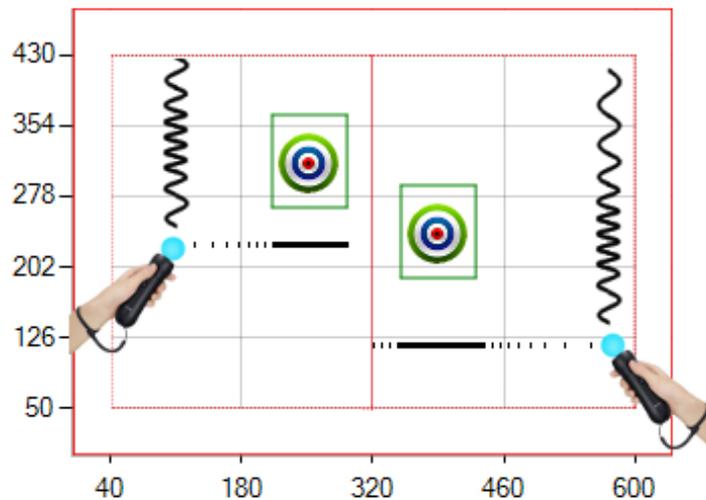
For each search we created trace graphs for the controller's position. Figure 2.5 shows a typical trace for each technique. For multilinear scanning, though users would perform the correct initial motion to find the direction to the target on all three axes (see Figure 2.5:left), they would typically scan to target's *X* and *Y* before scanning to it's *Z*-coordinate. Only for some of the last targets some users were able to scan to the target on all axes simultaneously. For projected scanning, we found that for larger target distance on the *P*-axis users would start to deviate on the *X* and *Y*-axis as following the projected axis *P* would become harder (see Figure 2.5:right).



**Figure 2.5:** Typical Scanning Strategies for *Multilinear* and *Projected Scanning*.

The red, blue, and green lines indicate the controller's movement along the *Y*, *X*, and *Z/P* axes. In both cases, the user located the *X* and *Y* coordinates of the target first, and then the *Z/P* coordinate.

Logs further show that for multilinear scanning users spent an average of .43 sec (SD = .04) searching for a target outside of the frustum, while for projected scanning users pointed their controller outside the frustum 2.18 sec (SD = 1.48) on average per target. This difference was statistically significant ( $t_{7.010} = 3.866$ ,  $p < .05$ ). For projected scanning this data was corrected for when the user starts scanning for the target and when the controller was outside the frustum. Closer analysis found that for multilinear scanning this sometimes occurred for targets close to the frustum's edge where users would move the controller through one of frustum's sides when scanning for the target's *Z*-coordinate. For projected scanning, pointing the controller outside the frustum predominantly occurred on the *Y*-axis, when users were acquiring the direction to the target.



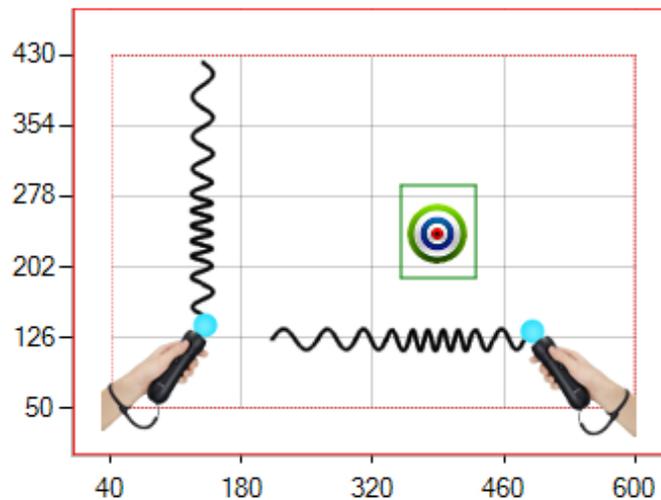
**Figure 2.6:** *Split Scanning* for 2D Target Acquisition

Using split scanning the user searches the left half of the display with their left hand and the right half of the display with their right hand. This effectively decreases the size of the search space per hand.

#### 2.1.4 Evaluation: Bimanual 2D Target Acquisition

Our second study extends prior work on 2D target selection [21] but extends it so as to explore bimanual use. While in the first study, only one of the two controllers was used to scan the search space, in this study we use both controllers for scanning. The goal of this study was to determine whether both arms could be used for target acquisition. Using both arms could possibly allow for faster target acquisition. The size of the search space consists of a vertical plane defined in front of the user, the size of which is determined by the length of the user's arm and the resolution of the camera. Two scanning strategies for bimanual 2D target acquisition were defined:

- **Split** scanning divides the available search space into two equal sized re-



**Figure 2.7:** *Conjunctural Scanning* for 2D Target Acquisition

Using conjunctural scanning the user only receives  $Y$  feedback on the non-dominant hand controller, and  $X$  feedback on the dominant hand controller.

gions, where each controller implements a display for each region (see Figure 2.6). We use the same haptic encoding scheme as in a previous study [21], i.e., multilinear scanning where different types of haptic feedback modulation are used to indicate the direction to the target on each axis, allowing for the user to search for the target's location on both axes simultaneously.

- **Conjunctural** scanning uses a single display where each controller indicates one of the targets' coordinates using haptic feedback modulation. Users can find the target's  $X$  and  $Y$  coordinates using one controller for each axis (see Figure 2.7). This is an asymmetric task performed synchronously. Upon finding the coordinates, the target can be selected by moving one controller to the intersection of the found  $X$  and  $Y$  coordinates.

The choice for these specific scanning strategies was motivated as they each evaluate one potential improvement in performance of bimanual operation. Split

scanning may be faster as each controller implements a smaller display that can be scanned through faster. Conjunctive scanning provides insight into whether users can use both controllers at the same time, which may be faster than using a single controller to find both targets' coordinates. Though bimanual operation could allow for multi target scanning, we restrict our study to single targets so that the identified strategies are equivalent for evaluation.

### **Instrumentation**

We used the same setup as for the first user study (see Section 3.1). For split scanning, pulse delay and frequency are used to indicate the direction of the targets  $X$  and  $Y$ . A short cue indicates in which region the target is rendered. For conjunctive scanning, we use frequency modulation to indicate the target's  $X$ -coordinate on the dominant hand controller and the  $Y$ -coordinate on the non-dominant hand controller. The same values as for the first study were used for frequency modulation, pulse delay modulation, and the target size. The game for study 1 was adapted to facilitate 2D scanning and was used to evaluate both scanning strategies. While targets were defined at random in the first study due to the relatively large search space, the restricted search space in this study made random target locations impractical. We therefore defined a grid in which targets appeared in order to ensure an even distribution of targets throughout the search space. The appearance order of target locations within this grid were randomized between trials.

## Participants

We recruited 16 participants (5 female, average age 25.7, SD = 3.53). All subjects were right handed and none had any self-reported impairments in tactile perception or motor control. We measured players' heights ( $M = 177.28$  cm, SD = 11.42) and arm lengths ( $M = 57.25$  cm, SD = 4.44).

## Procedure

Participants were randomly assigned into two eight-person groups (A and B) where group A played the game using split scanning and group B using conjunctive scanning. For the same reason as the first study, a between-subjects study design is justified to avoid interference effects, e.g., when users have mastered one scanning technique this may disrupt their ability to learn and use another. We used a similar procedure as in the first study (see Section 2.1.3). After calibration, both groups received instructions on how to scan for a target using their prescribed scanning technique. Users played the game briefly until they felt comfortable performing their scanning technique. The game was then be reset and users played the game until 20 targets were hit. All targets and all positions and orientations of the controller were recorded in a log file.

## Results

The average target search time was 7.07 sec (SD = 1.90) for split scanning (A) and 11.04 sec (SD = 3.15) (B) for conjunctive scanning. This difference was statistically significant ( $t_{14} = -2.854$ ,  $p < .05$ ). Unlike the first study, we did not analyze

**Table 2.2:** Average Search Time of Bimanual 2D Target Acquisition

| HAND  | SPLIT<br>ms/px (SD) | CONJUNCTIONAL<br>ms/px (SD) |
|-------|---------------------|-----------------------------|
| Left  | 6.00 (2.32)         | 6.72 (2.01)                 |
| Right | 4.64 (1.92)         | 7.41 (2.16)                 |
| Both  | 7.07 (1.90)         | 11.04 (3.15)                |

the search time corrected for distance, since for split scanning subjects would often lower the hand holding the controller that was not active. Upon becoming active, this would lead to very large distances causing an unfair comparison between scanning techniques. Table 2.2 lists the target search time per hand for each technique. For conjunctive scanning, only a few users were able to scan with both controllers along the axes at the same time, where the rest would scan for the coordinates sequentially. For conjunctive scanning, we therefore calculate search time from the moment the user begins scanning with that controller. No significant difference in search time was found between hands for split scanning ( $t_{14} = -0.657$ ,  $p > .05$ ) or conjunctive scanning ( $t_{14} = 1.276$ ,  $p > .05$ ) demonstrating that users were just as proficient with either hand. Logs further show that for split scanning users spent an average of .45 sec (SD = .32) searching for a target outside of the display and for projected scanning this was 2.59 sec (SD = 1.65) per target. This difference was statistically significant ( $t_{7.51} = -3.363$ ,  $p < .05$ ).

### 2.1.5 Discussion

Both studies show that tactile-proprioceptive displays are not particularly fast, but they do allow for communicating a type of information (a 2D/3D point in a space in front of a user) with a significantly large spatial resolution that would otherwise be difficult to communicate using conventional types of haptic feedback, such as

tactons [28].

Our first user study revealed no significant difference in performance between multilinear and projected scanning, which contradicts the previous study with 2D scanning [21]. Our study identified the advantages and disadvantages of using each scanning technique.

Projected scanning allows for more quickly finding the direction to the target (as rotating the controller is faster than moving the controller in the frustum), but users spend significantly more time searching for the target outside of the frustum as with multilinear scanning moving the controller within the frustum is physically constrained by the user's arm and therefore less likely to happen. Multilinear scanning allows a user to scan to the target directly, but in our study we rarely observed users being able to do this and instead they followed a two-step process similar to projected scanning where users first acquired  $X$  and  $Y$  simultaneously and then proceeded scanning the target's  $Z$ .

Similar to preliminary experiences, for some users, scanning along three axes simultaneously turned out to be too difficult to perform, which could indicate that we have run into a human limitation, as this was easier to perform for 2D scanning [21]. On the other hand a few users were observed to be able to do this for the last targets, which could indicate that it could also be a matter of practice. Due to the feedback variability limits of the vibrotactor in the Sony Move controller (only pulse delay and frequency can be varied, while amplitude cannot), we were required to use a second controller in the other hand to indicate the target's  $Z$ -coordinate. Users may have found it difficult to combine and interpret stimuli from both hands into a single sensation. However, if a third type of directional vibrotactile feedback is used, i.e., amplitude, simultaneous provision of three types

of haptic feedback using a single device could introduce the effect of frequency being perceived as the most dominant and this typically drowns out amplitude perception [46]. Therefore, for 3D target selection using two controllers may actually be the most optimal, as this interference problem will not occur.

For targets defined close to the frustum's back plane, projected scanning seemed more difficult to perform as users would easily deviate from the projected axis, which often requires the user to move the controller outside the frustum as to reacquire the target. The length of the user's arm and the resolution of the camera define the size of the frustum. As a result, the search space on the  $X$  and  $Y$ -axes is almost twice the size of the search space on the  $Z$ -axis, which does not really allow for a fair comparison of search performance between axes. For such a comparison, a uniform search space would be more suitable, but then users are more likely to move or point outside of the frustum. Though the arm length between users did not vary significantly, the volume of the frustum defined by the user's arm may vary significantly.

Our second study showed that the average search time for split scanning was 7.07 sec. In prior research [21], a scanning time of 7.7 sec was found, potentially indicating that using both hands is slightly faster. These findings are consistent with previous work on bimanual use of pointing devices [41], though the performance gain we found is much smaller. To an extent this is explained by the fact that users would typically lower the arm that was not actively scanning for a target, so it took longer to find a target, as it took some time for the user to raise their arm again. For conjunctive scanning, only a few users were able to scan with both controllers along each axis at the same time. This could have been a matter of convenience, though users were demonstrated how to do this, or it could

indicate that this was very challenging to perform, and that users would require more practice to master this. The primary reason why our research did not include a third study evaluating bimanual interaction in a three dimensional display, was the inability of users to effectively search with both hands simultaneously. Extending this to a three dimensional interaction would only provide a greater cognitive load, and very likely not facilitate bimanual interactions any more efficiently than was observed in a two dimensional display.

Reflecting on all studies with target acquisition using tactile-proprioceptive displays in 1D, 2D and 3D space, we observe the following. Ahmaniemi [5] found an average target search time of 1.8 sec for a 90° 1D display. In previous work [52], we found a scanning time of 7.7 sec for a 180° by 135° display. In our 3D scanning study we found a search time of 12.99 sec (multilinear) for a 180° by 135° by arm length display. Extrapolating these results to match the size of each display, we can observe that search time nearly doubles each time an axis is added to the search space, e.g., 3.6 sec ( $x$ )  $\rightarrow$  7.7 sec ( $x, y$ )  $\rightarrow$  12.99 sec ( $x, y, z$ ). However, because axes were not exactly of equal sizes between studies this finding should be further substantiated in subsequent research.

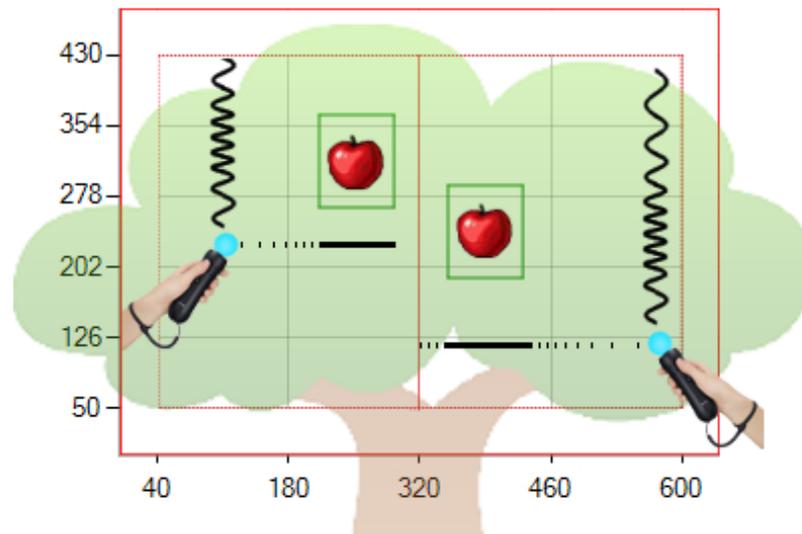
Our target selection studies were constrained to conveying a single target at a time, though for some applications, such as exergames [61], the rendering of multiple targets at the same time may be required, as to stimulate greater physical activity. For 3D target selection, rendering multiple targets is limited by technical constraints of the controller used. For 2D target selection users should be able to use both controllers to select two targets using split scanning (see Section 2.2). For an exergame, using this technique you could simulate punching targets in 2.5D, where targets are defined on the surface of a sphere whose size is defined by the

length of the user's arm.

Finally, our tactile-proprioceptive display relies on an external vision system to determine the 3D position of the user's controllers, but for mobile contexts, where ear and eye free interaction is most useful, we believe our display could be implemented using a wearable camera. Recent advances in 3D cameras may allow for the user to wear a small camera on their chest, allowing for accurate arm tracking where directional haptic feedback can be provided using a miniature haptic device [39]. This approach is different from how we evaluated our display as targets were defined in the frame of a fixed camera. Using a wearable camera, targets will be relative to a user and may be subject to interference from walking and moving.

## **2.2 Applying Haptic Target Acquisition to Create Interactions for Children who are Blind**

While our exploration into haptic display techniques in Section 2.1 exposed several design considerations for the creation of haptic proprioceptive displays for the blind, it only considered a comparison of the involved human performance factors between several proposed display techniques, and did not evaluate the usability of the considered techniques with the actual target population. In order to complete this research, we followed up the research presented in 2.1 with a study aimed at determining if the most efficient 2D display technique discussed above is usable by blind children.



**Figure 2.8:** Implementation of *Split Scanning* for ViAppleGrab.

The non-visual interaction space in front of the user is divided into a region for each of the user's hands. Vibrotactile feedback from the controllers encodes the target's X position with pulse delay and the target's Y position with frequency modulation.

### 2.2.1 Overview and Objectives

In this application of the previously discussed work (Section 2.1), we implement the most efficient of the explored 2D haptic target acquisition techniques and evaluate it within the context of allowing for children who are blind to play a gesture based game on a large display. We also investigate whether children who are blind can perform bimanual 2D scanning to select two targets at the same time as was proposed in the conclusions for Section 2.1.

In the previous section, we found that the most efficient scanning strategy in 2D target acquisition within the large non-visual display was the so-called *Split Scanning* strategy which involved splitting the interaction space into two sub-regions: one scanned by the right arm and one scanned by the left arm. Pulse delay and frequency were used to encode and communicate the X and Y positions of targets

to the user (see Figure 2.8).

For this study, we implement this proposed scanning technique in a computer game called VIAppleGrab. VIAppleGrab is themed to use the Split Scanning technique to direct the player to find apples (targets) hanging on an imaginary 2D tree in front of the user (the interaction space). Both music and audio feedback are included so as to increase the users' potential interest in the game. Additionally, the users score points based on the speed with which they are able to find targets, creating a competitive game atmosphere, which acts as a motivator incentivizing the children to play the game as best as possible. This study evaluates both the ability for children who are blind to interact with the proposed interface from Section 2.1 and whether the scanning strategy has the potential to be scaled up, i.e. finding multiple targets simultaneously.

### **2.2.2 Background**

Over the past several years a number of techniques have been developed to allow for gesture based interaction using non-visual means. Touch screens have been made accessible to users who are blind using: (1) speech feedback when browsing menus [22], (2) custom multitouch gestures that provide audio feedback [33], and (3) software overlays that convert 2D content into linear content using edge projection and speech output [34]. For upper body gestures, these techniques are difficult to apply seeing as, unlike touch screens, these gestures are made in the air and are not delimited by a physical surface.

For upper body gestures, a number of non-visual spatial interfaces have been developed. VI Bowling [51] is an exercise game for individuals who are blind

where players sweep a motion sensing controller to find the location of bowling pins as indicated using haptic feedback. This enables players who are blind to throw a virtual bowling ball at the sensed location. Virtual shelves [43] is a non-visual input technique where users who are blind can trigger shortcuts by positioning a motion sensing controller within a circular hemisphere in front of them. Imaginary interfaces [23] is a mobile interaction technique that defines virtual objects in a plane in front of a sighted user, which can then be manipulated using gestures. Airpointing [14] is a framework for non-visual spatial target acquisition. In [14] different 2D/3D pointing techniques are evaluated where subjects initially memorize targets using visual feedback. All these techniques largely rely on the user's visuospatial memory to memorize the location of objects, which may be challenging when there are many objects present and due to the fact that spatial memory tends to fade over time. To address this issue, Section 2.1 presented an interaction technique that appropriates a users arm using haptic feedback, provided using a motion sensing controller to point out the location of non-visual objects; thus enabling spatial interaction. User studies with sighted users evaluated different scanning strategies for acquiring a target in 2D and 3D.

### **2.2.3 Evaluation**

#### **Setup**

The large non-visual display evaluated in this study uses the same hardware as described in Section 2.1: a Sony PlayStation 3, two Sony Move controllers, the Sony Eye camera, the Sony Move.Me server application, and a laptop. As previously described, frequency is varied linearly between 91 and 200 Hz with respect to the

controller's distance to the target's  $Y$  location. Frequency is boosted from 200 Hz at the edge of the target region to 271 Hz when the controller enters the actual target  $Y$  region so as to facilitate an On-The-Target cue [55] that can significantly improve target acquisition. Likewise, pulse delay is zero when the controller enters the target's  $X$  region, 200 ms at the target's edge, and varied linearly by  $3 \text{ ms/px}$  with respect to the controller's distance to the target region. The non-visual display and the placement of targets within the display was defined in the same manner as in Section 2.1.

The game is organized into four levels, during each of which the user is required to obtain five targets per active controller (10 targets total per level during bimanual scanning, and 5 targets total per level during unimanual scanning). Background music is played in each level for aesthetic purposes. Additionally, audio cues are utilized for conveying game info to the players. Upon obtaining a target, a positive sound is played and the number of points scored on that target is spoken via speech synthesis. The user is also updated on the number of targets remaining in the level. If the user attempts to collect a target when the controller is not within the target region, a negative sound is played and the user is allowed to continue to search for the target.

The game was run on the laptop and the Sony Move.Me server application was used to gather controller information and send it to the laptop. The vibrotactile motors in the Move controllers allowed for haptic feedback, while the laptop speakers allowed for audio feedback.

## Participants

We recruited 8 participants between the ages of 11 and 15 (2 female, average age 12.5,  $SD = 1.2$ ) at a summer sports camp for children who are blind. One subject was left handed. All children were legally blind with no functional vision. None had any self-reported impairments in tactile perception or motor control. We measured users' heights ( $M = 148.9$  cm,  $SD = 12.12$ ) and arm lengths (Right:  $M = 26.11$  cm,  $SD = 1.29$ , Left:  $M = 26.21$  cm,  $SD = 1.49$ ).

## Procedure

Before each subject began the study, the subject was positioned approximately 8 feet from the Sony Eye camera and the system was calibrated to the length of the user's arms so that the interaction space was properly fit to each individual. The subject stayed in this position for the remainder of the study so that the system did not need to be recalibrated between phases of the study. To help children stay at this location we placed a piece of paper under their feet. The study was conducted in 2 main phases: unimanual and bimanual scanning. Each phase was preceded by a warm-up stage, conducted in an identical fashion to following study phase, in which the users were allowed to learn the scanning technique and become familiar with acquiring targets with one or both controllers. During the warm-up stage, the users were instructed in the scanning strategy and how to interpret the haptic and audio feedback. Additionally, users were instructed that the goal of the game was to gather apples (obtain targets) as quickly as possible. Once the user indicated that they were comfortable with the system, the game was reset in order to begin the main phase.

For the unimanual phase, the user obtained 20 targets with their dominant arm controller. All users received the same sequence of targets with the single exception that the locations of the left dominant arm targets were a mirror image (across the midline of the interaction space) of the right dominant arm controller's. Likewise, for the bimanual phase, the user obtained 20 targets with each of their arms. Targets were presented in pairs such that the user was required to obtain both the left and right targets before either of the next two targets would be displayed.

During execution of the study, results were recorded for each user, broken down into each phase of the study and each individual target within each phase. Controller positions and states were recorded every 100 ms. This data was stored in XML results files for post-processing analysis.

## Results

From the original group of eight participants, three users decided not to participate in the bimanual phase of the study due to the strenuous nature of target acquisition activity. Five participants completed all phases of the study to their full extent. Initial analysis of the users' results aimed at determining the scalability of the interaction system by comparing the unimanual phase to the bimanual phase of the study. On average, users required 12.00 sec (SD = 4.07) of scanning time to acquire targets in the unimanual phase and 22.83 sec (SD = 12.65) of scanning time to acquire targets in the bimanual phase. While there is a clear increase in time required to acquire targets between the two phases, large variability in the results data indicated by the significantly large standard deviation in the results set, resulted in no statistically significant difference being reported by a paired-samples t-test ( $p > 0.05$ ).

**Table 2.3:** Time Ratio Comparison of Bimanual to Unimanual Scanning Time

| Participant | Age | Bimanual to Unimanual Ratio |
|-------------|-----|-----------------------------|
| 1           | 13  | 2.311                       |
| 2           | 13  | 3.185                       |
| 3           | 14  | 2.503                       |
| 4           | 15  | 3.909                       |
| 5           | 12  | 1.531                       |

In order to account for the varying distances between targets which may have affected the required search time for targets, the results data for scanning time was corrected for the initial distance from each target to the associated controller, resulting in the velocity in pixels per second with which the users were able to scan for targets. Interestingly, this correction for distance revealed a statistically significant difference between the two phases, with bimanual scanning being significantly slower than unimanual scanning ( $t_3 = 4.962, p = 0.016$ ). These results were further corrected by the length of the users arms in order to find a true distance velocity ( $\text{mm/ms}$  rather than  $\text{px/s}$ ). The unimanual phase had an average true distance scanning velocity of  $0.073 \text{ mm/ms}$  ( $\text{SD} = 0.010$ ) and the bimanual phase has an average of  $0.028 \text{ mm/ms}$  ( $\text{SD} = 0.015$ ). A paired samples T-test found that the bimanual phase was significantly slower than the unimanual phase ( $t_3 = 5.093, p = 0.015$ ).

The relationship between the two phases was further investigated by determining the time ratio for each of the users between their unimanual and bimanual phases (see Table 2.3). This time ratio compares the difference between the time required for the user to obtain a single target in the unimanual phase and a pair of targets in the bimanual phase. A ratio of 1.0 would indicate perfectly simultaneous scanning of both arms in the bimanual phase, while a ratio of 2.0 or greater would indicate that the user scanned for each of the targets in a serial fashion (i.e. one after the other).

Finally, in order to determine the effect of scaling up the system with respect to the users' non-dominant arm performance, both the scanning time and controller velocity were compared between each user's arms for the bimanual phase. No significant differences in time or velocity were found between the users' dominant and non-dominant arms ( $p > 0.05$  in both cases).

## 2.2.4 Discussion

*Health benefits.* Previous studies have addressed both the concepts of large non-visual interfaces and potential scanning strategies, but have largely left open the question of the usability of these systems for children who are blind. The fact that three of the eight participants in this study were unable to complete the bimanual phase has a major implication towards the usability of large non-visual interfaces. It is clear that use of the system can be strenuous over extended use periods. Since this interface can be used to develop exergames, this may not be a weakness but a strength of the system. The fact that the interaction can be strenuous indicates that there could be some level of exercise value associated with this type of interaction, accommodating exergaming for children who are blind; however, the results from the study at hand are not substantial enough to justify this hypothesis, indicating that further research should examine the potential for applying this work in exergames for the blind.

*Limitations.* The primary limitation in our work is that we did not perform a qualitative analysis of our system. The study was designed in this manner so as to prevent the users from evaluating the game, rather than the actual interaction technique itself. The focus of this study was on the quantitative performance of

our approach; however, the fact that three users could not finish the study severely limited the strength of the conclusions which can be drawn from the quantitative data collected. That being said, there are several general conclusions which can be drawn from the observed data as discussed in the following section.

*Human limitations.* The significant difference observed between the controller velocities in the unimanual phase to the bimanual phase is not surprising in that the complexity of the target acquisition task is significantly increased when going from interpreting haptic feedback for a single target to two simultaneous targets. Hence, the decrease in velocities is a direct result of the increase in task difficulty. A better way to understand this interaction scaling issue is to examine the time ratios from Table 2.3. In Table 2.3, it can be seen that users fall into one of three cases: ratios greater than 3.0, ratios slightly greater 2.0, and ratios below 2.0. While those users with ratios above 3.0 clearly struggled with understanding or interpreting the two independent streams of haptic input at the same time, the users with ratios just above 2.0 achieved linear scaling of the technique. This linear scaling indicates that these users approached the increase in targets as a serial problem, first finding one target, and then the other. The final user was able to attain a ratio of less than 2.0. This indicates that this user was able to scan for both targets simultaneously, thereby accurately interpreting multiple streams of haptic feedback at the same time. It is possible that this variability in performance could be an effect of the rate of learning of each of the users. Clearly, these results, in addition to the fact that no differences were found between the performance of user's arms, indicate that the implemented scanning strategy has the potential to be scaled up to two simultaneous targets however, due to the variability in user performance, two targets is most likely an upper limit on the scalability. Further work should expand upon this analysis with a larger test group in order to verify the general

conclusions drawn here.

Altogether, this study demonstrates that children who are blind can effectively interact with large non-visual displays using the split scanning technique, however the efficiency of this interaction varies greatly between users. Additionally, the interaction method is scalable to multiple targets, although presenting multiple targets at once appears to be cognitively challenging for some children.

*Future work.* An interesting question which should be answered in future work is whether or not extended practice of this system generates positive health effects in the children who originally struggled with the physicality of the interface. Additionally, it would be interesting to adapt this interface into other use domains outside of the area of exergames, such as potentially creating a spatial information interface which helps children who are blind to learn spatial information about rooms or buildings which they have never entered before. Finally, it would be interesting to perform a longer term study, allowing the users more time to practice and become accustomed to the interaction, as well as providing an opportunity for a qualitative analysis of the technique's integration into an exergame, and a quantitative analysis of potential health benefits of regular use of the interface.

## 2.3 Potential Applications

In addition to complementing existing nonvisual mobile spatial interfaces [23, 42] (as discussed in Section 2.1.2), useful applications of our tactile-proprioceptive display techniques could include developing whole-body exercise games for individuals who are blind [51] (in a similar fashion to the game described in Section 2.2), as this typically involves punching and kicking virtual targets that are defined

in a 3D space around the player. Though scanning for a 3D target with the arm stretched out is some form of physical activity, it is unlikely to engage a player into levels of physical activity that are high enough to be considered healthy. Targets could be defined in 2D and the size of the display could be reduced to allow for rapid gestures. Additionally, as was indicated in Section 2.2, searching for two targets simultaneously was observed to be strenuous for several children, indicating that increasing the number of target, and thereby increasing the amount of user activity, could potentially engage greater levels of physical activity. Alternatively, a rehabilitation or yoga like game could be facilitated using our technique where finding 3D targets using both arms would force the user into a particular position, e.g., both arms extended to the user's sides.

In addition to exergames, tactile-proprioceptive displays could be useful for allowing blind users to access information presented on large interactive displays much in the same way that screen readers allow blind users to access information presented on standard desktop-based displays. Modern large interactive displays are highly visual, preventing blind users from being able to experience the presented information (e.g. interactive table displays in museums, or flight arrival boards at airports). Tactile-proprioceptive displays could potentially be developed which act as sensory substitution interfaces between the blind user and the large interactive displays. These interfaces would require the blind user to carry around some form of a mobile controller which can convey tactile information, however this could simply be implemented using a blind user's smartphone, since smartphones have become ubiquitous in today's society and nearly all smartphones are equipped with vibrotactors.

Finally, another application area of our technique could be human navigation

systems. Several tactile-proprioceptive techniques have already been developed that use the users arm to point out the direction towards an object of interest [58, 47] but they do not tell the user how far away the object of interest is. Our technique could enhance these existing techniques by using the  $Z$ -coordinate of a target's location to convey the relative distance to the point of interest. For example, if the user has to stretch their arm completely to touch the target's  $Z$ -coordinate this could indicate that the object of interest is 10 m away and if it is close to the user's body 1 m. This allows for intuitively finding objects without requiring the user to look at a display or listen to audio, which could be useful, for example, to develop a search and rescue application. In the following chapter, we will discuss a related navigation system which utilizes modern technology and sensor substitution to guide users toward landmarks much in the same way that this chapter utilized sensory substitution to guide users towards targets in a nonvisual display.

## CHAPTER 3

### FACILITATING DIFFICULT SPATIAL NAVIGATION TASKS WITH WEARABLE TECHNOLOGY

Chapter 2 studied spatial perception problems as they relate to interactions with large displays for blind users. In contrast, this chapter looks at the spatial perception problem (for blind users) of navigating across large open spaces. To that end, this chapter presents a large open space navigation application called HeadLock. Section 3.1 describes the motivation behind this work, the objectives in the development of HeadLock, and an overview of the application's functionality. Section 3.1.2 presents the relevant background research in the field of blind navigation with regards to both traditional navigation applications using hardware-based localization systems and modern computer vision-based localization systems. Sections 3.1.3 and 3.1.5 discuss the design, implementation, and an evaluation of the application. Section 3.1.6 concludes the chapter with a discussion of the potential for future development on HeadLock.

#### **3.1 HeadLock: A Wearable Interface for Helping Blind Individuals to Traverse Large Open Spaces**

For many people with visual impairments, especially for those who are completely blind, living an independent life is a daily challenge. People who are blind must rely on sighted individuals for a range of tasks, including a large set of navigation tasks such as navigating unknown spaces for the first time. In order to support their independence, people who are blind use tools such as guide dogs, or more commonly, the white cane, in order to achieve tasks such as path following and

obstacle avoidance without the help of a sighted individual. Tools like the cane are essential for allowing blind people to live independent and healthy lives; however the usability of the white cane for spatial perception tasks is limited to only providing sensing information within a small radius around the person. This limits the use of the white cane to tasks dealing with nearby information, e.g. navigating memorized routes based on series of closely positioned or contiguous tactile landmarks such as the edge of a walkway or wall, and avoiding obstacles along these routes. While many blind people are adept at navigating memorized routes based on tactile landmarks, navigating large open spaces lacking easily perceivable landmarks is a particularly challenging task, often requiring blind people to resort to relying on the aid of friends and family. Seeing as how large open spaces are often traversed (i.e. building foyers and airport terminals), this is a serious issue for the average blind user. In this section, we address this issue directly by presenting a large open space navigation application called Headlock.

### **3.1.1 Overview and Objectives**

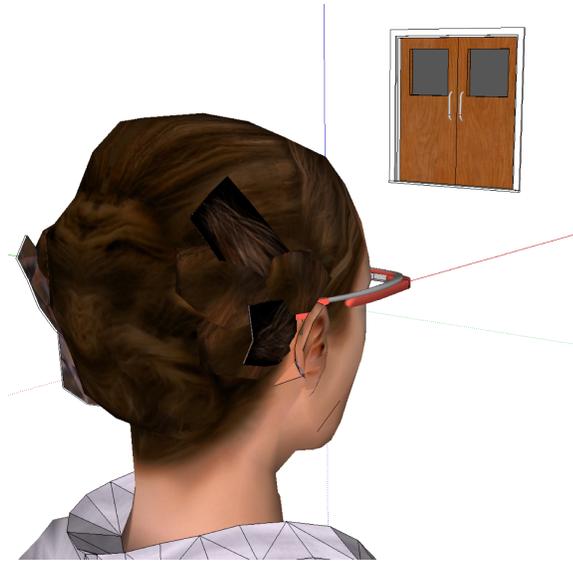
While many solutions have been developed to aid blind users during indoor navigation tasks (see Section 3.1.2), no solution to date has allowed for blind navigation across indoor large open spaces. Additionally, the indoor navigation solutions which have been proposed suffer from three primary limitations, each of which contribute to one of the objectives of the HeadLock project:

1. Many existing systems require the installation of distributed hardware systems such as WiFi routers or infrared beacons. This limitation prevents these systems from being used in unknown environments, and requires a signif-

icant monetary investment for installation. Therefore, HeadLock is specifically designed to require minimal hardware, and to allow for exploring previously unknown environments for the first time.

2. In addition the large hardware requirements, many of the related systems also have strong dependencies on a priori information. This limitation also prevents these systems from being used in unknown environments. With this in mind, HeadLock is designed to require minimal a priori information.
3. Finally, existing systems which run on mobile platforms (e.g. android smartphones) require blind users to properly aim a smartphone camera without being able to see the viewfinder for the camera. To address this issue, a primary objective of the HeadLock project is to utilize a wearable platform which does not require the user to aim the camera explicitly.

Within the context of these objectives, this work addresses the design and evaluation of the HeadLock system. Headlock is designed to run in real-time on a wearable computing platform with limited computational resources (e.g. Google Glass, see Figure 3.1). Headlock utilizes computer vision in order to remotely sense natural landmarks such as doors. The user interface allows blind users to scan for and lock onto one of these target landmarks across a large open space, and then provides feedback that directs the user to the location of the landmark. This feedback can either be provided as sonification or text-to-speech, both of which are designed to prevent the user from veering from their course and to provide the user with navigation task progress updates. The system is designed to be robust to accidental course deviation by the user. If the user loses track of the target landmark in the middle of a navigation task, the system can easily relocate the target and restart the navigational feedback. Finally, we present an evaluation of the



**Figure 3.1:** The HeadLock Platform

HeadLock runs on a wearable platform, allowing users to continue the use of their white cane for obstacle avoidance while receiving navigation guidance from the application.

HeadLock system consisting of a quantitative comparison of the sonification and text-to-speech feedback schemes and a qualitative analysis of the usability and utility of the HeadLock application.

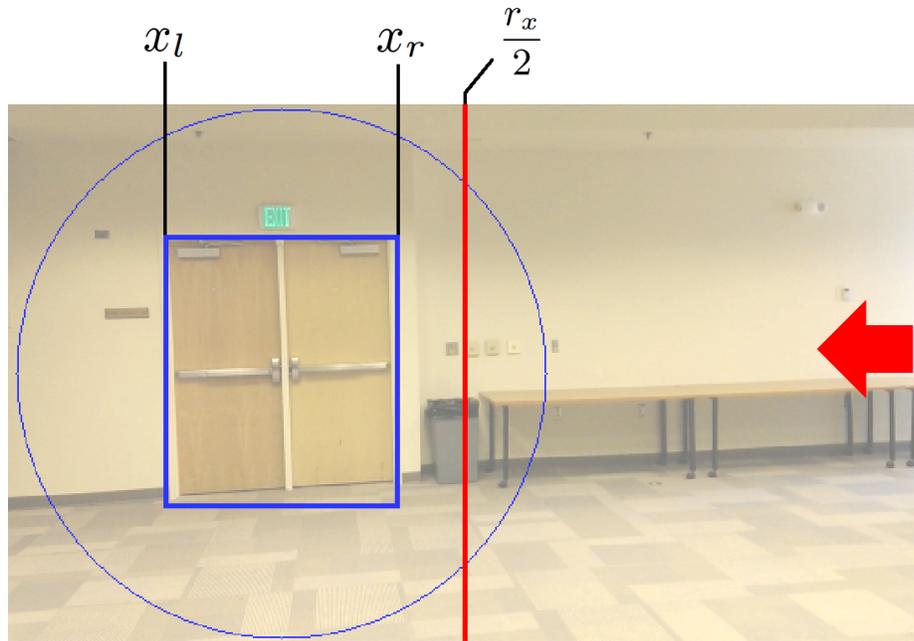
### 3.1.2 Background

The problem of solving blind navigation challenges with wearable technology is a relatively young problem, although it has given rise to the development of a significant number of wearable mobility aids. One of the earliest systems developed was that of Ertan et al. (1998) [18], in which, a wearable system conveyed navigation directions in the form of haptic feedback through an array of vibrational motors sewn into a vest worn by the blind user. While the preliminary results of this system were promising, the system was severely limited by the fact that it

required a large scale installation of infrared transceivers in order to localize the blind user within an indoor environment. Hub et al. (2003) [29] addressed this issue by utilizing a WLAN-based indoor localization technique which was more cost-effective and practical due to the fact that many indoor locations already have WLAN installations. In order to provide blind users with navigational information, Hub et al. augmented a cane with a stereo camera, a simple keypad, and a speaker. The stereo camera detected objects in front of the user and retrieved information regarding these objects from a 3D model of the test environment. The fact that the system relies on a priori information of the environment is limiting in that users cannot use this system to explore previously unknown environments. A related project by Schmitz et al. (2011) [59] eased the need to explicitly map the environment of the previous system by combining various navigational data sources already in existence such as street maps and lists of departure times in the Nexus Platform. While the Nexus Platform did not require users to explicitly map environments of interest, it was still dependent upon the existence of map information in order to generate navigational information for blind users. A similar map-based navigation system designed for indoor navigation is the Navatar system presented in Fallah et al. (2012) [19]. This system was unique in that it only required minimal hardware (a smartphone with an accelerometer) and allowed for highly accurate localization and navigation by updating particle filter location estimates with feedback from users upon reaching tactile landmarks, such as hallway intersections or doorways, in order to cull particles with poor localization estimates.

A commonality between each of these blind navigation systems is that they all rely on non-visual means of user localization (i.e. WLAN, infrared, or pedometry based localization) with the goal of allowing blind users to more efficiently wayfind through indoor environments. Unfortunately, each of these approaches

has a reliance on either a priori knowledge (in the form of maps) or large installations of hardware throughout the navigable area. As discussed by Manduchi and Coughlan (2012) [49], computer vision could be a better choice to reach this goal due to the fact that it is the natural technological parallel of the human vision system which normally handles wayfinding problems. Over the past several years, the use of computer vision on mobile platforms has increasingly been used to solve localization and navigation problems for blind users. Manduchi (2012) [48] presented a mobile computer vision system designed to detect and guide users towards artificial landmarks (i.e. fiducials). While this system requires the installation of a set of artificial landmarks, Manduchi argued that the system could easily be adapted to detect “natural” landmarks such as an elevator button or an informational sign. This approach to developing a blind navigation system was limited by that fact that it could only sense landmarks at a distance of 3.5 meters and it required users to aim a smartphone camera without being able to see the camera’s viewfinder. A similar system called VizWiz::LocateIt [8] allowed blind users to take a picture of a scene (e.g. a picture of a shelf of different cereals in a grocery store) and then receive feedback guiding them towards a nearby target (e.g. a box of Wheaties on the shelf). This approach employed both sonification and text-to-speech interfaces for guiding users towards the object of interest. Due to the fact that this system was developed for guiding a user towards an object in relatively close proximity, it is not well suited for long range navigation; however, blind users found the sonification and text-to-speech feedback useful for finding objects.



**Figure 3.2:** A Visual Interpretation of the HeadLock Interface

In order to generate guidance feedback for the blind user, HeadLock calculates whether the user is veering or not based on the position of the middle of the image (indicated by the red vertical line at  $\frac{1}{2}r_x$ ) relative to the nearest edge of the bounding box surrounding the target landmark (indicated by the blue box with vertical edges at  $x_l$  and  $x_r$ ). The red arrow pointing left indicates that HeadLock would generate feedback guiding the user to the left to correct for their right veering in this example.

### 3.1.3 Design

In [48], Manduchi poses his blind navigation problem as a *Discovery* phase followed by a *Guidance* phase. HeadLock adopts this problem decomposition and is designed within this context, however, where [48] was targeted at wayfinding by the use of a series of closely located fiducials, HeadLock is specifically concerned with the long range detection of natural landmarks in order to facilitate wayfinding across large open spaces.

## Use Context

A preliminary contextual analysis regarding the challenges of indoor blind navigation tasks showed the traversal of large open spaces can present a significant challenge for blind users. These challenges result from the manner in which many blind users localize themselves during navigational tasks. White cane users complete navigation tasks by first memorizing a mental model consisting of the series of tactile landmarks making up the navigation path (e.g. the edge of a curb followed by a door and then a wall, etc.) and then, during future attempts at the navigation task, aligning their stream of sensory input to this mental model. The problems arise when blind people are forced to cross large open spaces either lacking such tactile landmarks or having large distances between tactile landmarks. While the user may have landmarks memorized on either side of the open space, crossing from one landmark to the next without veering can be very difficult. As one user described what he does when confronted with this situation, “I cross my fingers and hope for the best”.

Therefore, HeadLock was designed with the motivation of enabling blind users to traverse large open spaces without veering, increasing blind users’ confidence in crossing such spaces and allowing them to live more independent lives. An important limitation in the design of HeadLock is that it is specifically targeted at large open spaces free of obstacles. While assuming obstacle-free paths for the purposes of our evaluation of HeadLock appears to be a strong assumption, it is justified due to the fact that the primary objective of HeadLock is to examine the potential to decrease user veering along long navigation paths. In order to ensure that HeadLock still allows for safe navigation of the blind outside of the carefully structured evaluation conditions in Section 3.1.5, HeadLock is designed to be used

in addition to a user's white cane. This allows the user to handle obstacle avoidance with their cane while using HeadLock to navigate towards target landmarks.

### **Landmark Selection**

As previously discussed, Manduchi's work [48] explored the use of artificial landmarks (fiducials designed for easy recognition) as navigational targets. Since one of the objectives of HeadLock is to facilitate navigation of unknown areas (i.e. areas which have not been previously explored, mapped, or augmented for the purposes of navigation), HeadLock is designed to detect natural landmarks, specifically doorways. Doorways are a practical natural landmark to use for navigation since they are easy to detect due to their uniform color and shape (within the context of a single building) and since blind users often cross large open spaces with the end goal of finding a particular doorway (e.g crossing a foyer to find the elevator door). There are certainly situations in which blind users could cross large open spaces where a doorway is not present to be tracked as a natural landmark; however, for the purposes of evaluating the HeadLock interface, the design is restricted to specifically dealing with situations in which a door is present. This could clearly be expanded with a less restricted landmark detection algorithm in the future to allow for tracking any sort of natural landmark.

### **Platform**

A major issue with previous navigation applications based on smart phones ([8, 48]) is that they required the blind user to properly aim the smart phone's camera without being able to see the viewfinder. This is a challenging task for many blind

users, prompting the development of assistive camera aiming applications [31]. Therefore, for the design of HeadLock, we opted for a wearable platform, Google Glass, which alleviates the need to aim a handheld camera. The camera on Google Glass is located on the right side of the frame with a front-facing orientation giving it a very similar perspective to what the human eye would perceive. This means that the user wearing Google Glass has an innate understanding of the direction in which the camera is aimed based on how their head is oriented. An additional benefit of using Google Glass as the hardware platform is that the user does not have to hold a smart phone in their hand, thereby leaving the user with one free hand.

### **Feedback Design**

A final crucial consideration for the design of HeadLock was in the form in which feedback should be presented to the user. Due to the fact that the navigation task data is quantitative in nature (i.e. measured in distance to target and degree of veering) it could easily be presented as an *auditory graph* [53] using sonification. However, while sonification methods such as the series of clicks used in VizWiz::LocateIt [8] have been shown to be effective for guidance, the fact that the data being presented to the user is navigational in nature and can therefore be posed as a series verbal instructions indicates that text-to-speech feedback may be a better method for generating feedback. There are potential advantages to each of these feedback representations; therefore, HeadLock is designed with two feedback modes: sonification and text-to-speech.

### 3.1.4 Implementation

The implementation of the HeadLock application is divided into two distinct phases. The first phase, *discovery*, allows for the user to remotely sense and lock onto landmarks (e.g. doorways). The second phase, *guidance*, provides the user with feedback guiding them towards the target landmark they locked onto and providing progress information indicating how much of the navigation task they have completed at any given time. Users proceed between phases by simply tapping the touchpad on the right side of the Google Glass. Additionally, users can switch between feedback modes by pressing the physical button on the top of the Google Glass, above the touchpad.

#### Discovery

The goal of the discovery phase is to use computer vision to locate landmarks and allow users to select a particular landmark to be guided towards. The computer vision algorithm implemented for doorway detection can be seen in Algorithm 1. Clearly, this color blob-based doorway detection algorithm is quite simplistic; however, this algorithm could easily be replaced with a more sophisticated algorithm for detecting other types of landmarks. This simple algorithm was accepted for HeadLock since the focus of this work is on the evaluation of the interface rather than the development of the supporting computer vision algorithms. In reference to the interface, both the sonification and text-to-speech modes are very similar for the discovery phase. Both modes provide feedback once every 2000 ms indicating whether or not a doorway has been found. In sonification mode, this consists of three short beeps to indicate a door has been found and a high pitch

---

**Algorithm 1** Basic Doorway Detection

---

**Input:**  $img \leftarrow$  Input image from Glass**Input:**  $upper \leftarrow$  Upper HSV limit**Input:**  $lower \leftarrow$  Lower HSV limit

```
1: //Gaussian blur
2:  $img \leftarrow$  PYRAMIDDOWN( $img$ )
3:  $img \leftarrow$  PYRAMIDUP( $img$ )
4:
5: //Change color space and filter by HSV value
6:  $img \leftarrow$  CONVERTTOHSV( $img$ )
7:  $mask \leftarrow$  INRANGE( $img, lower, upper$ )
8:  $mask \leftarrow$  DILATE( $img$ )
9:
10: //Find blobs
11:  $cntrs \leftarrow$  FINDCONTOURS( $mask$ )
12:
13: //Check the size and extent of the blob
14: for all  $cntr \in cntrs$  do
15:    $box \leftarrow$  FINDBOUNDINGBOX( $cntr$ )
16:
17:   if  $box.area \geq$  MinDoorArea then
18:
19:     if  $cntr.area \geq box.area * MinExtent$  then
20:       UPDATETRACKER( $box$ )
21:     end if
22:
23:   end if
24:
25: end for
```

---

beep immediately followed by a low pitch beep to indicate no door is visible. In text-to-speech mode, the feedback consists of the phrases “Door Found” and “No Door”.

## Guidance

The goal of the guidance phase is to provide the user with navigational feedback in order to guide them towards the target landmark. There are two essential metrics used to generate the guidance output by both feedback modes: a metric measuring the degree to which the user is veering, and a metric measuring the distance traveled by the user. The veering metric simply measures the distance from the closest edge of the bounding box surrounding the tracked landmark to the middle of the field of view:

$$v(x_l, x_r) = \max\left(\frac{r_x}{2} - x_r, x_l - \frac{r_x}{2}\right)$$

where  $x_l$  and  $x_r$  are the  $X$  coordinates of the left and right edges (respectively) of the bounding box surrounding the tracked landmark, and  $r_x$  is a constant representing the  $X$  resolution of the input image. The distance metric approximates the percentage of the navigation task the user has completed:

$$d(x_l, x_r) = \frac{\log\left(100\frac{(x_r - x_l) - w_i}{r_x - w_i} + 4\right) - 0.4}{1.5}$$

where  $w_i$  is a constant value set for each tracked landmark representing its width in pixels when first observed (this definition and the definitions above hold

for all equations in this section). This metric requires more explanation. Note that  $(x_r - x_l)$  is the current width of the observed landmark, meaning that the fraction

$$\frac{(x_r - x_l) - w_i}{r_x - w_i}$$

is a value ranging from 0.0 (when a landmark is first observed) to 1.0 (when the user reaches a point close enough to the landmark that the landmark fills the field of view of the camera). When plotted against the true distance to the door, it can be seen that this fraction generates a function which varies exponentially with the true distance; hence, the use of the logarithm to create a function which varies linearly with respect to the true distance to the door. The constant value of 4 is added to the fraction in order to ensure that the value inside the log is never zero. This value was determined experimentally. The remainder of the expression scales the potential values the logarithm to a range from 0.0 to 1.0, representing a relative measure of the approximate distance traveled during the navigation task as a percentage. This relative approximation could easily be replaced with a similar measure that determines a true distance approximation rather than a relative distance approximation, but this would require a priori knowledge of the actual width of the target landmark being observed. Since a goal of the HeadLock application is to limit the amount of required a priori knowledge, a relative distance metric was chosen.

*Sonification Feedback:* The sonification scheme employed by HeadLock can be described in terms of pulse delay and frequency of the generated tone. Pulse delay,  $p_s(x_l, x_r)$ , represents the delay in ms between each beep:

$$p_s(x_l, x_r) = \begin{cases} 200 + 800v(x_l, x_r) & : x_l > \frac{r_x}{2}, x_r < \frac{r_x}{2} \\ 0 & : x_l \leq \frac{r_x}{2} \leq x_r \end{cases}$$

When a user is veering ( $x_l > \frac{r_x}{2}, x_r < \frac{r_x}{2}$ ),  $p_s(x_l, x_r)$  varies the delay between beeps linearly from 200 ms to 1000 ms with respect to  $v(x_l, x_r)$ . When the user isn't veering ( $x_l \leq \frac{r_x}{2} \leq x_r$ ),  $p_s(x_l, x_r)$  returns a constant value of zero (a continuous tone, no beeping). The frequency of the generated tone,  $f(x_l, x_r)$ , represents the pitch of the feedback sound in terms of Hz.

$$f(x_l, x_r) = \begin{cases} 1710 & : x_l > \frac{r_x}{2}, x_r < \frac{r_x}{2} \\ 130 + 920d(x_l, x_r) & : x_l \leq \frac{r_x}{2} \leq x_r \end{cases}$$

When the user is veering off course,  $f(x_l, x_r)$  sets the output pitch to the constant value of 1710 Hz ( $G\sharp_6$ ). In the case that the user is not veering,  $f(x_l, x_r)$  varies the pitch of the feedback tone linearly from 130 Hz to 1050 Hz (3 octaves, from  $C_3$  to  $C_6$ ) with respect to  $d(x_l, x_r)$ . This informs the user of their progress. This mapping was chosen with the goal of making it easy to hear changes in pitch as navigation progresses by utilizing a large frequency range, and the goal of making it easy to distinguish the veering feedback from non-veering feedback by having a single pitch for indicating veering significantly outside of the frequency range indicating no veering.

*Text-to-Speech Feedback:* The text-to-speech feedback scheme can also be described with two equations. The first equation indicates the delay between each piece of text-to-speech feedback:

$$p_t(x_l, x_r) = \begin{cases} 200 + 800v(x_l, x_r) & : x_l > \frac{r_x}{2}, x_r < \frac{r_x}{2} \\ 2000 & : x_l \leq \frac{r_x}{2} \leq x_r \end{cases}$$

When the system detects the user is veering,  $p_t(x_l, x_r)$  varies in the same manner as  $p_s(x_l, x_r)$ , but is a constant 2000 ms when the user is not veering. This reduces the feedback redundancy that would be present if there were continuous speech feedback similar to the continuous sonification feedback described in section 3.1.4. Note that the system generates the feedback speech “Left” in the first case of  $p_t(x_l, x_r)$  above, “Right” in the second case, and “Straight” in the third case. Whenever the system generates feedback for the user to continue straight forward, it also generates speech for a number between 1 and 100 indicating a percentage of the navigation which has been completed. This value is simply a text-to-speech readable version of the distance metric  $d(x_l, x_r)$  discussed above:

$$d_{out}(x_l, x_r) = 100 * d(x_l, x_r)$$

This serves the purpose of providing the users with an understanding of their progress during the navigation task in the same manner that the tone of increasing frequency (i.e.  $f(x_l, x_r)$ ) does for the sonification feedback mode.

*Completing the Guidance Phase:* The metric used to determine when the user has completed the navigation task is the ratio of the perceived width of the door (in pixels) to the  $X$  resolution of the image. When this ratio surpasses 90% (i.e. when the tracked door fills 90% of the field of view), the task is considered complete. Regardless of the feedback mode, HeadLock plays the same “success” sound effect to signal to a user that they have completed their navigation task.

## **Error Handling**

HeadLock is designed to be robust to errors in the tracking of landmarks by providing a simple error recovery method. Whenever a landmark that was being tracked is lost, the user is provided with text-to-speech feedback indicating the direction in which they should turn to bring the landmark back into view, and the discovery phase is automatically restarted. This allows for landmarks to be relocated quickly, resuming guidance without having to restart the entire system.

### **3.1.5 Evaluation**

Our evaluation of the HeadLock system involved both a quantitative portion to determine which of the two proposed feedback modes results in the least amount of veering in blind users, and a qualitative portion to determine the usability and utility of the system. The qualitative portion of the study consisted of having users answer a series of questions in a brief survey immediately after completing the qualitative portion of the study. For the quantitative comparison between feedback modes, precise localization data was gathered using an installation of the StarGazer RS 1.0 robot localization system from Hagisonic. Users wore a belt with the StarGazer camera attached at the back and a backpack with a laptop in it for collecting data from the StarGazer camera. This system is capable of localizing the user within  $\pm 1$  cm of error.

## Participants

Six blind participants were recruited from the local chapter of the National Federation of the Blind (2 Female, average age 43.2, SD = 12.7). All participants were completely blind with no residual light perception, with the exception of two participants who could perceive small amounts of light. The participants wore a blindfold during the user studies in order to ensure that their performance was not unduly affected by their ability to perceive small amounts of light. None of the participants reported any cognitive or motor impairment.

## Procedure

Our user study of the HeadLock application took place in a large empty conference room which none of the participants were familiar with beforehand. All furniture was removed from the room prior to the study to ensure that the users would not encounter any obstacles during the navigation study as was discussed in Section 3.1.3. The user study was organized into five parts: an unaided navigation task without HeadLock (serving as a control for qualitative comparison), an initial tutorial of the HeadLock interface, trials with the first feedback mode, trials with the second feedback mode, and a brief survey eliciting qualitative feedback from the users. All participants were randomly divided into two groups (A and B). For group A the first feedback mode was sonification and the second was text-to-speech, while group B had text-to-speech first and sonification second. This counterbalancing ensured that the results of the study were not biased towards one feedback mode due to a learning effect.

*Unaided Navigation:* In order to compare navigation without the HeadLock ap-

plication to navigation with the HeadLock application, users were taken into the conference room, positioned on one end of the room and asked to locate the door on the other end of the room. In a real-world situation, it is unlikely that blind people who are faced with the task of having to cross a large open space would be initially oriented in the exact direction they need to move to make it to their desired destination unless they enlisted the aid of a sighted individual. Therefore, users were positioned facing the correct side of the room, but were not oriented in the exact direction to the target doorway. Qualitative observations of the manner in which users accomplished this goal were recorded. Additionally, users were asked several open-ended questions regarding how they typically deal with similar situations. Upon completing this task, users were returned to the hallway outside of the conference room in order to begin the HeadLock tutorial. Note that a quantitative analysis of unaided navigation versus aided navigation was not justified due to the significantly different manners in which users accomplished the navigation tasks with and without the aid of HeadLock.

*HeadLock Tutorial:* The participants were familiarized with the HeadLock application through the use of a tutorial explaining the use of Google Glass and the functionality of the HeadLock application. The tutorial provided examples of each piece of audio which the user could encounter within the HeadLock interface and their corresponding explanations. Users were allowed to ask questions regarding the operation of the HeadLock interface. Each group was instructed on the sonification guidance feedback before beginning the sonification trials and the text-to-speech guidance feedback before beginning the test-to-speech trials (i.e. not both at the same time). All tutorials were conducted in a hallway outside of the conference room where trials took place so as to ensure that users were not able to become oriented within the conference room. This ensured that the users were unfamiliar

with the environment they were attempting to navigate. Before each feedback trial, after users had been instructed in the use of the HeadLock application, they were allowed to use the application in the hallway outside of the conference room until they felt comfortable that they understood how to use the application. At this point, they were guided into the conference room to begin the feedback trials.

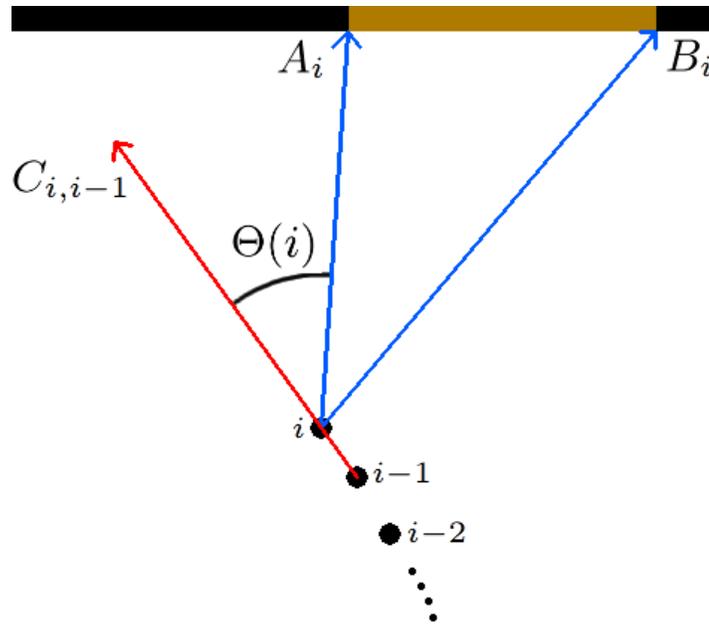
*Feedback Trials:* Trials for each of the feedback modes were conducted in an identical manner. Upon entering the conference room, users were guided by a sighted observer to five randomly selected points in the room before being guided to the starting location on the north end of the room. The point of this obscure path to the starting position was to disorient the users, ensuring that they did not start the discovery phase of the HeadLock application with an a priori understanding of where the target doorway was located. The starting position remained the same for all users and trials. They were positioned facing either the east, or west side of the room (randomly selected), while the target doorway was located on the south wall of the room, ensuring that the user would have to scan the room to discover the doorway. The HeadLock application was then started by the observer and the user was instructed to discover the doorway and then follow the application's guidance to navigate to it. The StarGazer system tracked the users' exact path of navigation and the time required for the user to complete each phase was recorded. Upon completing a trial (arriving at the target doorway), the procedure was repeated by again guiding the user around the room to five randomly selected points before returning the to the starting position. Each user completed six trials (three trials starting with the user facing east and three with the user facing west) for each of the two feedback modes. Immediately following the completion of all twelve trials, users were asked to respond to a series of prompts regarding the usability of the HeadLock application on a standard five-level Likert scale with 1 being "Strongly

Disagree” and 5 being “Strongly Agree”.

## Results

*Unaided Navigation:* All users followed very similar strategies for finding the doorway located on the opposite side of the conference room. Each user began this task by setting out in the general direction of the doorway, using their canes to sense for obstacles (e.g. walls). Several of the users veered significantly and found the east wall first, while the remainder of the users veered less significantly and found the south wall first. After finding a wall, all users followed the wall until finding the doorway. One user commented that this is the technique she uses on a regular basis for finding particular rooms in an unfamiliar environment (e.g. finding a particular store in a mall). Multiple users commented that when faced with situations such as this, they use both sounds (e.g. echoes off of walls and sounds from air vents) and smells (e.g. the scent of the food court in a mall) to help orient themselves in large open spaces.

*Feedback Trials:* In order to understand the efficiency of the sonification mode as compared to the text-to-speech mode, the average time required to complete the discovery and guidance phases for each mode was calculated for each user, and the averages were compared using a single factor ANOVA. The average times of one user were found to be significant outliers and were therefore not included in this analysis. It was found that text-to-speech required 12.56 s for discovery (34% faster than sonification) and 21.85 s for guidance (28% faster than sonification). Both of these results were found to be statistically significant ( $F_{1,8} = 6.438, p < 0.05$  and  $F_{1,8} = 5.318, p < 0.05$  respectively). Next, since HeadLock was designed to prevent blind users from veering during navigation, veering was analyzed across all users’



**Figure 3.3:** Explanation of the Veering Metric  $\Theta(i)$

$\Theta(i)$  represents the angular deviation of the user's trajectory from point  $i - 1$  to point  $i$  from the optimal range of trajectories between  $A_i$  and  $B_i$ . The veering cost function sums this measure of veering at each point along a user's path, weighted by the magnitude of  $C_{i,i-1}$ . Note that the vector  $C_{i,i-1}$  has been extended in this illustration for the purpose of indicating the measure  $\Theta(i)$ . Additionally, the brown region represents the doorway while the black region represents a wall.

navigation paths for each feedback mode. In order to measure veering, two gradients were defined at every point in the mapped area: the vectors between the every point and the points of the left and right sides of the doorway (labeled as vectors  $A_i$  and  $B_i$  in Figure 3.3 respectively). The current trajectory at every point along the users' path was calculated as the vector between point  $i$  and point  $i - 1$  (labeled vector  $C_{i,i-1}$  in Figure 3.3). These vectors were then used to calculate the users' degree of veering,  $\Theta(i)$ , at each point along their paths. If  $C_{i,i-1}$  was angled between  $A_i$  and  $B_i$  ( $\angle AC < \angle AB$  and  $\angle BC < \angle AB$ ), the section of the path between point  $i$  and  $i - 1$  did not exhibit veering (i.e.  $\Theta(i) = 0$ ). On the other hand, if  $C_{i,i-1}$  was angled outside of the area between  $A_i$  and  $B_i$ , then the degree of veering was accounted for as the angle between  $C_{i,i-1}$  and the closest of either  $A_i$  or  $B_i$  (i.e.  $\Theta(i) = \min(\angle BC, \angle AC)$ ). This

measurement of veering was then used to create a cost function assigning a value to each path representing the average veering which occurred along each path  $P$ :

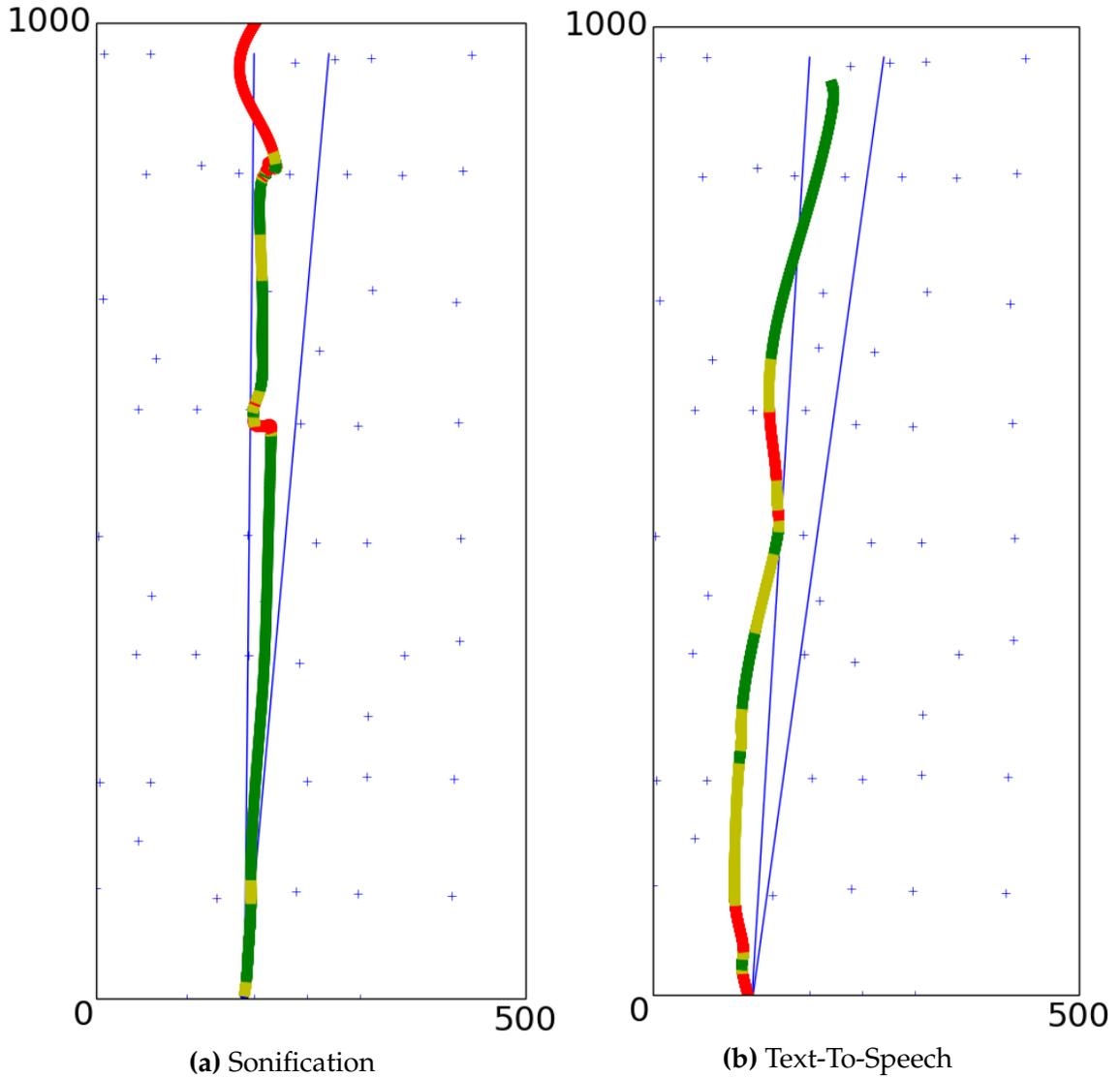
$$Cost_{veer}(P) = \frac{\sum_{i=1}^n \|C_{i,i-1}\| \Theta(i)}{\sum_{i=1}^n \|C_{i,i-1}\|}$$

The cost of all paths were calculated and averaged for each feedback mode for each user. These average cost values were compared between feedback modes with a single factor ANOVA. No significant difference was found between the veering exhibited in either mode ( $p > 0.05$ ).

*User Feedback:* The responses from the usability questionnaire were aggregated and the averages and standard deviations were calculated for each prompt. Users responded with the highest values (closest to “Strongly Agree”) to the statements: “The navigation system allowed me to navigate to a location efficiently” (M = 3.9, SD = 0.7), “The text-to-speech feedback was easy to understand” (M = 4.3, SD = 1.1), and “I liked using this navigation system” (M = 4.0, SD = 1.0). In contrast, the users responded with the lowest values (closest to “Strongly Disagree”) to the statements: “The sonification feedback was not sufficient to properly navigate to a doorway” (M = 1.8, SD = 1.2), and “The text-to-speech feedback was not sufficient to properly navigate to a doorway” (M = 1.4, SD = 0.4).

### 3.1.6 Discussion and Future Work

The initial results of our evaluation of HeadLock are quite promising, although they do indicate several points of the HeadLock design which should be improved



**Figure 3.4:** Examples of Typical Navigation Paths

Sonification feedback (a) tended to result in periods where the user would stop moving in order to correct their direction before continuing, while Text-To-Speech feedback (b) tended to result in users correcting their direction while continuing forward rather than stopping. Green indicates regions of the navigation path in which the user was not veering. Yellow indicates regions in which  $\Theta(i) < 10$ . Red indicates regions in which  $\Theta(i) > 10$ . Both axes are represented in cm. The blue lines represent the vectors  $A_0$  and  $B_0$  (see Figure 3.3).

and further evaluated. The significant difference between feedback modes in the time required to complete the navigation task indicates that the text-to-speech mode is the most efficient of the two feedback modes, even though there was no significant difference observed in the degree of veering between modes. From observation of the users completing navigation tasks, it appears that the sonification feedback may cause users pause momentarily to find the optimal path whenever veering feedback is provided (see Figure 3.4). On the other hand, text-to-speech feedback appears to result in smoother navigation paths, in which users continue to progress towards the target landmark while receiving veering feedback. This observation is most likely due to the fact that most blind users are very familiar with text-to-speech feedback due to their frequent use of screen readers, while most users have never used a sonification-based interface before. This potential explanation is supported by fact that users reported finding the text-to-speech interface the easiest to understand through their responses to the usability questionnaire.

While users commented positively on the user interface as a whole, there were several issues which were repeatedly mentioned in user feedback. The most prominent of these issues was that multiple users felt that the system could not be effectively used in a real world situation until the system can detect obstacles such as poles or people who are in the way. As was mentioned in Section 3.1.3, obstacle detection was specifically not included in this preliminary version of HeadLock due to our focus on determining the optimal interface for this system. However, future development of HeadLock should include obstacle detection with a specific goal of allowing for providing guidance around obstacles without losing track of the final goal destination.

With respect to the HeadLock interface, one potential area of future work could be in expanding the discovery phase to allow for the disambiguation of multiple landmarks whenever the system recognizes more than one landmark (i.e. how does the system choose the appropriate door if it observes three similar doors?). In relation to this question, it would be interesting to expand the user interface to allow for speech-based queries or interactions (i.e. HeadLock: "I see three doors", User: "Choose the door on the right"). Essentially, this would be the expansion of HeadLock in order to provide the ability for the system to synthesize contextual information. As one user commented, "a navigation application is only as useful as the context it provides." It would be interesting to explore how HeadLock could maintain its objective of requiring minimal a priori information while still being able to generate contextual information for the user (e.g. HeadLock: "I see three doors. *The door on the left goes to the men's room*", User: "Choose the door on the left").

Another interesting set of questions which should be addressed in future work involve a deeper analysis of the computer vision challenges underlying the HeadLock system. For instance, rather than restricting users to discovering only fixed landmarks, it would be interesting if users could instead discover and follow continuous landmarks such as patterns (e.g. lines on the floor). Additionally, future work should address the ever-present computer vision challenge of obstacle detection and avoidance.

Finally, it would be interesting to integrate our large open space traversal application with a general purpose indoor navigation system such as Navatar [19] in order to get both the open space navigation benefits of HeadLock and the general navigation (hallway based) benefits of a system like Navatar. This type of de-

velopment could represent a middle ground in which blind users could reap the benefits of both the computer-vision only navigation approach of Manduchi, and the plethora of non-computer vision navigation approaches previously discussed.

## CHAPTER 4

### CONCLUSION

The use of modern technologies as augmentative tools for solving spatial perception problems is a field with potential for major social impact. This thesis has presented two projects as examples of technology being used to augment the spatial perception capabilities of blind users; however, these projects do not represent a covering of the potential applications of research in assistive technology dealing with spatial perception. In this final chapter we focus on potential future work in this field by posing several spatial perception related problems. Both Chapters 2 and 3 present conclusions specific to the particular project discussed therein; therefore, we conclude this chapter with simple summary of the research projects discussed at length in Chapters 2 and 3.

#### 4.1 Future Work

A particularly interesting question related to the projects presented in this thesis is as follows: what role does spatial perception play in social interactions and how does a lack of spatial perception effect the ability of blind users to have social interactions? We often take for granted the fact that simple spatial perception tasks, such as recognizing that a particular person is approaching you in a hallway, are integral for carrying out basic social interactions (e.g. acknowledging the persons existence as they approach you). Given the dependence of this particular problem on visual perception, it is conceivable that a very similar approach to that of Head-Lock could be created to aid blind users in social situations by perceiving spatial information. In a more general sense, it is clear that social interactions rely on a complex set of interleaved cognitive tasks (visual perception, spatial perception,

long and short term memory access, etc.), so it may be interesting to explore how wearable technology can be leveraged to aid blind users with the cognitive tasks related to vision while providing an interface that is unobtrusive (i.e. an interface which does not get in the way of the user completing the remaining cognitive tasks without assistance).

Finally, an area of future work which holds the potential to revolutionize the field of assistive technology is the use of RGB-D (Red-Green-Blue-Depth) cameras as a means of perceiving spatial information. The advent of RGB-D sensors such as the Microsoft Kinect [40] opened an entirely new field of research allowing assistive systems to be developed which can directly relay accurate depth-based perception data to users [37]. However, the initial depth sensors like the Kinect which have been released over the past several years are all large and non-mobile. Up until recently, this has limited the use of RGB-D systems in wearable assistive solutions; however, new depth sensors are currently being developed [3] which are small enough to be integrated into mobile platforms such as smartphones. Use of this advanced hardware could potentially enable significant advances in computational spatial perception in order to aid blind users. This should be explored further as smaller sensors are released for development.

## 4.2 Summary

Chapter 2 presented a series of two research projects focused on the development of tactile-proprioceptive displays for blind users. Tactile-proprioceptive displays allow blind users to perceive the spatial layout of interface elements by substituting the vision required by traditional user interfaces with the sense of touch and

the user's sense of the position and orientation of their joints. The research discussed in Chapter 2 presented four methods for developing tactile-proprioceptive displays and presented quantitative studies determining the efficiency of each method. The most efficient two-dimensional method was then implemented in an exergame for blind children and the usability of the interface was discussed. While these techniques clearly apply to the development of exergames for blind users, a series of other potential applications for these tactile-proprioceptive display techniques were discussed.

Chapter 3 took a look at a different spatial perception related problem: the problem of blind users navigating large open spaces. To address this problem, we presented HeadLock, an application for a wearable platform (Google Glass) designed to guide blind users to natural landmarks such as doorways using sonification and text-to-speech feedback. HeadLock is unique when compared to related blind navigation systems in that it does not depend on a priori information about the navigable area, it does not require the user to aim a smartphone camera, and it is not reliant upon costly hardware installations throughout the navigable area. A user study of six blind users was presented in which a qualitative assessment of the system was performed in addition to a quantitative comparison of the two proposed interface feedback modes.

## BIBLIOGRAPHY

- [1] Freedom scientific - products for the visually impaired, jaws screen reader, <http://www.freedomscientific.com>.
- [2] Sony Playstation Move controller, <http://us.playstation.com/ps3/playstation-move/>, access date: 10-10-2010.
- [3] Startup looks to bring kinect-like camera to the ipad, <http://allthingsd.com/20130917/start-up-looks-to-bring-kinect-like-camera-to-the-ipad/>.
- [4] American foundation for the blind (2002), statistics and sources for professionals, [http://www.afb.org/info\\_document\\_view.asp?documentid=1367](http://www.afb.org/info_document_view.asp?documentid=1367). 2008.
- [5] Teemu Tuomas Ahmaniemi and Vuokko Tuulikki Lantz. Augmented reality target finding based on tactile cues. In *Proceedings of the 2009 international conference on Multimodal interfaces, ICMI-MLMI '09*, pages 335–342, New York, NY, USA, 2009. ACM.
- [6] Fred Attneave and Braddie Benson. Spatial coding of tactual stimulation. *Journal of Experimental Psychology*, 81(2):216, 1969.
- [7] Paul Bach-y Rita and Stephen W Kercel. Sensory substitution and the human-machine interface. *Trends in cognitive sciences*, 7(12):541–546, 2003.
- [8] Jeffrey P Bigham, Chandrika Jayant, Andrew Miller, Brandyn White, and Tom Yeh. Vizwiz:: Locateit-enabling blind people to locate objects in their environment. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 65–72. IEEE, 2010.
- [9] Lorna M. Brown and Topi Kaaresoja. Feel who's talking: using tactons for mobile phone alerts. In *CHI '06 extended abstracts on Human factors in computing systems, CHI EA '06*, pages 604–609, New York, NY, USA, 2006. ACM.
- [10] David Brown<sup>1/2</sup>, Tom Macpherson, and Jamie Ward<sup>ô</sup>. Seeing with sound? exploring different characteristics of a visual-to-auditory sensory substitution device. *Perception*, 40:1120–1135, 2011.
- [11] Michele E Capella-McDonnall. The effects of single and dual sensory loss on symptoms of depression in the elderly. *Int J Geriatr Psychiatry*, 20(9):855–61, Sep 2005.

- [12] Ning Cheung and Tien Y Wong. Obesity and eye diseases. *Surv Ophthalmol*, 52(2):180–95, 2007.
- [13] B Chua and P Mitchell. Consequences of amblyopia on education, occupation, and long term vision loss. *Br J Ophthalmol*, 88(9):1119–21, Sep 2004.
- [14] A. Cockburn, P. Quinn, C. Gutwin, G. Ramos, and J. Looser. Air pointing: Design and evaluation of spatial target acquisition with and without visual feedback. *International Journal of Human-Computer Studies*, 69(6):401 – 414, 2011.
- [15] Julie M Crewe, Nigel Morlet, William H Morgan, Katrina Spilsbury, Aqif Mukhtar, Antony Clark, Jonathon Q Ng, Margaret Crowley, and James B Semmens. Quality of life of the most severely vision impaired. *Clin Experiment Ophthalmol*, Nov 2010.
- [16] Mary Czerwinski, George Robertson, Brian Meyers, Greg Smith, Daniel Robbins, and Desney Tan. Large display research overview. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '06, pages 69–74, New York, NY, USA, 2006. ACM.
- [17] Reinhard Eckhorn, Marcus Wilms, Thomas Schanze, Marcus Eger, Lutz Hesse, Ulf T Eysel, Zoltán F Kisvárdy, Eberhart Zrenner, Florian Gekeler, Helmut Schwahn, Keisuke Shinoda, Helmut Sachs, and Peter Walter. Visual resolution with retinal implants estimated from recordings in cat visual cortex. *Vision Res*, 46(17):2675–90, Sep 2006.
- [18] S. Ertan, C. Lee, A. Willets, H. Tan, and A. Pentland. A wearable haptic navigation guidance system. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*, pages 164–165. IEEE Comput. Soc, 1998.
- [19] Navid Fallah, Ilias Apostolopoulos, Kostas Bekris, and Eelke Folmer. The user as a sensor: navigating users with visual impairments in indoor spaces using tactile landmarks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 425–432, New York, NY, USA, 2012. ACM.
- [20] Alexander Fiannaca, Tony Morelli, and Eelke Folmer. Haptic target acquisition to enable spatial gestures in nonvisual displays. In *Proceedings of the 2013 Graphics Interface Conference, GI '13*, pages 213–219, Toronto, Ont., Canada, Canada, 2013. Canadian Information Processing Society.
- [21] Eelke Folmer and Tony Morelli. Spatial gestures using a tactile-proprioceptive

- display. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, TEI '12, pages 139–142, New York, NY, USA, 2012. ACM.
- [22] T. Guerreiro, P. Lagoa, H. Nicolau, D. Goncalves, and J.A. Jorge. From tapping to touching: Making touch screens accessible to blind users. *Multimedia, IEEE*, 15(4):48–50, oct.-dec. 2008.
- [23] Sean Gustafson, Daniel Bierwirth, and Patrick Baudisch. Imaginary interfaces: spatial interaction with empty hands and without visual feedback. In *Proceedings of the 23rd annual ACM symposium on User interface software and technology*, UIST '10, pages 3–12, New York, NY, USA, 2010. ACM.
- [24] Sean Gustafson, Christian Holz, and Patrick Baudisch. Imaginary phone: learning imaginary interfaces by transferring spatial memory from a familiar device. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 283–292, New York, NY, USA, 2011. ACM.
- [25] Zohar Habet-Wilner and Michael Belkin. Obesity is a risk factor for eye diseases. *Harefuah*, 144(11):805–9, 821, Nov 2005.
- [26] Chris Harrison and Scott E. Hudson. Minput: enabling interaction on small mobile devices with high-precision, low-cost, multipoint optical tracking. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 1661–1664, New York, NY, USA, 2010. ACM.
- [27] Chris Harrison, Desney Tan, and Dan Morris. Skinput: appropriating the body as an input surface. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 453–462, New York, NY, USA, 2010. ACM.
- [28] Eve Hoggan and Stephen Brewster. New parameters for tacton design. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '07, pages 2417–2422, New York, NY, USA, 2007. ACM.
- [29] Andreas Hub, Joachim Diepstraten, and Thomas Ertl. Design and development of an indoor navigation and object identification system for the blind. *ACM SIGACCESS Accessibility and Computing*, (77-78):147, September 2003.
- [30] Ali Israr, Olivier Bau, Seung-Chan Kim, and Ivan Poupyrev. Tactile feedback on flat surfaces for the visually impaired. In *Proceedings of the 2012 ACM*

*annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts*, CHI EA '12, pages 1571–1576, New York, NY, USA, 2012. ACM.

- [31] Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P Bigham. Supporting blind photography. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, pages 203–210. ACM, 2011.
- [32] Elizabeth J Johnson. Obesity, lutein metabolism, and age-related macular degeneration: a web of connections. *Nutr Rev*, 63(1):9–15, Jan 2005.
- [33] Shaun K. Kane, Jeffrey P. Bigham, and Jacob O. Wobbrock. Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, Assets '08, pages 73–80, New York, NY, USA, 2008. ACM.
- [34] Shaun K. Kane, Meredith Ringel Morris, Annuska Z. Perkins, Daniel Wigdor, Richard E. Ladner, and Jacob O. Wobbrock. Access overlays: improving non-visual access to large touch screens for blind users. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 273–282, New York, NY, USA, 2011. ACM.
- [35] Gertrudis I J M Kempen, Judith Ballemans, Adelita V Ranchor, Ger H M B van Rens, and G A Rixt Zijlstra. The impact of low vision on activities of daily living, symptoms of depression, feelings of anxiety and social support in community-living older adults seeking vision rehabilitation services. *Qual Life Res*, Nov 2011.
- [36] Jaakko Keränen, Janne Bergman, and Jarmo Kauko. Gravity sphere: gestural audio-tactile interface for mobile music exploration. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 1531–1534, New York, NY, USA, 2009. ACM.
- [37] Vinitha Khambadkar and Eelke Folmer. Gist: A gestural interface for remote nonvisual spatial perception. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 301–310, New York, NY, USA, 2013. ACM.
- [38] Jungsoo Kim, Jiasheng He, Kent Lyons, and Thad Starner. The gesture watch: A wireless contact-free gesture based wrist interface. In *ISWC '07: Proceedings*

of the 2007 11th IEEE International Symposium on Wearable Computers, pages 1–8, Washington, DC, USA, 2007. IEEE Computer Society.

- [39] Keehoon Kim, J. Edward Colgate, Michael A. Peshkin, Julio J. Santos-Munné, and Alex Makhlin. A miniature tactor design for upper extremity prosthesis. In *Proceedings of the 2007 Frontiers in the Convergence of Bioscience and Information Technologies, FBIT '07*, pages 537–542, Washington, DC, USA, 2007. IEEE Computer Society.
- [40] Kinect. Kinect for windows, <http://www.microsoft.com/en-us/kinectforwindows/>, url access date: 5-4-2012.
- [41] Celine Latulipe, Craig S. Kaplan, and Charles L. A. Clarke. Bimanual and unimanual image alignment: an evaluation of mouse-based techniques. In *Proceedings of the 18th annual ACM symposium on User interface software and technology, UIST '05*, pages 123–131, New York, NY, USA, 2005. ACM.
- [42] Frank Chun Yat Li, David Dearman, and Khai N. Truong. Virtual shelves: interactions with orientation aware devices. In *UIST '09: Proceedings of the 22nd annual ACM symposium on User interface software and technology*, pages 125–128, New York, NY, USA, 2009. ACM.
- [43] Frank Chun Yat Li, David Dearman, and Khai N. Truong. Leveraging proprioception to make mobile phones more accessible to users with visual impairments. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility, ASSETS '10*, pages 187–194, New York, NY, USA, 2010. ACM.
- [44] Joseph Luk, Jerome Pasquero, Shannon Little, Karon MacLean, Vincent Levesque, and Vincent Hayward. A role for haptics in mobile interaction: initial design using a handheld tactile display prototype. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 171–180, New York, NY, USA, 2006. ACM.
- [45] I Scott MacKenzie. *Human-computer interaction: An empirical research perspective*. Newnes, 2012.
- [46] Karon Maclean and Mario Enriquez. Perceptual design of haptic icons. In *In Proceedings of Eurohaptics*, pages 351–363, 2003.
- [47] Charlotte Magnusson, Miguel Molina, Kirsten Rasmus-Gröhn, and Delphine Szymczak. Pointing for non-visual orientation and navigation. In *Proceedings*

of the 6th Nordic Conference on Human-Computer Interaction: *Extending Boundaries*, NordiCHI '10, pages 735–738, New York, NY, USA, 2010. ACM.

- [48] Roberto Manduchi. Mobile vision as assistive technology for the blind: An experimental study. In *Computers Helping People with Special Needs*, pages 9–16. Springer, 2012.
- [49] Roberto Manduchi and James Coughlan. (computer) vision without sight. *Commun. ACM*, 55(1):96–104, January 2012.
- [50] Pranav Mistry and Pattie Maes. Sixthsense: a wearable gestural interface. In *ACM SIGGRAPH ASIA 2009 Sketches*, SIGGRAPH ASIA '09, pages 11:1–11:1, New York, NY, USA, 2009. ACM.
- [51] Tony Morelli, John Foley, and Eelke Folmer. VI-Bowling: a tactile spatial exergame for individuals with visual impairments. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility (ASSETS'10)*, pages 179–186, Orlando, Florida, USA, 2010.
- [52] Tony Morelli and Eelke Folmer. Twuist: A discrete tactile-proprioceptive display for eye and ear free output on mobile devices. In *Proceedings of Haptics Symposium 2012 (HAPTICS'12)*, pages 443–450, Vancouver, Canada, 2012.
- [53] Michael A Nees and Bruce N Walker. Listener, task, and auditory graph: Toward a conceptual model of auditory graph comprehension. In *Proceedings of the International Conference on Auditory Display (ICAD2007)*, pages 266–273, 2007.
- [54] S R Nyman, M A Gosney, and C R Victor. Psychosocial impact of visual impairment in working-age adults. *Br J Ophthalmol*, 94(11):1427–31, Nov 2010.
- [55] Tal Oron-Gilad, J. L. Downs, R. D. Gilson, and P. A. Hancock. Vibrotactile guidance cues for target acquisition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 37(5):993–1004, 2007.
- [56] J Paillard. Cognitive versus sensorimotor encoding of spatial information. In *Cognitive processes and spatial orientation in animal and man*, pages 43–77. Springer, 1987.
- [57] Ivan Poupyrev, Shigeaki Maruyama, and Jun Rekimoto. Ambient touch: designing tactile interfaces for handheld devices. In *UIST '02: Proceedings of*

*the 15th annual ACM symposium on User interface software and technology*, pages 51–60, New York, NY, USA, 2002. ACM.

- [58] Simon Robinson, Parisa Eslambolchilar, and Matt Jones. Sweep-shake: finding digital resources in physical environments. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, pages 12:1–12:10, New York, NY, USA, 2009. ACM.
- [59] Bernhard Schmitz, Susanne Becker, Andre Blessing, and Matthias Großmann. Acquisition and Presentation of Diverse Spatial Context Data for Blind Navigation. In *2011 IEEE 12th International Conference on Mobile Data Management*, volume 1, pages 276–284. IEEE, June 2011.
- [60] Steven D Schwartz, Jean-Pierre Hubschman, Gad Heilwell, Valentina Franco-Cardenas, Carolyn K Pan, Rosaleen M Ostrick, Edmund Mickunas, Roger Gay, Irina Klimanskaya, and Robert Lanza. Embryonic stem cell trials for macular degeneration: a preliminary report. *Lancet*, Jan 2012.
- [61] Jeff Sinclair, Philip Hingston, and Martin Masek. Considerations for the design of exergames. In *Proceedings of the 5th International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia*, GRAPHITE '07, pages 289–295, New York, NY, USA, 2007. ACM.
- [62] K. Squire. Changing the game: What happens when video games enter the classroom? *Innovate: Journal of Online Education*, 1(6), 2005.
- [63] Daniel Wigdor and Dennis Wixon. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition, 2011.