Understanding and Designing Motion Gesture Interfaces for People with Visual Impairments

Nem Khan Dim

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Engineering

Kochi University of Technology

2016

Supervisory Committee:

Ren Xiangshi

Hamamura Masanori

Takata Yoshiaki

Hoshino Yukinobu

Yoshida Shinichi

Abstract

Although some interfaces have been developed for people with visual impairments, this demographic remains unable to fully access technology. This is because there is still very little understanding of user capabilities that would facilitate interactions with technology, appropriate design approaches, and guidelines for designing successful technology for this user group. Lack of these fundamental knowledge results in the existing technologies less efficient or unable to engage users.

To increase the interaction bandwidth, and to address the aforementioned problems, this dissertation proposes motion gesture interfaces for visually impaired people. Motion gestures interfaces enable users to interact with a device, in three dimensional space by translating or rotating the device, or by moving the hands, or other body parts without holding any devices. We studied (1) user gestures and their rationale, (2) user capabilities to perform motion gestures, (3) participatory design approaches for gesture design, and (4) suitable vibration feedback for motion gestures.

Based on understanding of user gestures and capabilities, we developed and evaluated two interaction techniques: (i) motion gesture interfaces and (ii) motion marking menus in smartphones.

The outcomes of this dissertation are:

1. Understanding of user gestures, gesture rationale and taxonomies
2. New interaction techniques
3. A hybrid participatory design approach
4. Theoretical guidelines

In summary, this dissertation contributes to the field of assistive technologies for visually impaired people, in gesture-based interactions. The conclusions drawn and methodologies proposed will benefit future research studies that explore gesture-based interaction techniques and scientific foundations of assistive technologies for people with visual impairments.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

LIST OF TABLES

ACKNOWLEDGMENTS

Thank God for these amazing three years. Thank mom and dad for inspiring my beliefs and choices in life.

I would like to express my sincere thanks to Kochi University of Technology and my advisor, professor Ren Xiangshi, for giving me the opportunity and supporting everything I would need for my study. Especially, I thank professor Ren for helping me grow as a research scientist. Your advice on both research and my career have been priceless, and they will be lifelong.

I would like to thank my committee members, professor Hamamura Masanori, professor Takata Yoshiaki, professor Hoshino Yukinobu and professor Yoshida Shinichi for serving as my committee members even at hardship. I also thank you for your brilliant comments and feedback to this project.

I would like to thank Dr. Kibum Kim, you have been a tremendous mentor for me throughout the project.

I am grateful to my friend Chaklam Silpasuwanchai for your smart and caring support. I am blessed to have a friend and colleague like you. I would like to thank my best friend, Mizobata Ryo, for making a friendly environment for me on my first days in Japan. You helped me so much for my study and personal life. Thank you so much. I am also grateful to John Cahill for helping my English writings and giving me so much spiritual support.

I would like to thank Kawagoe Chiyo and Morio Yoko for helping my first experiments and your continuous support. I would like to thank my amazing friends, Miyashita Yumi, Matsuoka Hitomi and Hatakenaka Kyoko. You have always been there to support me. You have been precious blessings to my life and my study.

I am grateful to my amazing friends and coworkers, Handityo Aulia Putra and Sayan Sarcar for being a wonderful part of this project and giving me helping hands whenever I am in need. Thank you so much. I would like to express my heartfelt thanks to Wang Zhenxin, Qi Fang and Chen Guanghui for your strong technical supports. I am so blessed to have you.

I would like to thank my research group members, Taiga Machida, Hamada Mitsushiro, Matsuoka Kotaro, Obata Masaki, Ishikawa Ai and Sakaue Takuma. Thank you so much for your contributions to this project. I would not complete this work without you.

I would like to express my special thanks to all Ren Lab members for making a best place to stay here and being an amazing part of my life in Japan.

I would like to thank my participants who always made time and actively participated in my study. Special thanks to you.

I am grateful to IRC staff members of Kochi University of Technology for your kind support.

I would like to express my heartfelt thanks to my family, my friends in Myanmar and Japan, and my relatives. Words cannot express how grateful I am for your prayers and caring love. Your support for me was what sustained me thus far.

At the end, this study has been partially supported by the Grant-in-Aid for Scientific Research of Japan.

DEDICATION

to every person with visual impairments.

Chapter 1

INTRODUCTION

1.1 Research Background

It is estimated that around 39 million people are living with blindness and 285 million people are living with low-vision around the world (WHO, 2014). The population of visual impairments is continuing to increase. In 2002 when 161 million people were living with visual impairments (of whom 124 millions with low vision and 37 millions with blindness), it was predicted that the number of blind and visually impaired people will double by 2030 [59]. Increased aging population in developed countries and the growth of the population in developing countries are one of ongoing causes of concern.

Technology plays an important role in achieving more independent livings of this demographic. Many customized devices have been developed to help visually impaired people perform simple daily tasks. These devices include watches for the blind, braille, tactile tags for clothing colors and others [85]. Such aiding devices can help visually impaired persons to some extent for more independent livings. However, there remain many unfulfilled needs for their satisfactory and productive livings.

While technology is growing rapidly, empowering our daily lives from leisure to productivity, these technologies are not primarily developed for people with visual impairments. This results in unequal access of this demographic to technology. In recent years, interaction styles have largely changed in a form that visually impaired persons cannot easily interact with technology, for example, the use of the Graphical User Interface (GUI). Rapid changes in interaction styles left visually impaired users behind, using old-fashioned and some customized devices.

More recently, as the next generation of user interfaces, interactions are transformed into form of *natural user interfaces (NUI).* NUI refers to a user interface which is effectively invisible to its users, and based on natural elements [103]. Gestures are one key feature of NUI given natural and efficient input styles of gesture-based interfaces. *"A gesture is a motion of the body parts that contains information"* [53].

To increase the interaction bandwidths of visually impaired users, this dissertation focuses on designing motion gesture interfaces of visually impaired people. Motion gestures enable users to interact with a device, in three dimensional space by translating or rotating the device, or by moving the hands, or other body parts without holding any devices. Motion gesture interfaces are particularly desirable for people with visual impairments for three reasons: (1) they are promising eyes-free interfaces, (2) motion gestures provide natural and easy interactions in various situations, (3) motion gestures are one key feature of emerging natural user interfaces.

Despite this potential, very few studies have been done on understanding and designing motion gesture interfaces for people with visual impairments. This dissertation seeks to fill this research gap.

1.2 Research Issues and Objectives

Although some interfaces have been developed for people with visual impairments, this demographic remains unable to fully access technology. This is because there is still very little understanding of user capabilities that would facilitate interactions with technology, appropriate design approaches, and guidelines for designing successful technology for this user group. Lack of these fundamental knowledge results in the existing technologies less efficient or unable to engage users.

To increase the interaction bandwidth, and to address the aforementioned problems, this dissertation proposes motion gesture interfaces for visually impaired people.

The objectives of this dissertation are:

1. To understand user gestures, their rationale and taxonomies
2. To understand user capabilities that will facilitate motion gesture-based interactions

3. To propose new interaction techniques in order to increase interaction bandwidths of visually impaired people
4. To determine effective design approaches for visually impaired users

1.3 Thesis Overview

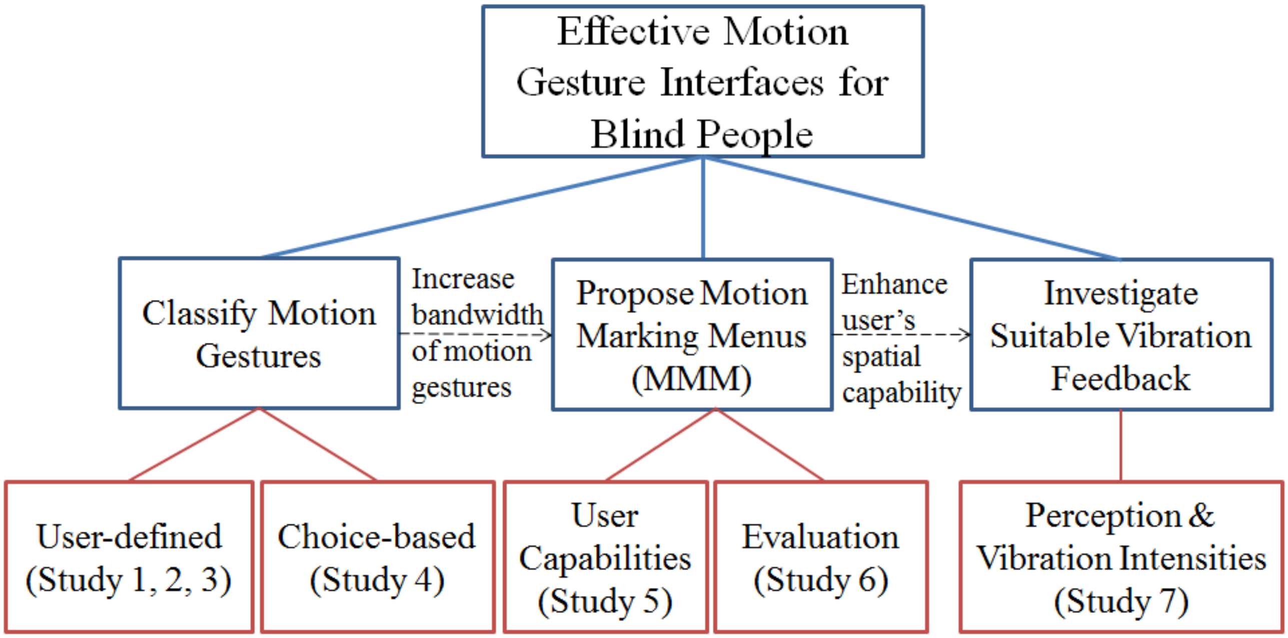


Figure 1.1: Dissertation outline.

1. To understand user gestures and their rationale, we explored the design space by conducting elicitation studies. First, we performed user-defined studies to better understand user gestures. Using the gesture set and design heuristics from the user studies, we implemented and evaluated motion gesture interfaces in mobile phones (Study 1, 2 and 3). From the user-defined studies, we found that users had difficulties defining their own gestures for some commands. Thus, to help users reach their commonalities of gesture preferences, we proposed a choice-based design approach, and demonstrated

the use of a hybrid design approach that leverages user-defined and choice-based design approaches (Study 4).

2. Motion gestures were found to be desirable interfaces for blind people in many situations. However, the bandwidth of motion gestures is somehow limited. To increase the bandwidth of motion gesture interfaces, we proposed motion marking menus, and studied user capabilities to perform motion marking menus. Then we developed and evaluated a motion-based marking menu system (Study 5 and 6).

3. Throughout the studies, we found a need to enhance existing user capabilities (spatial ability) for performing motion gestures, and vibration feedback is a promising feedback for our purpose. Thus, we investigated suitable body positions for wearable vibration feedback. We also investigated suitable vibration intensities for each vibration position (Study 7).

Chapter 2

LITERATURE REVIEW

In this chapter, we discuss (i) definitions of blindness and visual impairments, (ii) assistive interfaces for people with visual impairments, and (iii) the current research gap.

2.1 Blindness and Visual Impairments

The degree of visual impairment varies from total blindness (no light perception) to a slight blurring of vision, with different extent of sight in between [30]. Visual impairments are different in terms of etiology, visual acuity and extent of the visual field [101]. The term *legally blind* refers to a set of criteria for visual impairments based on either low acuity or restriction of the visual field. Legal blindness refers to having a visual acuity of 20/2001[1] or less, or having a visual field of 20 degrees or less. The term low vision is used to refer individuals having a serious visual impairment, while still having some useful vision. Thus, *visually impaired* refers to all the legally blind [101].

The age of onset of blindness also impacts a person's perception, mental representation of spaces and attitude towards blindness. Congenitally blind or early blind refers to a person who has been blind from birth. Adventitious or late blind refers to a person who has lost sight later in life [30].

Visual impairment is due to a variety of causes. The main causes include glaucoma, macular degeneration, cataract, optic nerve atrophy, diabetic retinopathy and retinitis pigmentosa [59]. The prevalence of blindness is much higher for the elderly people than younger people [30]. It was reported that more than 80% of people with visual impairments are with

---

[1] A visual acuity of 20/200 refers to the ability to read at 20 feet what someone with normal vision could read at 200 feet.

ages over 60 years. The aging of the population in developed countries and the growth of the population in developing countries are also causes of concern. It is predicted that the population of visual impairments will be around 320 millions in 2030 [59].

While visual impairments are estimated to continuingly increase, the current trend in technology also allow us to predict the potential of technology aids for this demographic. Thus, there is a pressing need of research studies which concern accessibility of visually impaired people.

2.2 Assistive Interfaces for People with Visual Impairments

We conducted literature reviews on assistive interfaces for people with visual impairments based on two main aspects: (1) inputs, and (2) outputs. Input and output modalities are categorized into several subcategories (see Table 2.1 and Figure 2.1).

2.2.1 Inputs

Input modalities for visually impaired users are classified into seven categories. (i) Surface gestures are where user inputs are performed as gestures on 2D touchscreen devices [9, 19, 31, 112]. Surface gestures were used as input in tactile graphs, media player and other touchscreen devices. (ii) Motion gestures are where users interact with a device, in 3D space by translating or rotating the device, or by moving the body parts without holding any devices. Motion gestures were used for menu selections on large wall-mounted displays [48]. (iii) Body movement inputs are where user movements are treated as inputs to the system [11, 67, 84]. These input modalities were used in navigation systems and drawings for bind users. (iv) Braille inputs are where users perform inputs to systems using braille tactile buttons [5, 23, 87]. Braille inputs were used for mobile public transit systems and haptic Sudoku games. (v) Speech inputs are where users say the desired commands to the system [6, 93, 105]. Speech inputs were used for mobile interactions. (vi) Environmental context inputs were where information from the environment (e.g., road information, weather) were used as inputs to the system [27, 32, 51, 91, 112]. These inputs were mainly used for

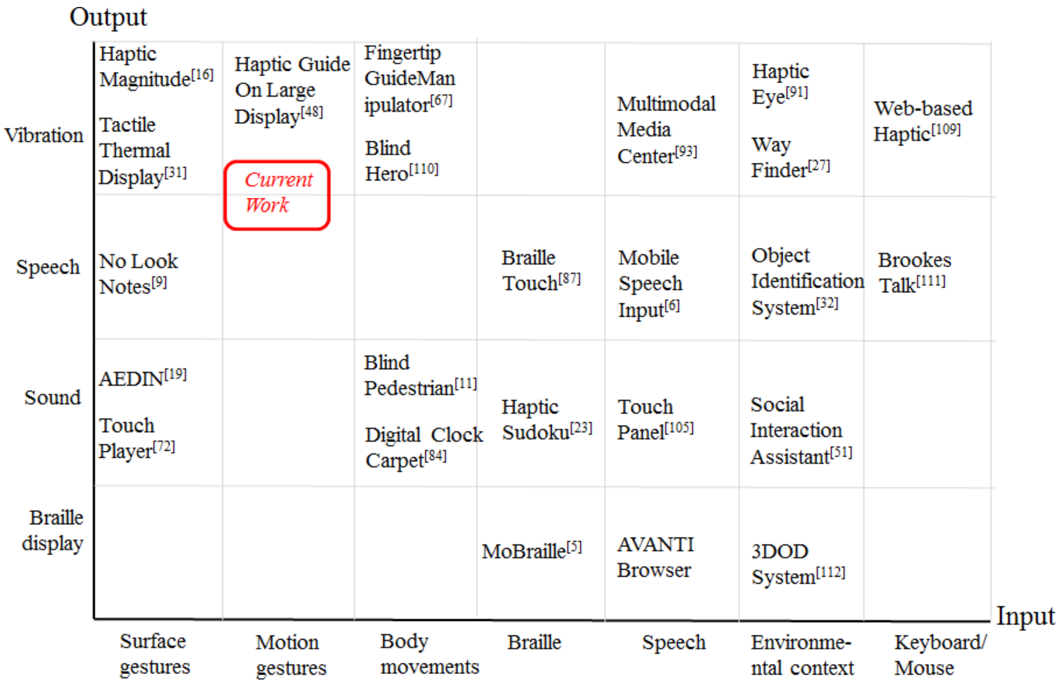
Category	Subcategory	Description
Input	*Surface gestures*	Users input by performing gestures on touchscreens.
	Motion gestures	Users input by moving the device or body parts in 3D space.
	Body movements	Users' movements (e.g., walking) are treated as inputs to the system.
	Braille	Users input by using braille tactile buttons.
	Speech	Users input by saying the commands.
	Environmental context	Environmental context (e.g., road information, weather) are treated as inputs to the system.
	Keyboard/Mouse	Conventional input devices (keyboard/mouse) are used for user inputs.
Output	*Braille display*	Users receive the system response by reading braille.
	Sound	System respond to user inputs through non-verbal sounds (e.g., beep).
	Speech	Users receive system response through verbal speech.
	Vibration	Users receive system response through vibrations which encode information in different patterns.

Table 2.1: Categories of input/output modalities for people with visual impairments.

mobility and orientation aids for visually impaired people. (vii) Keyboard and Mouse are current conventional devices. To be usable these input devices for visually impaired users, systems require non-visual aids and feedback to users [109, 111].

2.2.2 Outputs

Output modalities are classified into four categories. (i) Braille displays are where system responds are displayed through tactile braille [5, 77, 112]. (ii) Sound outputs are where the system responds to users through non-verbal sounds such as beeps [11, 19, 23, 51, 72, 84, 105]. (iii) Speech outputs are where users receive system responses through verbal speech



	Surface gestures	Motion gestures	Body movements	Braille	Speech	Environmental context	Keyboard/ Mouse
Vibration	Haptic Magnitude[16] Tactile Thermal Display[31]	Haptic Guide On Large Display[48] *Current Work*	Fingertip GuideManipulator[67] Blind Hero[110]		Multimodal Media Center[93]	Haptic Eye[91] Way Finder[27]	Web-based Haptic[109]
Speech	No Look Notes[9]			Braille Touch[87]	Mobile Speech Input[6]	Object Identification System[32]	Brookes Talk[111]
Sound	AEDIN[19] Touch Player[72]		Blind Pedestrian[11] Digital Clock Carpet[84]	Haptic Sudoku[23]	Touch Panel[105]	Social Interaction Assistant[51]	
Braille display				MoBraille[5]	AVANTI Browser	3DOD System[112]	

Figure 2.1: Input and output techniques for visually impaired people.

[6, 9, 32, 87, 111]. (iv) Vibration outputs are where users receive system response through vibrations which encode information in different patterns [16, 27, 31, 67, 91, 93, 109, 110].

2.3 *Summary of Research Gaps*

The literature reviews showed that considerable number of studies have been done on assistive interfaces for visually impaired people. Despite this, current technologies are limitedly accessible and less engaging to visually impaired users. This is due to the need of better understanding of users, which will guide to designing and developing more efficient and engaging technologies for them. Our reviews also revealed that very few work has been done on understanding users and user capabilities which will facilitate their interactions with technology. This dissertation aims to address this research gap, and propose new interaction techniques for visually impaired users which are based on understanding users.

Chapter 3

CLASSIFYING MOTION GESTURES

3.1 Introduction

To understand user gestures and their rationale, we explored the design space by conducting elicitation studies. First, we performed user-defined studies to better understand user gestures. Using the gesture set and design heuristics from the user studies, we implemented and evaluated motion gesture interfaces in mobile phones (Study 1, 2 and 3). From the user-defined studies, we found that users had difficulties defining their own gestures for some commands. Thus, to help lead users to reach their commonalities of gesture preferences, we proposed a choice-based design approach, and demonstrated the use of a hybrid design approach that leverages user-defined and choice-based design approaches (Study 4).

3.2 Related Work

Studies related to our first study include the classification of human gestures, gesture interactions for blind people, user-defined studies and choice-based elicitation study.

3.2.1 Classification of Human Gestures

Gestures are spontaneous movements of hands and body parts that people do to express their mental images and thoughts. Gestures can convey communicative information as speech [74], and they are performed in parallel with verbal expression at the conceptual level of communication [75]. Gestures serve some cognitive functions for communication, and they are performed regardless of individual abilities or the impacts produced by the gestures generated. Even individuals who have been blind from birth and have thus never seen anyone elses gestures spontaneously express themselves in gestures. Studies have shown that blind people gesture even when they are conversing with other blind people [35].

Poggi (2001) proposed a procedure for the generation of gestures. Gestures are generated taking into account the meaning intended and the cognitive construction of the gesture to be made. When people gesture to communicate some meaning, codified gesture is first considered. Codified gestures are gesture-meaning pairs constantly represented in the mind and standardized by repeated use, regulations and social conventions. Such gestures are spontaneously shared and understood by everyone. If a codified gesture is not readily found in memory, creative metaphoric gestures are resorted to. Evidently, these metaphoric gestures are generated by mimicking daily human actions (i.e. biological) or by the similarity in visual resemblances (i.e. iconic). If the intended meaning is information in a humans mind, gestural mind makers that can represent beliefs, goals and emotions of the human mind are invoked. Finally, gestures are arbitrary when signals and meaning are not linked by a relationship of similarity nor by any other kind of relationship. In the literature of Human Computer Interaction, gesture-based interactions have become a leading trend in natural user interface development [46]. Studies including [104] and [78] investigated user gestures for different applications. Despite the general classification of human gestures discussed above, the gesture taxonomies presented by user studies varied depending on the applications and users. Thus, in order to design motion gestures for mobile interactions that are usable by blind people, it is vital to understand the gesture producing mechanisms and gesture taxonomy of this user group.

3.2.2 Gesture Interactions for Blind People

Previous studies on defining suitable gestures for blind people mainly focused on touchscreen surfaces in terms of enhancing hardware to provide access to touch screen kiosks (e.g.,[55, 99]) and exploring interaction techniques for mobile touchscreens ([9, 21, 43, 81, 83]). Kane et al. [45] performed a touchscreen gesture elicitation study and compared the performance of sighted with the performance of blind people. They found that blind people have different gesture preferences, and their ways of performing gestures are significantly different (e.g., the gestures of blind people tend to be noticeably exaggerated).

A few studies also explored motion gestures to interact with mobile devices. Ruiz et al. [78] presented an elicitation study of motion gestures for mobile interaction. The study reported a consensus of motion gestures to invoke commands on a smartphone and presented a taxonomy of motion gestures. However, the study did not consider gestures of people with visual impairments. More recently, Dim and Ren [15] studied motion gestures for blind people in mobile phone interactions. They found that blind people mostly perform metaphorical gestures reflecting actions from their daily activities. Thus, such gestures had a high consensus for familiar activities related to daily activities such as *making phone call* and *voice search*, but they had a low consensus for unfamiliar commands such as *zooming*.

Although some studies have been done for touchscreen gestures and motion gestures in mobile phones for blind people, very few works have been done on suitable mid-air gestures. Thus, we conducted user studies to determine suitable mid-air gestures for blind people, especially for TV interactions.

3.2.3 User-defined Studies

User-defined study has been a recommending and maturing practice in human-computer interaction research. The core idea of this human-based approach is that users must be understood so that the system can be adapted to the users instead of requiring the users to adapt to a given interface. The major benefit of a user-defined study is the higher likelihood of designing interfaces that are easy to perform and to remember.

Many user-defined studies have been conducted especially for gesture-based natural interactions. Wobbrock et al. [104] presented a user-defined study where participants were shown the effect of a gesture and then asked to perform gestures for commands in surface computing. Inspired by the contributions of this study that include gesture taxonomy and implications for surface computing, many similar studies have been performed for various applications and computing environments. These include user-defined studies for device to device interactions [50, 52], surface and motion gestures for 3D manipulation of objects through mobile devices

[62], human-robot interaction [70], free-hand TV control [100], deformation -based interface [58], augmented reality [73], and gesture sets for people with communication difficulties [42].

Kane et al. [45] presented a gesture elicitation study where touch screen gestures performed by blind and sighted people were compared. This study found that blind people have different gesture preferences to sighted people, and reported design guidelines for accessible touch screen interfaces. The inclusive design guidelines presented by this study are specifically accommodated to touch screen devices. On the other hand, Ruiz et al. [78] presented an elicitation study of motion gestures for mobile interaction. This study reported a consensus of motion gestures to invoke commands on a smartphone and presented taxonomy of motion gestures. However, the study did not cover people with visual disability even though blind people are one of the largest potential user groups of motion gesture interfaces. Blind and sighted people have different visual experiences and daily activities that can affect the expression of their mental images as gestures. We identified the differences between gestures performed by the two user groups through gesture taxonomy and user-defined gesture sets, based on the previous study [78].

3.2.4 Choice-based elicitation Method

To our knowledge, only a little study has employed choices for gesture design, with the exception being the work of Silpasuwanchai and Ren [86]. They employed choices to understand how gamers perform simultaneous gestures during intense gameplay. However, there was no deep discussion regarding the use of choices for gesture design.

Where users have little or no idea of the range of possibilities for suitable gestures, employing designer-selected choices can serve as one way to non-threateningly introduce designer ideas into the design process without making the users feel compelled. In addition, allowing users to select their preferred choices can reduce the chance of designers introducing overly-complicated gestures, as suggested by Morris et al. [64]. This also increases the chance of settling on a gesture that is suitable in both the view of designer and that of the users. Informed by this potential benefit, we adopted the choice-based elicitation approach for our study.

In summary, while designing suitable gestures for blind people is an ongoing challenge, our study leverages the benefits of both user-elicitation and choice-based elicitation to determine the most suitable mid-air gestures for blind people.

3.3 Study 1: User-defined Motion Gestures in Mobile Phones

For people with visual impairments, mobile devices have become indispensable, empowering them for both leisure and more independent living. Mobile phones are the most commonly carried devices by blind people in their daily lives [44]. However, most people in this user group prefer old-fashioned mobile phones with familiar layouts and tactile buttons to modern smartphones. Touch screen interfaces in smartphones primarily require users to look-for-interaction sensitive areas on the screen and that is a major challenge for blind people because the interfaces on the screen are invisible to them. Existing assistive solutions such as screen readers, vibration, sound and speech output, are still not adequate or efficient enough to deliver all smartphone affordances to blind users. These interfaces require the blind user to memorize and browse soft-buttons on the screen. Nowadays, motion gestures have been gaining attention as more natural and intuitive interfaces that support distracted inputs and require less visual attention. We believe that motion gestures designed to logically map the users' mental model can offer more learnable and accessible interfaces for blind people.

Most blind people are not smartphone users and they are not aware of the affordances available in smartphones nor of the potential interactions available through motion gestures. Furthermore, they have less or no chance to see motion gestures performed by other people. We were thus motivated to find the best practices to design the most usable motion gestures for blind people. With respect to this research goal, we conducted a user-defined study where the participants were asked to define their own gestures to invoke some common tasks in a smartphone, and to mention the rationale and heuristics for the gestures they performed. We set three research questions for the user-defined study: 1) Are motion gestures usable as mobile interactions for blind people? 2) What unique motion gestures do the special characteristics of blind-users inspire for mobile interactions? 3) What are the heuristics of gestures produced by blind people, and how can these heuristics be described as principles of gesture design for blind users.

User-defined studies are conducted to elicit user behaviors to enhance the design process. During the study, we explored motion gestures from 13 blind participants. The participants were asked to perform motion gestures that could be used to command the available functions on a smartphone. We presented 15 tasks to the participants. The participants were asked to use a think-aloud method and to perform gestures in think-aloud protocol . They were also asked to supply subjective preference ratings for each gesture they performed.

3.3.1 Participants and Apparatus

For the experiment, we recruited 13 blind people from a local blind association (9 males and 4 females). The ages ranged from 25 to 77 (Mean = 61, SD = 16.91). Three of them could see light, two could see objects but none of these were able to distinguish between objects. The rest were totally blind. Two of our participants were smartphone users. All the participants were right-handed. Each was paid $10 for their participation.

We used a Samsung Galaxy smartphone to define the participants motion gestures. Participants were video recorded while performing gestures and two experimenters took detailed notes for the think-aloud data

3.3.2 Experimental Tasks

The tasks were categorized into action and navigation. Each was subcategorized into phone application (e.g. answering a phone call or switching to a previous application) and particular applications (e.g. navigating a map). We paid specific attention to explaining some tasks like zooming functions in map navigation. For example, just saying the function zoom in as for enlarging objects would be unreasonable for our participants because they cannot see objects. Instead, we made them understand that enlarging an object (an onscreen menu or a location point on a map) on a screen can help them more easily target or select that object.

Furthermore, two of our participants with low vision used a magnifying glass; this encouraged us to include these tasks. The experimental tasks we presented to each participant are described in Table 1.

3.3.3 Procedure

We started each experimental session by explaining the purpose and procedures of the study and also the think-aloud method to the participants. Then the participant was handed a smartphone and asked to perform a gesture for each task. The tasks were grouped into three sets of similar tasks. For example, tasks for normal use of the phone such as calling, answering and muting a call were grouped into the same set. Because our participants could not be presented with any visual description, an experimenter read aloud the descriptions and explanation about each task carefully.

Because repeating every defined gesture for confirmation was tiring, we didnt ask the participants to repeat their gestures. Instead, we carefully captured the expressed gesture and tried to confirm each gesture and rationale. After each group of tasks, the participants were asked to rate the gestures they performed using a 7-point Likert scale to indicate their agreement on the criteria (1 for strongly disagree and 7 for strongly agree):

- The gesture I made is a good match for its intended use.
- The gesture I made is easy to perform.

In order to assess the impression on using motion gesture interactions of our participants, we added an item, I would often use this gesture if it existed on a seven-point scale ranging from never (1) to very often (7). To conclude each session, we asked the participants to suggest additional use cases. If they had any suggestion, we encouraged them to perform gestures for the tasks suggested. The interview ended with the experimenter asking the participants if they had any questions, suggestions or comments. We recorded every comment or suggestion of the participants for later analysis.

3.3.4 Results

After detailed analysis of the data we collected from our study, we presented the study results including a user-defined gesture sets, motion gestures of the blind, physical characteristics of the gestures, subjective responses and open-ended use cases.

Agreement Scores and User-defined Gesture Set

From the gestures collected, we grouped identical gestures for each task and selected the largest group as the user-defined gesture for the task. We adopted Wobbrock et al.s method [6] in order to investigate the extent of agreement for each task. We calculated agreement scores for each task using the formula:

$$A_r = \sum_{P_i} \left(\left| \frac{P_i}{P_r} \right| \right)^2 \tag{3.1}$$

where P_r represents all gestures performed for event r and P_i is a subset of identical gestures from P_r. A_r ranges 0 to 1. Gestures were considered to be identical when they had similar trajectories and poses. Figure 3.1 shows the agreement score for each task.

The user-defined gesture set is shown in Figure 3.2. Among use-cases suggested during the interview, we included gestures for *Pause* and *Power Off* in the gesture set because most of our participants performed gestures for these commands and there was obvious consensus among the participants. On the other hand, among the experimental tasks, gestures for zooming (i.e. *Zoom In* and *Zoom Out*) could not be presented in the user-defined gesture set. Although the participants used similar rationales for creating gestures, the gestures were still performed differently. Unlike with sighted people, blind people cannot have visual references that are shared with others. Thus, different gestures were generated being influenced by daily experiences.

Being encouraged to freely express the most inspired gestures, our participants had the same preferences of non-motion gestures for some tasks. For example, gestures for *Ignore Call*, *Pause* and *Power Off* were seen to cover or sweep the phone screen by hand with the

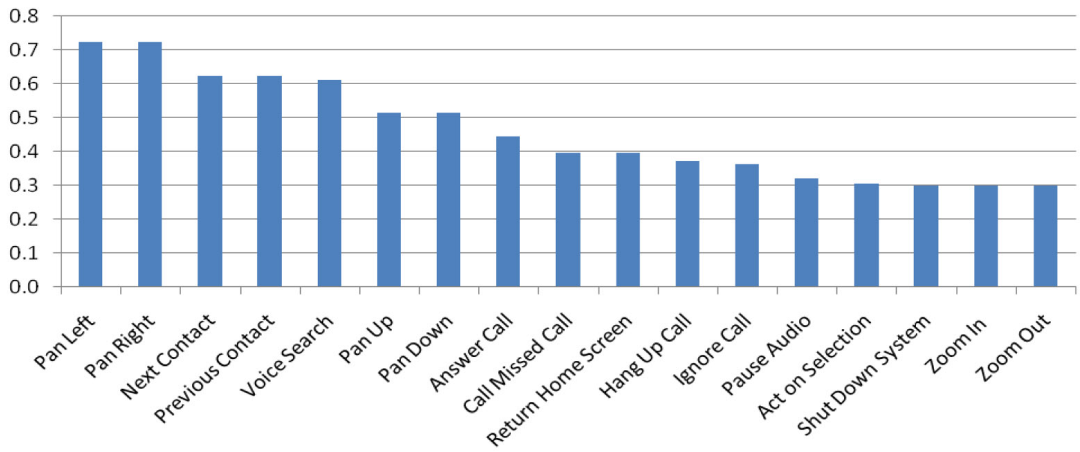


Figure 3.1: Agreement score of each task, sorted in descending order.

common reason being ‘stop the sound’ or ‘finish’. This informed us that the most ergonomic interactions need combinations of different modals optimized for a specific context.

Motion Gestures of Blind People

We analyzed the motion gestures collected and grouped the gestures into four-fold taxonomic themes. The themes include natural and intuitive gestures, real-world metaphors, natural consistent mappings, and arbitrary gestures.

1) Natural and Intuitive Gestures. As noted in the literature reviews [2, 3], some codified gestures are constantly presented to peoples minds. Here the natural gestures we called are subsets of codified gestures. These gestures are more natural than every other gesture for the intended meaning, and the purposed meaning can be inferred without learning. For example, bringing the mobile phone to the ear for answering a call, or bringing the phone to the mouth for making a voice search are standardized by repeated use, thus intuitive motions to everyone. Evidently, most of the participants designed the same motion gestures for tasks that have such codified gestures. In our study, for making a voice search, 8 out of 13 participants designed their motion gestures by bringing the smartphone to the mouth.

Figure 3.2: A User-defined gesture set. Commands for zoom in and zoom out are not included due to the lack of agreement among the participants. Blue curve shows front-back movement of shake, black arrow indicates state change of the phone or direction of movement, and brown-bold arrow indicates the direction of rotations.

The common reason for choosing that gesture was described as ‘natural’.

2) Real-world Metaphors. For the tasks where codified gestures cannot readily be found, the participants tried to generate creative gestures. It is obvious that generating creative gestures is primarily influenced by real-world metaphors which occur in daily lives. Blind people are primarily influenced by what they do on a daily basis. Gestures performed by blind people are linked to their meaning by mechanical determinism (i.e. daily action) rather than by similarity of visual resemblance (i.e. iconic). For example, to navigate to the home screen, half of our participants designed their gestures to flip the phone screen back to front (i.e. undo). The most common reason for that motion gesture was ‘returning to the original place’. Also, being asked to perform gestures for zoom in, some of our participants designed their gesture to continuously rotate the phone up along the x-axis. Some others performed gestures by raising the smartphone tending to increase height of the device. For design reason, a majority of the participants shared the opinion of one of the participants: “Enlarging an object is like increasing the level of volume in a music player. It is increasing a level”.

We also noted that blind people tend to create gestures when mimicking the normal use of different devices that they use daily. For example, for panning tasks in map navigation, one of our participants mentioned his design rationale was to mimic the use of the cardinal directions in his compass.

3) Consistent Mapping. In general, consistent mapping is where opposed actions are achieved by reversed movements. For example, if a rotate flick to the right is gestured for *Next*, the gesture for *Previous* is a rotate flick to left. In our study, we found that consistent mapping was mentioned very often by the participants as a design rationale for the gestures they made. For example, while performing gestures for *Hang-up Call*, one of the participants flipped the smartphone forwards and stated his design rationale as: “I flip the phone backward and bring to the ear for answering a call. Then I will remove the phone from the ear and flip forward to hang-up the call”.

Again, some other participants were found to generate related gestures for *Ignore Call* and *Place Call*, *Pause* and *Power Off*, etc. We speculated that blind people are more likely to use consistent mapping than sighted people to arrange tasks in more memorable and accessible ways. This occurred more obviously when suggesting open-ended use cases and performing gestures for the suggested tasks. It is arguable that consistent mapping is a way of grouping or relating gestures but not a specific type of gesture in and of itself. However, appreciating this often-used methodology for creating gestures by our participants, we presented consistently mapped gestures as a specific gesture type for this context.

4) Arbitrary Gestures. Arbitrary gestures are those not linked to the meaning by either the similarity of real-world metaphor or by any other relationship. During the study, we found that most of the gestures performed by our participants were labeled with a relevant rationale that came from their daily experiences. However, in some cases, some of the gestures performed with a rationale were still difficult to infer without learning the reason. We treated those gestures as arbitrary gestures.

Figure 3.3 shows the taxonomic decomposition of 208 gestures collected during our study. We classified the user gestures by the gesture types discussed above. As seen in the figure, most gestures are metaphoric gestures generated from daily experiences. Consistently mapped gestures also made up a considerable percentage of the gestures produced. Most task suggestions during open interviews came from consistent mapping, including use-cases to *Pause* the screen reader or music player, *Resume*, *Close* and *Power Off* the system.

Physical Characteristics of Gestures

Regarding the physical characteristics of gestures, we found that our blind participants used large movements of the hands to produce gestures, tending to have high kinematic impulses. We paid attention to the physical characteristics of gestures performed by our participants and used the videos recorded during the study for close analysis. We defined motions performed using only the wrist as small gestures, and defined those performed using both the wrist and elbow as large gestures. Our participants were found to mostly perform large gestures. This can be partially due to the lack of feedback in our study which would have

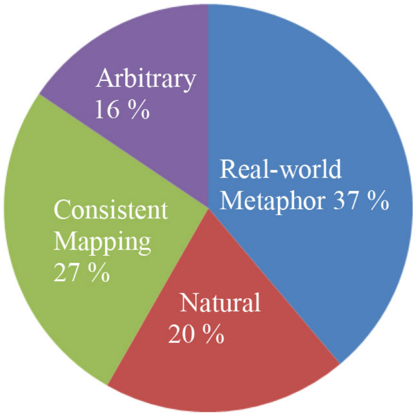


Figure 3.3: Percentages of gestures including in each gesture type.

confirmed that their gestures had been recognized. Despite this, we speculated that blind people are more likely to use large movements as they treated the movements themselves as feedback to their actions instead of visual feedback. In any case, it is still questionable whether the participants will still use large movements to create gestures when feedback is provided. We answer this question in the second study.

Because the physical characteristics of gestures are one of the main concerns when designing interfaces and supporting sensors, we also analyzed the physical characteristics of the gestures in terms of dimension and complexity. Here dimension means the number of axes involved in the movement while performing gestures. Single-axis motions include simple gestures like a flip or a flick. Motions that include a single rotation or translation are tri-axis gestures, and those including both translations and rotations of the device around its six degrees of freedom are defined as six-axis gestures. The complexity is concerned with whether the gesture is a simple single gesture or a compound gesture that is composed of more than one single gesture. Figure 3.4 illustrates the percentages of gestures in each category of dimension and complexity. The gestures tend to include more translations and rotations than simple single motions.

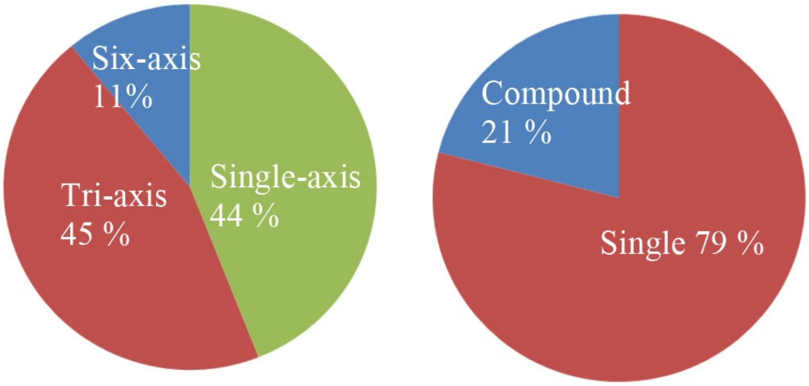


Figure 3.4: Physical characteristics of gestures.

Subjective Responses and Open-ended Use Cases

After each session of the experiment, the participants rated the gestures they performed in terms of goodness, easiness and frequency of use. Overall, the participants gave the gestures average scores of 6.14 (SD = 1.069) for goodness, 6.26 (SD = 0.41) for easiness and 6.10 (SD = 0.81) for frequency. All the ratings were found to be equally and relatively high. This convinces us that our participants were very receptive to motion gesture interactions.

After the experiment, we asked the participants to suggest possible tasks where motion gesture interactions can be effective for them. We received considerable suggestions from some of the participants. One of them suggested *Application Pause* mentioning that "Sometimes I want to pause the screen reader of my phone on the way. It is very annoying when I cannot easily find the button to do that". The same participant also suggested considering a motion gesture for *Power Off* stating the same reason i.e., difficulty in finding buttons. Another considerable use case was proposed by one participant who mentioned that, "It would be very great if I can define my preferred gestures for calling frequently used numbers. Sometimes, finding only 3 numbers for an emergency call is still difficult for me". Similarly, another participant mentioned that, "I use my phone more as a music player than making calls. It would be great if I can control the player using motion gestures".

Once we obtained tasks suggested by some participants, we asked them to perform gestures for the tasks they suggested. We then treated suggested use cases as additions for the next sessions asking the participants to define gestures in the think-aloud method, in order to find design heuristics for possible tasks. After performing a gesture for *Pause*, one participant stated, "If I can pause the audio player, motion gesture will be useful to resume the player". The same participant commented, "Activating the screen reader (i.e. Act on Selection) is like pressing the OK button or the Enter key. So, having Cancel and Close Program functions would also be useful".

As mentioned before, most of the use cases suggested came from relating similar or opposed functions, for example, *Pause* and *Resume.* Through the interviews for open-ended use cases, we noted that our blind participants were very alert to the possibility of motion gestures for more accessible mobile interaction. This awareness accordingly brings new potential and challenges to designers of smartphone applications and vendors of smartphone devices.

Differences from Gestures of Sighted People

Ruiz et al. [78] performed a user-defined study for motion gesture interaction that was applied in sighted people. Since blind people and sighted people have different visual experiences, we hypothesized that they will have differences in the gesture generation process and in the gestures themselves. Because we could not have full access to the data set of the previous study by Ruiz et. al, it is difficult to do a detailed comparison between the gestures of blind people and sighted people. However, it is worth highlighting some significant differences through user-defined gesture sets and gesture taxonomy.

1) Gesture Taxonomy Ruiz et al. [78] presented four-fold taxonomic themes for motion gestures in their study with sighted people: real-world metaphor, physical (direct manipulation), symbolic (visually depicting a symbol) and abstract gestures. Indeed, symbolic gestures and physical gestures primarily rely on visual capability. Our blind participants did not perform symbolic and physical gestures, and we could not include them in the taxonomic

themes of gestures. Instead, we added natural gestures and consistently mapped gestures in the taxonomic themes of our study.

2) Gesture Generation Process In Ruiz et al.'s study [78], for *hanging up the call*, the user-defined gesture of sighted people was found to remove the phone from ear and rotate the screen like hanging up the phone receiver on a telephone. None of our participants mentioned rotating the screen for hanging up the call because they rarely consider whether the screen is facing back or front. Rather, the majority agreement was just removing from the ear and putting somewhere. This implies that blind people may not be aware of some visually demanding actions that sighted people do. Also, sighted peoples gestures were found to mimic the use of a magnifying glass for *Zoom in/ Zoom out* tasks, and shaking the phone to *Return to home screen* with the reason of clearing current contents. On the other hand, gestures performed in our study were linked to the meaning by mechanism determinism, for example, drawing circles for zooming and flipping the phone front and back for returning to home screen. Again, Ruiz et. al reported that most of the gestures performed by their participants were slight flips because the participants concerned about the visual feedback on the screen. On the other hand, the physical characteristics of gestures in our study were large movements including more rotations and translations. This implies that the differences in visual capability, daily experiences, daily device uses and expected feedback make differences in the gestures performed by blind and sighted people. These differences are worth taking into consideration when designing motion gesture interfaces.

3.3.5 Discussion

In this section, we discuss some broader implications of our results for motion gesture interactions and mobile interactions. For blind users, one of the primary reasons for accessibility problems in smartphones is the difficulty they have learning the interfaces on an invisible screen. For the sake of learnability and memorability of motion gestures for blind people, designers should consider consistent mapping where related or reverse gestures are available for similar or opposed tasks. For example, designing relative gestures for *Pause*, *Close* and

Power Off can make the interfaces more learnable because more logical to the blind.

As noted, gesture generation by blind people was primarily influenced by metaphors from daily life, thus we argue that the most usable and memorable motion gestures are those designed to best reflect real-world metaphors from the users' daily lives. This design implication also indicates the need of the participation of representative blind users in design processes. Also, designers should pay specific care not to include gestures that are unexpected by these users. As discussed earlier, blind people are sometimes not aware of the visual-based actions that sighted people perform (for example, a gesture for hanging up the call). Designers should avoid gestures including these kinds of actions. Wherever possible, symbolic gestures and the direct depiction of visual objects should also be avoided.

Regarding the physical characteristics of gestures, we found that our participants used large movements to generate gestures so that their gestures were undoubtedly recognizable enough. With respect to the physical characteristics of gestures performed by blind people, demand for motion accuracy should be reduced. This means that the gesture recognition and the supporting sensors should allow flexible freedom of movement to perform gestures for interactions.

The user-defined gesture sets of our study also informed us that the most ergonomic interactions need combinations of different modals optimized for a specific task or context. Thus, various sensors integrated with todays smartphones should be used to support multimodal inputs and outputs for the most intuitive and natural interactions. Recalling suggestions in open-ended interviews, gesture customization is a very acceptable and beneficial interface options for blind users. More customizable motion gestures should be available for simple tasks on smartphones.

3.4 Study 2: Motion Gesture Interfaces in Mobile Phones

In this study, we investigated the usability of motion gesture interfaces implemented by gestures suggested in Study 1. In this study, the participants used both a smartphone with a motion gesture interface and a feature phone with a button interface, to browse contacts and make calls. The study was motivated by three research questions: 1) do motion gesture interfaces provide more efficient use of smartphones compared to traditional feature phones? 2) what do blind users actually experience using motion gesture interfaces? 3) what design implications can be learned for smartphone assistive interfaces?

Besides the user defined motion gestures set, Study 1 also provide some design insights to consider when developing interfaces for the second study. We identified three common guidelines for designing the motion gesture interfaces for smartphones. First, feedback was provided to every gesture the participant made. In the first study, gestures were found to be created using large movements. It was questionable whether the lack of feedback affected the gestures performed and if the participants still use large movements for gestures when feedback are provided. The interfaces in the second study were thus designed to provide vibration feedback or speech feedback to each gesture input. Second, motion gestures were designed to minimize the need for motion accuracy. This is related to the first consideration. The physical characteristics of gestures performed in the first study suggested that we should allow users more freedom of movement for doing gestures. Third, motion gestures were designed following consistent mapping. Understanding the users' reliance on consistent mapping in the first study, we designed the motion gestures to consistently map wherever possible. Figure 3.5 describes the interactions used, and Table 3.1 describes feedback provided to each step of experimental task.

Following the design guidelines above, we developed a set of interactions that allowed users to browse a contact list and make a call non-visually. The task selection for this study was based on the fact that browsing a phone book and making a call are two of the most common and fundamental functions available on a smartphone. Although there are advanced functions available on smartphones, even the most basic functions such as receiving and making a call still impose limitations on visually impaired people. These limitations cause other smartphone affordances to be out of reach for visually impaired users. We were motivated to investigate how motion gesture interactions enable this user group to access the most basic smartphone functions.

Figure 3.5: Motion gesture interfaces of making a call. (1) A flick gesture to browse phone book, (2) Flip motion to right is used to browse contact, (3) Flip motion to left is used to browse previous contact, (4) Flip backward to select the contact and make call, (5) Flip forward to hang-up the call.

3.4.1 Participants and Apparatus

Eight participants (2 females, 6 males) were recruited. The ages ranged from 22 to 49 years (Mean = 28.37 years). Three of the participants were blind and 5 of them were blind-folded during the study. All of the blind participants had participated in the first study. 2 of them are totally blind and one can see light. All participants are right-handed. Each was paid $10 for their time and effort for the experiment.

For the study, we used a Samsung Galaxy Nexus smartphone and Panasonic EZ180 feature phone. For the motion gesture interface in smartphone, custom software was running to recognize the participants' gesture and perform corresponding functions. The custom software read the value of accelerometer sensor on the smartphone and the sensor values were mapped to the functions. The system was developed in Java Eclipse IDE. All the experimental sessions were video recorded.

3.4.2 Procedure

Each experiment session started with a practice session where the two systems in smartphone and the feature phone were explained and demonstrated.

In each trial, the participant started by browsing the contact list. Contact lists were

Step	**Feedback**
Phone Book	Vibration to confirm the participants that the gesture is recognized
Next/Previous	Speech reading out the contact names
Make call	Vibration
Hang-up call/Home screen	Vibration

Table 3.1: Feedback provided to each step of the experimental task.

set up in the same order in smartphone and feature phone. The participants selected the contact name instructed by the experimenter by going next and previous. When the contact has been selected, the participant made a call and hold the phone for 5 seconds. They hung up the call after 5 seconds. The procedures occurred with the same steps and in the same order in both systems.

The participants were allowed to practice on the systems until they could successfully perform the tasks. At the end of the practice session, the participant was handed the smartphone and asked to perform the experimental task. After completing the experimental task with the smartphone system, the participant was handed the feature phone and asked to perform the same task. The participant performed two trials for each system.

During the study, task completion time and errors were recorded for each trial. Task completion time was the time elapsed from the moment the participant started browsing the contact list until they hungup the call. After all trials, participants were asked to complete a questionnaire about the two systems. Participants indicated their agreement with three statements about each system using 7-point Likert scale (1 for strongly disagree, 7 for strongly agree) based on ease of use, user satisfaction and fatigue to use the system.

The experiment ended with the experimenter collecting all subjective comments and suggestions from the participants.

3.4.3 Results

Task completion time, errors and subjective ratings were analyzed.

Task Completion Time and Errors

Task completion time and errors were recorded during the study. The participants completed the tasks faster with the motion gesture interfaces than with the button-based interfaces (F(1,7) = 8.76, $p < 0.05$). The mean time for task completion using the motion gesture interface was 16.56 seconds (SD = 2.35) while the mean time for button interface was 24.75 seconds (SD = 9.17). During the study, half of the participants started using the smartphone system while the other half started using feature phone. However, there was no learning affect or significance occurred in performance of the participants. Figure 3.6 shows task completion time using each system.

Overall, very few errors were made using either system. The difference in error rates between the two systems was not statistically significant. The mean values of errors were 0.25 for the motion gesture interface (SD = 0.462) and 0.50 for the button interface (SD = 0.93) respectively. Blind-folded participants made more errors with button interface because they had less experience using the feature phone. But blind participants were able to use both systems without any errors.

Subjective Assessment

After the experiment, the participants completed questionnaires about the two systems. Among the three questions, 'ease of use' was found to have statistical significance (F(1,7) = 6.818, $p < 0.05$). Participants indicated that using the motion gesture interface was easier than using the button interface. The questions for 'ease of learning' and 'less fatigue' revealed no significant differences between the two systems. Figure 3.7 shows the subjective responses of participants to the two systems.

Also, following all the experimental sessions, participants answered the questions about

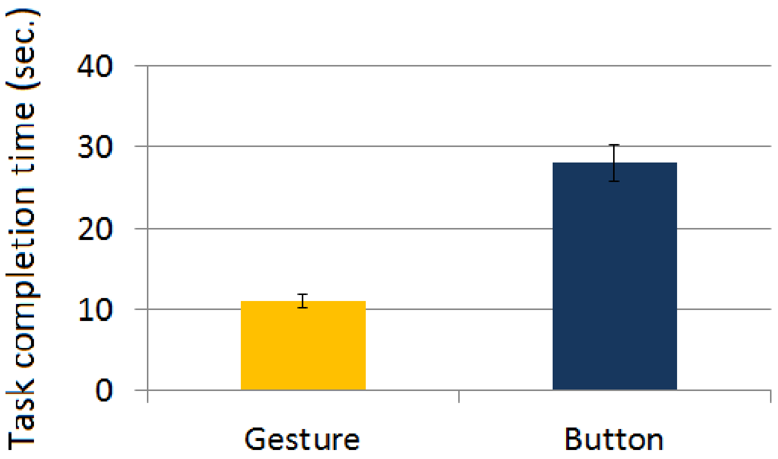


Figure 3.6: Task completion times for the two systems..

their experiences using the two systems and any suggestions.

Motion Gesture Interfaces. Participants commented positively on the motion gesture interface. They mentioned that the motion gesture interface was more natural, and easier to remember and to use. As expected, blind-folded participants who are smartphone users preferred the motion gesture interface to the button interface. Of the three blind participants, two preferred the motion gesture interface. One of them was neutral or positive about both systems, because she was familiar with feature phone. One negative comment about the motion gesture interface was that it would become tiring while browsing many contacts. Furthermore, one of the blind participants commented that grouping contacts may be useful to reduce browsing time. He also commented that the touch-fling gesture for opening the contact book was very comfortable. One of the blind participants also commented that he was never confident using a touch screen device, but motion gestures would be useful for blind users.

Button Interfaces. Most of the participants agreed that the button interface delivered slow response times compared to the smartphone system, and they had difficulty finding the buttons. Blind participants who are feature phone users mentioned that many buttons were arranged on a small space and it was difficult to distinguish the buttons.

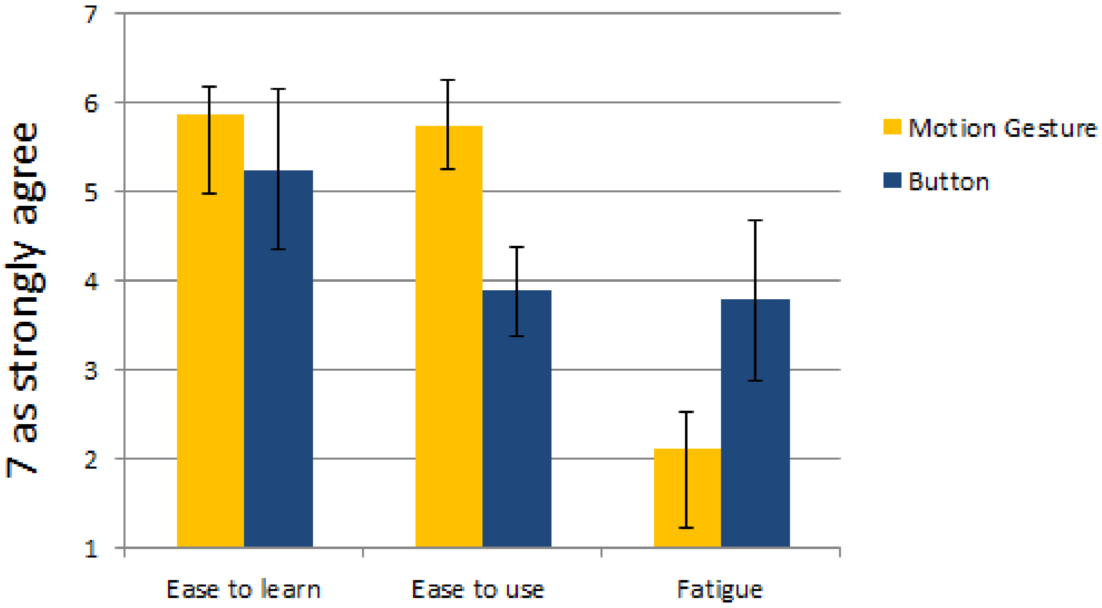


Figure 3.7: Subjective assessment of users on two interfaces.

Feedback. One noticeable comment from all participants was that vibration feedback was very understandable and comfortable. In the study, vibration feedback was provided to every action that had no speech output. The participants mentioned that vibration feedback made them sure that the gesture they performed was recognized by the system.

3.4.4 Discussion

The study results convinced us that motion gesture interfaces can offer a successful mobile interactions that enable non-visual interaction for blind users. The positive comments about the motion gesture interface primarily focused on its simplicity and quick access. However, according to analysis of the qualitative feedback, we must declare that motion gesture interfaces are effective mostly for discrete tasks such as *Making call*, *Hang up call*, *Enter/act on selection*, *Cancel*, *Resume*, etc. Therefore, for continuous tasks such as navigating and scrolling, motion gesture interfaces should be combined with other interaction techniques that are efficient for supporting continuous tasks.

Positive comments about vibration feedback also provided us with insights for non-visual output in mobile contexts. The use of vibrations can be extended to convey rich of non-visual information to blind users. Various kinds of information can be encoded in different patterns of vibration. However, today smartphones are usually equipped with only one vibration motor. We hope that smartphone vendors will equip more vibration motors that provide richer interactions in the future.

Regarding the issue of feedback and gesture size from the first study, we paid attention to the participants' gestures when provided with vibration feedback and speech feedback. For this purpose, we performed close analysis of each participant profile using the videos recorded during the study. We found that most of the participants performed small gestures when speech output was provided (i.e. browse next or previous contact). On the other hand, most participants (6 out of 8) performed large gestures when only vibration feedback was provided (e.g., when making and hanging up a call). Therefore, designers should pay attention to the feedback type when designing motion gestures.

Finally, we learned that a successful assistive interface is not only about usability and performance. The users' confidence and motivation to use the system, and the users' experiences while using the system make the interfaces most effective. In our study, we found that the participants were very pleased with the simplicity and performance of the interfaces we proposed. During the training session of experiment, one of the participants stated that *"It's very interesting, I will use a smartphone if I can control it this way!"* Although motion gestures are not always the best in every situation, our study met a good match between task and interaction techniques that suited the users' capabilities. This encouraged the users to use the system. Therefore, instead of building the interfaces only around knowledge about the users' disabilities, it is important to identify the users abilities and to find the best interaction technique for a given situation.

3.5 Study 3: User-defined Mid-air Gestures for TV Control

Mid-air gesture interfaces have become increasingly popular due to its intuitive and natural control. However, only a very few studies [48] have explored the use of mid-air gestures for blind people. Consequently, blind people have not enjoyed equal access to the benefits provided by such technologies. One promising use of mid-air gestures for blind people is their use for TV interactions because TV is one of the most common and essential activities of many blind people [3]. We have confirmed this fact through our preliminary interviews with blind people where we found that they used TV for more than 7 hours per week. Furthermore, the interviews informed us that blind people often have difficulty finding their remote controls or figuring out the button layout. Such problems suggest that employing mid-air gestures for interaction with TVs is a useful alternative method.

To design usable and accessible gestures, we determined that the first important step was to understand the mental mapping of blind users when performing mid-air gestures. Thus, we investigated the mid-air gesture preferences of blind people for TV interaction using a user-elicitation approach. With the user-elicitation approach, we were able to formulate a classification of gesture types and the frequency of various body parts usage. However, we found two primary limitations with the user-elicitation approach. First, our participants experienced much difficulty and stress trying to come up with their own gestures for some commands because they did not have adequate understanding of the design alternatives and possibilities. Consequently, there was a gesture consensus only for commands that were familiar to users. Second, for unfamiliar commands, our participants suggested many abstract/random gestures which seemed to be rather inefficient and unsuitable for gesture inputs.

To address these limitations, we conducted a second study which adopted a choice-based elicitation approach [86] as a complementary approach where the participants were given a predetermined range of possible gestures from which to choose. All participants reported that they appreciated the gesture choices as the choices helped them to discover more suitable gestures. We also found that the choice-based elicitation approach can help confirm the preferences of certain gestures, while guiding the participants to reach a consensus for unfamiliar commands. We conclude our work with some generalized implications for mid-air TV gesture interaction design, as well as some discussion regarding the use of the choice-based elicitation approach.

Our contributions include:

1. We analyzed 180 defined gestures from which we developed a classification of gesture types for blind people and determined the frequency of body parts usage.
2. We leveraged the choice-based elicitation elicitation approach to address commands that had little consensus among participants and about which blind participants had little understanding or awareness.
3. We present a complete gesture set of mid-air TV gestures for blind people.
4. We present design guidelines for mid-air gestures for blind people by analyzing patterns of user gestures and their mental mappings.
5. We discuss the usefulness and general guidelines of the choice-based elicitation approach.

3.5.1 Participants

Twelve blind participants (7 males, age range = 26 to 78 years, Mean = 53.9 years, SD = 12.64) were recruited. One of the participants could see light, two of them could see the presence of objects but not able to distinguish shapes. One could see objects. The other participants were totally blind. Ten of the participants were ‘early blind’ (age 0 to 3 years), and the rest two were ‘late blind’ (one at 6 years and the other at 17 years). All participants regularly ‘listen’ TV for more than 7 hours each week. None had experience using mid-air gestures for TV control or any other devices. Eleven participants were right-handed and one was ambidextrous. Each participant was paid $15 for the experiment.

3.5.2 Procedure

Our study design used a user-elicitation approach similar to that of [78, 104]. As with previous elicitation studies, we did not want participants to take recognizer issues into consideration, i.e., to remove the *gulf of execution* [33] and to observe the users' best possible gestures.

At the start of the experiment, participants were asked to define gestures for 15 TV commands in randomized order. To identify the most preferred gesture and reduce any ordering effects, participants were allowed to reuse the same gesture for multiple commands.

Each command name was communicated by the experimenter, along with the description of the command's effect (e.g., "*Open TV*" followed by "*This command will power-on your TV*").

During gesture defining, participants were instructed to think-aloud while performing their gestures, confirming the start and end of their performed gesture and describing their corresponding rationale. To determine how well they imagined the gestures, we recorded gesture defining times, i.e., the time between command instruction and gesture execution.

After each group of commands was defined (each group contains approximately 5 commands), participants were asked to evaluate their gestures, using similar evaluation to that of [104]: *"The gesture I performed is a good match for its purpose"*; *"The gesture I performed is easy to perform"*; *"The gesture I performed is tiring"*.

The experiment took around 1.5 hours, and it was audio and video recorded for qualitative data analyses.

3.5.3 Commands

We identified commonly used TV commands. We also selected commands that are specifically used by blind people, e.g., *Voice Guide*. A total of 15 commands were selected for the experiment (see Table 3.2).

No.	**Command**	**Description**
1	Open	Power on TV
2	Close	Power off TV
3	Channel	Change channel
4	Favorite Channel	Go to saved channels
5	Next	Go to next channel/menu
6	Previous	Go to previous channel/menu
7	Volume Up	Increase volume
8	Volume Down	Decrease volume
9	TV Guide	Check channel list and time
10	Pause	Pause
11	Play	Play
12	Menu	Open menu
13	Yes	Answer Yes to system question
14	No	Answer No to system question
15	Voice Guide	Activate voice guidance for blind users

Table 3.2: TV commands used in Study 1 and 2.

3.5.4 Results

Our results include a gesture taxonomy, body parts used, agreement score, gesture defining time and subjective ratings.

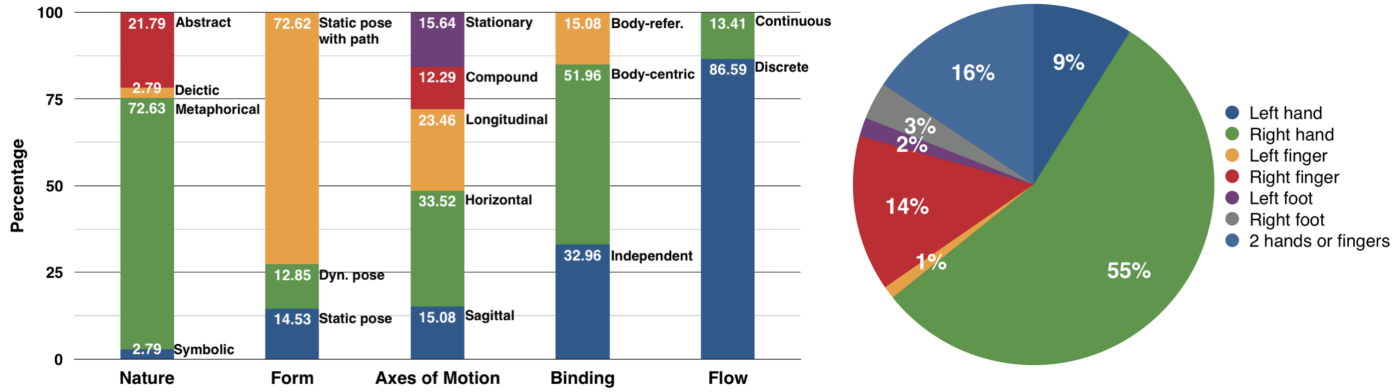
Classification of Mid-Air Gestures

As noted in related work, gesture classifications have been developed for surface gestures [104] and mobile interactions [78]. However, no work has established a taxonomy of mid-air gestures for blind people. We classified gestures collected from the study along five dimensions: *nature, form, axes of motion, binding* and *flow*. Each dimension was further classified into multiple subcategories (see Table 3.3 and Figure 3.8a). Gesture taxonomies obtained in our study and those in [104] and [78] were different for two main reasons: (1) the gestures were performed for different interactions (i.e., surface gestures, motion gestures in mobile phones and mid-air gestures), and (2) they were performed by users with significantly different capabilities (i.e., sighted users and blind users).

Nature dimension is a classification of gesture-meaning relationships.(1) Symbolic gestures are depictions of symbols, for example, drawing a letter 'V' for the *Voice Guide* command. As we expected, very few symbolic gestures were performed in our study, given that these gestures require visual references. (2) Metaphorical gestures linked to their meaning by mechanical determinism (i.e. linked to the logical meaning of daily actions or feelings, not linked to their visual similarities). For example, some of our participants put their hands near by the ears for the *Voice Guide* command indicating that "I want to hear". Metaphorical gestures reflecting common daily actions were the most frequently performed gestures by our participants. Among the metaphorical gestures, we found some conventional gestures. For example, four of our participants performed 'OK gesture' for the *Yes* command. (3) Deictic gestures refer spatial information of objects. For example, when asked to perform a gesture for the *Menu* command, one of our participants said, "I would put the *Menu* command at the upper right corner of the screen and point at it". It is important to note that, unlike

Dimension	**Category**	**Description**
Nature	*Symbolic*	Gesture depicts similarity of symbols.
	Metaphorical	Gesture indicates a metaphor.
	Deictic	Gesture refers to spatial information.
	Abstract	Gesture-command mapping is arbitrary.
Form	*Static pose*	Hand pose is held in one location.
	Dynamic pose	Hand pose changes in one location.
	Static pose with path	Hand pose is the same but the hand is moving.
Axes of motion	*Sagittal*	User moves the hand/leg forward and backward.
	Horizontal	User moves the hand/leg left and right.
	Longitudinal	User moves the hand upward/downward.
	Compound-axes	User hand movement includes more than one direction.
	Stationary	User performs the gesture keeping the hand or fingers at one location.
Binding	*Independent*	Gesture requires no information about object positions or body parts.
	Body-centric	User moves the hand/leg using spatial reference of the body.
	Body-referenced	User moves the hand with respect to other body parts.
Flow	*Discrete*	Response will occur after user performs the gesture.
	Continuous	Response will occur while user is performing the gesture.

Table 3.3: Taxonomy of mid-air gestures based on 180 gestures.



a) b)

Figure 3.8: Percentage of gesture types and body parts used.

sighted people who are able point at visual on-screen contents, our participants performed pointing gestures only relying on their proprioception. That is, they could only refer the content position to hand movements in distinct and specific directions which are related to their body (i.e. upper left, upper-right, etc.) rather than by accurate reference to the display or on-screen contents on the display. (4) Abstract gestures are linked to the meaning neither by the similarity nor by any other relationship that allows one to infer the meaning from the gesture. Arbitrary gestures were the second most performed gestures in our study. For example, when asked to perform a gesture for the *Voice Guide* command, one of our participants came up with the gesture which showed a finger stating no clear gesture-meaning relationship.

In the *form* dimension, (1) static pose is where the hand pose is held in one location, for example, showing the palm for the *Stop* command or showing ‘OK’ for the *Yes* command. (2) Dynamic pose is where the hand pose changes within one location, for example, the *blink gesture* (close and open the palm) for the *Open TV* command or the *double tap* by the foot for the *Yes* command. (3) Static pose with path is where user moves the hand but hand pose is not changed, for example, moving the hand from left to right for the *Next* command or moving the hand upward for the *Volume Up* command. These gestures were the most frequently performed gestures. One interesting observation is that our participants mostly used distinctly exaggerated movements, that is, moving their entire forearm or arm instead of only the wrist and the hand when performing hand or finger gestures. When asked about the reasons, our participants commented that they were not sure of the boundary of interactions, thus making large movements seemed to reassure them and give them confidence that they had performed the actions in a manner that made them clearly recognizable. While gestures with *static pose with path* were the most frequently performed in our study, it is noticeable that there can be cases where the user moves the hand position and also changes the hand pose (i.e. *dynamic pose with path*). However, none of our participants performed such gestures in our study.

The *axes of motion* dimension shows the direction of user movements of the hand, foot or fingers. (1) Sagittal gestures are those where users move the hand forward and backward. Examples include moving the hand backward to represent ‘undoing’ for the *Pause* command and pushing the finger forward as ‘pushing a button’ for *Open TV*. (2) Horizontal gestures are those where users move the hand or foot left and right, for example, moving the hand left and right for *Next/Previous.* (3) Longitudinal gestures are those where users move the hand upward and downward, for example, moving the hand upward for *Volume Up.* (4) Compound-axes gestures are those where user movements include more than one spatial dimensions. For example, some of our participants drew a circle in the air for *TV Guide.* Another example is moving the hand in the upper-right direction where our participant stated, *“I would put the Menu command at the upper-right corner of the screen and point at that direction.”* (5) Stationary gestures are those where users perform their gestures keeping their hands or fingers in one location. With respect to the *form* dimension, stationary gestures are *static* and *dynamic gestures* while sagittal, horizontal, longitudinal and compound-axes gestures are *static gestures with path.*

In the *binding* dimension, (1) independent gestures require no information about the world or the body. Examples are showing ‘OK’ sign for the *Yes* command and showing the palm for *Stop.* These gestures can be performed anywhere regardless of their relation to the user's body or objects on the screen. (2) Body-centric gestures are performed by users by referring spatial information to the body, for example, moving the hand to the left of the body or by moving the foot forward. These gestures were the most performed gestures by our participants. While it is understandable that *body-centric gestures* are specifically selected and performed for certain commands, such as *Next/Previous* commands (moving the hand left and right), we found that our participants performed *body-centric* gestures in many other cases as well. Examples are waving the hand left and right for the *Play/Pause* commands, drawing the hand backward to represent undoing for the *No* command and moving the hand in an upper-right direction to represent skipping for the *Favorite Channel* command. (3) Body-referenced gestures were performed by referring other body parts, for example, putting the hands near by the ears for *Voice Guide*. Another example was that our participants performed *body-referenced gesture* by opening the hands (clasp two hands together and then separate them) for commands such as the *Open TV, Play* and the *Voice Guide* commands.

In the *flow* dimension, (1) discrete gestures are performed and responded to as one event, for example, the *blink gesture* for the *Open TV* command. (2) Continuous gestures require ongoing recognition. One example is *raising the hand* for the *Volume Up* command. *Discrete* or *continuous gestures* were performed depending on the users' perceptions of the system responses to their gestures for the commands. For the *Open TV* command, it was expected that the system would respond by opening the TV instantly after the gesture; the participants performed *discrete gestures* for that command. On the other hand, for the *Volume Up* command, the participants continuously moved the hand upward as if progressively increasing the volume to certain desirable level. Similarly, for the *TV Channel* command, the participants rotated the hand steadily as if to continue until they found the desired channel.

Use of Body Parts

Figure 3.8b shows the use of body parts. One-handed gestures with the right hand were the most preferred, followed by one finger gestures, two hand/finger gestures, and foot gestures.

Hand gestures were commonly performed in the form of movements, such as waving the hand left and/or right for *Next* and *Previous*, waving the hand left and/or right for *Select Channel*, and moving the hand up and down for *Volume Up/Down*. Finger gestures commonly featured the index finger to draw a circle to activate *TV Guide* or to pointing in order to *Open/Close* the TV. Fingers were also used to perform 'OK' gesture for the *Yes* and the *crossed fingers* gesture for *No*. Foot gestures were used by some participants when they wanted to do the same gesture they had already performed using the hands or fingers. For example, one of our participants who moved the right foot to right for the *Channel* command mentioned, "I would like to use my leg because I have done the same gesture for *Next* using the hand." Although the participants were allowed to suggest the same gesture for different commands, they often used the feet reasoning that they wanted to use different gestures from the previous gestures they had already performed. We observed that, our participants performed their foot gestures mostly for abstract gestures where they could not suggest any gesture-meaning relationship. For example, one participant moved the right foot to the right for the *Menu* command stating, "When I have no idea what gesture to use, I just change the body part".

Agreement Score

To determine the gesture consensus among participants, we calculated the agreement score using a formula from [104]. The calculation of agreement score is as follows:

$$A_r = \sum_{P_i} \left(\left| \frac{P_i}{P_r} \right| \right)^2 \tag{3.2}$$

where P_r represents all gestures performed for event r and P_i is a subset of identical gestures from P_r. A_r ranges 0 to 1.

Figure 3.9 shows the agreement score of each command. We found that user-defined gestures in our study had a slightly lower agreement score (0.24) compared to past studies (0.28 and 0.32 in [104]). There were some commands that had relatively high agreement scores such as *Next* and *Previous*. However, for more abstract commands such as *Voice*

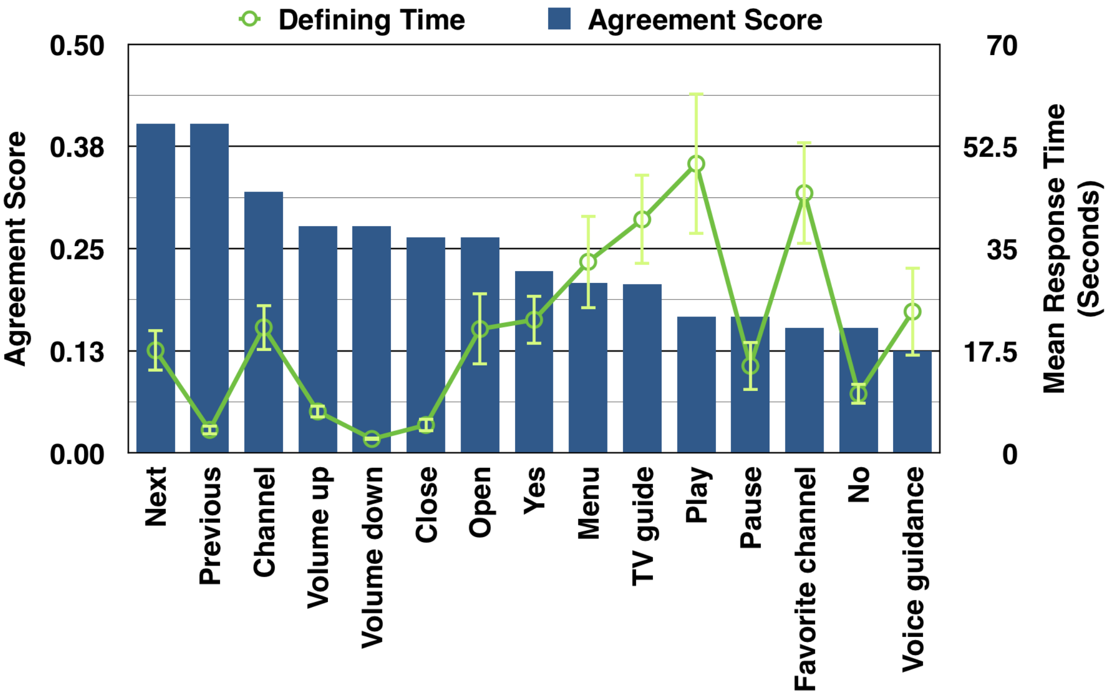


Figure 3.9: Agreement scores and gesture defining times of the user-elicitation approach.

Guide and *Favorite Channel*, participants reported that they had difficulty imagining the gestures. As a consequence, most participants came up with arbitrary gestures about which they had no clear rationale as to why they defined the gesture in that particular way. Some participants reported that they felt tired as the elicitation process was cognitively demanding.

Defining Time

Gesture defining time is the time between command instruction and gesture definition. It can measure the speed and ease of defining of commands which can help inform how well participants can imagine their gestures. The mean defining time of all commands was 22.92 seconds (SD = 12.02). Figure 3.9 shows the mean defining time for each command. Some commands showed consistency between agreement scores and defining times. For example, *Previous* and *Volume Down* both had high agreement scores and low defining times.

Similarly, there were commands such as *Play* and *Favorite Channel* that participants had difficulty imagining and consequently there was low consensus among the participants. On the other hand, there were also some commands with low agreement scores but also low defining times, e.g., *Pause* and *No*, suggesting that there were certain commands where, although each participant had some clear idea of the gestures, they shared little commonality between participants' preferred gestures.

Subjective Ratings

Unlike past works, we found that the highest agreed gestures in our user-elicitation study had no significantly higher good match rating compared to other gestures. This indicates that, in our case, popularity did not necessarily identify better gestures over the worse ones. This was largely due to the fact that our participants had a lot of difficulty imagining suitable gestures and partly due to their lack of experiences with gesture interactions.

3.5.5 Discussion

The classification of gestures revealed that most gestures performed by our participants came from their daily experiences, for example, raising the hand for the *Open TV* command which mimics switching on the light and pushing the hand forward for the same command which mimics pushing buttons on their household appliances and devices. From the *binding* dimension, we observed that our participants preferred *body-centric* gestures, that is, moving hands or fingers to directions with respect to the body (left, right, forward, backward, etc.). Given that our participants did not perform gestures combining dynamic poses with path gestures, we observed that they preferred gestures with simple hand movements. When performing those gestures, user movements could include any directions (of left/right, forward/backward and upward/downward) and most of the gestures occurred at the *frontal plane* of users. Also, for the gesture recognition, it is important to note the inclusion of compound gestures, for example, moving the hand to draw circle in the air (i.e. the hand moves to more than one spatial dimension in a single gesture). Regarding the physical movements when performing gestures, gesture recognition should be developed to accommodate exaggerated hand movements which were preferred by our participants. Gesture recognition should also be developed with attention to user expectations regarding the system's response. For example, *continuous gestures* such as gestures for *Volume, Select Channel* would require on-going (sustained) recognition of gestures.

Regarding the use of body parts, we observed that right and left hands, right fingers, and 2 hands and fingers were mostly used for the *metaphorical, deictic and symbolic gestures*. The left finger, left foot and right foot were not recommended for gesture design, given that these body parts were not commonly used and we observed that they were mostly used for abstract commands.

Some commands in the user-elicitation study achieved low-agreement scores, given that our participants had no clear idea of mid-air gesture interfaces and design. Defining gestures for the given TV commands was very hard and tiring for our participants resulting in low agreement scores, especially for abstract and unfamiliar commands such as *Voice Guide, No, Favorite Channel, Play, Pause, TV Guide* and *Menu* commands. The higher defining time revealed the fact that our participants had much higher difficulty imagining suitable gestures in the user-elicitation approach. Thus, to address these gaps, we recruited the same participants and conducted a choice-based elicitation study where participants were asked to select their preferred gestures from a predefined list.

To minimize any possible effect of the user-elicitation approach on the choice-based elicitation approach, we set sufficient time gap (around 1 month) between Study 1 and 2.

3.6 Study 4: Choice-Based Elicitation Approach

Study 1 allowed us to develop an understanding of user gestures and their mental mappings through the detailed analyses and classification of gestures. However, we found that our participants had difficulty suggesting suitable gestures for certain TV commands, for example, *Voice Guide, No, Favorite Channel, Play, Pause, TV Guide* and *Menu*. The objective of Study 2 was to adopt the choice-based elicitation approach to better understand the preferences of gestures in commands where users had less idea of gesture design. In choice-based elicitation, participants were asked to select their preferred gestures from a list of designer-selected choices. The choice list contained two data columns, one for gesture choices, and another for commands, with the relationship as many-to-one, respectively.

3.6.1 Selecting Choice Candidates

We regard the process of selecting which choices to include in the choice list and how choices are introduced to participants to be key to the effectiveness of the choice-based elicitation approach. Four measures were considered.

First, there is abundant evidence to show that too many or too few choices may decrease user engagement and can possibly confuse users [36]. However, there exist no simple rules to decide the suitable number of choices. To decide the number of choices, we adhered to two grounds - (1) modern working memory studies show that humans can only hold about 4 chunks at a time in their memories [14]; (2) consistent with the memory capacity theory, our initial pilot study with two blind people shows that they preferred a range of 2 to 4 choices. Based on these two grounds, we kept our number of choices to no more than 4.

Second, in order to ensure that each gesture choice is reasonable, there is a need to select gesture choices that are considerably intuitive. We recognized that there are many different ways to construct the choice list and that this depends largely on designer preferences. To demonstrate one possible way, we derived our choices from three places (Study 1, related work, and designers) - (1) we selected the highest-agreed gestures for each command from Study 1 (26.67% of gestures in the choice list). We also selected any gesture from Study 1 that we intuited to be potentially intuitive and suitable (28.33% of gestures in the choice list); (2) we derived some gesture choices from existing works (e.g., [100]); (3) we designed some gesture choices based on our own experience with blind people. Also, gesture classification derived from Study 1 also helped us to design the most suitable gestures for Study 2, for example, selecting the body parts, avoiding symbolic gestures wherever possible. The final list of gesture choices achieved a high interrater reliability among authors and one independent rater (Kappa = 0.919, $p < 0.001$).

Third, to ensure that each choice was equally exposed and judged, experimenters asked participants to perform each choice, instead of only passively listening to the different choices. This also enabled the participants to weigh each gesture based on how comfortable it was to use.

Last, to prevent any possible ordering effect of choices (i.e., due to the recency or primacy effect on choices), we decided that all choices should be presented to participants in randomized order.

3.6.2 Experimental Design

To understand how choices can help us better understand users, the same 12 blind participants in Study 1 were recruited. We used the same set of commands as in Study 1. The procedure was also conducted in a similar manner to Study 1, except that the participants were provided with a list of choices and were asked to select their most preferred choices. Each command name was communicated by the experimenter, along with the description of the command effect. The experimenter then provided the first gesture choice, i.e., the experimenter explained how to perform the gesture, followed by asking the participants to perform the gesture. For each command, the experimenter went through all choices in a similar manner. Gesture choices were presented in a randomized order. The experimental session took around 1.5 hours.

3.6.3 Results

Agreement score, gesture defining time,subjective ratings and the mid-air gesture set were presented.

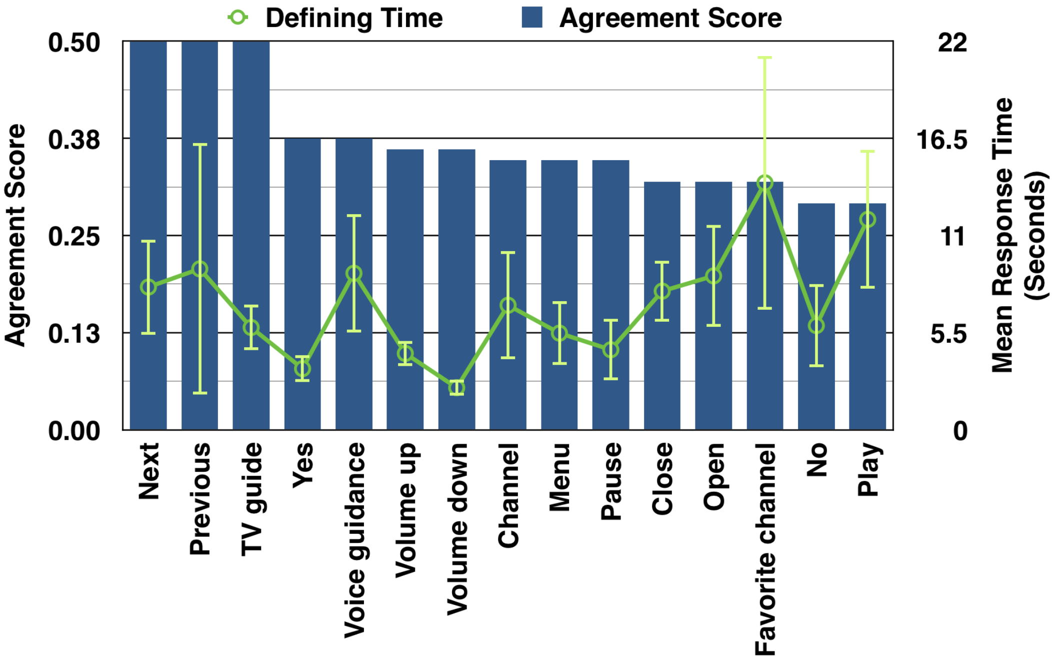


Figure 3.10: Agreement scores and response times of the choice-based elicitation approach.

Agreement Score

The choice-based elicitation approach achieved a noticeably higher agreement score (0.37) (Figure 3.10). This result was initially expected because choices reduced the possible variations between participants. Nevertheless, the choice-based elicitation approach was found to be useful, particularly for commands where participants have no clear ideas of their own, e.g., *Favorite Channel*, *Voice Guide*, *TV Guide*. Participants reported that the supplied gesture choices enabled them to learn, to feel confident, and to discover more suitable gestures, particularly for difficult commands.

Defining Time

To understand how long it took the participants to select their choice, we measured the time between the choice presentation time and decision time. The mean defining time of all commands was 7.16 seconds (SD = 3.13). Figure 3.10 shows the mean defining time for each command. We found that the gesture choices can guide users to discover better gestures.

Subjective Ratings

We found that the highest agreed gestures had significantly higher good match rating (Z = -2.612 , $p < 0.01$) compared to other gestures. This indicates that, in the choice-based elicitation study, popularity does identify better gestures over the worse ones. We were thus confirmed that the choice-based elicitation approach can help confirm the preferences of certain gestures, and thus increase users' confidence on their choices.

Mid-Air Gesture Set

Based on the highest agreed gestures, we developed the mid-air gesture set for TV commands (see Figure 3.11). Nine of the 15 commands shared similar highest-agreed gestures in both user- and choice-based elicitation approaches. These nine commands includes Open/Close, Volume Up/Down, TV guide, Yes, Previous/Next, and Voice Guide. Another six commands were mainly based on the choice-based elicitation approach which include Play/Pause, Change Channel, Favorite Channel, Menu, and No. Using the choice approach, we found a good average consensus on these six commands (0.32).

3.6.4 Discussion

Gesture agreement scores for all commands increased in the choice-based elicitation study. Higher agreement scores were expected given that the choice-based elicitation approach has fewer choices than the user-elicitation approach. However, we observed the effectiveness of the choice-based elicitation approach especially for abstract or unfamiliar commands which had quite low agreement scores in the user-elicitation study. The agreement scores for those commands noticeably increased in the choice-based elicitation study: *Voice Guide* (increased by 192.31%), *TV Guide* (138%), *Favorite* (113.33%), *Pause* (105.88%), *No* (93.33%), *Yes*

(72.73%), *Play* (70.59%) and *Menu* (66.67%) respectively. Gestures for commands that had the highest agreement scores in the user-elicitation study (*Next, Previous, Channel, Volume Up, Volume Down, Close* and *Open*) had the highest scores in the choice-based elicitation study and in the same order .

The choice-based elicitation approach also helped confirm that there was indeed some commonality in the participants' mental mappings to the gestures, and this commonality could be more likely observed when all participants were equally informed about the different possibilities within the design space.

*Highest-agreed gesture in both user-elicitation and choice-based approaches

Figure 3.11: Highest-agreed gestures in choice-based elicitation approach.

3.7 General Discussion

We have explored the use of the user- and choice-based elicitation approach, and our studies showed promising results. This section discusses (1) mental model observations, and (2) user-elicitation and choice-based elicitation.

3.7.1 Mental Model Observations

Our quantitative data, qualitative data and video recordings enable us to capture user mental models that will benefit the design of mid-air gestures for blind people. Guidelines are summarized in Table 3.4.

Symbolic vs metaphors

We observed that the mid-air gestures of blind people were quite different from those of sighted people in a way that sighted people performed a lot of *symbolic gestures* [100] and our participants performed very few (fewer than 3%) of those gestures. Instead, *metaphoric gestures* were mostly performed by our participants. For example, while the gesture for *Menu* was drawing letter 'M' in [100], the gesture for the same command in our study was the "*Open Book*" gesture (Figure 3.11). Most of our participants chose that gesture stating that, "*This gesture is like opening a menu book at a restaurant*". Thus, it is important for designers to understand which daily actions of blind people can be exploited as potential gestures in interactions with digital devices.

Also, for the *Volume Up/Down* commands, gestures of sighted people included having the non-dominant hand as a visual reference of measurement while moving the dominant hand upward/downward increasingly with reference to it [100]. However, none of our participants performed gestures that referenced measurement. Thus, for commands like *VolumeUp/Down*, gesture recognition should be developed so that users can rely only on their non-visual skill via proprioception, that is, the system should recognize the fine self referenced movements of users as measures of degree (increase/decrease).

Visual resemblances vs logical mapping

It is important to understand how blind people perceive the meaning of certain commands. For example, the highest agreed gesture in user-elicitation study for *TV Guide* was drawing a circle in the air stating, "*the meaning of a circle is 'all' and a TV Guide also shows all TV channels.*" This gesture was also the most agreed gesture for *TV Guide* in the choice-based elicitation. During the choice-based elicitation study, although the participants were given other gesture choices for the same command such as moving the hand from up to down (as if to indicate a channel list), none of our participants chose that gesture. This informed us that the imagination of our participants for certain commands can be very different from that of sighted people (i.e. visual-based imagination of the *TV Guide* as a channel list vs logically mapping as *showing the whole*). Thus, designing gestures that are related to visual resemblances should be avoided as these gestures may not be intuitive to blind users.

Body-centric

As discussed in earlier sections, our participants mostly performed *body-centric* gestures (i.e. moving the hands or fingers to left/right, forward/backward, and so on). From our semi-structured interviews, we also observed that spatial references relating to their body was a great aid in many situations (for example, saying *"coffee cup at the left and mobile phone at the right"*). Thus, it is clear that *body-centric gestures with simple movements* are desirable for blind users.

Also, both sighted and blind users may simply want to select contents on the screen when they have no gesture ideas for triggering the commands. In our study, *deictic gestures* were performed by our participants stating that they would like to place the menu items and select them. However, finding and selecting the on-screen contents will be different between sighted and blind users, given that blind users have no visual cues to help them find the content on the screen. For such cases, gesture recognition should allow users to rely on non-visual skills such as proprioception, for example, users can select the content just by moving their hand in certain directions (upper-left, right, etc.) with reference only to their own bodies.

I want to use old things that I have learned

During the study, our participants often came up with gestures that they commonly used to interact with appliances and devices in their daily lives. For example, the highest agreed gesture in both studies for the *Select Channel* command was rotating the hand left and right which mimicked rotating the channel dial on old televisions. Similarly, the participants selected a "*Push button* gesture in mid-air for the *Open/Close TV* command which mimicked pushing buttons in their mobile phones. Thus it is important for designers to understand the commonly-used actions used by blind people when interacting with their commonly used devices and appliances. Although the influence of daily device usage may also be applicable for sighted users, the devices used and the interaction experiences can be quite differently perceived between blind and sighted people (e.g., smartphone vs. feature phones). Thus, designers should pay close attention to those differences when designing gestures that mimic the use of older devices.

Reversible gestures are more memorable

We observed that our participants used consistent mapping of gestures for many commands, for example, *Next/Previous, Volume Up/Down, Play/Pause*, and so on. Notably, reversible gestures were very important for blind users for gesture learnability and memorability because mapping similar and opposite things was an efficient learning aid for them in many cases. During Study 1, one of the participants stated, *"I do an open hands gesture (i.e. claps the two hands then separate them) one time for Play, so I would do the same gesture 2 times for Open TV because they are similar commands."* This implies that even for commands that are not directly dichotomous, our participants preferred mapping similar or opposite gestures where possible. Thus, it is a good idea to design reversible gestures wherever possible.

Big movements are more recognizable

All our participants used noticeably larger movements when they performed gestures. This can be partially due to the lack of feedback during gesture performance in our study. We speculated that by exaggerating their movements our participants compensated for the lack of feedback intending to eliminate uncertainty regarding the perception of their gestures and ensuring that their movements were clear enough to be recognized by the system. Thus, gesture recognition should be developed to allow freedom of movement. Also, time-based gesture recognition should be avoided because blind people may perform gestures with big movements at a different pace to sighted people.

3.7.2 User-Elicitation and Choice-Based Elicitation

We observed that the choice-based elicitation approach can work well when the user has no clear idea of gesture possibilities. Nevertheless, we cannot disregard the results from user-elicitation. In particular, we found that the results from the user-elicitation proved to be very useful. They enabled us to understand the gestures that users were familiar with and the commands that they had little idea about and those that they had difficulty imagining gesture(s) for. Conducting only the choice-based approach may not allow us to obtain such insights. By combining the information from both the user-elicitation and the choice-based elicitation approaches, we gained useful insights for designers, as demonstrated in previous section. We thus suggest that it would be beneficial to use both approaches during gesture design processes.

Participants commented on the benefits of the choice-based elicitation approach in three ways. First, it helped users understand what is possible within the design space. By learning through the choices, participants commented that they understood better about the different possibilities in the design space. This understanding was reflected by one participant who stated that, "*For some commands, it was just impossible to imagine, I really appreciated the choices, as I could learn through examples.*" Second, the choice approach made participants to think actively about all the possibilities presented to them. Participants commented that the process of selecting their preferred choices caused them to think actively. One of the participants commented that it was fun as they felt like they were 'wearing the hat of a designer'. In addition, by carefully considering all possibilities, the choices allowed them to discover better gestures than they had thought of themselves. Third, the choice approach enhanced their confidence and engagement during the participatory design process. By giving choices, our blind participants felt very confident when they discovered that the gestures they had preferred were also in the choice list. All blind participants reported that they appreciated the gesture choices as they helped guided them to decide on their preferred gestures. One participant stated, "*The choice list helps me to express what I really want.*", another commented, "*This is like a game and it's fun!*".

For designer convenience, we summarized our experience of using the choice-based elicitation approach in the form of five generalized guidelines.

1. We should consider the number of choices based on the memory and attentional capacity pertaining to different demographics. For example, old people may prefer a fewer choices, compared to young people.
2. It is essential for users to actually perform each gesture choice instead of just passively to look at or listen to a description of a gesture. By actually performing the choice, users can better estimate the suitability of the gesture.
3. The criteria for selecting candidate gestures for the choice list should be specified as objectively as possible, e.g., high-comfort, high-speed and/or low-error rate, etc.
4. It is important to carefully consider the issue of learnability in the choice-based elicitation approach, as the selected gesture may not have originated from the user's mind.
5. Although we found that the choice-based approach is effective for our blind participants, user-defined and choice-based approaches have their own respective advantages and they can well complement one another limitations.

Guidelines	**Criteria**
Metaphors	Gestures should be designed mimicking metaphor in users' daily lives. Avoid symbolic gestures such as drawing letters (e.g., 'M', 'X'). Avoid gesture that needs visual references (e.g., sizes, distances). For measurements, system should recognize the fine self-referenced movements of users as measures of degree.
Logical Mappings	When mapping gestures to daily metaphors, *actions* (e.g., gesture that represents opening a menu book for the *Menu* command) should be used. Mimicking *visual similarities* (e.g., gesture that represents the *Menu* as a list of items) should be avoided.
Body-centric	Movements in performing the gestures should be easily relatable to user body (e.g., moving left and right, backward and forward). If pointing are required to trigger a menu, users should be able to rely on reference only to their own bodies (e.g., moving the hand upward, upper left). The need of accurate reference to the display or on-screen contents on the display should be avoided.
Commonly Used Device	Gestures should be designed to mimic the use of users' older devices whereas possible. Attention should be paid to differences of commonly used devices of blind and sighted people (e.g., smartphone vs feature phone).
Reversible gestures	Reverse gestures should be designed for dichotomous commands. Reverse or related gestures should also be designed for similar commands (e.g., *Pause, Close TV*) where possible.
Exaggerated user movements	Gesture recognition should allow freedom of movements for performing user gestures. Expect slow pace due to big movements in gestures performed by blind users. Avoid time-based gesture recognition.

Table 3.4: Design guidelines and criteria.

Chapter 4

MOTION MARKING MENUS

4.1 Introduction

Motion gestures were found to be desirable interfaces for blind people in many situations. However, the bandwidth of motion gestures is somehow limited. To increase the bandwidth of motion gesture interfaces, we proposed motion marking menus, and studied user capabilities to perform motion marking menus. Then we developed and evaluated a motion-based marking menu system in smartphones.

Even though some accessibility issues remain with smartphones, reliance of people with visual impairments on smartphones has been increasing [106]. Although smartphones support screen readers, such as VoiceOver^{TM} and TalkBack^{TM} [102], and voice commands, these features can be inefficient in noisy environments and inappropriate in quiet public environments. These systems enable users to browse menu items on touch screens using speech feedback. However, they require users to perform long sequences of touch gestures to browse the menus. This results in increased user fatigue and dissatisfaction. There is an increasing need for more efficient interactions as supplements or alternatives to the accessibility features that are currently available for users with visual impairments.

Literature has shown that marking menus are readily adapted to mobile devices because they offer fast and eyes-free interaction [69]. This indicates that marking menus can offer significant benefits to people with visual impairments for eyes-free mobile interactions. In mobile situations, these users mostly have one hand occupied with a cane or a guide dog [106]. Thus, we propose marking menus working together with motion gestures to provide users with fast access to smartphones using only one hand. Adequate motion sensors are now available on most common mobile devices [66], and we developed motion gesture-based

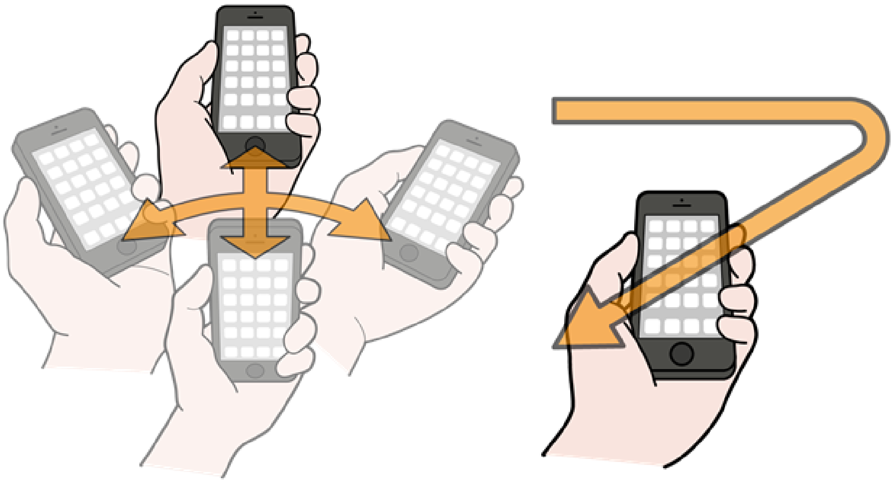


Figure 4.1: Motion Marking Menu (MMM) interfaces in smartphones.

marking menus in 3D space, which is called Motion Marking Menus (MMM) in order to enable people with visual impairments to access smartphone menus (Figure 4.1).

In spite of the potential, the capability of people with visual impairments to perform motion marking menus has not been investigated. A thorough understanding of such capabilities will help in the design of more efficient and accessible interfaces. With the motivation of providing implications in designing motion marking menus, this study seeks to investigate the spatial ability of people with visual impairments to perform with motion marking menus.

Research questions include: 1) How many menu items can there be at each level? In other words, we wanted to determine how many directions people with visual impairments can distinguish to successfully perform with marking menus. 2) How deep or how many hierarchic levels can people with visual impairments go before marking menu performance drops off?

4.2 Related Work

Related work includes marking menus for mobile devices, mobile device accessibility for people with visual impairments and mobile spatial interactions.

4.2.1 Marking Menus on Mobile Devices

Kurtenbach et al. [54] proposed marking menus that allow users to perform menu selection by either selecting an item from a pop-up menu or drawing marks in the direction of the desired menu item. Recently, marking menus have been adapted to mobile devices. Jain and Balakrishnan [37] developed a marking gesture based mobile text entry system which requires less visual attention from users. pieTouch [17] is a marking gesture based vehicle information system designed to reduce visual demand. Francone et al. [20] presented a touch-based marking menus for navigating data hierarchies on mobile phones. Although not focused on accessibility, Oakley and Park [69] demonstrated a marking menu based eyes-free menu system with 3D rotational strokes. This system deviated from the traditional marking menu system in that it was one dimensional and the system involved dividing a 90 degree portion of rotational space into three targets of 30 each. Bauer et al. [7] presented a marking menu system for eyes-free interactions with large displays using smartphones and tablets. In their study, a marking menu was placed on the touch screen of smartphones and tablets so that the users could remotely interact with large displays.

These studies have demonstrated that marking menus are beneficial for fast and eyes-free interactions in mobile devices. Despite the fact that marking menus promise eyes-free input, questions about the capability of people with visual impairments to perform marking menus have not been investigated.

4.2.2 Mobile Device Accessibility for People with Visual Impairments

Kane et al. [43] presented a specialized touch-based interface for menu selections called Slide Rule. Slide Rule is a set of audio-based multi touch interaction techniques that enable people with visual impairments to access smartphone functions including making phone calls, mailing, and music performance functions. Zhao et al. [113] developed EarPod using touch input and sound feedback for eyes-free menu selections. Audio-based text entry systems were also developed by Sáncheznchez and Aguayo [82] and Yfantidis and Evreinov [108]. These systems used multi-tap and directional gestures and audio feedback, to enable users to enter text on touch screens.

The literature has demonstrated several accessibility features on mobile devices that utilized touch-based gestures and speech feedback. In mobile situations, people with visual impairments mostly have one hand occupied with a cane or a guide dog [106], while touch-based interfaces may require users to use both hands, one hand for holding the phone and one hand for performing gestures. Although touch-based interfaces enable one hand operation using the thumb, user performance in thumb-based interactions greatly relies on several factors such as surface size, hand size and hand posture [8]. For the purpose of use “on the go”, 3D motion gestures are suitable because they provide fast access and enable users to operate the system with only one hand [66]. Also, a previous study has reported that motion gesture interfaces were efficient and receptive to users with visual impairments [15]. Thus, we propose motion-based marking menus as an efficient way for these users to interact with smartphones.

4.2.3 Mobile Spatial Interactions

Recently, there has been a growing interest in research regarding input techniques in mobile devices that allow spatial inputs. For example, SideSight [12], Abracadabra [24], Minput [25], Skinput [26] and HoverFlow [49] provided spatial inputs in mobile devices. Peephole displays [107] offered spatial input that maps physical movements of a device to movements in a virtual world. VirtualShelves [60, 61] extends these techniques by treating the space around the user as a discrete set of regions (shelves), so that the user can access contents on these virtual shelves. Gustafson et al. [22] presented Imaginary Interfaces i.e. spatial interactions that occur only in the user’s imagination.

Several studies have demonstrated spatial interactions in mobile devices as a promising eyes-free input modality. Our study seeks to investigate the spatial ability of people with visual impairments to perform with motion marking menus for eyes-free input in mobile phones.

4.3 Preliminary Interviews

Before the experimental studies, we conducted a preliminary interview with 10 participants with visual impairments (ages range from 27 to 78 years). 5 of them were totally blind, 2 of them could distinguish between light and dark, and 3 of them could see objects but none were able to distinguish between objects. The interview for each participant took around one hour. Each participant was paid $10 for their participation.

The purpose of the interview was to investigate (i) current problems with mobile phones that they were using, (ii) the potential of marking menus and spatial interactions as eyes-free interfaces, (iii) the spatial awareness of people with visual impairments in their daily lives.

4.3.1 Current problems with mobile phones

Seven of the participants used feature phones and the other three used smartphones. When asked about their current mobile phone usage, the feature phone users commented mostly on the limited features in their mobile phones and the need for faster operation. All of them mentioned that they would like to have more access to more utilities such as GPS, calendar, weather, etc. The smartphone users mostly commented on the fatigue they experienced using their smartphones. Current accessibility features in smartphones support flick gestures and speech output to browse menus on the screen. One of our participants stated, *"Many times (when using the smartphone), I want to jump the cursor to the function I want to select. Sometimes the guiding voice is frustrating in public places."* This encouraged us to test the efficacy of the motion marking menu system as a solution to problems with current accessibility features.

4.3.2 Potential of marking menus and spatial interactions

Through the interview study, we learned that many interactions in the daily lives of people with visual impairments are facilitated by spatial awareness and kinesthetic memory. All participants consistently mentioned that performing their daily tasks was mostly facilitated by touch, relocation of objects that they use each day and by memorizing habits that they learned through repetition. When asked about daily tasks that they could successfully perform by repetition and habit, one of the participants replied, *"I can put an appropriate amount of water into a jug to make coffee even though I cannot see it."* The same participant mentioned, *"It is not that difficult to do daily tasks. But I am in trouble if someone has moved things that I usually use."*

The important point to note with marking menus is that the physical movements performed when selecting a menu item are rehearsed and embedded in the user's muscle memory. Thus, we were convinced that marking menus could be used as eyes-free interactions for people with visual impairments.

4.3.3 Spatial awareness

When asked about their awareness of directions, all participants expressed difficulty in figuring out directions in terms of cardinal directions (i.e. North, West, etc.). Most of the participants were not familiar with directions either in terms of a compass (N, S, W, E) or in terms of the twelve divisions on the dial of a clock or watch (e.g. "at 2 oclock"). One of the participants mentioned, *"I can roughly say directions of hours on a clock because I used a tactile watch before. For example, 8 o'clock would be at the lower-left of my body."* Another participant stated, *"I have never used a compass though a compass with sound or texture would be usable."* The same participant also mentioned, *"Nowadays, it is very easy to use digital clocks with sound, so, I am not familiar with the directions on a clock."* Instead, they all agreed that directions relative to their body (i.e. left, right, etc.) were easy to understand and that they frequently used those directions to arrange items they used each day. One participant mentioned, *"I put my daily items around my chair where I can easily access them, for example, my mobile phone at my right side, my radio or charger at my left side."* All of the participants mentioned that they preferred saying directions using left/right because they are constant (related to the body). The answers regarding the spatial awareness of our participants suggested that their familiarity with compass or clock layouts depends on the individuals' training in visual thinking (e.g., whether they were exposed to and taught to read a braille clock or a raised compass).

Being informed by the interview study about the potential of marking menus and spatial interactions for people with visual impairments, we conducted user studies to investigate the extent of efficiency and comfort with which people with visual impairments can perform marking menus.

4.4 Study 5: User Capabilities to Perform Motion Marking Menus

Study 1 was conducted to answer the questions, 1) How many menu items can there be at each level? 2) How deep or how many hierarchic levels can people with visual impairments go before marking menu performance drops off? 3) Is wider breadth better than deeper depth in marking menus, or vice versa?

4.4.1 Experiment Design

Experimental trials to establish the number of items (breadth) and number of levels (depth) were designed as within-subject repeated measurements. The participants performed menu selections in 16 menu configurations of 4 breadths (angular width 90°), 6 breadths (angular width 60°), 8 breadths (angular width 45°) and 12 breadths (angular width 30°) crossed with depths from 1 to 4 levels. The schematic diagram of menu breath and depth is shown in Figure 4.2. The rationale for selecting the number of menu breadths and depths was based on findings from our interview study and the experiment design from a previous marking menu study [54]. We also added 6-item menus because we hypothesized that this menu configuration could be a good option when menu selections in 8-item menus were found to be too error prone for our participants.

In each menu configuration, three different menu selections were presented. Each selection was presented three times. Three different menu selections were configured to include both easy and difficult target menus were included. Easy menus were those that existed along vertical and horizontal axes (i.e. left, right, up, down). Difficult menus were those that

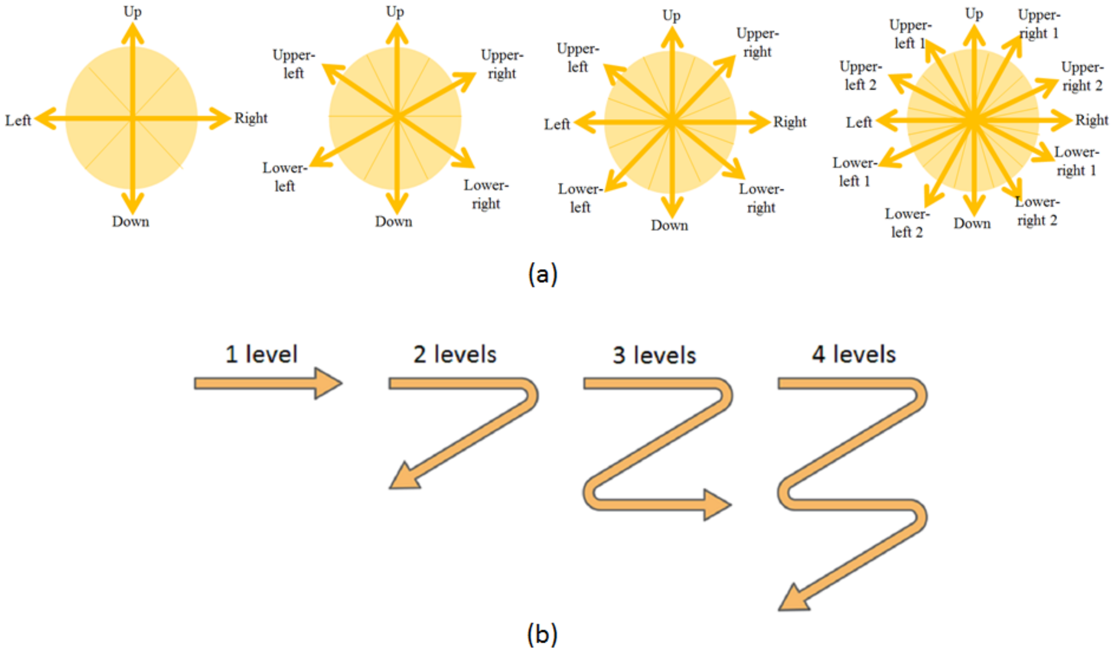


Figure 4.2: Schematic diagram of menus used in the experiment. (a) Menu items (breadth), (b) Menu levels (depth).

existed in off-axis positions (i.e. upper-left, upper-right, etc.). We paid attention to ensure that menu selections in our study included easy, moderately difficult and difficult menus. Each participant performed 432 (16 menu configurations × 3 menu selections × 3 repetitions × 3 blocks = 432) trials in total.

The order of the menus was counterbalanced using a Latin Square. The occurrences of menu configurations were randomized among the participants.

4.4.2 Participants

Thirteen participants (2 female and 11 male), with ages ranging from 26 to 78 years, participated in the experiment. One of the participants could see the light and three of them could see objects, but none were able to distinguish between objects. The rest were totally blind. All the participants were right-handed. Each was paid $10 for their participation.

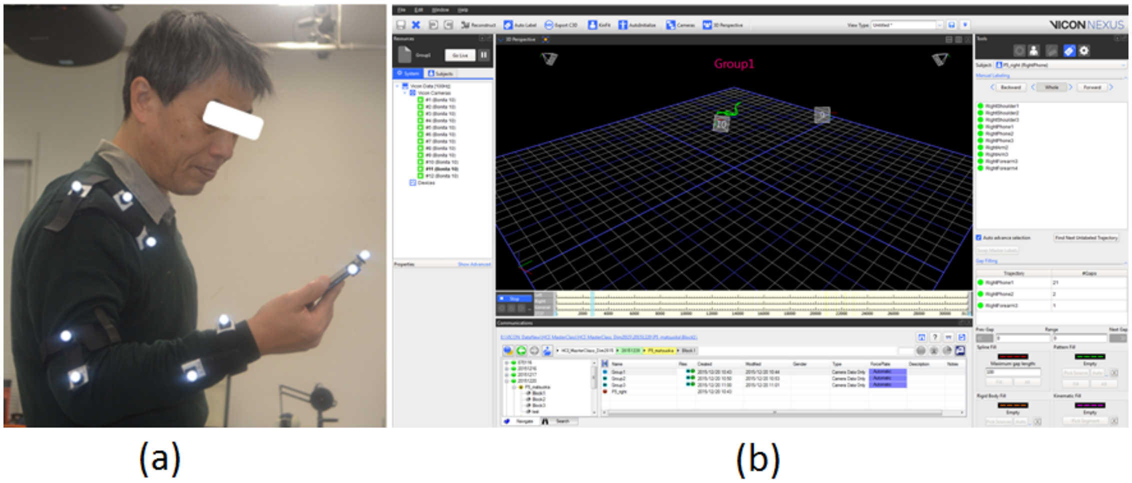


Figure 4.3: Experimental setup. (a) Participant performing menu selection, (a) Interface in Vicon Nexus 2.1 while performing gestures.

4.4.3 Apparatus

For capturing user hand movement in menu selections, we used 12 Bonita 10 cameras (frame rate: 250 fps, resolution: 1 megapixel (1024 × 1024), lens operating range: up to 13 m angle of view wide (4mm): 70.29° × 70.29°, angle of view narrow (12mm): 26.41° × 26.41°).

The participants hold an Alcatel OneTouch (Dimension - 5.37 × 2.74 × 0.30 in, Weight - 1.6g) smartphone for performing motion gestures. 10 markers (4mm wide) were attached on the smartphone, the participants wrist, forearm, elbow, arm and shoulder.

For the Vicon system, Nexus 2.1 motion capture software was used. A custom software was also developed to track user hand movement data from the Vicon system and trigger real-time instruction to the participant. The system was written in C#.Net. The software systems and their interfaces are shown in Figure 4.3.

4.4.4 Building Models

Before the experiment started, we asked the participants their comfort hand side for holding and using a mobile phone. During the experiment, the participants used their comfort hand (left or right hand) for performing the gestures. For this purpose, user models were built for both left and right hands. For each hand side, the model included 4 segments: shoulder, arm, forearm and the Phone. Three markers were used for the shoulder segment, two for the arm, two for the forearm and three for the phone.

One marker on the shoulder, one on the elbow and one on the phone were used for connecting two segments that is, connecting the hand and forearm, connecting the forearm and arm, and connecting the arm and shoulder.

To enable the system differentiate the left and right hand with no errors, markers on the arms were put in different positions at the left and right side. That is, markers on the left arm were put higher than those on the right hand so that the two ratios would be different and the system made little no confusion between the left and right arms.

4.4.5 Procedure

The participants were introduced to the purpose of the study and the study procedure. Then, we got biographic information participants' biographic information and a signed consent form. The participants were asked about their comfort hand side for using a mobile phone. Then markers were put on the participants and they were also handed a smartphone on which three markers attached.

Then the participants were trained with the menu layouts. To help the participants learn the menu layouts better, we put tactile patterns of directions on the wall. The participants were exposed to the tactile directions and allowed to practice until they felt that they were familiar to all directions and labels of the directions.

After training and practice session, the experiment started. Each experimental trial occurred as follows. The participants stayed at relax state that is, put their hand side against the body. Then system said the participants to stay at 'ready' state. The participants responded by moving their hand to anywhere that they felt comfortable to start the menu selection. Then, the system read out the menu name which the participant had to select. The participant responded by moving their hand to select the menu. For menu selection with more than one menu level, the system read out the next menu to select while the participant is performing gesture for the previous menu level (see Figure 4.4). This procedure was repeated until all menu levels were completed.

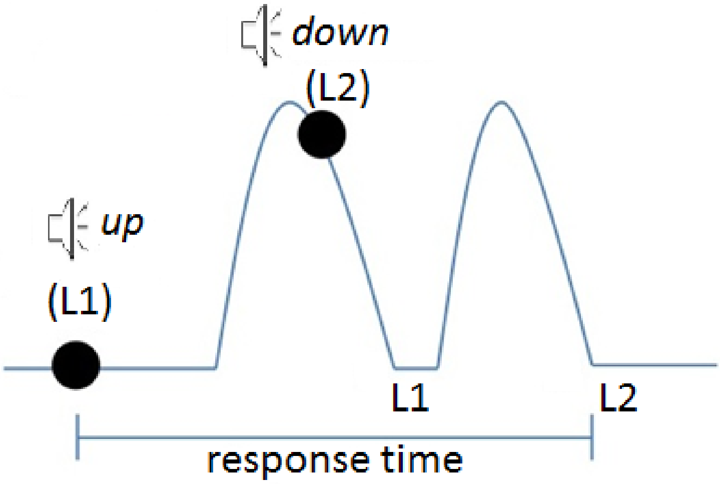


Figure 4.4: Menu selection with 2 menu levels which included ‘up’ and ‘down’ directions.

Once the participants stopped their hand movement as they have performed the menu selection, the system instructed the participant to stay back at ‘relax’ state. The next menu was read out by the system with 3 to 5 seconds time-interval. Menu selection time and errors were recorded. Menu selection time was the time elapsed between the times when the participants started hand movement from the ‘ready’ state to the time when they stopped their hand movement as they have performed the menu selection. Errors were angle errors of the participants; hand movement in a particular menu layout, with respect to their starting ‘ready’ point. Angular errors were calculated for each gesture stroke (each menu level). Thus, if one of the gesture stroke with a particular menu level had angle error, that menu selection was regarded as error menu selection.

All experimental trials were divided into 3 blocks. All the experimental trials were video recorded. After the experiment, the participants were asked about their subjective comments on each menu configuration, directions that were particularly easy or difficult for them to figure out and perform the gestures.

We measured errors, menu selection time and selection accuracy. Errors were angle errors of the participants; hand movement in a particular menu layout, with respect to their starting 'ready' point. Angular errors were calculated for each gesture stroke (each menu level). Thus, if one of the gesture stroke with a particular menu level had angle error, that menu selection was regarded as error menu selection. Response time is the time elapsed between the times when the system started saying the first command to the time when participants ended gestures. We we collected the participants' subjective assessments based on the statements, "The direction is easy to figure out", "The gesture is easy to perform", and "The gesture is tiring". The experiment for each participant took around 1 hour and 40 minutes.

4.4.6 Result

The data of interest were error rates and response times. Error rates were measured by the percentage of incorrect selections out of 9 trials in a particular menu layout. Response time was defined as the time elapsed between the display of the menu until the completion of the gesture by the participant.

Before analyzing the data, we removed errors caused by "mental slip" [68]. For example, sometimes when instructed to select the menu on the left, participant would accidently move the hand to the right. We did not include these types of errors in the data sets because they were not caused by selection inaccuracies. During the experiment, the participants were given real-time instructions using the custom software. The software was developed to give the instructions based on thresholds set for vector speeds of participants hand movements (e.g., start moving the hand for gesture, stop moving the hand after gesture, etc). In a few cases, the software could not give the instruction and log the movement data because the participants' hand movements were not matched to the system's thresholds (e.g., the participants moved slower that the threshold speed). For those cases, we recorded the trial number of those movements and also removed those data from the data set.

A two-way repeated-measure ANOVA (analysis of variance) was used with the two factors of the number of items (breadth) and the number of levels (depth). All tests were run at a significant level of alpha (α) = 0.05.

Errors

Figure 4.5 shows error rates in each menu layout. All participants were able to select a menu from a 4-item menu without any errors up to 2 levels. Error rates were affected by both the menu breath ($F(3, 30) = 65.96$, $p < 0.05$) and by the depth ($F(3, 30) = 51.07$, $p < 0.05$). There was a statistically significant interaction between breadth and depth to affect the error rates ($F(9,90) = 3.95$, $p < 0.05$). Post hoc analysis with Bonferroni correction indicated that error rates were not significantly different among 4-item, 6-item and 8-item menus, until 2 levels. However, error rates become significantly higher in 6-item and 8-item menus than in 4-item menus starting from 3 levels ($p < 0.001$). No significance was found between 6-item and and 8-item menus at any level. Error rates in each menu layout are shown in Table 4.1.

Breath	Depth			
	1	2	3	4
4	0.0 (0.0)	0.0 (0.0)	1.35 (0.56)	6.05 (2.70)
6	0.76 (0.76)	3.79 (2.60)	11.36 (3.03)	19.69 (3.23)
8	1.51 (1.02)	5.30 (2.58)	6.06 (2.27)	19.72 (3.24)
12	17.49 (2.78)	24.89 (3.74)	36.33 (2.38)	44.40 (3.16)

Table 4.1: Mean error rates. Standard errors are shown in parentheses.

We divided the menus into three groups: one group consisted of entirely on-axis menus, another group consisted of a mixture of on-axis and off-axis menus, and the last group consisted of entire off-axis menus. Then we analyzed user performance in the three groups. As we hypothesized, axis had significant effect on errors ($F(2, 22) = 16.2$, $p < 0.05$). We also found that axis and menu levels had significant interactions on errors ($F(8, 88) = 30.56$, $p < 0.05$). This was particularly reflected in 6-item menus. In 6-item menus, errors become significantly higher in level 3 (up to 11.36). As the number of menu levels increased, more off-axis items were included in the combination because most of the menus existed "off-axis" (60, 120, 240 and 300 degrees) in 6-item menus. Similarly, in 8-item menus, errors increased from 6.06 up to 19.72 in level 4.

To evaluate the affect of motion directions on performance, we compared performance in mirror pairs of menus (i.e. left/right, up/down, down-left/down-right and up-left/up-right). In general, more errors were found for right and up-right directions when compared to left and up-left. However, no significant effect of directions on performance was found for any directions.

We analyzed the learning effect by analyzing the error rates after each block. Error rates decreased over blocks. Significant differences in error rates were found between block 1 and block 3 ($p < 0.05$) and between block 2 and block 3 ($p < 0.05$). No significant difference was found between block 1 and block 2 ($p = 0.265$). This result is shown in Figure 4.7a.

"Mental slip" errors are accidental errors that occur in any human actions [68]. In our study, the participants reported that left and right directions were easily mistaken when performing marking menus. We performed detailed analysis of errors caused by "mental slip", and confirmed that 80% of errors caused by "mental slip" were made for left and right directions.

Response time

Response time was affected both by the menu breadth ($F(3, 33) = 14.24$, $p < 0.05$) and by the depth ($F(3,33) = 165.91$, $p < 0.05$). Breadth and depth also interacted to affect the response time ($F(9, 99) = 3.04$, $p < 0.05$). Post hoc analyses with Bonferroni correction indicated that

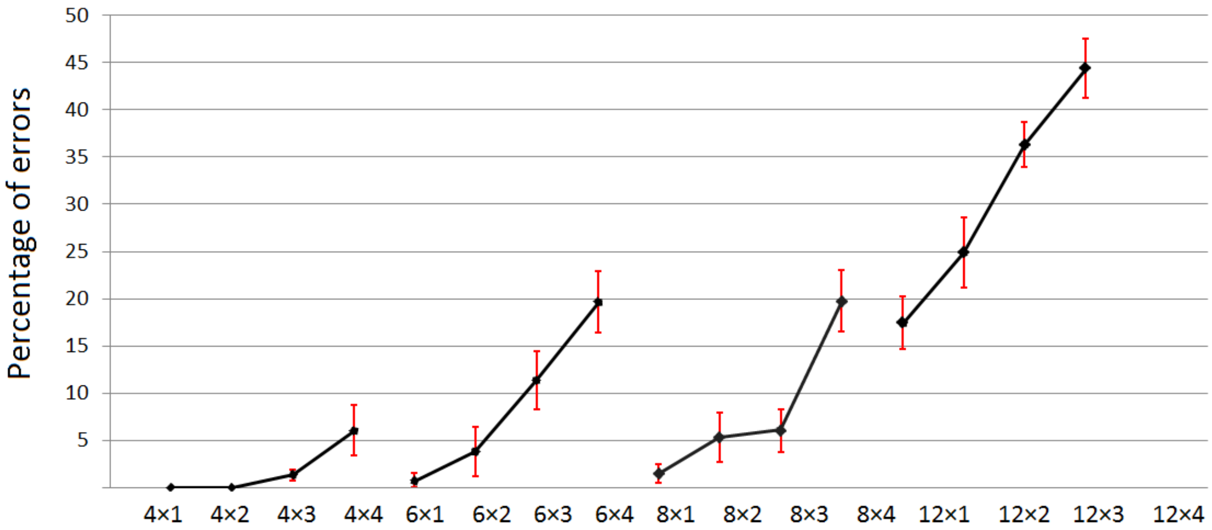


Figure 4.5: Percentages of errors in each menu layout.

4-item menus were significantly faster than 6-item menus, 8-item menus and 12-item menus. No significant differences were found between 6-item and 8-item menus. Response times increased linearly as the number of menu levels increased (Figure 4.6), and the differences were significant at any level ($p < 0.001$). Response times in each menu layout are shown in Table 4.2.

We analyzed the learning effect by analyzing the response times after each block. Response times decreased over blocks. Significant differences in response times were found between block 1 and block 3 ($p<0.05$) and between block 2 and block 3 ($p<0.05$). No significant difference was found between block 1 and block 2 ($p=0.328$). This result is shown in Figure 4.7b.

4.4.7 Discussion

The experiment results and qualitative data collected in the interview enabled us to answer our research questions relating to the marking menus of people with visual impairments.

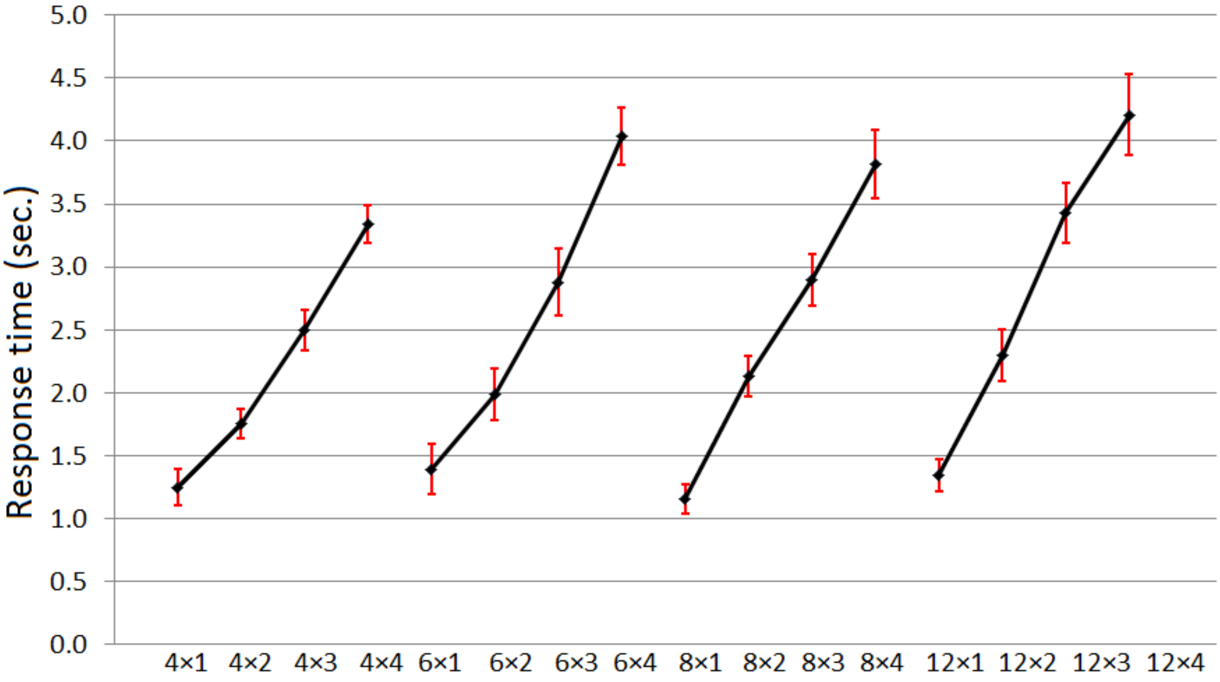


Figure 4.6: Response times in each menu layout.

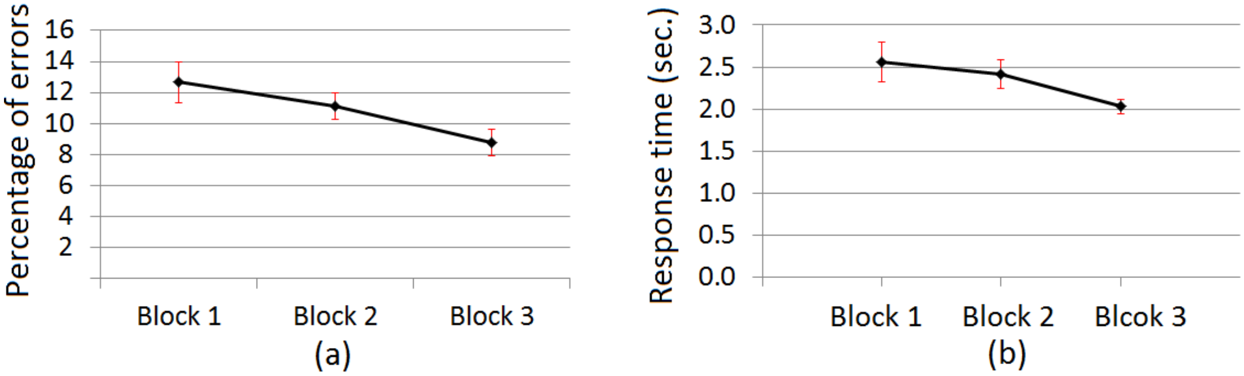


Figure 4.7: Average error rates and response times after each block. (a) Error rates, (b) Response time.

1) How many menu items can there be at each level?

From our study and preliminary interviews, we observed that people with visual impairments are less aware of directions in terms of cardinal directions. On the other hand, they were

Breath	Depth			
	1	2	3	4
4	1.25 (0.15)	1.75 (0.12)	2.50 (0.16)	3.34 (0.15)
6	1.39 (0.20)	1.99 (0.21)	2.88 (0.27)	4.04 (0.23)
8	1.16 (0.11)	2.13 (0.16)	2.89 (0.20)	3.82 (0.27)
12	1.35 (0.13)	2.29 (0.20)	3.43 (0.24)	4.21 (0.32)

Table 4.2: Mean response times. Standard errors are shown in parentheses.

able to successfully perform marking menus for each number of menu items (i.e. 4, 6, 8) used in our experiment. This is because we labeled the menu directions according to the human body, i.e. left, right, upper-right, etc. Thus, in general, we can suggest that people with visual impairments can perform directional gestures in up to 8 directions. According to the data analyses, people with visual impairments are able to perform motion gestures in 8 directions until level 3 (error rate = 6.41%, SD = 6.93) and up to 4 directions in level 4 (error rate = 3.15%, SD = 5.28) with error rates of less than 10%. However, in 6-item menus, the error rates became 10.25% (SD = 9.71) in level 3. Most of the menu items in 6-item menus existed “off-axis”, resulting in more errors made by the participants because more “off-axis” menu items were included in the combination as the number of menu levels increased (e.g., a menu selection with 3 levels may include more “off-axis” items such as up-right, down, up-left). Thus, we can suggest that our participants could successfully perform gestures in up to 6 or 8 directions until 3 levels, when the combination of menus to be selected included moderate numbers of “on-axis” and “off-axis” items.

When paid close attention to menu selections performed by participants with different levels of sight and different times when they became blind. We found no significant differences in performance by the participants. This suggests that the participants could distinguish the menu directions equally well when the directions were labelled in relation to the human body (left and right). The literature has also reported that, although people who are congenitally blind have more difficulty in representing spatial information allocentrically, tasks requiring egocentric frames of reference (relating to one's body) were similarly performed by early blind (early onset of blindness), late blind and sighted people [34].

2) How deep or how many hierarchic levels can people with visual impairments go before marking menu performance drops off?

Error rates are the main limiting factors for the number of hierarchic levels in marking menus [54]. According to analyses of our data, we can suggest that people with visual impairments can perform hierarchical marking menus up to 4 levels in 4-item menus (error rate = 3.15%, SD = 5.28), up to 3 levels for 8-item menus (error rate = 6.41%, SD = 6.93), and up to 2 levels for 6-item menus (error rate = 3.20%, SD = 8.01) with error rates less than 10% as shown in [54].

In terms of response time, menu selection time took longer in 8-item menus than in 6-item menus because of "cognition and choice time (time taken to make a decision from a number of choices)" [28]. A predictive model of menu selection also suggested that the length of marks increased as the number of items increased, resulting in longer pointing times [1]. However, the main limiting factor for reliable menu selection would be error rates because response times are significantly reduced after a number of practices [1, 54].

A previous marking menu study which was performed with sighted people for pen and mouse interactions recommended using menus with a breadth of 4 up to a depth of 4 (maximum error rate 5.10%, SD = 4.20) and menus with a breadth of 8 up to a depth of 2 (maximum error rate 8.82%, SD = 4.62)[54]. In our study, the nearest error rate to 8.82% was found when the participants performed gestures for 8-item menus up to a depth of 3 (error rate 6.41%, SD = 6.93). Comparing results from our study and those from [54], our

Criteria	**Usable menu layouts**
Error<1%	4×1 (0.00), 4×2 (0.00), 6×1 (0.76)
1% <Error <3%	4×3 (1.35), 8×1 (1.51)
3% <Error <6%	6×2 (3.79), 8×2 (5.30)
6% <Error <10%	4×4 (6.05), 8×3 (6.06)

Table 4.3: Criteria of acceptable error rates and usable menu layouts.

study results indicated lower error rates in all menu configurations. Thus, it is questionable whether people with visual impairments have any advantage over sighted people in spatial ability to perform hierarchic marking menus.

We speculated that the difference in performance from our study and that in [54] is because of the different input modalities. In [54], the participants drew marks on the screen using a pen or a mouse. In our study, the participants performed markings using motion gestures. Physical movement with motion gesture is an expressive channel which has six degrees of freedoms such that the direction of movement can be more easily related to the human body by applying proprioception. Moreover, menu direction in our study was labeled in relation to the human body (left, right) which allowed the participants to rely on kinesthetic cues (awareness of object positions in space with respect to one's body). Studies have also demonstrated that interactions took advantage of proprioception and kinesthetic cues over interaction in other forms of interaction [63, 89].

4.5 Study 6: Evaluating Motion Marking Menus in Smartphones

Study 1 confirmed that marking menus are usable as eyes-free input for people with visual impairments. Study 2 was intended to investigate user performance and subjective assessment using the motion marking menu system. Study 2 was intended to answer the following research question: How receptive are people with visual impairments to motion marking menu systems for mobile interactions? To better understand the relative efficiency

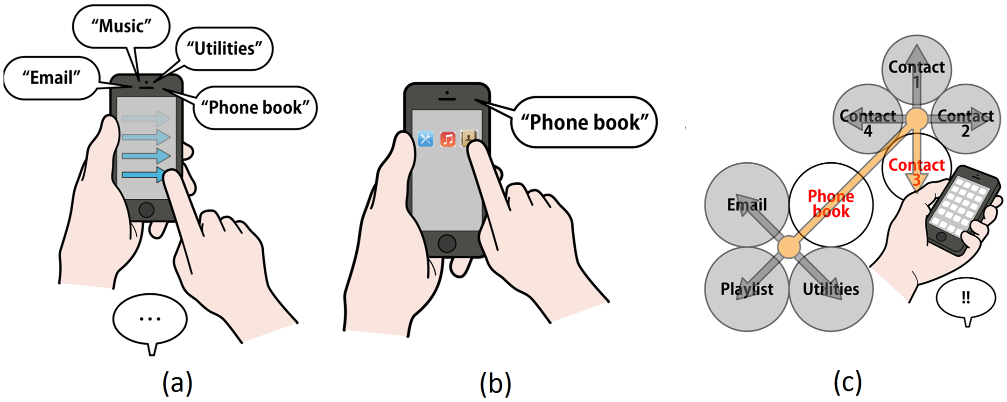


Figure 4.8: Menu systems used in experiment. (a) TalkBack - linear gestures, (b) TalkBack - spatial localization, (c) Motion Marking Menus.

and users assessments of the marking menu system, we compared our prototype to TalkBack, a commercial menu system on Android devices for users with visual impairments. TalkBack enables users to browse menus on touch screens using directional flick gestures and speech output.

4.5.1 Experiment Design

We developed a proof-of-concept prototype for motion marking menus with 4 × 2 (i.e. 4 items with 2 levels) menu configuration. Two mobile phone applications, phone book and music, were assigned for the high-level menu items. Four subtasks were assigned to each high-level menu. For example, four contact names were assigned in the phone book menu. The menus used in the study are shown in Figure 4.8. To help design the experiment and optimize the experimental tasks, a pilot study was conducted with one participant with visual impairments and two blind-folded participants. We chose four by four menu selections in order to minimize the cognitive load of the participants when remembering the menu layout and to optimize the time taken for the experiment.

The participants selected each submenu three times. The same eight menu items were assigned in the TalkBack system. Therefore, each participant performed 48 (2 menu systems

× 8 menu selections × 3 repetitions = 48) menu selections in total.

Experiments for selection techniques (i.e. marking menus and TalkBack) were conducted using a within-subject design. Task completion time and error rates were measured.

4.5.2 Participants and Apparatus

12 participants (3 female, 9 male) took part in Study 2. Nine of them participated in Study 1. The ages ranged from 26 to 78 years. Four of them were smartphone users. All were right-handed. Each was paid $10 for their participation.

For both menu systems, an Alcatel OneTouch Idol 2S smartphone (Processor - Quad-core 1.2 GHz, Dimension -5.37×2.74×0.30 in, Weight - 1.6g) was used. The motion marking menu system was implemented using a built-in accelerometer senor, in Java language. For the marking menu system, both novice and expert modes were implemented. In novice mode, the system read out all menu items at each menu level. In expert mode, only the final target menu was read out once it was selected. Novice mode was used to let the participants discover the menu items and practice menu selections. After practicing in novice mode, the participants used only the expert mode in the experimental trials.

(a) Participant using the TalkBack system

(a) Participant using the motion marking menu system

Figure 4.9: Participant performing menu selection tasks.

4.5.3 Procedure

The experimenter demonstrated each application and explained how to perform each task. The participants were allowed to practice with each system. The practice session took around 15 minutes. Once the practice session was completed, the experiment began. The order of menu systems was randomized among the participants. Half of the participants started with TalkBack, and the other half started with the marking menu system (Figure 4.9). The participants performed 24 trials for each of the following applications: (i) phone book (ii) music. Menu selection times and errors were recorded. All the experimental trials were video recorded.

After the experimental trials, the participants were asked to rate the respective systems in terms of ease of use, satisfaction with the systems and fatigue when performing the menu selections, using a 7-point Likert scale. They were also asked to rate on the statement, "I would like to use them if marking menus and motion gesture interfaces were available on smartphones." using a 7-point Likert scale (1 = strongly disagree and 7 = strong agree). After the experimenter interviewed the participants a set of questions, the experiment ended with the participants making questions or comments if they had any. We recorded all comments of the participants for qualitative data analysis. The experiment took around one hour.

4.5.4 Results

After analyzing the data collected in Study 2, we presented menu selection times, error rates and the subjective comments of our participants using the two menu systems.

Selection Times and Errors

A one-way repeated measures ANOVA test showed that the menu systems have significant effect on the selection times ($F(1,11) = 70.63$, $p < 0.05$). The average selection time for marking menus was 2.37 seconds ($SD = 0.57$) while selection time for TalkBack was 14.19

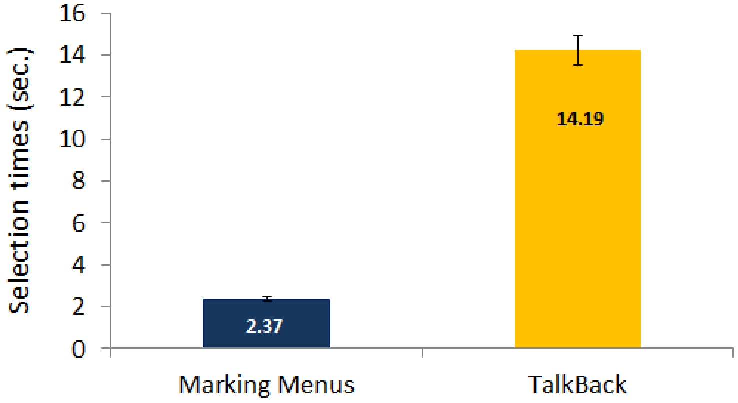


Figure 4.10: Selection times using the two menu systems.

seconds (SD = 4.46). The menu selection times in the two systems are shown in Figure 4.10.

Selection times both in TalkBack and marking menus decreased over blocks. For TalkBack, significant differences in selection time were found between block 1 and block 2 ($p < 0.05$) and between block 1 and block 3 ($p < 0.01$), but no significant difference was found between block 2 and block 3 ($p = 0.224$). For marking menus, significant differences were found between block 1 and block 3 ($p < 0.005$) and between block 2 and block 3 ($p < 0.05$), no significant difference was found between block 1 and block 2 ($p = 0.502$). Selection times over blocks were shown in Figure 4.11.

The average selection time in TalkBack in block 3 was 11.9 seconds while the selection time in the marking menus was 1.87 seconds. We remark that, in TalkBack, the speech feedback after every gesture or spatial localization was one cause of slow selection time, because users need to wait for voice feedback after each gesture. To make selection times in the two menu systems more comparable, we also analyzed the data after discarding the audio feedback time in TalkBack. Also, we analyzed the selection times ranging from the best cases to the worst cases, that is, where the target menus (main menu and submenu) located at the top in the menu list, and vice versa. Table 4.4 shows selection times in TalkBack with the voice feedback times discarded. For this purpose, we analyzed the data from block 3 as we regarded that users are the most familiar with and expert in using the TalkBack system

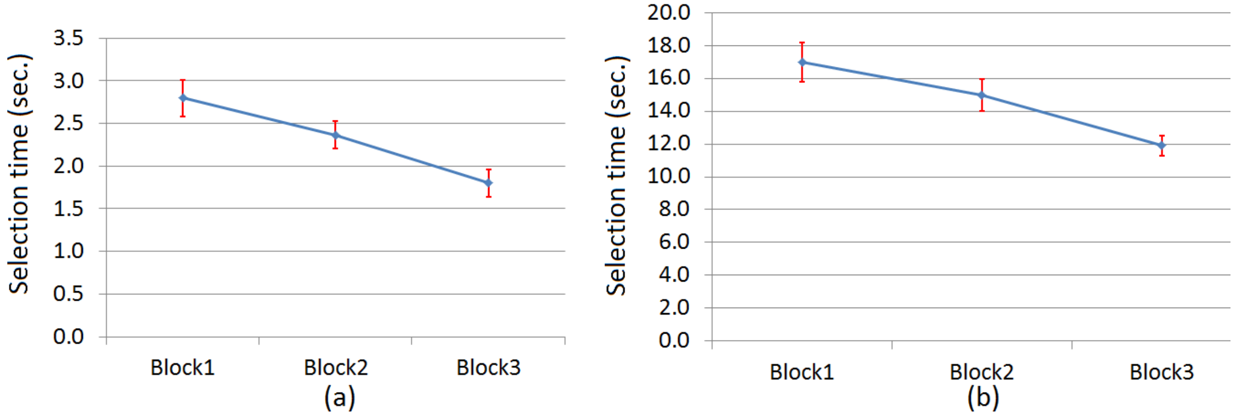


Figure 4.11: Selection times decreased over blocks in the two menu systems. (a) Motion Marking Menus, (b) TalkBack.

in block 3.

Regarding selection errors, in general, participants made more errors with marking menus than with TalkBack but there was no statistically significant effect on error rates. Error rates were 4.56% in the marking menu and 1.66% in the TalkBack system.

Subjective Assessment

Figure 4.12 shows the subjective ratings of the participants. The participants scored better for the motion marking menu system in all dimensions. A further Wilcoxon Signed-rank test showed significant differences in satisfaction ($z = -2.75$, $p < 0.05$) and fatigue ($z = -2.03$, $p < 0.05$), there was no significant difference for ease of use. When asked to rank on the statement regarding how much they want to use motion-marking menus, 7 participants rated to strongly agree and 3 participants rated to agree. Two of the participants preferred to use the traditional feature phones that they were currently using.

Most of the participants made positive comments about the marking menu system. One of the participants who was a smartphone user commented that using the motion marking menu was like using shortcuts and it was very efficient. The participant also mentioned, "Sometimes, the continuous voice guidance in current smartphones is frustrating in public

Main menu	**Submenu**	**Selection time (sec.)**
Phone	*Contact 1*	2.92 (0.73)
	Contact 2	3.31 (0.98)
	Contact 3	4.20 (1.02)
	Contact 4	5.20 (1.56)
Music	*Playlist 1*	3.73 (0.98)
	Playlist 2	4.64 (1.64)
	Playlist 3	5.20 (1.56)
	Playlist 4	6.43 (1.31)

Table 4.4: Mean selection time in each menu level after discarding speech feedback times.

places." Another participant who did not use a smartphone mentioned that both TalkBack and marking menus were great, really enabling people with visual impairments to use smartphones. Ten out of twelve participants preferred the motion marking menu system to the TalkBack system.

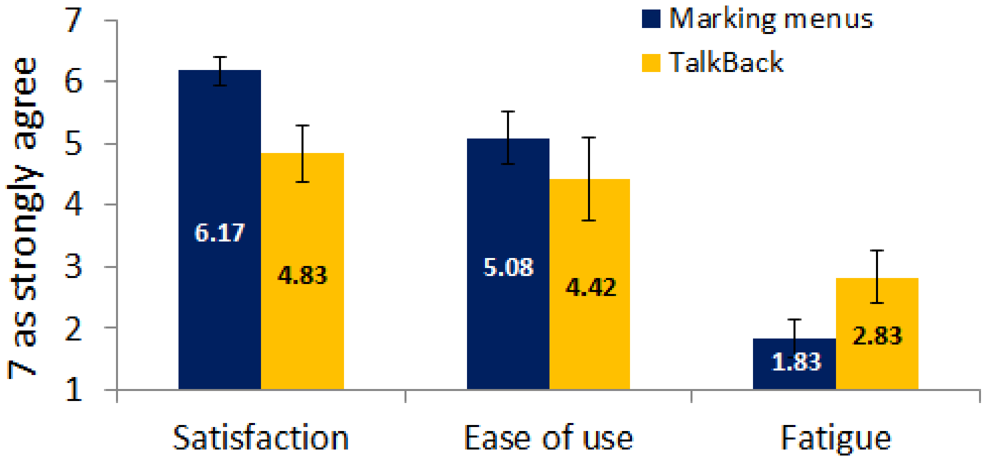


Figure 4.12: Subjective assessment on the two menu systems.

The participants' comments also showed that there are learnability issues in motion marking menu systems. Our participants commented that it was hard to remember all menu items in a short time, for example, one of the participants stated, “The gestures are easy to perform but it’s difficult to remember menu positions. But it would be different if I could arrange the menus by myself.” The same participant commented that customizing gestures for calling frequently called contacts would be very useful. Another participant mentioned that the marking menu system was more difficult to learn compared to the TalkBack system.

When doing the gestures for menu selections, most of the participants used small movements (i.e. moving only the wrist). Using small movements, gestures to left were particularly difficult for the participants. One of the participants stated, “Leftward movements are very difficult. I should move my arm (i.e large movement using elbow) for gestures to left.” Another participant stated, “I want to use only one stroke motion for frequently used functions like answering calls.” Comments from the participants indicated that the participants were very alert to the potential benefit of using motion marking menu systems.

4.5.5 Discussion

Recalling research question 4, “How receptive are people with visual impairments to motion marking menu systems for mobile interactions?,” both quantitative and qualitative data from Study 2 convinced us that these interfaces are efficient and desirable for people with visual impairments. In particular, the subjective comments of our participants confirmed that people with visual impairments are receptive to motion marking menus on smartphones as a way of interacting with smartphone devices. In addition, marking menus would be easier for people with visual impairments to learn and use because these users are less aware of other forms of gesture such as iconic gestures.

While the key advantage of motion gesture interfaces is that they enable one-handed operations for users, we also observed two important features of motion gesture interfaces throughout the study. (1) The first is the familiarity of the interaction space. For visually impaired users, familiarity of the interaction space is particularly important. Compared to motion gesture interfaces, our participants had much more difficulty to get familiar with and perform the tasks on touchscreens. Tasks on touchscreens such as gestures and taps were particularly difficult for some of our participants. (2) They provide more freedom of interaction boundary. That is, users can start the interactions anywhere in 3D space, not limited to the space of a device (e.g., touchscreen).

Nevertheless, we have to declare that motion gestures are not the most suitable for all cases. For example, users may not remember the menu items which are not frequently used. In those cases, navigating the menus on touchscreens will be easier for users than recalling those menus from a motion marking menu system. Thus, we regard that currently available menu systems and motion marking menus can be good alternatives and supplements.

Throughout the studies, we collected the participants' subjective comments about gesture preferences, menu selections and preferred menu layouts, etc. Based on comments from the participants, we provided design guidelines for motion marking menus in mobile devices.

Preferred Gestures

In Study 1, all participants commonly stated that diagonal directions (i.e. upper-left, upper-right, etc.) are difficult to instantly understand. Because all our participants were right-handed, they mentioned that gestures to lower-left and upper-left directions were the most tiring gestures to perform. Thus, designers should avoid gestures in the opposite direction of the dominant hand for frequently used functions. The participants also mentioned that downward gestures were particularly easy to perform because they followed the force of gravity. Regarding "mental slip" errors caused in left and right directions, menus in the left and right positions should be designed such that the consequences of operations are close (e.g. Contact Book and Phone menus in mobile phones) or left and right directions should be avoided for functions where the consequences of making mistakes are serious (e.g., answering a call and declining a call).

Menu Layout

Our participants suggested that one motion (i.e. single stroke gesture) was desirable for functions such as answering or declining a call. Also, gesture customization should be supported wherever possible. The subjective comments informed us that end-user customization is particularly desirable for functions such as Phone (arranging contact names) and Media Player (customizing playlists). Although customization can be allowed for menu levels up to the extent of efficiency and comfort with which people with visual impairments can perform (as discussed earlier), for default menus assigned by designers, menu items of 4, 6, 8 with menu up to 2 levels deep would be the most preferable when considering the learnability and memorability of menu layouts for users with visual impairments.

Menu Learnability

There remains a need to consider the learnability of the entire menu layout in novice mode for motion marking menus. One advantage of the traditional marking menu is that it helps users make efficient transitions from novice to expert modes and skill levels because it provides visual information about the entire menu layout in novice mode. Once transited to expert mode, marking menu works like providing users with a shortcut for the target by continuous strokes. For motion marking menus, information about the entire menu layout should be provided to users using speech, before the user switches to the expert mode. Once switched to the expert mode, users could select the target by a series of continuous motions. Although participants could learn menu layout and transit to the expert mode easily in our experiment, more learnability would be facilitated if menu items were arranged as close as possible to the users' expectations, e.g., place the most frequently used menus in on-axis positions, and sort the menu items in meaningful ways. This is particularly important because mental mappings are a great aid in learning of visually impaired people [94]. End user personalization of menus would also be helpful for learnability and memorability of menu layouts.

Gesture Recognition

When performing menu selections in Study 2, we found that the participants had different preferences for movement to select a menu. Movements were different in length and velocity, etc. It was difficult for the system to recognize gestures performed using movements with very low velocity. Gesture recognition should be implemented to allow more freedom of movement to the users. However, it is not always preferable to have gesture recognition that is too sensitive. Thus, in real applications, it would be a good idea to allow end user customization of thresholds for gesture recognition (e.g. slow movement, small gesture, etc.).

Feedback

For successful menu selections, feedback was provided with the menu names read aloud. Feedback and guidance should also be provided when menu selections are not successful. For example, users may not perform enough movement necessary to trigger menu selection. In that case, vibrations or speech guidance should be issued to notify required actions to the users. Also, it would be desirable to give more feedback options. Users may prefer a non-verbal sound such as vibration feedback to speech, once they become more expert at using the systems.

Chapter 5

DESIGNING WEARABLE VIBRATION FEEDBACK

5.1 Introduction

Research has shown that motion gestures were the most effective when vibration feedback was provided. In particular, spatial vibration feedback can enhance user's spatial ability to perform motion gestures. Vibration feedback has unique affordances to support interactions of people with visual impairments in two ways: (i) enhancing user capabilities for performing motion gesture inputs, and (ii) providing non-visual outputs. Researchers have also shown that vibration feedback in different body positions are more effective for providing spatial information to users [2]. Nowadays, wearable vibration devices are gaining interests for interactions in various applications, indicating that wearable vibration applied in different body positions can be beneficial both for input and output for people with visual impairments.

Despite the potential benefit of wearable vibration devices, understanding about suitable bodily positions is lacking. Furthermore, important factors such as context of use, user comfort, wearability and mobility, appropriate vibration intensities which are important factors in designing wearable vibration devices for navigation have not been adequately explored. Thus, it is important to find the most suitable vibration positions to convey directional information in real situations including home and out; these should also be positions that are practically desirable for users to wear in real contexts.

Thus, we conducted user studies to determine the most suitable body sites for wearable vibrations and to provide design guidelines. For the study, we proposed 15 vibration positions (ears, neck, chest, waist, wrist, hand, finger, ankles and feet). We also studied the most suitable vibration intensity for different posture (static, walking and fast walking). User perception ratings and errors were recorded. The studies were conducted in real mobile environments. We addressed three specific research questions:

Q1: Among the body sites of interest, which positions offer the best vibration perception in static and walking conditions respectively?

Q2: Which vibration intensities are suitable for those body parts?

Q3: Which vibration positions are the most preferable for users to use in the real mobile setting?

The study results indicate that the ears, fingers, neck and waist are the most sensitive to vibrations, followed by the hand, chest and writs. The feet and ankle were the least sensitive to vibrations. Subjective ratings indicated the writs, waist, hand and neck are the most preferable vibration positions for users with visual impairments.

The contributions of our study are:

- Assessment of vibration perception on different body sites when users are walking in the real mobile environment.
- Identifying the most suitable body sites for walking navigation.
- Suggesting suitable vibration intensities for different body part.
- Concrete design guidelines for wearable vibration devices for people with visual impairments.

5.2 Related Work

The literature reports several applications of wearable vibration devices. There is also a substantial body of psychophysical studies regarding haptic perception of the human body. For our study which is focused on vibration feedback for walking navigation, we highlight the most relevant literature in human haptic perception, wearable vibration and walking navigation.

5.2.1 Human Haptic Perception

Haptic perception across different body sites varies depending on the size and adaptation of the receptors in each body site [39, 40]. Many researchers have attempted to examine vibration sensitivity of particular body sites including the back, waist, chest, wrist and thigh [41, 47, 95]. Among the body sites of interest, the wrist was proved to be the most sensitive to vibration; thighs and feet are the least sensitive, followed by the waist, arms and chest. Lederman et al. [57] who reported the sensitivity of different body sites to haptic stimuli based on two-point and point-localization threshold methods, noted that the fingertips had the highest haptic acuity. Notably, the more distal parts of the body have higher haptic spatial acuity [56].

Animal and human experiments have also shown that the cortical responses to cutaneous stimuli are profoundly diminished during movement [13, 38, 80]. [4] demonstrated correlations between sensory suppression and movement speed in detection rates of the index finger tip. Other researchers studied the same effect using vibrotactile simulation[65, 71, 76].

The studies consistently showed that movement during physical exertion had a significant effect on the perception of tactile simulation. However, there is no clear understanding about the relative sensitivity of these body sites of interest when exposed to motion in walking in the real mobile settings.

5.2.2 Wearable Vibration

Van et al. [97] presented a multi-purpose tactile vest for astronauts that supported the astronaut's orientation awareness. The location of vibration on the torso indicated the direction of the standard International Space Station (ISS) orientation. Rukzi et al. [79] developed a guidance system that used vibration on a mobile phone to represent directions on a public display. The user's phone vibrated when direction on the public display matched the user's route direction. Wearable haptic devices were also evaluated for driving support systems. Ho et al. [29] examined directional vibration feedback in a driving simulation. The study reported that encoding directional information of an oncoming car on the torso (i.e., front vs back stimuli) was promising. Other researchers [88] compared tactile and audio feedback for pedestrian collision warning applications for drivers. For the tactile feedback, tactors were attached to the drivers' left and right biceps. The study indicated that tactile feedback on the drivers' biceps was more effective than audio signals for collision warnings. Other researchers [96] demonstrated applications of tactile displays on the body for sports and showed that tactile feedback improves rowing efficiency compared to traditional feedback systems.

Some researchers also demonstrated the use of wearable vibration feedback for non-visual feedback for navigations, for example, directional vibration feedback on the waist belt [27, 92, 98]. Other researchers [18, 90] also applied directional vibration feedback on the users' back for navigation. The studies showed that tactile feedback can be used to provide non-visual information to help users find their way around. However, differentiating directional vibration patterns on a single body site is confusing and probably increases the users' cognitive load in overloaded mobile environments. This assumption was confirmed by [2] who found that it was difficult for participants to perceive spatial vibration patterns on particular body sites. Bosman et al. [10] introduced GentleGuide, a two wrist-mounted vibration feedback system for indoor navigation. The study showed that two vibration devices, one on each wrist is more effective than using a single output device on one location. Consistent with this finding, we proposed that optimal tactile interfaces for walking navigation can be achieved by an effective combination of several body sites.

All these studies consistently indicate the potential utility of vibration feedback for future wearable computing. However, the most suitable bodily positions of vibrators for particular applications were not adequately understood.

In summary, human perception of haptic sensation varies across human body sites, and it was confirmed that movement degrades performance in haptic perception. The literature has also shown several applications of wearable haptic sensors on different body sites. However, factors remaining unanswered are (a) relative interference of movement in the real mobile context on vibration detection of body sites, and (b) the most effective body sites for vibration feedback in walking navigation. Our study focuses on effective body sites and design guidelines for vibration feedback in walking navigation.

5.3 Study 7: Suitable Vibration Positions and Vibration Intensities

5.3.1 Design

The study was conducted in a public park. We investigated the participants' vibration perception in three walking conditions: static, walking and fast walking, and three vibration levels: weak, medium and strong vibrations. Walking speeds for each condition for all participants were recorded. The average walking speeds were 4.36 km/h for normal walking and 5.28 km/h for fast walking. As for the vibration levels, we conducted a pilot study to determine suitable vibration intensities. Informed by the pilot study, we used 1200 rpm for weak vibration, 2200 rpm for medium vibration and 3200 rpm for strong vibration.

Regarding the proposed vibration positions, our design was based on careful consideration as to whether vibrations are easier to perceive in particular positions, whether vibrators can be easily integrated into clothing or accessories, and whether vibrations in those positions are wearable in the long-term in real-world settings. Thus, the vibration system was made so that the vibrations could be wearable and felt through the users' clothing or accessories.

The vibration positions, walking condition and vibration levels were with-subject. Thus each participant performed 3 motion conditions × 3 vibration levels × 15 vibration positions × 3 repetitions = 405 trials in total.

5.3.2 Participants and Apparatus

Twelve participants (11 male and 1 female, ages ranged from 27 to 60, height ranged from 150cm to 180cm, Mean height = 166.91cm, weights ranged from 45kg to 80kg, mean weight = 36kg) were recruited. One of the participants could see light, two of them could see objects but unable to distinguish between objects, two of them could see objects. The rest were totally blind. All participants were right-handed. The participants were paid $20 for their participation.

Fifteen LilyPad vibrators (rated speed - 12000 rpm, frequency - 200 Htz, vibration amplitude - 0.8G, dimension - 20mm outer diameter and thin 0.8mm PCB, weight - 1.2g) were used. The motor controller in Experiment 2 consisted of a vibrator controller and a remote controller. For the motor controller, Arduino Uno Adafruit 16-channel motor driver shield and PWM driver circuit was used to drive 15 vibrators. The motor driver shield was powered by an SG 9V battery which drives the vibration motors at 6V. For the remote console, Arduino Mega and 15 push buttons were used. 4 AA (1.5V) batteries were used to power the remote controller. The vibration controller and the remote console were connected using two Xbee wireless modules. Each Xbee module was connected to the Arduino Uno and Arduino Mega using Xbee shield (see Figure 5.1).

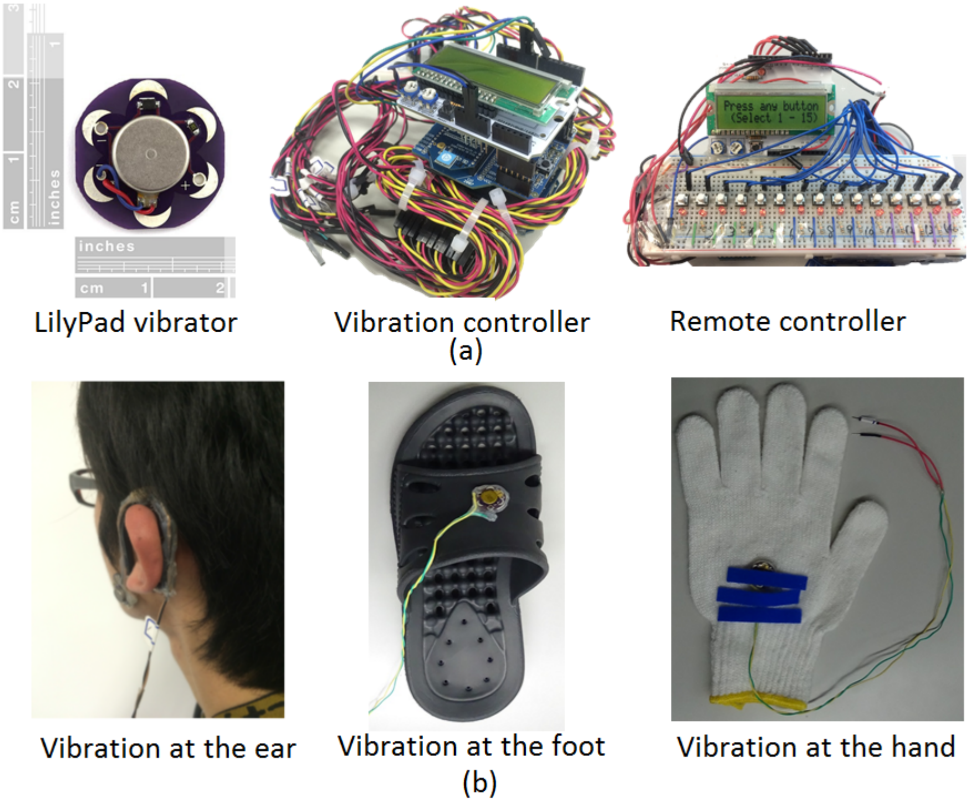


Figure 5.1: The vibration used in the experiment. (a) Vibration system, (b) Examples of vibrators attached to accessories.

Figure 5.2: Participant in static condition.

5.3.3 Procedure

Before the experiment, participants were asked to change into the experimental clothing (see Figure 5.2). The vibrations were randomized across the body sites. The interval between vibrations was randomized to between 3 and 5 seconds. The participants were instructed to say aloud the vibration position when they received a vibration. Errors in the perceptions of vibration positions were recorded.

The participants performed the experimental tasks in 3 conditions: static, normal walking and fast walking. After completing the experimental tasks for one condition, the participants were asked to rate the perceptibility of each vibration position on a 7-point Likert scale (1 = very difficult, 7 = very easy to perceive). Using the same scale, the participants were asked to rate the statement "I would like to wear the vibrator in this position, in real-world usage" for each vibration position (1 = strongly disagree and 7 = strongly agree).

5.3.4 Results

Data of interest were participants' rating on vibration perception and preferences of vibration positions for real-world applications.

A three-way repeated-measure ANOVA (analysis of variance) was used with the two factors of the number of vibration positions, motion conditions and vibration levels (depth). All tests were run at a significant level of alpha (α) = 0.05.

Vibration Perception

Vibration level and vibration positions interact to affect the vibration perceptions ($F(28, 308) = 16.37$, $p<0.05$). Motion condition and vibration positions also had significant interaction to affect the perceptions ($F(28, 308) = 32.47$, $p <0.05$).

Motion condition significantly affected the participants' vibration perception ($F(2,22) = 4.63$, $p <0.05$). The vibration level ($F(2, 22) = 13.55$, $p <0.05$) and vibration positions ($F(14, 154) = 85.28$, $p <0.05$) also had a significant affect on perceptions. The perception ratings for each vibration position in the motion conditions are shown in Table 5.1. Perception ratings in vibration levels are shown in Table 5.2. Overall, vibration on the ears and the finger had the highest perceptibility and vibration on the chest offered the lowest perceptibility in all conditions. Perception scores of the vibration positions in each motion condition are shown in Figure 5.3. Perception scores of the vibration positions in each vibration level are shown in Figure 5.4.

Errors

Vibration level and vibration positions interact to affect the errors ($F(28, 308) = 2.03$, $p <0.05$). Motion condition and vibration positions also had significant interaction to affect the errors ($F(28, 308) = 1.63$, $p <0.05$).

Motion condition significantly affected the participants' vibration errors ($F(2,22) = 4.63$, $p<0.05$). The vibration level ($F(2, 22) = 13.55$, $p<0.05$) and errors ($F(14, 154) = 8.10$,

Vibration Position	Motion Condition			Mean
	Standing	Walking	Fast Walking	
Ear	6.61 (0.14)	6.61 (0.15)	6.22 (0.20)	6.48
Neck	6.28 (0.16)	5.89 (0.26)	5.67 (0.26)	5.94
Chest	6.00 (0.21)	5.69 (0.28)	5.25 (0.26)	5.65
Waist	6.25 (0.25)	5.97 (0.36)	5.58 (0.40)	5.94
Wrist	5.67 (0.28)	5.57 (0.26)	5.25 (0.31)	5.50
Hand	5.69 (0.28)	5.82 (0.25)	5.46 (0.30)	5.66
Finger	6.11 (0.26)	6.24 (0.21)	6.03 (0.24)	6.13
Ankle	5.26 (0.39)	5.17 (0.44)	4.75 (0.45)	5.06
Foot	5.04 (0.44)	4.99 (0.47)	4.69 (0.46)	4.91

Table 5.1: Mean perception ratings (with standard errors) of the participants for each vibration position and motion condition.

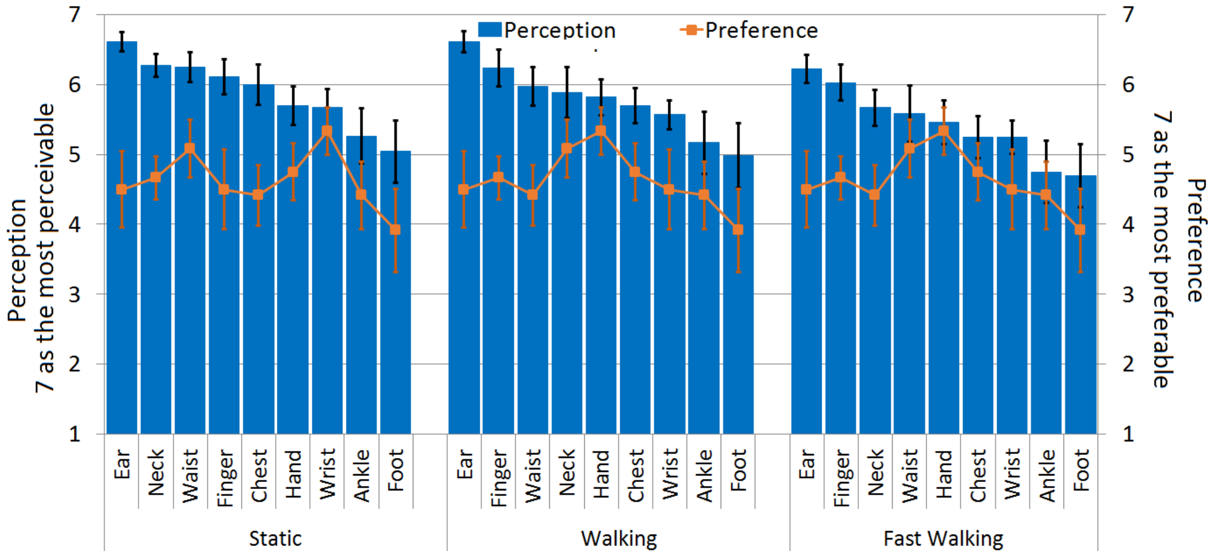


Figure 5.3: Perception ratings (with standard error bars) in each motion condition.

Vibration Position	**Vibration Level**			**Mean**
	Level 1	**Level 2**	**Level 3**	
Ear	6.19 (0.27)	6.47 (0.18)	6.78 (0.07)	6.48
Neck	5.31 (0.29)	6.08 (0.20)	6.44 (0.19)	5.94
Chest	5.08 (0.33)	5.78 (0.29)	6.08 (0.25)	5.65
Waist	5.36 (0.38)	6.14 (0.30)	6.31 (0.34)	5.94
Wrist	4.86 (0.26)	5.65 (0.30)	5.97 (0.32)	5.50
Hand	5.14 (0.30)	5.71 (0.28)	6.13 (0.28)	5.66
Finger	5.76 (0.32)	6.13 (0.22)	6.49 (0.18)	6.13
Ankle	4.31 (0.47)	5.21 (0.46)	5.67 (0.40)	5.06
Foot	4.01 (0.57)	5.18 (0.45)	5.33 (0.39)	4.91

Table 5.2: Mean perception ratings (with standard errors) of the participants for each vibration position and motion condition.

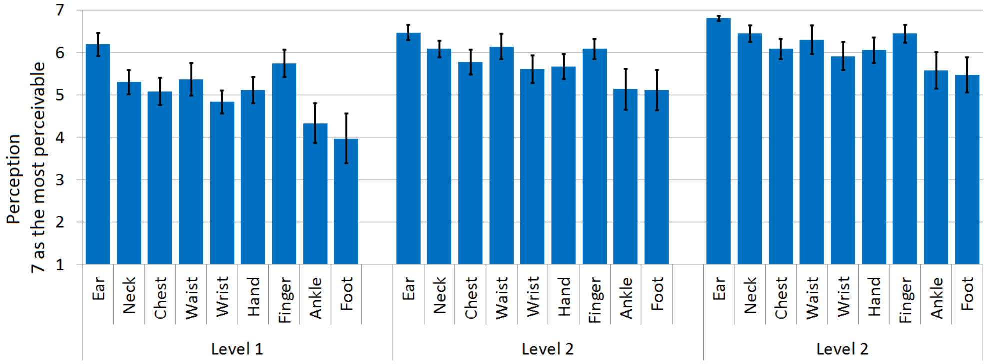


Figure 5.4: Perception ratings (with standard error bars) in each vibration level.

$p<0.05$) also had a significant affect on perceptions.

Post hoc test with Bonferroni indicated that for all conditions, the wrist and the feet had significantly higher error rates than other positions. Error rates for each vibration position are shown in Table 5.3.

Vibration Position	**Level 1**	**Level 2**	**Level 3**
Ear	0%	0%	0%
Neck	0%	0%	0%
Chest	2.78%	0%	0%
Waist	0%	0%	0%
Wrist	4.17%	2.78%	2.78%
Hand	0.46%	1.39%	0.46%
Finger	0.93%	0.46%	0%
Ankle	3.24%	1.85%	0.46%
Feet	2.31%	0%	0%

Table 5.3: Percentage of errors in each vibration position.

User Preference

The wrist, waist, hand and finger had the highest scores of preferences. Most of the participants stated that vibration on the chest is the least comfortable. The choice of vibrator positions for our study was based on careful consideration as to whether vibrators can be easily integrated into clothing or accessories in those positions in real-world settings. Thus, the best perceivable vibrator positions and the most preferable positions are not much different. The participants' subjective ratings for each body part as a suitable vibrator positions are shown in Figure 5.3.

5.3.5 Discussion

Informed by the study results, research questions were addressed as follows.

Among the body sites of interest, which positions offer the best vibration perception in static and walking conditions respectively?

In term of vibration perceptions, the ears, neck, waist and fingers offer the best perceptions, followed by the wrists and hands. The feet, ankles and the chest offer the lowest

Vibration Position	**Suitable Vibration Level**
Ear	Level 1
Finger	Level 1
Waist	Level 1
Neck	Level 2
Chest	Level 2
Wrist	Level 2
Hand	Level 2
Foot	Level 2
Ankle	Level 3

Table 5.4: Suitable vibration levels for each vibration position. (Level 1 = 1200 rpm, Level 2 = 2200 rpm, Level 3 = 3200 rpm).

vibration perceptions.

Which vibration intensities are suitable for those body parts?

Different vibration positions require different vibration levels to maintain reasonable perceptions. While errors and perceptions were affected by motion conditions, we found that discrepancy in performances were largely compensated by vibration levels. Thus, it is important to know suitable vibration levels which maintain reasonable performance for each position. Table 5.4 shows recommended vibration levels for each body part.

Which vibration positions are the most preferable for users to use in the real mobile setting?

In terms of wearability and user comfort, the wrist, waist and the neck were the most preferable, followed by the hand and fingers. The feet, ankles and chest were not preferable both in term of perceptions and users' comfort. Despite the highest perceptions at the ears, the ears are not desirable for wearable vibration, given that visually impaired people largely rely on hearings for their safety in mobile situations. The chest were also found not to be comfortable for users in addition to low perceptions.

Chapter 6

CONCLUSIONS

6.1 Summary of Dissertation

This dissertation presents four studies which focused on understanding and designing motion gestures for people with visual impairments. Study 1 and 2 are concerned with understanding user capabilities, user gestures, gesture rationale and taxonomies. In these two studies, we presented gesture taxonomies and theoretical guidelines. We also proposed two interaction techniques which were evaluated by visually impaired users. The proposed interaction techniques achieved desirable results both in term of user performance and subjective assessments.

In Study 3, we proposed a hybrid participatory design approach, based on our experiences and insights gained from Study 1 and 2. We proposed a design approach which leveraged the user-elicitation and choice-based elicitation approaches. We were confirmed that this hybrid approach was a promising design approach especially when designing gestures for visually impaired users.

From gesture studies, we observed the need of feedback to support user capabilities to perform the gestures. Thus we proposed the use of wearable vibrations to support users' spatial ability in performing motion gestures. For this purpose, we assessed and presented suitable body positions for spatial vibrations. We discussed several design guidelines of wearable vibrations with respect to vibration perceptions, wearability, suitable vibration intensities and user preferences.

The outcomes of this dissertation are:

1. Understanding of user gestures, gesture rationale and taxonomies
2. New interaction techniques

3. A hybrid participatory design approach
4. Theoretical guidelines

In summary, this dissertation contributes to the field of assistive technologies for visually impaired people, in the view of gesture based interactions. The conclusion drawn and methodologies proposed will benefit future research studies that explore gesture-based interaction techniques and scientific foundations of assistive technologies for people with visual impairments.

6.2 Future Direction

Throughout the project, the main design insight we have achieved is that instead of building the interfaces only around knowledge about users' disabilities, it is important to identify users' capabilities and to provide suitable interaction techniques for a given situation. This will improve users' confidence and motivation to use certain technologies. This also informed us a pressing need of ways to increase users' existing capabilities. Thus, in the future we will conduct research studies on increasing user capabilities such as spatial ability and cognitive abilities.

BIBLIOGRAPHY

[1] Ahlström, D., Cockburn, A., Gutwin, C., and Irani, P. (2010). Why it's quick to be square: modelling new and existing hierarchical menu designs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1371–1380. ACM.

[2] Alvina, J., Zhao, S., Perrault, S. T., Azh, M., Roumen, T., and Fjeld, M. (2015). Omnivib: Towards cross-body spatiotemporal vibrotactile notifications for mobile phones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2487–2496. ACM.

[3] American Foundation for the Blind (1997). Blindness statistics. Available: `http://www.afb.org/info/blindness-statistics/statistical-snapshots-archive/technology-related-statistics-archive/technology-video-description/2345`, Accessed on June 30, 2015.

[4] Angel, R. W. and Malenka, R. C. (1982). Velocity-dependent suppression of cutaneous sensitivity during movement. *Experimental neurology*, 77(2):266–274.

[5] Azenkot, S. and Fortuna, E. (2010). Improving public transit usability for blind and deaf-blind people by connecting a braille display to a smartphone. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, pages 317–318. ACM.

[6] Azenkot, S. and Lee, N. B. (2013). Exploring the use of speech input by blind people on mobile devices. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, page 11. ACM.

[7] Bauer, J., Ebert, A., Kreylos, O., and Hamann, B. (2013). Marking menus for eyes-free

interaction using smart phones and tablets. In *Availability, Reliability, and Security in Information Systems and HCI*, pages 481–494. Springer.

[8] Bergstrom-Lehtovirta, J. and Oulasvirta, A. (2014). Modeling the functional area of the thumb on mobile touchscreen surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1991–2000. ACM.

[9] Bonner, M., B. J. A. G. and Edwards, W. (2010). No-look notes: Accessible eyes-free multi-touch text entry. In *Proc. Pervasive 2010*, pages 409–427. Springer.

[10] Bosman, S., Groenendaal, B., Findlater, J.-W., Visser, T., de Graaf, M., and Markopoulos, P. (2003). Gentleguide: An exploration of haptic output for indoors pedestrian guidance. In *Human-computer interaction with mobile devices and services*, pages 358–362. Springer.

[11] Bousbia-Salah, M., Fezari, M., and Hamdi, R. (2005). A navigation system for blind pedestrians. In *16th IFAC World Congress*.

[12] Butler, A., Izadi, S., and Hodges, S. (2008). Sidesight: multi-touch interaction around small devices. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, pages 201–204. ACM.

[13] Chapman, C., Bushnell, M., Miron, D., Duncan, G., and Lund, J. (1987). Sensory perception during movement in man. *Experimental Brain Research*, 68(3):516–524.

[14] Cowan, N. (2010). The Magical Mystery Four: How Is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science*, 19(1):51–57.

[15] Dim, N. K. and Ren, X. (2014). Designing motion gesture interfaces in mobile phones for blind people. *Journal of Computer Science and Technology*, 29(5):812–824.

[16] Douglas, S. A. and Willson, S. (2007). Haptic comparison of size (relative magnitude)

in blind and sighted people. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, pages 83–90. ACM.

[17] Ecker, R., Broy, V., Butz, A., and De Luca, A. (2009). pietouch: a direct touch gesture interface for interacting with in-vehicle information systems. In *Proceedings of the 11th international Conference on Human-Computer interaction with Mobile Devices and Services*, page 22. ACM.

[18] Ertan, S., Lee, C., Willets, A., Tan, H., and Pentland, A. (1998). A wearable haptic navigation guidance system. In *Wearable Computers, 1998. Digest of Papers. Second International Symposium on*, pages 164–165. IEEE.

[19] Ferati, M., Mannheimer, S., and Bolchini, D. (2011). Usability evaluation of acoustic interfaces for the blind. In *Proceedings of the 29th ACM international conference on Design of communication*, pages 9–16. ACM.

[20] Francone, J., Bailly, G., Lecolinet, E., Mandran, N., and Nigay, L. (2010). Wavelet menus on handheld devices: stacking metaphor for novice mode and eyes-free selection for expert mode. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pages 173–180. ACM.

[21] Guerreiro, T., L. P. N. H. G. D. and Jorge, J. (2008). From Tapping to Touching: Making Touch Screens Accessible to Blind Users. *IEEE Multimedia*, 15(4):48–50.

[22] Gustafson, S., Bierwirth, D., and Baudisch, P. (2010). Imaginary interfaces: spatial interaction with empty hands and without visual feedback. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 3–12. ACM.

[23] Gutschmidt, R., Schiewe, M., Zinke, F., and Jürgensen, H. (2010). Haptic emulation of games: haptic sudoku for the blind. In *Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments*, page 2. ACM.

[24] Harrison, C. and Hudson, S. E. (2009). Abracadabra: wireless, high-precision, and unpowered finger input for very small mobile devices. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, pages 121–124. ACM.

[25] Harrison, C. and Hudson, S. E. (2010). Minput: enabling interaction on small mobile devices with high-precision, low-cost, multipoint optical tracking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1661–1664. ACM.

[26] Harrison, C., Tan, D., and Morris, D. (2010). Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 453–462. ACM.

[27] Heuten, W., Henze, N., Boll, S., and Pielot, M. (2008). Tactile wayfinder: a non-visual support system for wayfinding. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*, pages 172–181. ACM.

[28] Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4(1):11–26.

[29] Ho, C., Tan, H. Z., and Spence, C. (2005). Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(6):397–412.

[30] Hollins, M. (1989). *Understanding blindness: An integrative approach.* Lawrence Erlbaum Associates, Inc.

[31] Hribar, V. E. and Pawluk, D. T. (2011). A tactile-thermal display for haptic exploration of virtual paintings. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, pages 221–222. ACM.

[32] Hub, A., Diepstraten, J., and Ertl, T. (2004). Design and development of an indoor navigation and object identification system for the blind. In *ACM SIGACCESS Accessibility and Computing*, number 77-78, pages 147–152. ACM.

[33] Hutchins, E., Hollan, J., and Norman, D. (1985). Direct Manipulation Interfaces. *Human-Computer Interaction*, 1(4):311–338.

[34] Iachini, T., Ruggiero, G., and Ruotolo, F. (2014). Does blindness affect egocentric and allocentric frames of reference in small and large scale spaces? *Behavioural brain research*, 273:73–81.

[35] Iverson, J. M. and Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature*, 396(6708):228–228.

[36] Iyengar, S. (2011). *The Art of Choosing.* Twelve.

[37] Jain, M. and Balakrishnan, R. (2012). User learning and performance with bezel menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2221–2230. ACM.

[38] Jiang, W., Chapman, C., and Lamarre, Y. (1991). Modulation of the cutaneous responsiveness of neurones in the primary somatosensory cortex during conditioned arm movements in the monkey. *Experimental brain research*, 84(2):342–354.

[39] Johnson, K. O. (2001). The roles and functions of cutaneous mechanoreceptors. *Current opinion in neurobiology*, 11(4):455–461.

[40] Johnson, K. O. and Hsiao, S. S. (1994). Evaluation of the relative roles of slowly and rapidly adapting afferent fibers in roughness perception. *Canadian journal of physiology and pharmacology*, 72(5):488–497.

[41] Jones, L. A., Kunkel, J., and Piateski, E. (2009). Vibrotactile pattern recognition on the arm and back. *Perception*, 38(1):52.

[42] Jung, Y. H. and Qin, S. (2011). User-defined gesture sets using a mobile device for people with communication difficulties. In *Automation and Computing (ICAC), 2011 17th International Conference on*, pages 34–39. IEEE.

[43] Kane, S. K., Bigham, J. P., and Wobbrock, J. O. (2008). Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, pages 73–80. ACM.

[44] Kane, S. K., Jayant, C., Wobbrock, J. O., and Ladner, R. E. (2009). Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, pages 115–122. ACM.

[45] Kane, S.K., W. J. and Ladner, R. (2011). Usable gestures for blind people: Understanding preference and performance. In *Proc. CHI 2011*, pages 413–422. ACM.

[46] Karam, M. et al. (2005). A taxonomy of gestures in human computer interactions.

[47] Karuei, I., MacLean, K. E., Foley-Fisher, Z., MacKenzie, R., Koch, S., and El-Zohairy, M. (2011). Detecting vibrations across the body in mobile contexts. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 3267–3276. ACM.

[48] Kim, K., Ren, X., Choi, S., and Tan, H. Z. (2016). Assisting people with visual impairments in aiming at a target on a large wall-mounted display. *International Journal of Human-Computer Studies*, 86:109 – 120.

[49] Kratz, S. and Rohs, M. (2009). Hoverflow: expanding the design space of around-device interaction. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, page 4. ACM.

[50] Kray, C., Nesbitt, D., Dawson, J., and Rohs, M. (2010). User-defined gestures for connecting mobile phones, public displays, and tabletops. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, pages 239–248. ACM.

[51] Krishna, S., Colbry, D., Black, J., Balasubramanian, V., and Panchanathan, S. (2008). A systematic requirements analysis and development of an assistive device to enhance the social interaction of people who are blind or visually impaired. In *Workshop on Computer Vision Applications for the Visually Impaired*.

[52] Kurdyukova, E., Redlin, M., and André, E. (2012). Studying user-defined ipad gestures for interaction in multi-display environment. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, pages 93–96. ACM.

[53] Kurtenbach, G. and Hulteen, E. A. (1990). Gestures in human-computer communication. *The art of human-computer interface design*, pages 309–317.

[54] Kurtenbach, G. P. (1993). *The design and evaluation of marking menus*. PhD thesis, University of Toronto.

[55] Law, C. and Vanderheiden, G. (2000). The development of a simple, low cost set of universal access features for electronic devices. In *Proc. CUU 2000*, page 118. ACM.

[56] Lederman, S. J. (1991). Skin and touch. *Encyclopedia of human biology*, 7:51–63.

[57] Lederman, S. J. and Klatzky, R. L. (2009). Haptic perception: A tutorial. *Attention, Perception, & Psychophysics*, 71(7):1439–1459.

[58] Lee, S.-S., Kim, S., Jin, B., Choi, E., Kim, B., Jia, X., Kim, D., and Lee, K.-p. (2010). How users manipulate deformable displays as input devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1647–1656. ACM.

[59] Leonard, R. (2005). Statistics on vision impairment: A resource manual. arlene r. gordon research institute of lighthouse international, 2002.

[60] Li, F. C. Y., Dearman, D., and Truong, K. N. (2009). Virtual shelves: interactions with orientation aware devices. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, pages 125–128. ACM.

[61] Li, F. C. Y., Dearman, D., and Truong, K. N. (2010). 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*, pages 187–194. ACM.

[62] Liang, H.-N., Williams, C., Semegen, M., Stuerzlinger, W., and Irani, P. (2012). User-defined surface+ motion gestures for 3d manipulation of objects at a distance through a mobile device. In *Proceedings of the 10th asia pacific conference on Computer human interaction*, pages 299–308. ACM.

[63] Mine, M. R., Brooks Jr, F. P., and Sequin, C. H. (1997). Moving objects in space: exploiting proprioception in virtual-environment interaction. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 19–26. ACM Press/Addison-Wesley Publishing Co.

[64] Morris, M. R., Wobbrock, J. O., and Wilson, A. D. (2010). Understanding users' preferences for surface gestures. In *Proc. GI 2010*, pages 261–268. CIPS.

[65] Morrison, A., Knudsen, L., and Andersen, H. J. (2012). Urban vibrations: Sensitivities in the field with a broad demographic. In *Wearable Computers (ISWC), 2012 16th International Symposium on*, pages 76–79. IEEE.

[66] Negulescu, M., Ruiz, J., Li, Y., and Lank, E. (2012). Tap, swipe, or move: attentional demands for distracted smartphone input. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 173–180. ACM.

[67] Nomura, Y., Yagi, Y., Sugiura, T., Matsui, H., and Kato, N. (2007). A fingertip guiding manipulator for mental image creation of multi-stroke drawings. *Microsystem technologies*, 13(8-10):905–910.

[68] Norman, D. A. (1981). Categorization of action slips. *Psychological review*, 88(1):1.

[69] Oakley, I. and Park, J. (2007). A motion-based marking menu system. In *CHI'07 Extended Abstracts on Human Factors in Computing Systems*, pages 2597–2602. ACM.

[70] Obaid, M., Häring, M., Kistler, F., Bühling, R., and André, E. (2012). User-defined body gestures for navigational control of a humanoid robot. In *Social Robotics*, pages 367–377. Springer.

[71] Pakkanen, T., Lylykangas, J., Raisamo, J., Raisamo, R., Salminen, K., Rantala, J., and Surakka, V. (2008). Perception of low-amplitude haptic stimuli when biking. In *Proceedings of the 10th international conference on Multimodal interfaces*, pages 281–284. ACM.

[72] Pirhonen, A., Brewster, S., and Holguin, C. (2002). Gestural and audio metaphors as a means of control for mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 291–298. ACM.

[73] Piumsomboon, T., Clark, A., Billinghurst, M., and Cockburn, A. (2013). User-defined gestures for augmented reality. In *Human-Computer Interaction–INTERACT 2013*, pages 282–299. Springer.

[74] Poggi, I. (2002). From a typology of gestures to a procedure for gesture production. *Lecture notes in computer science*, pages 158–168.

[75] Poggi, I., Pelachaud, C., and Caldognetto, E. M. (2004). Gestural mind markers in ecas. In *Gesture-Based Communication in Human-Computer Interaction*, pages 338–349. Springer.

[76] Post, L., Zompa, I., and Chapman, C. (1994). Perception of vibrotactile stimuli during motor activity in human subjects. *Experimental brain research*, 100(1):107–120.

[77] Prescher, D., Weber, G., and Spindler, M. (2010). A tactile windowing system for blind users. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, pages 91–98. ACM.

[78] Ruiz J, L. Y. and E., L. (2011). User-defined motion gestures for mobile interaction. In *Proc. CHI 2011*, pages 197–206. ACM.

[79] Rukzio, E., Müller, M., and Hardy, R. (2009). Design, implementation and evaluation of a novel public display for pedestrian navigation: the rotating compass. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 113–122. ACM.

[80] Rushton, D., Rothwell, J., and Craggs, M. (1981). Gating of somatosensory evoked potentials during different kinds of movement in man. *Brain: a journal of neurology*, 104(3):465–491.

[81] Sanchez, J. and Aguayo, F. (2006). Mobile messenger for the blind. In *Proc. ECRIM 2006*, pages 369–385. Springer.

[82] Sánchez, J. and Aguayo, F. (2007). Mobile messenger for the blind. In *Universal Access in Ambient Intelligence Environments*, pages 369–385. Springer.

[83] Sanchez, J. and Maureira, E. (2006). Subway mobility assistance tools for blind users. In *Proc. ECRIM 2006*, pages 386–404. Springer.

[84] Sánchez, J., Sáenz, M., and Garrido, J. M. (2010). Usability of a multimodal video game to improve navigation skills for blind children. *ACM Transactions on Accessible Computing (TACCESS)*, 3(2):7.

[85] Shop, O. (2015). Royan national institute for the blind.

[86] Silpasuwanchai, C. and Ren, X. (2015). Designing concurrent full-body gestures for intense gameplay. *International Journal of Human-Computer Studies*, 80:1 – 13.

[87] Southern, C., Clawson, J., Frey, B., Abowd, G., and Romero, M. (2012). An evaluation of brailletouch: mobile touchscreen text entry for the visually impaired. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 317–326. ACM.

[88] Straughn, S. M., Gray, R., and Tan, H. Z. (2009). To go or not to go: Stimulus-response compatibility for tactile and auditory pedestrian collision warnings. *Haptics, IEEE Transactions on*, 2(2):111–117.

[89] Tan, D. S., Pausch, R., Stefanucci, J. K., and Proffitt, D. R. (2002). Kinesthetic cues aid spatial memory. In *CHI'02 extended abstracts on Human factors in computing systems*, pages 806–807. ACM.

[90] Traylor, R. and Tan, H. Z. (2002). Development of a wearable haptic display for situation awareness in altered-gravity environment: Some initial findings. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002. HAPTICS 2002. Proceedings. 10th Symposium on*, pages 159–164. IEEE.

[91] Tsetserukou, D. (2011). Flextorque, flextensor, and hapticeye: exoskeleton haptic interfaces for augmented interaction. In *Proceedings of the 2nd Augmented Human International Conference*, page 33. ACM.

[92] Tsukada, K. and Yasumura, M. (2004). Activebelt: Belt-type wearable tactile display for directional navigation. In *UbiComp 2004: Ubiquitous Computing*, pages 384–399. Springer.

[93] Turunen, M., Hakulinen, J., Melto, A., Hella, J., Rajaniemi, J.-P., Mäkinen, E., Rantala, J., Heimonen, T., Laivo, T., Soronen, H., et al. (2009). Speech-based and multimodal media center for different user groups. In *Tenth Annual Conference of the International Speech Communication Association*.

[94] Ungar, S. (2000). 13 cognitive mapping without. *Cognitive mapping: past, present, and future*, 4:221.

[95] van Erp, J. (2005). Vibrotactile spatial acuity on the torso: effects of location and timing parameters. In *Eurohaptics Conference, 2005 and Symposium on Haptic Interfaces*

for Virtual Environment and Teleoperator Systems, 2005. World Haptics 2005. First Joint, pages 80–85. IEEE.

[96] Van Erp, J. B., Saturday, I., and Jansen, C. (2006). Application of tactile displays in sports: where to, how and when to move. In *Proc. Eurohaptics*, pages 105–109.

[97] Van Erp, J. B. and Van Veen, H. (2003). A multipurpose tactile vest for astronauts in the international space station. In *Proceedings of eurohaptics*, volume 2003, pages 405–408. Dublin, Ireland: ACM, Press.

[98] Van Erp, J. B., Van Veen, H. A., Jansen, C., and Dobbins, T. (2005). Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception (TAP)*, 2(2):106–117.

[99] Vanderheiden, G. (1996). Use of audio-haptic interface techniques to allow nonvisual access to touchscreen appliances. In *Proc. HFES 40*, page 1266. Springer.

[100] Vatavu, R.-D. (2012). User-defined gestures for free-hand tv control. In *Proc. EuroiTV 2012*, pages 45–48. ACM.

[101] Warren, D. H. (1978). Perception by the blind. *Perceptual ecology*, 10:65.

[102] WebAIM (2015). Screen reader user survey #4 results.

[103] Wikipedia (2015). Natural user interfaces. https://en.wikipedia.org/wiki/Natural_user_interface/.

[104] Wobbrock, J. O., Morris, M. R., and Wilson, A. D. (2009). User-defined gestures for surface computing. In *Proc. CHI 2009*, pages 1083–1092. ACM.

[105] Yamashita, A., Kuno, S., and Kaneko, T. (2011). Assisting system of visually impaired in touch panel operation using stereo camera. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 985–988. IEEE.

[106] Ye, H., Malu, M., Oh, U., and Findlater, L. (2014). Current and future mobile and wearable device use by people with visual impairments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3123–3132. ACM.

[107] Yee, K.-P. (2003). Peephole displays: pen interaction on spatially aware handheld computers. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1–8. ACM.

[108] Yfantidis, G. and Evreinov, G. (2006). Adaptive blind interaction technique for touchscreens. *Universal Access in the Information Society*, 4(4):328–337.

[109] Yu, W., Kangas, K., and Brewster, S. (2003). Web-based haptic applications for blind people to create virtual graphs. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2003. HAPTICS 2003. Proceedings. 11th Symposium on*, pages 318–325. IEEE.

[110] Yuan, B. and Folmer, E. (2008). Blind hero: enabling guitar hero for the visually impaired. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, pages 169–176. ACM.

[111] Zajicek, M., Powell, C., and Reeves, C. (1998). A web navigation tool for the blind. In *Proceedings of the third international ACM conference on Assistive technologies*, pages 204–206. ACM.

[112] Zeng, L., Prescher, D., and Weber, G. (2012). Exploration and avoidance of surrounding obstacles for the visually impaired. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*, pages 111–118. ACM.

[113] Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., and Baudisch, P. (2007). Earpod: eyes-free menu selection using touch input and reactive audio feedback. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1395–1404. ACM.