UNIVERSITY OF CALIFORNIA,
IRVINE

Exploring the Potential of Using Spatial Audio to Improve Web Accessibility for Screen Reader Users

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Information and Computer Science

by

Tao Wang

Dissertation Committee:
Professor David Redmiles, Chair
Professor Gary M. Olson
Professor Virginia M. Richards

2018

DEDICATION

To

my parents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor and committee chair, Professor David Redmiles, for his support and guidance. It would have been truly impossible to get to this point without him. In the past three years, he always had faith in me when I had none. I could always find courage and strength talking to him. Thank you!

I would also like to thank my other committee members, Professor Gary Olson and Professor Virginia Richards. They provided not only valuable feedback on my research but also insight and inspiration. They are the role models that every graduate student dreams of having. I am very lucky to have them! Thank you!

I also owe a debt of gratitude to Professor Donald Patterson. He played a particularly important role in my graduate school career. He came into my life when everything was about to fall apart. He was the reason that I could stay focused and continue on this journey. I admire everything about him. I hate that I could never be as talented as he is but find solace in being able to call him a mentor and count on him for advice. Thank you!

I also want to thank Professor Alfred Kobsa. He helped me start this whole adventure and navigate the unfamiliar world of academia in the early years. Without him, there would be no graduate school and I would not have become the person I am today. Thank you!

A special thank-you also goes to Braille Institute Anaheim Center, whose friendly staff are like family to me. They gave me the most positive experience working there. Their knowledgeable assistive technology instructors also provided constructive feedback for my research and critical help in my user recruitment. Thank you!

Finally, I want to thank every fellow student, researcher, and friend with whom I had crossed path, shared a quick conversation, or engaged in a prolonged debate. There are too many names to list here. These interactions collectively transformed me from an uninitiated graduate student to a researcher. Thank you!

CURRICULUM VITAE

Tao Wang

2005 BSc in Computer Science, University of Auckland, New Zealand

2009 BSc(Hons) in Computer Science, University of Auckland, New Zealand

2018 Ph.D. in Information and Computer Science, University of California, Irvine

FIELD OF STUDY

Accessibility, Human Computer Interaction, Informatics

ABSTRACT OF THE DISSERTATION

Exploring the Potential of Using Spatial Audio to
Improve Web Accessibility for Screen Reader Users

By

Tao Wang

Doctor of Philosophy in Information and Computer Science

University of California, Irvine, 2018

Professor David Redmiles, Chair

The web has the potential to greatly improve the quality of life for people with visual impairments. Unfortunately, there are still many web accessibility issues that prevent this vision from being realized. In this research, I explore the potential of using spatial audio feedback to improve screen reader users' experiences browsing web pages. There are three main components to this research. They represent three key elements in problem solving: identifying causes, designing solutions, and evaluating outcomes.

First, I used a text analysis study to determine the specific communication barriers between sighted and blind web users. By analyzing written instructions produced by 48 sighted people, I identified that the use of spatial terms, among other things, was a main challenge for screen reader users. The result led to a proposal of using spatial audio cues to convey layout information to screen reader users.

Second, to learn the most effective spatial audio design patterns, I designed two lab experiments to uncover what spatial properties contribute to positive recognition of spatial audio cues. Based on more than 4000 data points collected from 18 sighted participants, I

obtained an intimate understanding of how users perceived and interacted with audio cues spatialized in the horizontal audio space.

Finally, informed by both studies, I developed a proof-of-concept screen reader prototype that provides additional spatial audio feedback to convey web page layout information. To evaluate the design, I conducted a user study with 20 blind participants. Participants completed layout-related tasks during study sessions. Based on a range of data collected, including task performances, interviews, surveys, and keystroke logs, I learned that participants responded positively to the design. They believed that spatial audio feedback enabled them to deal with spatial terms and the newly acquired layout information could help them handle unexpected accessibility incidents more effectively. Participants also provided feedback on the prototype's usability and envisioned how similar technologies could assist them in other scenarios. Inspired by the results, I reflect on implications for assistive technology designs and accessibility trainings. I also chart the research agenda in this undeveloped area and discuss promising future research topics.

Chapter 1 Introduction

The power of the Web is in its universality.
Access by everyone regardless of disability is an essential aspect.
Tim Berners-Lee, W3C Director and inventor of the World Wide Web

1.1 Overview

British Scientist Tim Berners-Lee wrote a paper [9] proposing the idea of an "information management" system on March 12, 1989 [91]. This document provides the conceptual and architectural structure for the World Wide Web (or "the web"). This date has often been credited as the birthday of the web. In less than thirty years, internet access has become an essential element of our daily lives. Based on a series of surveys conducted over the span of 15 years, Pew Research Center found that 84% of American adults used the internet in 2015 [128]. Though there were still adoption gaps based on age, class, racial group, or community, the rate of adoption increased over the surveyed period steadily and neared saturation in some groups. Consequently, the web has touched every corner of our lives, including how we search for information, how we shop, and even how we network with other people. The web is also a place to conduct everyday business, for example, a majority of U.S. adults bank online [35]. Many government agencies make their services available online. In many situations, it is easier and even cheaper to perform a task online.

Being able to access the web has become an important factor for one to participate fully in this society.

Therefore, it is particularly important to make sure that everyone has access to the web. For people with disabilities, performing various tasks online also has the added benefits of improving their independence and quality of life. By completing tasks using computers in the comfort of their own homes, they do not have to deal with many of the physical obstacles associated with traveling. However, in reality web accessibility has lagged behind the pace of technological development since the beginning, evidenced by early web accessibility evaluation work [31,58]. With new services and technologies taking shape online constantly, the progress of their accessibility seems to be playing catch up forever.

Research on web accessibility is a sub discipline within the broader Human Computer Interaction (HCI) discipline. Over the years, researchers have conducted web accessibility evaluations in various domains; they have raised concerns and proposed recommendations. Informed by often alarming results, researchers, information workers, and assistive technology vendors have also actively pursued new technologies and studied their potential for making the web more accessible to users with various disabilities. While this progress should not be overlooked, the outlook of an accessible web is still less than desired. There is still more to be done.

One area that does not receive enough attention is the interaction between people with visual impairments and sighted people. Most accessibility improvement efforts focus on inventing techniques to make content accessible to blind web users, for example, by automatically inferring and creating alternative text for inaccessible images [10]. Such research implicitly holds an assumption that a challenge exists between a visually impaired user and a computer. The goal is

to help the user understand what is on the computer screen or make the computer convey information to the user non-visually. However, no one lives in a vacuum. In the real world, completing any task often involves other people. The interaction with other people is even more critical when a person with visual impairment performs a task in the workplace.

This dissertation presents my research addressing one known interaction issue. Prior work has reported communication barriers between sighted and blind web users [87]. Blind users expressed problems communicating with sighted people, presumably because they have different mental models of the same web page due to how they access the content: whereas sighted people process web page content visually, blind users use screen readers that have an innate linear nature. Unfortunately, research has stopped short of identifying what exactly the communication barriers are and in what specific context these issues occur. If the relevant details were known, it might be possible to design technologies to mediate the process and alleviate the barrier.

In this work, I took on this accessibility issue. My research comprises a serial of three studies. The first study identified the specifics of the communication barriers based on text analysis of written instructions produced by sighted people. With a better understanding of what factors contribute to the communication barriers between sighted and blind web users, I proposed a few possible solutions. Among others, one solution is to use non-speech spatial audio to convey web page layouts. This led to the question of how such spatial interfaces should be designed. Though HCI researchers have studied audio interfaces before, synthesized spatial audio design pattern has not received much attention. To learn the most effective design pattern, I designed two lab experiments to evaluate how various spatial properties influence spatial audio recognitions. The experiments generated design insights and informed the development of a proof-of-concept screen reader prototype with stationary and moving spatial audio feedback. Using this prototype, I conducted a

user study with blind screen reader users to evaluate the design concept. Their feedback validates some design assumptions, adds more details about the context of communication barriers between sighted and blind web users, provides implications for spatial audio feedback designs, and explores the potential of using spatial audio in a broader space.

1.2 Dissertation Outline

Chapter Two provides a background for web accessibility and screen readers. Though this research focuses on a specific accessibility problem, understanding how accessible the web is in general today helps put the problem in perspective. Human beings are capable of seeking alternative solutions and reasoning based on partial information. Therefore, one might question whether or not it is possible for blind users to derive layout information based on other clues. In reality, most web pages in existence today have some kind of accessibility issues. Blind users often have to deal with inconsistent or missing information when browsing the web. Getting a sense of the current level of web accessibility helps us understand the challenges that blind people face.

Screen readers are central to how blind users access web pages. Therefore, it is necessary to provide a brief review of how screen readers function, especially features that blind users rely on to acquire a web page's structure. Three popular screen readers are reviewed as well as a summary of how blind users use screen readers.

Chapter Three presents a text analysis study that was designed to gain better understanding of the communication barriers between sighted and blind web users. Forty-eight written instructions of web-based tasks were collected from the Amazon Mechanical Turk platform. By analyzing the language used and identifying consistent patterns, the study uncovered a few issues that would cause problems for blind screen reader users and contribute to communication barriers. The issues

include inconsistent use of web element names and frequent use of spatial terms. The study also found positive language patterns, including the use of clear action verbs and quantitative-oriented written structures. This chapter ends with a discussion of possible solutions to counter the identified issues, including a proposal of using spatial audio feedback to convey layout information to blind screen reader users, which drives the other research activities featured in this dissertation.

Chapter Four answers the question of how to design effective spatial audio interfaces. Audio designs have been an area of interest in HCI since the beginning of the discipline. However, spatial audio has not received much attention. There are technologies or prototypes of research that have featured spatial audio elements in their interfaces, but the research was often on the overall usability, rather than determining the effectiveness of spatial audio designs. Chapter Four reports two experiments aiming to learn best design patterns concerning stationary and moving spatial audio cues. The two experiments produced a dataset of more than 4000 data points. Statistical analysis revealed a few interesting patterns, including spatial audio locations that are easier to localize, and how spatial properties, such as direction and length, help the recognition of moving audio cues. This chapter ends with a few recommendations on how to design spatial audio features.

Chapter Five and Six takes the research back to blind screen reader users. These two chapters describes a user study exploring what roles spatial audio feedback could play in reducing the communication barriers between sighted and blind web users. Chapter Five focuses on the planning of the user study. It describes the development of a proof-of-concept screen reader prototype with spatial audio features. It also provides details of the user study including the protocol and instruments implemented. Chapter Six reports the user study results in detail. User feedback is organized to answer three main questions: does spatial audio feedback help screen reader users interpret common spatial terms? Can screen reader users perceive a web page's overall

layout based on spatial audio feedback? How do screen reader users feel about the spatial audio feedback regarding our prototype specifically and in general? Both quantitative and qualitative data are analyzed and presented. The chapter finishes with a discussion of implications in terms of spatial audio feedback designs.

Chapter Seven draws from the studies and reflects on how spatial audio could help improve web accessibility. Based on the study results, three main contributions are concluded. They are assisting blind screen reader users to interpret spatial terms, helping users deal with and recover from unexpected accessibility incidents, and creating spatial audio-based interface models for access learning purposes. It concludes the dissertation with a discussion of limitations and future work.

1.3 A Note About Contributors

I am the principle investigator of this entire research project. However, like the context in which this research is situated, we do not live in a vacuum and it is never truly one person's work.

Ms. Bev Jensen from Braille Institute Anaheim Center has helped me code the data in Chapter Three. Professor Donald Patterson and Professor David Redmiles have provided much feedback and close guidance to the experiment study reported in Chapter Four. Professor David Redmiles has also guided me through the design and implementation of the user study presented in Chapter Five and Six. And, of course, my committee has provided critique and suggestions throughout the research. I owe a debt of gratitude to all of them. To respect their contributions, I used "we" throughout the whole dissertation.

Chapter 2 Background

This chapter provides a necessary introduction to the general topic, web accessibility. It starts with a brief background of web accessibility. After a review of the current status of web accessibility, it delves further into research that tries to understand why web accessibility is still in a poor state.

This chapter also provides a summary of screen readers. Many of the challenges that this research addresses and the design of user study prototype are related to how screen readers function. Therefore, it is necessary to learn how screen readers work in order to understand the context of the problems. Section 2.2 first introduces a few visual impairment terms. Then it provides brief summaries of the three most popular screen readers. It should be noted that the summaries are not designed to provide a full tutorial of how to use the respective screen reader. The main focus of the review is on navigational features available in each screen reader, as these features are central in a user's experience when browsing web pages.

2.1 Web Accessibility

2.1.1 What is Web Accessibility?

"Web accessibility means that people with disabilities can use the web" [55]. "Use the web" can be further interpreted as perceiving, understanding, navigating, and interacting with web pages. If a web page is developed properly, it should allow all users, regardless of their abilities, to access information and functionalities. A narrow understanding of web accessibility concerns users with

impairments in vision, hearing, motor skills, speech, cognition, etc. However, it is also important to accommodate people with changed abilities due to aging. In addition, accessible web pages also benefit people with "temporary disabilities", such as a broken arm, or stiff fingers on a cold winter day.

The web has become a portal of resources for many aspects critical to one's life, such as education, healthcare, employment, commerce, etc. Therefore, having equal access to the web is essential for providing equal opportunities to everyone. The UN Convention on the Rights of Persons with Disabilities recognizes access to information and communication technologies as a basic human right [129]. Around the world, many countries and regions have enacted laws that mandate web accessibilities, for example, Section 508 of the US Rehabilitation Act of 1973 in the United States [130], Accessibility for Ontarians with Disabilities Act in Canada [96], and Disability Discrimination Act 1992 in Australia [131].

To help the production of accessible websites, W3C, the governing body of the World Wide Web, has formed the Web Accessibility Initiative (WAI) [56] to create accessibility standards and guidelines. WAI produced guidelines for all components necessary to make the web accessible, including web page authoring, designing, and how user agents should render web pages. The most influential guideline is the Web Content Accessibility Guideline (WCAG), which has guided the development of assessment tools and is incorporated into laws around the world [109]. WCAG 1.0 was released in 1999 [132], which includes 14 guidelines describing accessibility principles. Each guideline includes one or more check points that could be used to access a website's accessibility. In 2008, WAI released WCAG 2.0 [133] that superseded WCAG 1.0. This version includes 12 guidelines under four principles. Each guideline includes testable success criteria. Notable improvements of WCAG 2.0 are that it is more technology neutral, and its success criteria are

more testable than the previous version. There are three levels of conformance to either WCAG 1.0 or 2.0: A, AA, and AAA. A is the lowest level of conformance and each incremental level of conformance requires full conformance of the previous level(s).

WCAG 2.0 breaks down the accessibility of a website to four perspectives: perceivable, operable, understandable, and robust. For a web page to be perceivable, information on the web page needs to be presentable to users who might perceive it in different ways. For example, non-text content on a web page should be annotated with text alternatives. When blind users who cannot view a web page visually visit the page, screen reading software can pick up the alternative text and communicate it to users. An operable web page should make sure users can operate and navigate among interface components regardless of how they access web pages. One example is to enable keyboard only interactions in addition to mouse-based interactions, as the use of a mouse requires sufficient vision and motor abilities. In addition, since the main goal when a user visits a web page is to navigate and find desired content, to be operable, web designers should provide ways to assist such actions, for example, by helping users determine where they are and how to move to their targets. An understandable web page should make text content easy to understand and the interactions predictable. It should also help users prevent or recover from errors. Finally, being robust means a web page could be used via a wide range of user agents and assistive technologies. The key step to achieve this goal is to follow markup syntax correctly, including annotating elements with proper names, roles, etc.

It should be noted that conforming to web accessibility guidelines does not guarantee usable websites for users with disabilities. Power et al. conducted a study [102] where 32 blind users evaluated 16 websites manually. In total, blind users identified 1,383 instances of user problems. However, only 50.4% of the problems uncovered by blind users were covered by WCAG 2.0

Success Criteria. Among those covered by WCAG 2.0, 16.7% of websites implemented the recommended solutions, but the problems were not solved completely. The researchers argued that even though some of the problems could be classified as usability issues rather than accessibility issues, they nonetheless caused accessibility hurdles and should be addressed responsibly.

To determine a website's accessibility, there are a few commonly used evaluation methods, including automated tools, expert inspection, user study, etc. Automated tools, such as WAVE [134], AChecker [43], and Hera-FFX [40], can check accessibility issues that could be determined programmatically, for example, missing alt tags or links without text. Automated tools have the advantage of being fast and inexpensive, and requiring no accessibility expertise from operators to conduct. However, automated tools cannot assess some issues that can only be evaluated manually [17], for example, whether or not the alternative text for an image is meaningful (e.g., not simply "image 1"). There are also two common methods to conduct manual evaluation. One method is expert inspection. An expert with expertise in accessibility [16] can review a website and identify accessibility issues. Another manual method is usability testing. Auditors recruit subjects from targeted user populations, i.e., users with disabilities, and ask them to complete a series of tasks. These two methods have their own strengths and weaknesses. Expert inspection tends to focus more on technical accessibility and leave usability issues undiscovered. In contrast, usability testing might not ensure compliance with the law or guidelines since the subject is unlikely to touch all components required for conformance [69]. Compared to automated tests, properly conducted manual accessibility evaluations require many resources and much time. To achieve the best accessibility evaluation results, it is recommended to use a hybrid of automated and manual tests [17], use multiple human inspectors [70], and conduct multi-stage human inspections involving screen readers [78]. Additionally, prior research has also shown that a

website's homepage often represents the best accessibility of the entire site [89]. Therefore, if time and resources are of concern, it is acceptable to conduct accessibility evaluation with the homepage only.

2.1.2 How Accessible is the Web?

Over the years, researchers have conducted studies to assess the accessibility of websites in various domains. The following research, mostly done in the past ten years after WCAG 2.0 was released, provides a close look at how accessible the web is for people with disabilities.

One important space online is government agency websites. These websites often host critical civic services. Therefore, their accessibilities can have a direct impact on people's lives. In addition, in many countries, government websites are required by law to conform to certain accessibility standards. Making sure their websites are accessible is more than just a voluntary desire. However, even driven by such legal and practical obligations, the accessibility of government websites is still less than ideal. Goodwin et al. evaluated national government portals and ministry websites from 191 (i.e., all but North Korean) United Nation member states [48]. For each website, they applied 23 tests, derived from WCAG 1.0, and calculated a score in percentage based on the fraction of the number of detected barriers over the number of applied tests. The best score from their calculations was 1.28% (for Germany based on five websites) and the worst is 66.75% (for United Republic of Tanzania based on six websites). Only 12 countries scored 10% or less.

There are more accessibility evaluations for specific countries or regions. When exploring the connection between website accessibility and website's accessibility statements, Olalere and Lazar conducted accessibility evaluation of over 100 US federal government websites against Section 508 of the Rehabilitation Act, which has incorporated most WCAG 1.0 and 2.0 items [95]. They

used both automated tools and human evaluator. Their human evaluation, which captured additional accessibility issues that are hard for automated tools to assess, found that only four of the 100 website homepages were free of accessibility violations. Among the violations, the most frequent one was the lack of alternative text for images and other non-text elements, and the lack of content bypassing features (i.e., skip navigation link).

Lazar et al. evaluated 25 Maryland government homepages in 2012 [74]. Multiple evaluators manually evaluated the web pages against the Maryland State Guidelines for accessibility, which are similar to Section 508 of the Rehabilitation Act. They found that most of the web pages evaluated (23 out of 25) had one or more violations of the Maryland IT Non-Visual Access Guidelines. Among the violations, the most frequent one was the lack of alternative text for graphical content. The corpus of web pages evaluated included 15 homepages that had been evaluated using the same method in 2009. By comparing results from both studies, the researchers found that there was a slight improvement in accessibility (decrease from 2.5 paragraphs of the law violated in 2009 to 2.1 paragraphs of the law violated in 2012). They believed that the improvement was due to the introduction of a state-wide government homepage template.

Hanson and Richards studied the accessibility of over 100 high-traffic and government websites over 14 years (1992 to 2012), totaling 952 web pages from high-traffic websites and 231 web pages from government websites, from the US and the UK [52]. Their assessment was based on six unambiguous WCAG 2.0 Level A Success Criteria (that's the lowest level of conformance) that can be tested using software reliably (note: in some cases, the Success Criteria were applied to web pages predating WCAG 2.0). They found violations across categories from all websites. The violations include some straightforward, easy to implement items. For example, providing alternative text for images is well-understood and it has high priority in both versions of WCAG.

However, the researchers still found that 42% of high-traffic websites and 24% of government websites violated this measure. They also found some positive trends. For example, the accessibility of all websites increased over the years. And in comparison, government websites had higher accessibility compliance than non-government websites. They also found evidence that the improved accessibility was probably due to adoption of new web development technologies.

Kuzma evaluated websites of 130 randomly selected Members of Parliament in the UK [68]. The researcher used an online accessibility evaluation tool and analyzed the websites' compliance to both WCAG 1.0 and 2.0. They found that none of the websites were free of accessibility issues. 30 websites met the conformance level required by law under WCAG 1.0 and 7 websites met the conformance level required by law under WCAG 2.0.

Similar accessibility evaluations of government websites have also been conducted in other countries, such as Saudi Arabia [2], India [84], China [104], and Bangladesh [6].

Another important sector online is education. Kane et al. evaluated the accessibility of 100 top international universities [61]. They applied both automated tests and manual inspections to each university's homepage against WCAG 1.0 checkpoints. They found on average 4.68 errors per page. Two out of 100 web pages were free of any errors; 36 had no priority 1 errors. Ringlaben et al. evaluated special education department websites from 51 American universities [107]. They found that 97% of the pages evaluated had accessibility problems, including 39% severe errors. In recent years, Massive Open Online Courses (MOOCs) have played an increasingly important role in learning. Al-Mouh et al. evaluated ten courses on a popular MOOC platform, Coursera.org, against 105 WCAG 2.0 testable success criteria [3]. They identified a minimum of 23 and a

maximum of 34 open accessibility issues in every course. No course has achieved full conformity in any priority level.

Libraries are an important source for information during learning. Lazar et al. evaluated 24 public library websites in Maryland against Section 508 guidelines using expert inspections [69]. They found that all websites had violated at least two or more paragraphs. Two websites had the highest number of violations, six paragraphs. They reported that the most common violations were providing alternative text for graphical elements, labelling form fields properly, and providing a method to skip navigation links. Oud evaluated 64 Ontario university, college, and public library websites against WCAG 2.0 A and AA priority [97]. Different from other similar studies, their corpus included the homepage and up to 30 web pages randomly selected from each website. The total web pages evaluated were 1,860. Using an automated test, they found on average 14.75 accessibility problems per web page. If markup and contrast errors, which are not always included in other similar studies, were excluded, the average error rate was 5.68 per web page. They concluded that there was much work for Ontario library websites to do to reach the accessibility level required by local law. Conway studied 29 public library websites in West Australia [26]. The researcher used both automated tools and expert inspections to evaluate various accessibility conformance. Conway found that none of the websites had met the lowest conformance level using either WCAG 1.0 or WCAG 2.0. The most common errors included incorrect HTML markup and lack of alternative text for non-textual elements. Conway et al. also studied nine Australian national and State/Territory libraries [27]. They found that none of the websites met the WCAG 2.0 Level A compliance.

Though accessibility regulations often are only applicable to government agencies and public institutions, private businesses can also benefit from providing accessible websites to their

customers. However, a study conducted by Goncalves et al. [47] painted a bleak outlook for this sector. They evaluated the accessibility of the top 250 largest enterprises of the year 2009. The 250 enterprises were selected according to Forbes listing of "The Global 2000". Due to conflicts between some websites and the automated tool they used, eventually their evaluation included only 236 enterprises. For each website, they applied automated tests against WCAG 1.0, WCAG 2.0, and Section 508 requirements. They found that 86% of the websites had more than 500 WCAG 2.0 level A errors, 58% of the websites had between zero and 30 WCAG 2.0 level AA errors, and 63% of the websites had between 60 and 300 level AAA errors. They concluded that the accessibility for top enterprises was far from desired.

Andres et al. conducted a study with websites of 108 international (USA, France, Germany, and Spain) large listed non-financial firms [4] to understand the factors that influence the implementation of accessible websites. They used Web Accessibility Barrier [126], which produces a score based on WCAG 1.0 checkpoints. They found that six websites had no accessibility barrier, whereas 11 websites had severe accessibility issues. All other websites had various degrees of accessibility violations. Martinez et al. conducted a similar study with 49 leading European banks [79]. They used automated tools to evaluate each website based on three different metrics. They also applied manual evaluation to ensure validity and reliability. They found that 13 out of 49 websites had acceptable accessibility levels. However, none of them were free of all accessibility issue. 18 banks had troublesome scores that indicated the existence of many accessibility issues.

In summary, web accessibility around the world has improved since the inception of the web. However, the current accessibility is still less than desired. This is the same for government

agencies and private enterprises. In many cases, the accessibility issues are so severe that people with disabilities are not likely to be able to use the hosted services independently.

2.1.3 Why is Web Accessibility Not Better?

Given the fact that many identified issues were easy to fix, one can't help but wondering why web accessibility is in such a dire state. The literature suggests a network of factors involving multiple actors.

One major factor is the lack of awareness and training among professionals who are responsible for creating websites. Freire et al. conducted a web-based survey with people involved or with experience in web development [39]. They received 613 valid replies from all states in Brazil. They found that less than 25% of their responders knew how blind users browse the web and knew how they could build web pages to accommodate these users. Almost 70% of responders knew little or nothing about WCAG. Less than half of them said they had taken any accessibility development training. However, the accessibility evaluation methods mentioned suggested that, even for those who had performed some accessibility related testing, they were mainly doing web standard conformance and little manual inspection or user testing. For people who have considered accessibility in their work, they listed personal motivation, trying to reach more users, and web standards as main motivations. When asked about ways to improve web accessibility, most of their respondents thought increasing awareness was important. Lazar et al. surveyed webmasters for their perceptions of web accessibility [72]. They received 175 responses from all over the world and it seemed that most webmasters were familiar with the concept of web accessibility and were aware of the legal framework around it. They knew the availability of online accessibility evaluation tools and had used them on some occasions. However, only a minority of them had

used non-web based tools or used screen readers to check their websites. There was also a false belief about the conflict between aesthetics and accessibility. Some webmasters believed that by making a website accessible, it would become less visually appealing. They hesitated to enforce accessibility as it might interfere with designers' work. Putnam et al. studied how User Experience and HCI professionals consider accessibility in their work [103]. They collected 173 survey responses from professionals around the world. By analyzing their answers to open-ended questions, the researchers found that 83% of respondents considered accessibility important for their work. However, the perceived scope and their actions were limited. They also found that empathy and professional experience were correlated with how accessibility considerations were reported. It should also be noted that accessibility expertise is also important even when automated tools are available, as lack of expertise could lead to misuse of the tools, incorrect interpretation of issues, and ineffective evaluation performance [121].

There are some indications that the demand for web accessibility is on the rise. In a study of 100 web development company websites in the UK, Gilbertson and Machin found that 46 websites in their dataset had listed accessibility as their skills and 23 websites had accessibility statements on their websites [45]. The accessibility level of the website, unfortunately, did not correlate with the existence of these features, i.e., companies with or without the mentioning of accessibility in their websites were equally likely to have accessibility issues. Nonetheless, increased market demand could provide motivations for investment in education and training.

Another factor contributing to poor web accessibility is a lack of shared understanding and coordination from multiple people. In Lazar's study [72], respondents had different views on who should be responsible for creating accessible websites. The top three roles mentioned were webmasters, system analysts, and programmers. Some also believed that upper management

played the most important role as they could mandate implementation. This point was also reflected in Putnam's study [103]. Their participants said that the decision of implementing accessible websites was often out of their control. They needed people with higher positions in their organizations or clients to justify investing time and effort in accessibility. Among the participants in Freire's survey [39] who were not involved in projects where accessibility was considered, two of the top three reasons were that it was not a requirement either from the organization or from the clients.

This leads to another factor, the prioritization of accessibility. Respondents of Putnam's study [103] discussed having to sacrifice accessibility for time, budget, and the needs of their organizations and clients. Every project is bound by its budget and resources. An accessible website takes extra time to develop and evaluate. When many requirements are competing for resources, it would often come down to priorities associated with the features. A feature's value is one important consideration for prioritization. If stakeholders who have deciding power over resource distribution do not appreciate the importance of web accessibility, it is unlikely for them to feel the urge to invest in web accessibility. In Lazar's survey [72], some of their respondents said web accessibility was often perceived as trivial and non-important, therefore they considered conveying the value of web accessibility to shareholders the biggest challenges of making a website accessible. Feature prioritization is also driven by user needs. The participants in Lazar's survey [72] mentioned that knowing a big portion of the users they serve have disabilities would influence their accessibility implementation plans. Since there is no easy and reliable way to detect if an assistive technology is used to browse web pages, the proportion of a website's users with disabilities is driven by assumption. In Freire's study [39], they also found that among people who did not consider accessibility in their project, lack of requirements from organizations or users was

the main reason. Finally, government actions can also influence the prioritization. Participants in Lazar's survey [72] proposed a few more specific examples, such as mandating accessibility by creating policies or laws, or creating financial incentives by giving tax breaks.

Finally, inadequate tools also contribute to poor web accessibility. In a study of the web accessibility of 50 websites, Lazar et al. found that the web accessibility testing tools used by some web development companies were flawed and inconsistent [71]. In addition, many tests required large number of manual checks, which made it hard for many developers to do. Lopes et al. surveyed more than 400 people whose roles were relevant to creating accessible technologies and they learned developers had little time to spare for learning accessibility skills. Therefore, providing tools integrated into their development environment was desired [76]. Trewin et al. surveyed 49 IBM web developers to understand the challenges of developing accessible technologies [116]. They learned that design, testing, and finding technology workarounds were the most difficult aspects of implementing accessible websites. To reduce time and effort, developers used tools to help them assess, understand, and revise accessibility issues. However, many tools did not reliably report errors. When asked what features developers consider valuable in accessibility evaluation tools, the top three features were itemizing problems detected, providing explanations for problems, and being able to pinpoint problems in DOM object or in source code. These features are important because not only do they help developers identify potential accessibility issues effectively but also allow developers to understand general accessibility problems and increase their expertises during the process.

It should be noted that the legal framework did help corporations to take accessibility seriously. In 2005, the National Federation for the Blind (NFB) reported a number of web accessibility issues on the website target.com to the Target Corporation. When Target did not make any

accommodations that would help screen reader users access their site, the NFB filed a lawsuit claiming that Target was violating the Americans with Disabilities Act (ADA) and California legislature governing accessibility in businesses. After Target's motion to dismiss was refused, they settled with the NFB and paid $6 million dollars to members of the class action. [57].

2.2 Visual Impairments

Visual impairments are common conditions around the world. The International Classification of Diseases defines four levels of visual functions: normal, moderate visual impairment, severe visual impairment, and blindness [135]. The term "low vision" refers to both moderate visual impairment and severe visual impairment. In other word, "low vison" and "blind" cover all visual impairments. According to a 2014 Word Health Organization (WHO) report, an estimate of 285 million people worldwide has some kind of visual impairments. Among them, 246 million are low version and 39 million are blind [136]. It should be noted that blind can be further classified as either totally blind or legally blind. Total blindness refers to having no sight with either eye. Legal blindness refers to the condition where the better eye with the best possible correction has a central visual acuity of 20/200 or less, or a visual field of 20 degrees or less. Though people diagnosed as legally blind often still has some usable vision, this term is defined by law to determine eligibility for social programs [137]. The same WHO report also reports that 82% of people living with blindness are aged 50 or above [136]. This implies a further challenge as one's hearing and cognitive abilities decrease after the age of 50. It becomes harder for this population to adopt new technologies.

2.3 Screen Readers

Many software products have been developed to assist people with visual impairments to use computers. They can be classified into two groups based on the population they serve. The first

group of software caters to users with low vision. Since people with low vision have impaired but functional vision, such software programs are still visual in nature. They provide features such as screen magnification, color replacement, etc. Though they also provide audio output, the information communicated in audio is limited to textual content. Notable products in this market are ZoomText, Supernova, and Magic [138]. Though these products function similarly to screen readers in some scenarios, they are more commonly referred to as screen magnifiers. They are not the main focus in this research.

The other group of software, which is the concern of this research, caters to blind users. These products do not assume that their users have any vision. Screen readers convey an interface to the user entirely over synthesized speech, including both textual content and semantic information. For example, when reading a header on a webpage that is coded as "<h3>Sports</h3>", a screen reader would read it as "header level 3, sports". When encountering visual elements, such as a video or an image, screen readers would read the alternative text if provided. Screen readers also provide many fast navigation features that assist users to navigate an interface. Notable products in this market are JAWS, VoiceOver, and NVDA [138].

In addition to these two main groups, there is a new emerging group of software that targets users who are currently low vision but are expected to lose sight gradually and eventually become blind. The software assists this transition by providing both screen magnifier features and screen reader features. This allows their users to stay with the same product, instead of learning different systems when their vision deteriorates. One such product is ZoomText Fusion.

In the rest of this section, I will provide a review of common screen reader navigation features. In 2015, WebAIM, a web accessibility advocacy group, conducted a survey of screen reader users

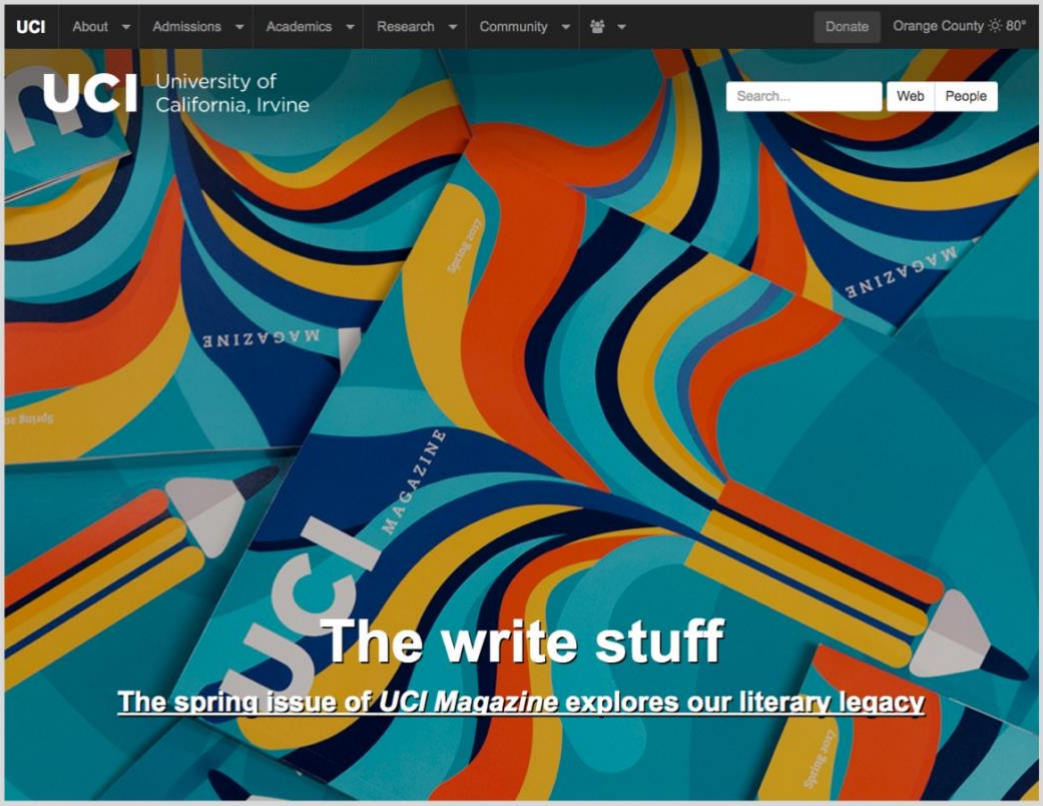
[139]. They received 2,515 valid responses from all over the world. Based on the data, JAWS was the most popular screen reader (30.2%). The second most popular screen reader was NVDA with 14.6% market share. VoiceOver was the third most used with 7.6% market share. In 2016, GOV.UK also conducted a survey of its website visitors on their assistive technology use. They found similar results: JAWS (38.5%), VoiceOver (21.2%), and NVDA (12%). Therefore, the following review will cover only these three screen readers.

2.3.1 JAWS

JAWS, short for "Job Access With Speech", was developed by Freedom Scientific [140]. Its first version compatible with Microsoft Windows was released in 1995. The latest version is version 18, released in late 2016.

JAWS provides a rich set of options that can be customized to suit a user's needs. For example, users can configure JAWS' synthesizer to use different speech rates (experienced screen reader users often set the speech to a faster rate to save time. Compared to the regular reading speed of about 300 words per minute, screen reader users often set it to up to 500 words per minute [13]), change verbosity levels, set punctuation pronunciation, etc. Users can also apply different speech rates, speakers, pitches, or volumes for different types of information encountered in a document or web page. For example, they can use a different speaker for French words when reading a web page primarily written in English or use a high-pitched sound to announce an editable text box instead of reading the word "edit".

After a webpage is loaded, JAWS first announces some overview information about the web page, such as its title, total numbers of regions, headings, and links. Then JAWS automatically executes the "Say All" command, which reads the rest of the page continuously. If not interrupted, this



Home | UCI - Mozilla Firefox
Home | UCI
Skip to navigation landmark, link, Skip to main content
Primary navigation, navigation landmark, link, UCI homepage
list with 6 items
link, About, menu button, collapsed, submenu, Toggle dropdown: About
link, Admissions, menu button, collapsed, submenu, Toggle dropdown: Admissions
link, Academics, menu button, collapsed, submenu, Toggle dropdown: Academics
link, Research, menu button, collapsed, submenu, Toggle dropdown: Research
link, Community, menu button, collapsed, submenu, Toggle dropdown: Community
menu button, collapsed, submenu, Toggle dropdown: Find information for...
out of list
link, Donate to UCI
Orange County 80∞
banner landmark, visited link, graphic, University of California, Irvine
clickable, Search, edit, has auto complete
button, Web
button, People
The write stuff
link, the spring issue of UCI Magazine explores our literary legacy

Figure 2.1 A Screen Reader Output Example

process will read all page elements one by one following the tab order. If web authors do not mark specific tab orders, the default order moves from left to right and from top to bottom. Users can use shortcut keys to rewind, fast forward, replay the current element, or pause the screen reader during the "Say All" process.

JAWS provides many fast navigation features. One commonly used feature is to navigate by a specific kind of web element, such as headings, links, etc. When using this feature, the user presses the designated key to move to the next web element of interest. In most cases, the user can press the SHIFT key at the same time to move in reverse order, i.e., to the previous web element of interest. Users can use this feature to navigate most types of web elements, including radio button, button, combo box, edit box, form control, graphic, heading, list item, list, frame, paragraph, table, check box, etc. Users can also navigate among web elements with logical criteria. For example, users can move to the next element that is of a different type than the current one, non-link text, unvisited/visited link, or clickable element. If a web page uses ARIA landmarks [32] or HTML5 sectioning elements (e.g., <main>, <nav>, <header>) to identify the role of regions, users can also navigate based on them. At any point, the user can stop and change to a different navigation method, such as using "Say All" to listen to content listed under a desired heading.

Another way in JAWS to navigate among a particular kind of web elements is to bring up a list of such elements in a pop-up window. The shortcut keys are often the INSERT key and one of the function keys, for example, INSERT+F6 for a list of all headings, INSERT+F7 for a list of all links. The user can then use arrow keys to move up or down on the list. Once the desired content is located, she can press the ENTER key to move to it on the web page.

There are two features that assist JAWS users to browse frequently visited web pages more easily. The first is PlaceMarkers, which are bookmarks for specific locations on a web page. When learning a web page for the first time, a user often spends a long time to figure out its components and functionalities. During this process, she can insert PlaceMarkers to locations that carry useful information or functions. Next time she needs to use a function or seek certain information, she could move to the PlaceMarkers directly or navigate among a list of PlaceMarkers. An alternative, but based on a similar concept, is to navigate by line numbers. If a web page does not change much and a user memorizes the line number of the content of interest, she can jump to that particular line with shortcut keys directly.

Though screen reader users primarily interact with the browser and screen reader via keyboard, in some situations the information on a web page is only accessible with a mouse, or in some situations mouse cursor events are necessary to interact with a web page. To accommodate these scenarios, JAWS also provides a Cursor mode that allows users to use shortcut keys to move the mouse cursor around on the interface, simulate mouse left click or right click, or perform more complex actions such as drag and drop.

2.3.2 NVDA

NVDA, short for Non-Visual Desktop Access, is developed by Australia-based NV Access [141]. NVDA provides many features similar to JAWS. However, while JAWS costs thousands of dollars to acquire, NVDA is completely free. The development of NVDA was initiated by two blind developers in 2006. They have released the software as an open-source project, which has allowed developers around the world to contribute to its development.

NVDA targets Microsoft Windows operating system. It includes generic features on using software and browsing the web. It also includes features specifically designed for some popular software, such as Microsoft Office, Skype, etc. Like JAWS, NVDA also offers many options that could be customized by users to suit their preferences. For example, users can choose different synthesizers, use different rate, pitch, or volume for a different kind of information. One feature specific to its default synthesizer, Espeak NG, is to configure the inflection for the speech generated.

NVDA shares many similar functions with JAWS. One minor difference is that it does not announce overview information about the web page once loaded. It automatically starts reading from the top of the web page. It also follows the tab orders. If tab orders are not explicitly provided, the default order moves from left to right and from top to bottom. During this process, users can use shortcut keys to fast forward, rewind, pause, or stop the speech.

NVDA supports both navigating using one-letter quick key navigation and navigating through a list of a particular kind of element. One-letter quick keys include all main web elements, such as headings, landmarks, forms, tables, buttons, lists, etc. When adding the SHIFT key, the navigation moves in the reversed order, i.e., moves to the previous element of interest. In NVDA, to bring up an element list, users do not use key combinations specific to the type of the web element. Instead, users can use one shortcut key combination to bring the element list interface. The interface includes multiple filter options. Users can navigate forward or backward among them by pressing the TAB or SHIFT and TAB keys. Users can choose one of three types of web elements: links, headings, and landmarks. Users can navigate to the Type selection area and use arrow keys to select which web element list is needed. Users can also enter keywords to filter the list. Upon

typing, the list is updated immediately to include only the ones that contain the characters entered. To go to a desired location, users need to select the target in the list and press ENTER.

NVDA provides a unique option to learn the mouse's location on the screen based on audio cues. When this option is enabled, the user would hear tones when she moves the mouse. The tone uses horizontal panning to indicate the mouse's horizontal locations, i.e., the tone is closer to the left ear if the mouse is closer to the screen's left edge. Vertical location is communicated using pitch, i.e., higher pitched tone corresponds to a higher location on the screen. This feature requires wearing a headset or stereo speakers to work. It is useful if the user has some vision and can use the mouse to navigate a web page (NVDA has a feature that follows the mouse cursor and reads the content around the cursor). In the situation where the user is lost on the web page, this tone could help her find the mouse again.

2.3.3 VoiceOver

VoiceOver is a screen reader built into Apple's macOS [142]. It is available for all Apple products, including laptop, desktop, iPad, and iPhone. However, the interactions are quite different depending on the platform. VoiceOver on handheld devices is mainly based on gestures, whereas VoiceOver on laptop or desktop can be used via keyboard or gestures performed on the trackpad. Since VoiceOver is built into Apple's operation system, it does not require the user to purchase it separately. This is a great advantage over JAWS as it can cost more to acquire JAWS than purchase a new computer.

Using VoiceOver Utility, users can configure many aspects of how VoiceOver functions. For audio output, it supports setting speech voice, rate, pitch, volume, and intonation. VoiceOver has a positional audio feature that is on by default. This feature uses audio cues to communicate an

item's location on the screen. However, the documentation does not provide details on what audio properties are used for this purpose. It seems that only an item's horizontal position is conveyed via stereo panning.

VoiceOver offers two modes of navigation on a web page: DOM or group. Under DOM mode, VoiceOver traverses web pages in an order similar to other screen readers. Users navigate by going to the next or previous item. By default, it moves from left to right and from top to bottom. Under group mode, related items on a web page are grouped together and users can navigate in any direction.

When a web page is loaded, VoiceOver only announces the title of the web page, but users can use a shortcut key to play the statistics of the web page. Next, it proceeds to read the first item on the web page and stop. Users can use shortcut keys to request reading the whole web page or navigate actively. Similar to other screen readers, VoiceOver offers navigation shortcut keys that allow users to easily move among a particular kind of web elements, such as links, headings, form controls, tables, lists, etc. Alternatively, users can use Rotor, which is similar to element lists from other screen readers. Rotor includes different kinds of list, such as headings, links, landmarks, etc. Once brought up, users can use Left and Right Arrow keys to change to different lists. Within a list, users use Up and Down Arrows to locate an item and use ENTER to move to it on the web page. VoiceOver also has a Quick Nav mode where users can perform navigations using just arrow keys. For example, a user can press Right Arrow and Up Arrow keys together to bring up the Nav Rotor and choose heading, then use Up Arrow or Down Arrow to move among headings.

VoiceOver uses "web spots" to identify useful locations on a web page. When loading a page, VoiceOver automatically evaluates the web page's visual design and identifies "auto web spots".

User can use Rotor or shortcut keys to move among them. In addition, users can also add their own web spots to bookmark content area. This is similar to the PlaceMarker feature provided by JAWS. Users can also designate one web spot as the "sweet spot", which would appear as the first item in the Rotor web spot list.

2.3.4 Screen Reader User Navigation Strategies

Borodin et al. provided a detailed overview of common screen reader browsing strategies [13]. In general, many screen reader users adopt fast navigation strategies when looking for information on web pages. One common strategy is navigating by headings. Users move from heading to heading. When they find a heading relevant to the information they look for, they will switch to listening details. If none of the heading sounds relevant, they will try other strategies. Exhaustive sequential navigation, i.e., starting from the beginning of the page and listening everything, is the least efficient method. So, it is often reserved as a last resort. Screen reader users also use exhaustive sequential navigation when they want to make sure not missing any information, such as when learning a new web page.

Screen reader users adopt different fast navigation strategies based on the web page. For example, navigating by heading is more appropriate for reading news websites where major news is often presented as headings. However, if a web page features primarily a form, it is more efficient for screen reader users to navigate by editable fields. If a web page serves as an index page, screen reader users can go through visited or unvisited link lists to find the link of interest.

When browsing a frequently used web page, memory and landmark play important roles. Screen reader users will memorize the positions of important web elements, for example, they might remember the third editable field is for user name. When they visit again, they can quickly locate

it by navigating through editable fields. The memorized web elements can also serve as landmarks, i.e., reference point to build other instructions. For example, on a shopping website, the Buy button might be right after the produce price. So, by using the Buy button as a landmark, screen reader users can quickly locate the price information. If the screen reader supports virtual landmarks, users can utilize this feature and save them from memorizing landmarks. When they visit the page again, they can simply bring up the list of virtual landmarks and jump to a desired landmark directly.

Another important strategy is to search for keywords. If the content of interest includes specific words, a quick way to find it is to do a search on the web page. For example, to find contact information, users can search for "contact", or "phone", or "email". The search feature is universal in almost all interfaces. Therefore, it can be very useful once a user has mastered the process.

2.4 Summary

In this chapter, we reviewed the current status of web accessibility and three popular screen readers available in the market. We also present a summary of screen reader user navigation strategies. These information sets the background to much of the design consideration for the work presented in later chapters.

Chapter 3 Understanding the Communication Barriers Between Sighted and Blind Web Users

In this chapter, we present our work that furthers the understanding of one particular web accessibility issue: the communication barriers between sighted and blind web users. Other researchers have reported this issue previously [87]. However, little is known about what the specific issues are. A necessary first step to resolve the communication barriers and improve the interaction between sighted and blind users is to gain a better understanding of where the problems lay. We achieved this goal by conducting a text analysis study based on 48 written web-based task instructions collected from sighted web users who had no accessibility training. We discovered consistent content patterns that would confuse screen reader users and lead to communication difficulties. With a better understanding of the issue, we also discussed design opportunities that could aid the production of accessible web-based task instructions, including features that are the focus of the research reported in later chapters.

3.1 Introduction

No man is an island. In mobility and orientation training, one important learning goal for people with visual impairments is to seek help safely and effectively [59]. This is useful as sighted people perceive the physical environment differently from blind people; consequently, when asked for help by blind people, sighted people do not necessarily know what information is helpful. Similarly, when encountering problems on the web, blind users also consider asking for help as a coping

strategy [119]. This extends to online help. Brady and Bigham studied how six popular companies used Twitter to engage customers around accessibility issues [15]. They found that users with disabilities used Twitter to interact with these companies' accessibility teams. Among responses sent by the companies, 29% were simple instructions on how to complete a task or use a feature.

However, there are no established guidelines or lessons for training blind web users how to solicit effective information concerning web-based activities. To make matter worse, prior work has found that blind users had communication barriers with sighted users. Murphy and Kuber conducted an empirical investigation to understand the difficulties experienced by web users with visual impairments [87]. They interviewed 30 participants with visual impairments. The interviews focused on five areas, including the participant's experience of navigating on the web, perception of various content, strategies of acquiring the web page overview, and unmet requirements. Their participants shared that it was easier to describe a website to another visually impaired person as they could use terms that they were familiar with. Presumably, it is because sighted users and blind users have different mental models resulting from accessing web pages in very different manners. Sighted users process web pages visually. The web page's layout and semantic information embedded within are perfectly preserved. In contrast, blind users browse web pages using screen readers, which read one piece of information at a time via audio. This linear process gives rise to mental models that best resemble one-column lists [1]. To train blind users how to seek help on web-based activities more effectively, we must first understand the communication barriers and develop mediating strategies.

Unfortunately, research on accessibility stops short of identifying the specific issues that may have contributed to the communication barriers between sighted and blind web users. Compared to other remaining accessibility challenges, effective communication with sighted people seems to be a

small problem, especially with more and more synchronous technologies becoming available where a blind user can simply ask for clarifications in real time. However, asynchronous communications are still necessary in many scenarios. For example, to reduce cost and improve efficiency, many websites require their users to search through documentation of common problems first and only put users in contact with customer service staff as a last resort. Even support hotlines or online chats are not necessarily synchronous communications anymore. Robot operators and chat bots have been deployed to act as intelligent agents to search existing documentation based on customer inquiries.

Ensuring effective communication between sighted and blind web users is also critical for the success of human-powered access technologies. Borodin et. al argued that though traditionally website owners had the responsibility of developing accessible websites, this objective was hard to realize due to lack of incentives. Given the volume of user-contributed content, it is also not realistic for site owners to guarantee proper accessible attributes for all content [14]. One promising solution is to develop human-powered access technologies that rely on social media users or crowd-sourced work forces [11] to complete tasks or provide services. Researchers have tested such ideas in some promising work. Takagi et al. developed a social accessibility system that allows external volunteers to contribute metadata authoring [112]. Burton et al. studied using remote volunteers to provide fashion advice to blind people via a mobile phone app [22]. Guy and Truong used the Amazon Mechanical Turk crowdsourcing platform to capture accessibilities of intersections in rural and suburban areas based on Google Street View images [51]. Hara et al. employed similar methods to identify more detailed street level accessibility problems [53].

So far, successful human-powered access technologies often comprise tasks that are highly structured or that require only simple answers. Human-powered access systems can play a greater

role if they can be used to answer more generic and free-formed questions from screen reader users. For example, many essential services provided by governments or businesses are expected to be performed online nowadays. Being able to successfully interact with these websites has become a critical requirement for one to fully participate in society. Given how ubiquitous accessibility problems are on websites, a system that accepts questions from screen reader users on how to carry out various activities on websites and provides accessible instructions could be very useful.

However, workers or volunteers behind human-powered technologies do not often go through accessibility or any other training systematically. If the communication barriers between sighted and blind web users is left unresolved, the aforementioned service would be rendered useless when a blind user receives an instruction that she cannot interpret effectively. Therefore, it is critical to provide accessibility guidelines that can be easily followed by untrained workers. In addition, if the communication barriers can be clearly defined, service providers may also implement automated workflows to guide untrained workers in following best practices and avoiding common problems.

In this study, we are motivated to investigate the communication barriers and complete this knowledge gap in accessibility. The communication barriers are rooted in the words that a sighted user uses. We hypothesize that when a sighted user and a blind user collaborate on a web page, the sighted user may reference information, such as visual attributes, which are not available to blind users due to differences between the perceived mental models. To evaluate this hypothesis, we collected 48 instructions for three web-based transactional tasks via Amazon Mechanical Turk. The workers were informed that the instruction would be used by blind screen reader users, but the workers received only basic accessibility tips. This configuration allowed us to gain a realistic picture of the information quality from untrained but willing helpers. We analyzed the written

instructions solicited and assessed their accessibilities for blind screen reader users, i.e., what particular information will screen reader users find problematic? What content patterns account for these problems? The rest of this chapter presents details of the study, including its design, results, and design implications.

3.2 Related Work

Research studying barriers that people with visual impairments experience often focuses on understanding what poses challenges for them to participate activities in physical settings. For example, O'Day interviewed 20 unemployed people with visual impairments and found both personal and social factors contributed to employment barriers [92]. Naraine and Lindsay studied social interaction in the workplace of 13 blind or low vision employees [88]. They found that visually impaired employees had problems integrating fully in the workplace: social and interpersonally communication challenges posed challenges on seeking and retaining employments; lack of non-verbal clues also contributed to difficulties of interacting with other employees, which led to the view of them unable to fit into the workplace. Researchers have tried to develop technological solutions to reduce these barriers. For example, Krishna et al. proposed a wearable social interaction assistant device to use advanced recognition technologies and provide users with visual impairments social clues in real time [67].

There is no existing work exploring the language patterns between sighted and blind users. However, there are a large group of work studying how visual impairments contribute to language development and use. Most of such studies focus on children. For example, Perez-Pereira and Castro studied a twin, one is blind and the other is sighted, when they were aged 2.5 to 3.5 [98]. They found that the blind child used languages that refer to her own actions with more routines,

calls, and repetitions. In contrast, the sighted child's language made more connections to external reality and social interchanges. Conti-Ramsden and Perez-Pereira studied the conversational interactions of three mother-infant pairs in which one infant is sighted, one is low vision, and one is blind [25]. They found that the mothers of sighted infant and low-vision infant had similar conversation patterns, whereas the mother of the blind infant talked more, used more directives, and the directives included more descriptions and occurred in clusters. Flowers and Wang conducted a study in which 41 sighted and 17 blind children listened six short music excerpts and described them orally [34]. They found no accuracy difference between two groups. Analysis of language usages shows that sighted children used more metaphors and emotional descriptors than blind children.

There are also some but fewer studies focusing on blind adults. One example is an experimental study of social interaction in blind people [64]. The study involves three conditions: interactions between two blind people, between two sighted people, and between a blind and a sighted people. Each condition has ten pairs. Researchers found that the blind pairs behaved very similarly to other pairs. The data also verifies the model of cuelessness, which predicts that the fewer the social cues, the greater the psychological distance; and in turn the content is more task oriented and depersonalized, the style is less spontaneous and the outcome is less favorable.

Researchers have also studied whether or not visual impairments influence how one perceives music. Park and Chong conducted a comparative study where 120 participants, 60 with visual impairments and 60 without, reported their emotion responses and the extent after listening 16 short music excerpts. They found that the two groups showed no difference in their music emotion identification abilities. However, the two groups show difference in their emotion responses, most

notably in “happiness” for the group with sighted participants and “sadness” for the group with visually impaired participants.

3.3 Study Design

3.3.1 Transactional Task Design

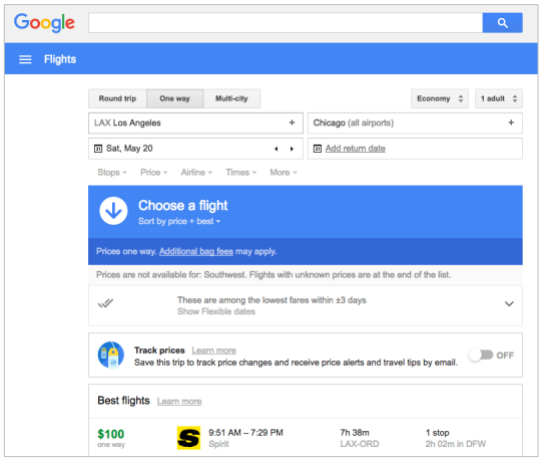
There are different types of web-based activities. In this study, we focused on transactional tasks. Broder developed a taxonomy of web search activities based on survey and query logs [19]. The taxonomy classifies web query activities into three types: navigational, informational, and transactional. A user performs a navigational query to reach a particular site that she has been to before or she assumes to exist. To find information, a user conducts informational queries. Here, the information is defined as in static form. Once found, the next action is simply reading the information. Transactional queries refer to reaching a site where further interaction will happen, such as shopping. Kellar et al. conducted a field study to identify various web-based information seeking tasks’ characteristics based on the log data of 21 participants over a week’s period [63]. They found that 46.7% of web-based tasks were transactional tasks, the most common type recorded in their data. Terai et al. evaluated differences between informational and transactional tasks on the web [114]. They found that transactional tasks were more complex than informational tasks. Users tended to visit more pages when completing a transactional task and the time spent on each page was shorter. In the context of web accessibility, ineffective instructions can be more damaging when transactional tasks are concerned. Confused blind users may spend more time on each page. Since transactional tasks involve more pages, users would spend much longer time over all and this could lead to a frustrating experience.

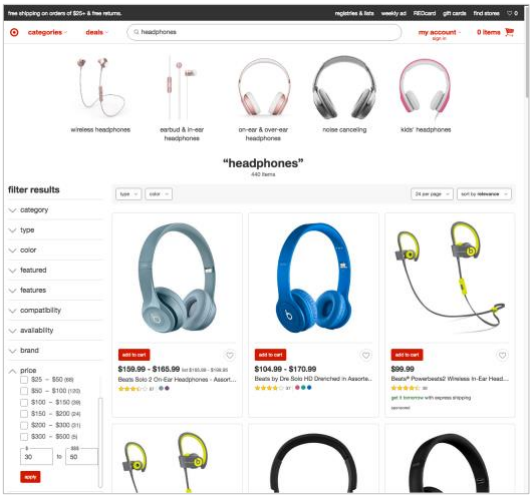
We designed three multi-step transactional tasks to be used for data collection in this study. A possible alternative approach is to mine existing written instructions from Q&A websites or social media websites. However, given the purpose of the study, we believe that analyzing existing instructions on the web would not be adequate. Our plan is to analyze written instructions for common patterns. Therefore, the instructions should be produced independently from multiple authors, but online Q&A websites often employ techniques, such as moderating or rating, to actively discourage repeated information. Responses submitted later often build on earlier ones by providing additional information. Such responses will not be suitable for this study.

The three tasks designed represent three common activities that people perform online (Figure 3.1):

- Booking a flight: searching for a one-way flight from LA to Chicago for a particular date on Google Flights
- Online shopping: searching for headsets between a customized price range of $30 and $50 on Target.com
- Finding how-to information: searching and loading the print view for a recipe on allrecipes.com

We performed a few checks to ensure the tasks had reasonable complexities. When piloting the tasks ourselves (i.e., sighted users), each task took between three to five steps to complete. We also consulted three people with visual impairments one-on-one for their input. Among them, one is legally blind and the other two are totally blind. They are all regular screen reader users and have one, 13, and 16 years of experience respectively, using JAWS, a popular screen reader software. None of them has used these particular web sites before. During each meeting, we first provided time for them to get familiar with the websites. We also described to them the web page's





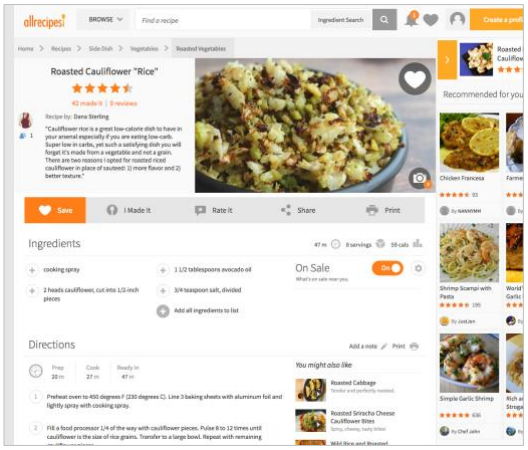


Figure 3.1 Web Pages Used in Transactional Tasks

interface design and layout. This is similar to how a screen reader user would learn how to use a new web page in daily life [13]. We answered any questions they raised until they were confident about using the web pages. At the end, they all agreed that the web tasks entailed reasonable complexities.

We also checked with these experienced screen reader users whether or not these specific webpages featured anything unusual that may have an impact on how they use screen readers when browsing the pages. We asked them to describe how to complete the tasks once they were familiar with the web pages. After they gave us the instructions for a task, we made sure to probe whether they could come up with other alternative procedures and which procedures were most effective. In the end, their preferred navigation paths were consistent. Their instructions mainly involved well-documented screen reader navigation techniques, such as navigating and interacting with edit boxes, i.e., editable text boxes, navigating by links to locate relevant sections, and finding desired areas by searching landmark text on web pages.

3.3.2 Data Collection

We created three projects on Amazon Mechanical Turk. Each project was based on one task. The projects included the same instructions and prompt. Only the task description itself was different. In the instructions, we briefly described the purpose of this task: help assist a blind user to complete a task online. We emphasized that the task would be done on a desktop or laptop computer, not on tablets or phones. We provided some, but very limited, background information about how blind users browse the web because we wanted the workers to be aware of the limitation of screen readers without tipping off how they should form their language. Though providing a bit more information could help workers produce better instructions, not doing so allowed us to determine

a quality base line based on the crowd workers' intuition alone. We also deliberately did not provide any clue of the hypothetical blind user's gender in case such knowledge would influence the crowd worker's response. The prompt used was as follows:

> *The blind user has no prior experience with this website and needs some simple information on how to do it and what to expect. Blind users use screen reading software to listen to webpages (i.e., no visual clues). If you don't know how the screen reader works, that's OK. Just be specific on your instructions and provide whatever you think would be useful. If the webpage has many similar pieces of information, such as in search results or product listings, please describe the key formatting patterns/elements of a result/listing so that the blind user knows where to find the information that could help locate the best choice.*

In each project, we asked the workers to provide written instructions using 100 to 300 words, and whether they had any experience or knowledge of how to browse the web with screen readers. In case of regional language differences having any impact on the language of instructions, we also asked where the workers grew up (if a native English speaker) or where they learned their English (if not a native English speaker).

We made sure the responses received were unique by applying a customized qualification to workers who had completed one of the projects. In subsequent projects, they were excluded from the potential worker pool.

3.4 Results

We solicited a total of 52 responses from Amazon Mechanical Turk. However, some of the responses did not provide instructions of how to complete the task online. For example, one gave information on how to call Target customer service. After removing responses that did not contain stepwise instructions, we ended up with 48 valid responses: 14 for the flight searching task, 16 for the headphone searching task, and 18 for the recipe searching task. Among them, seven responders reported that they had knowledge or experience with screen readers. However, the responses of five of them did not demonstrate sufficient knowledge of screen readers, as there were many visual descriptions in their instructions. In contrast, the other two responders clearly understood how screen reader users interact with web pages based on their instructions, i.e., featuring language such as tabbing, searching keywords, etc.

The responses were classified into information categories introduced by Ummelen et. al [117]. They defined two main categories for information in technical documentation: procedural and declarative. Procedural information refers to specific actions that must be executed to get a product working, whereas declarative information refers to all explanatory information other than action information, such as descriptions of the internal working of a system or its features. Research has concluded that procedural information plays a vital role in guiding users to perform tasks [80]. The role of declarative information is less conclusive. However, it is shown that users do spend time reading declarative information even when having the option to skip it [62], and declarative information can contribute to better performance [110]. It could be that declarative information helps users construct useful mental representations of systems [65], which influences how users interact with the system [90].

Two researchers, both with screen reader experience, independently classified the responses. During the coding process, researchers compared coding and discussed any disputes. In the end, all coding conflicts were resolved. Next, the text from each category of information, declarative or procedural, was analyzed to identify consistent usage patterns and their implications in terms of web accessibility. This phase of coding was open-ended and iterative. Axial coding was used to recognize structures. Although the dataset was small, some strong consistent patterns emerged at the end of this process.

3.4.1 Content Patterns in Procedural Information

There are two content patterns that we have observed from procedural information. Procedural information describes actions required to achieve a goal. Its grammatical composition is straightforward: often verb and object. Occasionally, there is also additional information clarifying the action or object.

Actions are Consistent

In the instructions produced, our responders used 29 unique verbs when describing what actions to take. These words were used 265 times in total. Among the 29 verbs, seven words were used more than 10 times: click (85 times), type (35 times), select (25 times), enter (21 times), press (16 times), choose (12 times), and scroll (12 times). Though screen reader users do not directly click or select with the mouse, they can understand easily most of the actions and know how to achieve the same effect. The only slightly troublesome verb is “scroll.” “Scroll” was used when the content occupied a large area, such as a vertical list, or a grid. Responders instructed users to “scroll up,” “scroll down,” or “scroll through” until reaching the desired landmark or information. Blind users do not perform “scroll up” or “scroll down” with screen readers. However, these terms imply that

the desired landmark or information is before or after the current position, which is information useful to screen reader users to take actions, e.g., navigating backward or forward. If such intentions are communicated more clearly, these terms would not create communication barriers.

Non-standard or Incorrect Web Element Names

While our responders used consistent verbs to describe actions, they were very inconsistent when describing the web elements that would receive the actions.

One main problem was using incorrect terms. For example, both the shopping task and the recipe finding task entailed the use of search fields. The search fields featured in both websites were quite standardized, i.e., a long text box on the top of the web page with a "search" button to its right. Most responders referred to them as the "search bar." Only a few used the terms "search field" or "search box." This might cause some confusion for screen reader users because screen readers describe web elements by their standard names, such as "combo box," "radio button," etc. If a blind user learns about web pages with a screen reader, she would be accustomed to these standard terms.

Though "search bar" is not too far a stretch from "search box," sometimes responders used completely incorrect terms. For example, in one instruction, when referring to an expandable menu on Target's website, the responder referred to the top-level menu as "tab." As "menu" and "tab" suggest very different interactions, this could cause real problems for screen reader users. Other examples include referring to a radio button as "circle," a link as "font," etc.

There were also some instances where the instruction did not refer to the object explicitly. For example, without describing the search field, one instruction simply stated "click search." Another instruction that described providing customized price ranges as a filter option read "input 30 into

min price and 50 into max price." Since objects implied in these instructions are unambiguous for sighted people, the responders might have written them in this manner without realizing that such descriptions would not provide enough clues for screen reader users to locate the web elements quickly.

3.4.2 Content Patterns in Declarative Information

In our dataset, declarative information mostly describes the web page interface relevant to actions or the outcome of an action. There is also information embedded within procedural instructions that clarifies the objects concerned. We analyze them together here.

Uses of Spatial Terms

Thirty-five responders used some sort of spatial terms in their instructions. Overall, we found 117 cases. They can be further classified into two groups: relative spatial terms and absolute spatial terms.

The most common relative spatial concept was below ("below," "slightly below," "under," "underneath," "beneath"), which occurred 27 times. This term was relative in nature since it was used with a referencing object to describe a target object, for example, "below this header, there is a link…," or "the boxes directly below the first rectangle box." Though screen reader users would not be able to make sense of the term directly, due to the sequential manner of how screen readers read information, it could inform the user that the target object is somewhere following the referencing object. However, the distance between these two objects is less certain. Unless specifically ordered, web page elements often have default tab orders from top to bottom and from left to right. Therefore, even if the target object appears directly below the referencing object, a screen reader might have to read all other elements to the referencing object's right before reaching

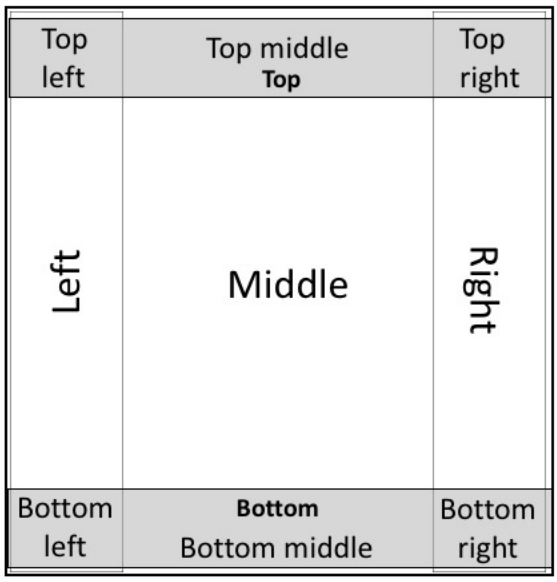


Figure 3.2 Spatial Positions Commonly Used in Instructions

the target object.

Another frequently used relative spatial concept was to the left or to the right of a referencing object. Such terms (e.g., “left side of,” “to the right”) appeared 20 times in total. Compared to the concept of “below,” this term can be better utilized by screen reader users. If web elements follow the default sequential order, a target object to the left of a referencing object could be reached by navigating backward, whereas a target object to the right of a referencing object could be reached by navigating forward.

There were in total 68 cases using absolute spatial terms. Positions in the horizontal dimension were mentioned 42 times, whereas positions in the vertical dimensions were mentioned 45 times (some terms that implied positions in both dimensions, such as “center” or “top left,” were counted for both). These terms shed light on how the responders divided web pages into regions. Figure

3.2 shows a sketch of the main regions identified. In summary, each dimension was divided into three sections, though each section did not necessarily have equal size.

There were a few cases where “far left” and “far right” were used. However, “far left” was only used once and it was referring to the filtering option column on the Target website, which was described as left by most others. “Far right” was used only for describing the position of the “print” button in the recipe searching task (Figure 3.1, bottom). On that web page, there were five buttons in a row located below the summary of the recipe, and “print” was the very last one. Three different responders used “far right” to describe its position. It could have been the existing five-button layout that prompted the responders to differentiate the positions more finely. However, perhaps due to the small dataset, we did not see the same term used in other pages where five items were arranged horizontally.

Limited Use of Visual Clues

Though responders used many spatial terms that were inherently visual, they did not use many other kinds of visual properties to describe objects. Throughout the dataset, we only found nine cases of visual descriptions from seven responders. The visual clues used concerned color, shape, and image content.

This might be attributed to the task prompt, which reminded responders that blind users listened to web page content via screen readers and there were no visual clues. However, since we did not have a dataset from a controlled condition that used the same tasks but without such warnings, we cannot be sure whether the prompt really made a difference. What makes this observation relevant is that even with the prompt, our responders still used many spatial terms. The contrast to the limited use of other visual clues suggests that spatial terms are hard to avoid.

Effective Structure Description

In general, most instructions would be hard to follow by screen reader users. We somewhat expected to see this. Therefore, we were particularly excited to see a few very effective instructions. None of these responders indicated having any experience with screen readers. This was reflected in their instructions as none had mentioned the use of the "tab" key or fast navigation strategies. However, one common quality of their instructions was that they all provided specific quantitative-oriented information, which helped communicate the structure of the relevant interface. For example, when describing the filter options from the Target website, one responder wrote:

> *There are several sub-headings that read "category," "type," "color" and so on. One of these sub-headings is "price" and it is the ninth sub-heading down.*

Another responder described the search result page on the Target website as:

> *The results are displayed in rows of three. There are eight rows on the first page.*

Such information is effective because it would give screen reader users clear information of what to expect. Since there were only a few such cases, we believed they were just by chance. However, it does show that experience with screen readers is not required for sighted users to produce effective instructions.

3.5 Discussion

The primary goal of this study is to identify specific communication barriers between sighted and blind web users. Our analysis is based on 48 written instructions collected from crowd-workers. The small dataset size does not allow us to uncover all communication barriers. However, we believe that we have identified two major accessibility issues that are critical to communication between sighted and blind web users.

First, sighted web users tend to use spatial terms in their descriptions. These terms are challenging or impossible for blind screen reader users to follow. It should be noted that here we refer to spatial terms specifically rather than visual terms more generally. Though we have seen other visual-oriented clues in the collected written descriptions, spatial terms seem be somewhat different. It is straightforward that blind users cannot comprehend visual information, at least not easily. We have reminded our responders about this fact. It is interesting that our responders used some, but not too many, visual terms concerning colors or shapes. In contrast, spatial terms, either referring to relative positions or absolute positions, were widely used in their instructions. This leads us to conclude that spatial terms entail some special linguistic attributes that make it hard for sighted users to avoid using them. It can be that sighted users perceive the world visually and there is no good alternative dimension to describe spatial concepts.

Spatial terms are problematic for blind users since screen readers do not convey spatial concepts that imply physical positions, such as a particular spot on a web page. However, screen readers do convey logical or sequential orders. Therefore, it is possible for screen readers to replace terms implying a higher position with “previous,” or terms implying a lower position with “next.” Other concepts that make sense to screen reader users include “the top of the page” (interpreted as the

beginning of the web page) or “the bottom of the page” (interpreted as the end of the web page). However, screen reader users would have a difficult time to comprehend any other spatial concepts. When one of the problematic spatial terms is used in a task instruction or a web page description, blind screen reader users are likely to experience communication barriers and will not be able to follow the instruction properly.

Second, sighted web users tend to use informal terms when referring to web elements. This is not a surprise since web users do not need to learn HTML before using the web. When referring to a web element, they take clues from the web element’s appearance. If it looks like a button and it is clickable, calling it a button is logical and others would understand it perfectly, unless they are blind screen reader users. Screen readers process web pages programmatically. Screen readers do not normally consider a web element’s appearance. If a link has a rectangular background and appears like a button, screen readers will still announce it as a link unless a different role is configured using HTML attributes. Consequently, if a blind user learns how to use the web with screen readers, she would learn to use the correct web element name since that is what screen readers provide. Using mismatched web element names can cause communication berries between sighted and blind web users. For example, a sighted user might instruct the blind user to click on a button, but the blind user can only find links on the web page. And vice versa, when a screen reader user refers to a link, a sighted user might have problems finding the link if she cannot find any web element that matches the description.

3.6 Implication for Design

With a better understanding of what communication barriers exist between sighted and blind web users, we can also propose some potential solutions to mediate the communication.

3.6.1 Conveying Spatial Terms

The first communication barrier identified concerns the use of spatial terms. The root of the issue is that screen readers do not provide layout information to blind users. The obvious solution is for screen readers to convey layout information to the user somehow. One possibility is that screen readers can simply add additional layout information to the end of the synthesized speech output. When browsers render a web page, they first create a Document Object Model (DOM) object based on the HTML document. Each web element is assigned a pair of numeric attributes indicating its horizontal and vertical position on the rendered web page. Screen readers can use this information in the speech output. However, a challenge associated with this approach is that users might have problems perceiving a position based on numeric values or comparing between web elements. A slight variation is for screen readers to first calculate the approximate region where the web element is located, such as "top left," "middle right," and announce the region in the speech output. This will make it easier for users to comprehend the information, but users will not be able to distinguish relative positions of two web elements positioned in the same region. A third variation is to use non-speech spatial audio to convey positions, similar to how humans can recognize spatial information based on sounds in real life. This requires advanced audio spatialization technologies. Usability studies are also needed to evaluate how screen reader users can process the clues encoded in non-speech audio cues.

Another approach is to utilize non-audio means to communicate spatial information. HCI researchers have explored using various modalities to provide feedback in interfaces [38]. For example, Crossan and Brewster have used a pen-like device with strong force-feedback to teach visually impaired students to recognize shapes [28]. Van Erp et al. used a belt-like haptic device to provide distance and orientation information to pedestrians [30]. Oh and Findlater have also

studied how users might feel about using part of the body as interface [93]. Based on a study with 12 visually impaired participants, they found that the least preferred locations for on-body interaction were face/neck and forearm, whereas locations on hands were preferred. Researchers can explore similar designs where a blind user can perceive the current position on the web page via haptic feedback. An advantage of this approach over an audio-based approach is that blind user's audio channel is already crammed with a lot of information and the use of another channel can bring relief. Using other senses in interface is still new in HCI. More researches are needed in order to provide spatial information to blind users via non-audio channels.

3.6.2 Describing Web Elements with Accessible Language

The second communication barrier is caused by sighted users using incorrect or inconsistent web element names. Reminding sighted people about this barrier can avoid much of the problem. This works better for professional settings, such as workplaces, where short trainings can be easily organized. For informal scenarios, such as crowd workers or a human-power access service's volunteers, it can be troublesome to organize trainings and ensure their quality. In this case, authoring tools can play a mediating role. For example, a text editor interface can provide dropdown lists of verbs appropriate for screen reader users and standard web element names. A sighted user can use a shortcut key to invoke the list when she needs to describe an action and pick the most suitable verb. The list can even be customized according to a specific screen reader's features so that blind users would receive the most appropriate instructions for their choices of screen readers.

Another useful design is to help sighted users check if the wording regarding a web element is accessible to blind users. For example, a browser extension can respond to a shortcut key and

generate an accessible description about the current web element. The sighted user can simply copy the wording into the instruction. This interface can also help incorporate specific, quantitative information into written instructions, similar to the good instructions that we have seen in our dataset. Quantitative information is easier for screen reader users to execute. However, it can be difficult or bizarre for sighted users to produce. But a script can achieve the task much more quickly. For example, the interface can count the web elements on the web page and describe the current web element as "the 17th button on this page" or "five links after the edit box search."

3.7 Limitation and Future Work

A main limitation of this study is the small dataset. The findings were based on the analysis of 48 written instructions collected from 48 individuals. We believe that this dataset has captured the two most prominent communication barriers. However, there might be other less common issues, which require a larger dataset to uncover.

This study focused on sighted people. We have found main language patterns that sighted people use that cause communication barriers without considering any specific context. The next step of this research should shift the focus to blind users. Researchers should investigate when communication barriers are most likely to occur and whether or not there are contextualized usage or interaction patterns. Answers to these questions would serve as the basis for developing technologies that can address the issues effectively.

3.8 Summary

In this chapter, we presented a text analysis study of instructions produced by sighted web users for blind web users. We reported findings regarding the communication barriers between these

users. In summary, we found that sighted users used consistent verbs when describing actions to be taken, but they often used inconsistent or incorrect terms when referring to objects of the actions. We also found that, though sighted users did not use many descriptive terms based on color, shape, or image content, they used many spatial terms in their instructions. This suggests that spatial terms are essential in describing webpage interfaces and may be hard to avoid. We believe that these two issues contribute to two main communication barriers between sighted and blind web users. Informed by the findings, we discussed technological solutions that could alleviate the barriers and mediate the process of generating written instructions accessible to blind screen reader users.

Among the technological solutions, using spatialized non-speech audio cues to convey spatial concepts is the most interesting one, as it is situated in a design domain that is less studied and the technology is readily available. This motivated us to implement a spatial audio interface and evaluate its potential in improving web accessibility for screen reader users. The following chapters will report the continuing work.

Chapter 4 Evaluating Web Audio API

In Chapter Three, we identified that the lack of layout information from screen readers' output had contributed to communication barriers between sighted and blind web users. The chapter ends with design implications including using non-speech spatial audio to convey layout information to users. This approach has much potential because it is simple to implement and audio synthesis technologies are readily available. However, there are no established design patterns for spatial audio interfaces. Some basic questions essential to good design remain to be answered.

In this chapter, we present two experiments that produce design insights. The first experiment is aimed at evaluate the effectiveness of using Web Audio API to synthesize spatial audio. The second experiment pursues a more general question of how dynamic moving audio feedback should be designed, i.e., what spatial properties make a dynamic moving audio more recognizable.

4.1 Introduction

In the past two decades, web pages have evolved from simple content pages to dynamic web applications. However, they are still primarily visual design products. Other than multimedia content or streaming services, audio features are rare on web pages. This is partially due to limited audio support by the web platform. Before HTML5, web pages relied on Flash or other plugins to load and play audio files. Web designers could control only some basic playback features. HTML5 introduced the new element <audio>. This new element allows browsers to provide native audio playback support [143]. Users do not have to deal with installing and updating external plugins to

enjoy audio content. But a more exciting design prospect was unlocked when W3C launched the process of creating a new specification, Web Audio API [144]. This high-level JavaScript library provides sophisticated features including modular routing, flexible handling of audio channels, etc. Although the specification is still a work in progress, its main features are already supported by all major desktop browsers. Google has also chosen Web Audio API as the supporting infrastructure to develop its Omnitone project, which aims to transform web browsers to full-fledged VR platforms [145].

Web Audio API is the ideal technology to implement designs for this study. First and foremost, this research proposes audio designs to communicate web page information. Since Web Audio API is supported by browsers natively, designs based on Web Audio API can be deployed to web pages seamlessly. In addition, Web Audio API, being part of a standard browser, is completely free. Though Virtual Reality or Virtual Augmented Reality systems are likely to produce better spatial sound quality, building the designs using Web Audio API ensures that adoption is affordable. Users do not need to purchase specially made hardware, such as headsets or goggles, or upgrade their computing environments to support intensive processing. A second advantage of using Web Audio API is that there is no requirement of any other third-party technologies. This reduces the complexity of adoption if released in the real world as users do not need to worry about the installation and maintenance of any external software. Finally, the web is an influential platform that could reach more users than any other technologies. Though the focus of this research is to improve web accessibilities, lessons learned here about audio designs could be applied to other design domains. Conducting this research using Web Audio API enables this possibility and increases the potential of this research's impact.

This study is most concerned with Web Audio API's spatialization features. These features, which are common in modern 3D games, allow placing audio sources in a 3D space and moving one or more of them in real time. Web Audio API supports two spatialization algorithms. The first is a simple *equalpower* panning. When configured to use this algorithm, an audio's elevation value is ignored. The resulting sound is placed somewhere on a line between the user's left ear and right ear, i.e., the user would feel the sound lying inside her head. Another rendering option is *HRTF*. HRTF is a function that describes how a sound at a specific external location will reach one's ears. By using HRTF, users could perceive the spatial location of a sound in a 3D space, i.e., the user would feel the sound outside her head. However, Web Audio API does not provide a way to use HRTF created for an individual. As shown by prior work [120], non-individualized HRTF often leads to less accurate spatialization. In addition to the sound's location, Web Audio API also supports direction configuration. A sound can be configured to be omnidirectional or directional. Directional sound is achieved by providing parameters for sound cones. When the sound cones are set up to point in a specific direction, the audio listener located outside the reach of the cone will hear reduced or no sound depending on the configurations.

However, evaluating Web Audio API is necessary before any design work takes place. As a new technology, Web Audio API has not received much attention. Its capacity has not been fully evaluated. There are two specific areas critical to spatial audio interface design.

First, how well can a user localize spatial sound produced using Web Audio API? Previously, researchers have conducted experiments to evaluate other similar technologies. In general, they have only found mediocre performance. Though their conclusions might be specific to the technologies used in their studies, their experiences serve as a cautious reminder that a careful evaluation is needed to fathom what Web Audio API can deliver. It is critical to have a realistic

expectation for the Web Audio API, as designing based on an expectation beyond the technology's capacity would only lead to failure.

Second, in addition to placing a sound in a stationary location, can more dynamic audio, such as moving audios, be used in interface effectively? If so, what properties can impact the effectiveness? Prior work has looked at the role of different sounds and specific sound qualities (such as pitch) in interface designs, but there is no work that has evaluated the various spatial properties a sound entails, such as its moving direction, movement pace, or distance. To explore a full range of possibilities and find the most effective design patterns, an evaluation of this specific topic is necessary.

In the rest of the chapter, we present our work answering these two questions. We first provide a background of sound localization and a review of related work featuring audio interface designs. Then, we describe the study design, data collection, and findings. This chapter ends with a discussion of design implications in the context of this research as well as in a broader design space.

4.2 Background of Sound Localization

Psychoacoustics is the study of sound perception. For decades, the human auditory system's ability to localize sound sources has been its subject. This section provides a brief overview of key psychoacoustic terms and findings. A more comprehensive introduction can be found is the book by Moore [86].

There are two focuses in the study of auditory system localization abilities. One is the recognition of a sound source's direction. The other is the human auditory system's resolution. The ability to

recognize direction is often explained using Duplex Theory, which is based on two cues [105]: interauraul time difference (ITD), and interaurual intensity difference (IID). ITD refers to the slight delay when receiving a sound between the left and the right ear, whereas IID is the difference between the intensity of signals that the left and right ear receive. When the intensity is described in decibels, IID can be referred to as interaurual level difference (ILD).

ITD and ILD work more effectively at different frequencies due to the physical nature of sounds. ITD works the best with low-frequency sounds. When a sound is located straight ahead (i.e., with azimuth 0°), ITD is minimized since there are equal distances from the sound to either ear. ITD is maximized when the sound is located directly opposite one ear (i.e., with azimuth 90° or -90°). The time difference can be calculated based on the path difference [33]. The approximate range is between 0 and 690 μs. Since an ITD is equivalent to a sinusoidal tone's phase difference between two ears (interaural phase difference, or IPD), for high-frequency tones with small periods, the auditory system cannot determine how the cycle one ear receives corresponds to the cycle received by the other ear. This ambiguity starts to occur when the frequency reaches about 725 Hz and becomes more confusing at 1500 Hz and over [86].

In contrast, ILD works better at localizing high-frequency tones. The wavelength of low-frequency sound is long compare to the size of the human head. They can diffract, or bend, around the head. Therefore, ILDs are negligible below 500 Hz. However, high-frequency sounds have shorter wavelengths compared to the size of the human head. With little diffraction happening, the head casts a "shadow" for the far ear. ILD can lead to as large as 20 dB at high frequency [33]. One exception is when the sound source is very close to the head, e.g., less than 1 meter. In that case, ILD is recognizable for low-frequency sound as well [21].

Duplex Theory works well for pure tones. However, most sounds we encounter in real life are complex sounds that have onsets and offsets. They contain a range of frequencies. Localizing such sounds relies on the interpretation of multiple clues including those generated by the environment. To understand how a particular factor affects sound localization, typical studies in this area evaluate the discriminable ITD or ILD thresholds under various scenarios in acoustic labs.

While ITD and ILD are effective for localization on the horizontal plane, they are not sufficient when vertical direction is involved [83]. Moving one's head can generate changes that help in such situations [54]. One's physical shape also helps. When we hear a sound in the real word, we receive not only the sound directly from the source but also sounds reflected off our torsos, heads, and pinnae (the external part of the ear) [7]. The human auditory system can pick up such subtle differences. The head-related transfer function (HRTF) captures the complex pattern of how sounds from different directions vary systematically for a particular person. It records ratios of the spectrum of a sound source and the spectrum of the sound reaching the eardrum [86]. Using HREF, sound delivered via headphone can be spatialized as if existing in the external environment, in contrast to lying somewhere on an imaginary line between two ears inside one's head. However, since different people have different head and ear shapes, HRTF based on one individual might not be as effective for a different individual [120].

Another aspect of localization is the perception of distance. As with azimuth and elevation, the human auditory system relies on multiple clues to recognize the distance from a sound source [125]. One important clue is the sound's intensity because it decreases as the distance increases [115]. In addition, differences between sound directly perceived and sound reflected off the environment also provide useful clues [125].

Audio resolution is another focus in human auditory localization research. The resolution of the auditory system is assessed by measuring the smallest detectable changes when shifting a sound source's position. For stimuli presented via loudspeakers, the minimum audible angle (MAA) is measured. In the horizontal plane (i.e., 0° elevation), azimuth of 0°, i.e. sounds from directly ahead, is the best reference point for low-frequency sinusoid sounds. A shift of merely 1° can be detected. MAA worsens when the reference azimuth moves away from 0° [82]. Sound frequency also influences MAA. The worst performance occurs around 1500-1800 Hz where ITD's effect wanes and ILD's effect is still not effective. In a natural setting, ITD and ILD happen at the same time. So, to study the effects of ITD or ILD alone, researchers have to conduct experiments with headphones where a subject hears stimuli with only ITD or ILD. These studies have produced findings that support the observed MAAs from experiments conducted in natural acoustic environments [123,124].

While MAA is relevant to stationary sound sources, minimum audible movement angle (MAMA) is a similar measure that assesses the changes necessary for the human auditory system to distinguish a moving sound source from a stationary source. In general, slower movements are easier to detect than faster movements. For example, MAMA of 8.3° has been measured for a movement at 90°/s, and it increases to 21.2° for a movement at 360°/s [99].

These findings allow modern software to synthesize spatial audio. However, users need to wear headsets to perceive the effect. Depending on the production method, the spatial location of the sound can be perceived as internalized or externalized. Internalized sounds are associated with the term lateralization where the perceived location of the sound lies on an imaginary line between two ears. In contrast, externalized sounds are perceived as out there in space. While stereo panning can produce internalized spatial audio (it also works if a user has dual speakers), HRTF is the

primary method to produce externalized spatial sound. Since non-individualized HRTF is ineffective to people who are not the one being modeled and software often relies on generic HRTFs, it is often challenging for users to localize spatial audio cues produced using software.

4.3 Related Work

Audio has long been studied in HCI as an interface design element. Most notables are auditory icons and *earcons*. Auditory icons exploit the natural associations between objects or events and certain sounds [42], whereas *earcons* rely on learned association between interface and abstract tones [12]. This study is different as we aim to understand the effect of spatial properties of an audio in interface design.

Spatial audio has been featured in many HCI research. For example, audio data is mapped to spatial positions in Dynamic Soundscape [66] and Nomadic Radio [108] to enable fast navigation and management. Other applications include positioning attendee audio in a teleconferencing system [24], communicating interface to drivers [111], or providing non-visual feedback when operating MP3 players [100]. To fully utilize the audio channel, many assistive systems designed for blind users have employed spatial audio. Zhao et al. presented geodata to blind users on virtual maps [127]. Geronazzo et al. improved non-visual exploration of virtual map by using spatial audio to convey anchors [44]. Ohuchi et al. used spatial audio to help blind users construct cognitive maps of physical environment [94]. Plimmer et al. provided spatial audio feedback to assist blind users learning cursive handwriting [101]. Lee et al. experimented five designs including spatialized audio tones to guide blind people reaching objects [75]. In these researches, the evaluations were focused on the overall interfaces' usability. The spatial audio design itself is not the subject of

investigation. Therefore, they did not help understand what design patterns make an interface more effective.

There are a small number of studies that do produce spatial audio design insights. Among them, the most relevant work is experiments reported by Goose et al. for their design of an audio browser [49]. They found that users could only accurately recognize stationary and moving sound on the horizontal axis. Their system projected synthesized speech moving along a semi-circle initially, but they found that the recognition at the extreme ends was reduced (the exact range was not reported). In addition, their users had difficulties recognizing an audio's movement when it moved too slowly. The sound localization evaluation conducted by Lorho et al. is another closely related work [77]. They tested five locations, i.e., left, mid-left, center, mid-right, and right, and they found an average 18.8% error rates. They observed that center and right locations have significantly lower error rates, and error rates are asymmetry. They also noticed that most differences between targeted positions and responses were no more than 2, which suggests better localization if only three locations, i.e., left, center, and right, were used. Vazquez-Alvarez et al. also evaluated the localization of five locations evenly distributed on a semi-circle surrounding the user [118]. Their participants were instructed to match an auditory pointer to stimulus. They found that participants could accurately recognize these locations.

A few other studies also reported findings that are not related to sound localization but contribute to audio interface design. Brewster et al. evaluated a spatial audio radial pie menu and found that an egocentric audio design where four menu items were spatialized around the user with 90° apart is more effective than other proposed designs [18]. Begault et al. conducted experiments to determine the optimal azimuth positions to broadcast important call signs during NASA shuttle launches [8]. They found that 60° and 90° on either left or right have 6-7-dB advantages. Tang et

al. described an assistive eyewear system that uses a RGB-D sensor and spatial audio to convey object locations to blindfolded users [113]. They found that users tend to select locations to the left of and below the target's true location.

In summary, there are only limited work on the effectiveness of spatial audio generated using audio synthesis technologies and they offer some basic understanding of how well users can localize spatial audio, but Web Audio API has not been studied yet. Spatialized moving audio is a subject that has not been investigated. We do not know how well users can interpret a moving audio, or what spatial properties make a moving audio more or less effective.

4.4 Study Design

The goal of this study is to learn design patterns relevant to potential spatial audio interfaces. In the physical world, spatial audio may have indefinite variations. However, in terms of synthesized spatial audio, prior psychoacoustic and HCI work have shown that the horizontal plane is more effective than other dimensions. So as a start, we decided to focus on the horizontal plane. There are two fundamental topics that need to be examined:

1. **Horizontal stationary audio localization**: the most straightforward way to apply audio spatialization technology is to position regular auditory icons or *earcons* in spatial positions. An assumption for such a design is that users can localize the spatial audio, not necessarily perfectly, but at least to a certain threshold. No matter how interesting a spatial audio design might seem in theory, it cannot work if users cannot localize key locations. Therefore, our first goal is to establish a realistic baseline for how accurately users can localize audio spatialized using Web Audio API.

Figure 4.1 Web Audio API Evaluation Experiment Interface

2. **Horizontal audio movement recognition**: a more complex spatial audio design can be audio moving on a webpage to guide or convey navigation actions. How well a user can recognize such movements is critical. Additionally, knowledge of how a moving audio's spatial properties, such as its path, direction, and distance, might influence its recognition can also aid in the production of better designs.

4.4.1 Experiments

We designed two experiments to address the aforementioned topics, i.e., horizontal stationary audio localization and horizontal audio movement recognition. To access Web Audio API, we implemented them on two webpages with the same layout: a horizontal bar is located in the middle of the page with some margins on either side (Figure 4.1). The bar is divided evenly into 13 blocks. Each block represents a different horizontal location in the audio space. In contrast to the five locations tested by Lorho et al. and Vazquez-Alvarez et al. [77,118], we used 13 locations as it allowed us to evaluate a finer audio resolution. With Web Audio API, the webpage constructs a virtual audio space. The horizontal bar is mapped onto a half-circle in the virtual audio space. The half-circle is centered on the listener with a radius of three and zero elevation. This minimizes the roll-off effect, i.e., ensures equal distance from all sound sources to the listener as in [49].

The first experiment evaluates the localization of stationary audio placed in one of 13 possible locations. Each location is tested ten times in a total of 130 trials. The order is randomized. During each trial, participants play the audio and select where they think the audio is on the horizontal bar. They can replay the audio or change selection before making final confirmation. The correct location is revealed during the training, but not in the experiment session. The webpage provides visual feedback and uses script to prevent errors, e.g., missing answers, and overlapped playbacks. This experiment takes about 20 minutes to complete.

The second experiment tests the recognition of audio movements. Participants hear a moving audio cue and select the starting and ending locations. To reduce fatigue and maintain reasonable experiment duration, we constructed 42 movements using seven out of the 13 locations (Figure

4.3). Each path is tested three times for a total of 126 trials. The order is randomized. The webpage provides visual feedback and includes an error prevention script. This experiment takes about 45 minutes to complete.

Stimulus

The audio used in both experiments is based on a clock "tick-tock" sound. It includes frequencies from 127Hz to 20 kHz. Prior study has observed that simple sine wave tone led to faster and more accurate performance, but their users found tonal sound annoying [81]. Since our experiment runs over an hour, we tried to use a sound that participants are familiar with in an effort to reduce irritation after long exposure.

The sound sample was edited so that it could be played continuously without noticeable gaps. In the stationary audio experiment, the script randomly selects one of the 13 locations, then the sound sample is positioned there and played for one second. In the moving audio experiment, the script randomly selects one path (i.e., a pair of starting and ending locations). Then the path is divided to 20 steps with even intervals. The sound's movement is constructed by playing the sound at these 20 steps consecutively. There is a gap of 250 ms between steps as prior study has shown that the perceptual analysis of a first sound is interrupted if a second sound is played less than 250 ms after the first one [29]. Overall, the playback takes five seconds to complete.

Apparatus

All experiments were conducted using Chrome browser (Version 49.0.2623) on an Apple MacBook Pro laptop (MD313LL/A, OS X Version 10.11.4). Participants were provided with a pair of Sony MDR-E828LP earbud (Frequency: 12-22,000 Hz, Sensitivity: 108.0 dB/mW). This

earbud is inexpensive, but popular based on Amazon reviews. We chose it over other more expensive high-performance headphones because we wanted to replicate an average user's computing audio environment.

4.4.2 Procedure

Each session starts with an introduction of spatial audio, Web Audio API, and the purpose of this research. Participants are encouraged to ask questions. It is followed by the stationary audio experiment, a brief break, then the moving audio experiment. The session finishes with a demographic survey.

In each experiment, a training session is presented first. The webpage is the same as used in later experiment, but it has only 13 trials. The correct answer is also revealed after each selection. The goal is to prepare participants with the test procedure and calibrate mapping audio to the visual representation. Participants can repeat the training if desired.

Participants can take breaks during experiments if necessary. We faced a dilemma when making this decision. On one hand, fatigue might decrease a participant's performance. On the other hand, participants might lose any subtle clues they have learned after a break. Finally, we decided to allow breaks. But when resuming the experiment, participants first hear all 13 sounds played one by one from left to right. We hope this helps participants recall previous learned clues.

4.4.3 Data Collection

We collect performance and interaction data. In each trial, we collect the final selection, the number of audio playbacks, the number of selection changes, and the trial completion time (from

when the audio is played for the first time until a final selection is confirmed). The demographic survey collects age, gender, and self-reported hearing conditions.

4.5 Results

We recruited 18 subjects (9F/9M). The average age is 30.6 (SD=5.24). None of the subjects reported any known hearing issues. All subjects participated both experiments. None of them has prior experience with Web Audio API.

4.5.1 Dataset and Analysis Overview

The stationary audio experiment generated a dataset of 2340 data points. The moving audio experiment dataset has 2268 data points. About 1% of trials took three standard deviations or longer than the average trial completion time. Since there was no irrational behavior observed, we kept them in the data analysis. On average, participants took 0.83 (SD=1.1) during the stationary audio experiment and 2.1 breaks (SD=2.6) during the moving audio experiment.

We used ANOVA or similar methods when comparing overall recognition rates. However, due to violation of assumptions required by parametric methods, we often cannot use them to explore more complex cases. Non-parametric methods could be used in some cases. But we would work with a small dataset with 18 aggregated data points for each condition. Eventually, we decided to use regression models to explore the connection between recognition and various spatial audio properties. These methods have no assumption and allow us to take advantage of the full dataset. The final models can also provide additional information on a predictor's influence.

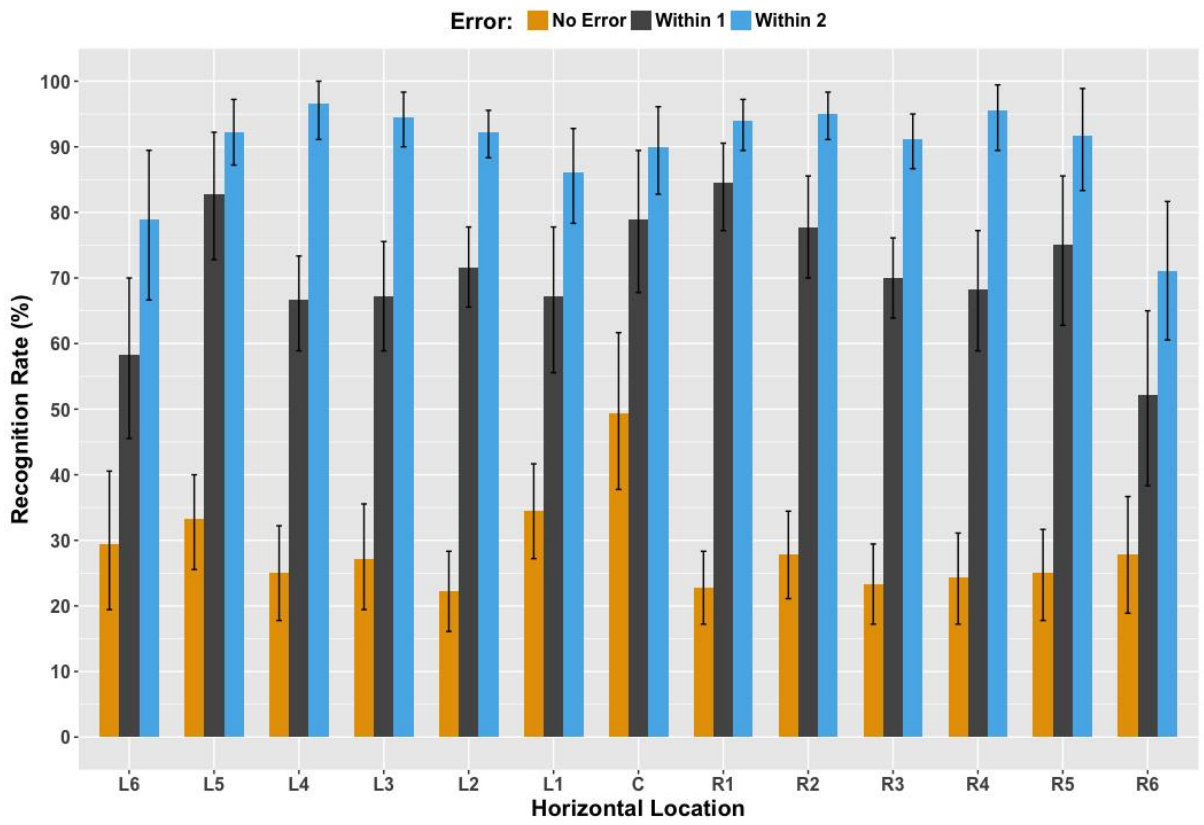


Figure 4.2 Stationary Audio Recognition Rates

4.5.2 Stationary Audio Recognition

To learn how accurate Web Audio API can spatialize sound, we calculated the recognition accuracy for each virtual location. We performed this calculation based on the stationary dataset only. The average recognition rate is 28% (SD=7%). We used repeated measure ANOVA to check whether any location has significantly different recognition rate and did not find any significant effect. However, most user answers are close to the correct answers: if we include answers one location away from the correct answer, the mean recognition rate increases to 71% (SD=9%).

There are significant differences between left 6 and left 5 ($p<.01$), left 5 and right 6 ($p<.05$), right 5 and right 6 ($p<.01$). If we include answers up to two locations away from the correct answer, the mean recognition rate increases to 90% (SD=7%). There are significant differences between left 4 and right 6 ($p<.05$), right 4 and right 6 ($p<.01$), right 5 and right 6 ($p<.05$). Figure 4.2 shows these three measures.

Allowing space for errors reduces the resolution as the space next to a target location cannot be used. We did an exhaustive search to identify locations with the best overall performance for a given resolution. We found that if the horizontal audio space is divided into two parts, the best recognition rate is 96% (left 4, right 4); if divided into three parts, the best recognition rate is 90% (left 5, center, and right 5); if divided into four parts the rate is 72% (left 5, left 2, right 1, and right 5); if divided into five parts, the rate is 52% (left 6, left 3, center, right 3, and right 6).

4.5.3 Error Distribution

When calculating recognition rates, we noticed that the direction of errors is not uniform. This informed us to explore further with a linear model. The independent variable of the model is location as a factor variable. Location center is the baseline. The dependent variable is the signed error distance, i.e., if the center is the correct answer, right 1 has an error distance of 1 and left 1 has an error distance of -1. When an independent variable has a negative coefficient, it reduces the value of the dependent variable. In this context, it indicates that the independent variable contributes to errors towards the left of the correct location comparing to the baseline; and vice versa.

Our final model is a multilevel linear model as subject is a significant random variable, $x^2(1)=42.57$, $p<.0001$. Location is a significant fixed effect variable, $x^2(12)=549.08$, $p<.0001$.

Table 4.1 Stationary Audio Recognition Error Model

Variable	b(SE)	CI
(Intercept)	-0.26(0.12)*	-0.49, -0.03
Left 6	1.74(0.14)***	1.47, 2.02
Left 5	0.52(0.14)***	0.24, 0.80
Left 4	-0.13(0.14)	-0.41, 0.14
Left 3	-0.25(0.14)	-0.53, 0.03
Left 2	-0.45(0.14)**	-0.73, -0.18
Left 1	-0.31(0.14)*	-0.59, -0.03
Right 1	0.07(0.14)	-0.21, 0.34
Right 2	0.24(0.14)	-0.04, 0.52
Right 3	0.54(0.14)***	0.26, 0.82
Right 4	0.39(0.14)**	0.11, 0.67
Right 5	-0.29(0.14)*	-0.57, -0.01
Right 6	-1.4(0.14)***	-1.68, -1.12

p-value: ***<.001 **<.01, *<.05

Left 6, left 5, right 5, and right 6 are special cases as errors could only go as far as where the edges are in one direction. For others (left 4 to right 4), all left locations have negative fitted coefficients and all right locations have positive fitted coefficients, though not all are significant (Table 4.1). This provides evidence that sound locations influence the direction of user errors. Another notable observation is the negative intercept. This suggests that our participants tend to make errors to the left for the center location.

4.5.4 Interaction with Stationary Audio

The number of playback, answer changes, and trial completion time tell us how participants behaved in each trial. These data indirectly shed light on different difficulties among locations. For example, for a location hard to recognize, participants might play it a few more times, change her selection due to uncertainty, and spend generally longer time to make up her mind.

We started with playback data. Participants played the stimulus once in 1301 (56%) trials. 590 (25%) trials were played twice; 273 (12%) were played three times; 176 (7.5%) were played four times or more. A multilevel linear model shows that subject is a significant random variable, $x^2(1)=1240.41$, p<.0001, and locations is a significant fixed variable, $x^2(12)=119.21$, p<.0001. The model shows that left 6 (p=.0000), left 5 (p=.0000), left 4 (p=.0000), left 2 (p<.01), right 4 (p<.01), right 5 (p<.001), and right 6 (p<.01) have less number of playbacks than the center.

Selection data shows that 2244 (96%) trials were confirmed without changes; 87 (3.7%) were changed one additional time; 9 (0.4%) trials were changed three times or more. We did not find any difference among locations.

Finally, we explored task completion time with a multilevel linear model. Controlling playback and selection change, we did not find any difference among locations.

In summary, the tests suggest that participants were more confident when recognizing sound placed at far left or right. When the sound is closer to the center, they were more careful (or confused) as they replayed more times. Once they made up their mind, they quickly moved on.

4.5.5 Moving Audio Movement Recognition

While the stationary audio experiment analysis focuses on the precise location recognition, we focus on movement recognition when analyzing the moving audio dataset. We divided the horizontal audio space into three regions: left, center, and right (Figure 4.3). A moving audio can start in one region and end in the same or a different region. Combined with direction, we have totally 10 movement types (Table 4.2).

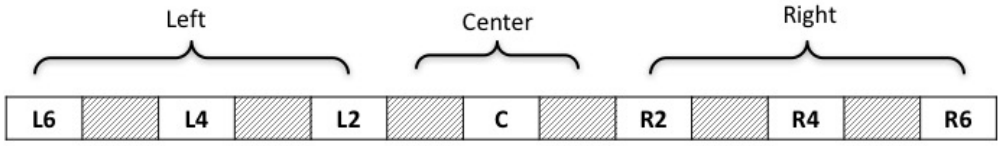


Figure 4.3 Moving Audio Movement Regions

Table 4.2 Movement Types

Movement Type Code	Direction	Starting	Ending
LR	Right	Left	Right
LC	Right	Left	Center
LLR	Right	Left	Left
CR	Right	Center	Right
RRR	Right	Right	Right
RL	Left	Right	Left
RC	Left	Right	Center
RRL	Left	Right	Right
CL	Left	Center	Left
LLL	Left	Left	Left

The overall movement recognition rate is 79.8% (SD=7.86). Figure 4.4 presents recognition rates for each movement type. We used non-parametric Friedman test (due to violations of distribution and variance assumptions) to test the differences among groups. The test shows that movement recognition rates are significantly different among movements, $x^2(9)=27.21$, $p<.01$. Post hoc test shows that the recognition of LR is significantly better than LLR, RC, and CL.

We employed generalized linear models to explore how a moving audio's properties can influence the correct recognition of its movement. The independent variables are properties of a moving

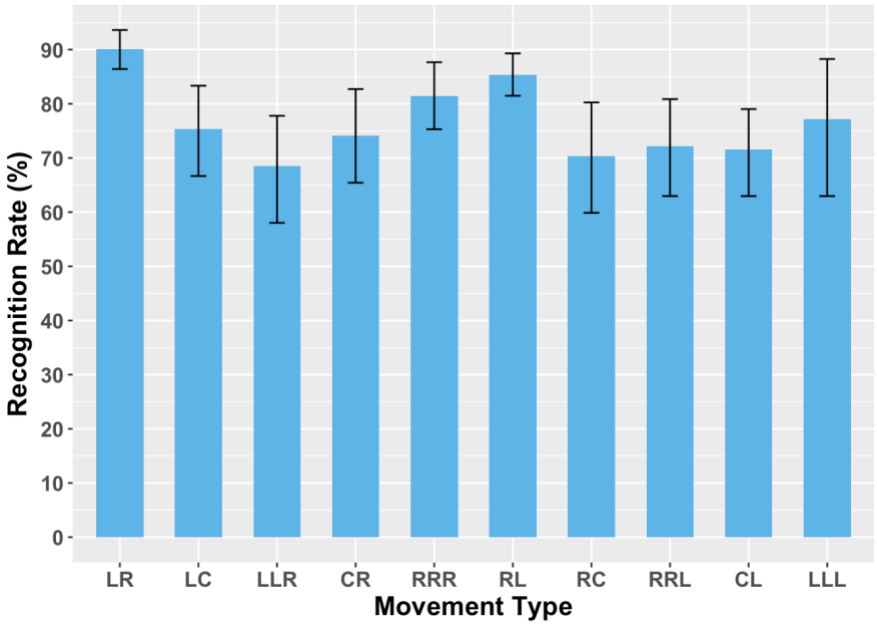


Figure 4.4 Moving Audio Movement Recognition Rates

audio that we are interested in: direction, path, and length. Direction and path are a breakdown of the movement. For example, LR and RL have the same path (between left region and right region), but different directions. The dependent variable of the model is the binary outcome of whether the movement is recognized correctly. Such a model describes how a parameter could change the odds of obtaining a positive outcome, i.e., movement being recognized correctly.

Our final model is a generalized linear mixed-effects model as subject has significant random effect, $x^2(1)=78.3$, $p<.0001$. All independent variables, length, $x^2(1)=101.09$, $p<.0001$, path, $x^2(4)=9.654$, $p<.05$, and direction, $x^2(1)=3.845$, $p<.05$, are significant. Table 4.3 presents the model. The baseline for path is movements that stay in the left region, and the baseline for direction is towards right. Based on the model, the corresponding movement with a length of 2 (minimum length used) has a probability of 73.5% being recognized. Longer length increases the odds of recognition (1.49 times for each increase of 2 units), and audio moving towards left is less likely

Table 4.3 Moving Audio Movement Recognition Model

Variable	b (SE)	95% CI for odds ratio		
		Lower	Odds ratio	Upper
(Intercept)	0.62(0.20)**	1.27	1.86	2.75
End at the center	-0.25(0.19)	0.54	0.78	1.12
Cross the center	0.03(0.22)	0.66	1.03	1.59
Start at the center	-0.25(0.19)	0.54	0.78	1.12
Right	0.22(0.19)	0.87	1.25	1.80
Length	0.20(0.03)***	1.14	1.22	1.30
Direction: left	-0.21(0.11)*	0.65	0.81	1.0

p-value: ***<.001 **<.01, *<.05

to be recognized (reduces the odds by 0.81). There is no significant difference among movement types.

Goose et al. reported that the accuracy of tracking audio movement decreases significantly at the extreme ends of a half circle audio space [49]. We examined this by building a model based on moving audio with length of 2. Our final model is a generalized linear mixed-effects model as subject has significant random effect, $x^2(1)=26.68$, p<.0001. Starting location is a significant fixed effect variable, $x^2(6)=36.34$, p<.0001. Ending location and direction are not significant. Based on the fitted model, we learned that moving audio starting from left 6 was significantly less likely to be recognized than all other starting locations except right 6; moving audio starting from right 6 was significantly less likely to be recognized than all but left 6 and center; moving audio starting from the center is more likely to be recognized than left 6, but less likely to be recognized than left 2.

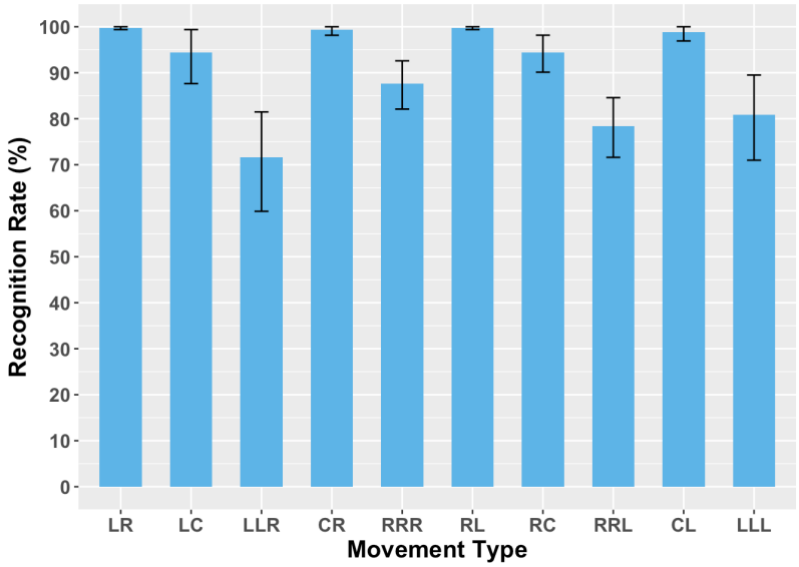


Figure 4.5 Moving Audio Direction Recognition Rates

4.5.6 Moving Audio Direction Recognition

We also analyzed the recognition of movement direction alone. The mean for overall direction recognition is 93.2% (SD=3.38%). Wilcoxon signed-rank test (chosen due to violation of distribution assumption) shows that there is no significant difference between two directions.

Figure 4.5 shows direction recognition rates for each movement type. We used Friedman test (due to violation of distribution and variance assumptions) to test differences among movement types and found significant differences among movements, $x^2(9)=27.21$, $p<.0001$. Post hoc tests show that the direction recognition of LLR is lower than those of LR, RL, CR, CL, and LC; the direction recognition of LLL is lower than those of LR, RL, and CR; the direction recognition of RRL is lower than LR, RL, CR, and CL. It should be noted that there is no difference among six movements that concern the center location, i.e., crossing, starting at, or ending at the center.

Table 4.4 Moving Audio Direction Recognition Model

Variable	b (SE)	95% CI for odds ratio		
		Lower	Odds ratio	Upper
(Intercept)	-0.17(0.30)	0.47	0.84	1.51
End at the center	1.24(0.29)***	2	3.45	6.22
Start at the center	3.12(0.60)***	8.25	22.72	94.16
Cross the center	2.81(0.78)***	4.39	16.69	110.82
Right	0.45(0.2)*	1.05	1.57	2.35
Length	0.55(0.10)***	1.43	1.73	2.12

p-value: ***<.001 **<.01, *<.05

Likewise, there is no difference among four movements that stay within either the left or the right region.

We also used generalized linear models to assess how a moving audio's properties influence its direction recognition. Subject is a significant random variable, x^2 (1)=28.93, p<.0001. Path (x^2 (4)=281.78, p<.0001) and length (x^2 (1)=36.767, p<.0001) are significant fixed effect predictors, whereas direction is not. The baseline for path is movements that stay in the left region. The fitted coefficients are presented in Table 4.4.

Based on the fitted model, we can calculate that the probability of recognizing the reference audio, which moves within the left region with a length of 2 (minimum length used), is 71.7%. Changing to any other moving path while maintaining the same length would increase the odds of being recognized correctly significantly (between 1.57 times and 22.7 times). Specifically, the probability of such an audio being recognized is 79.9% if it moves within the right region, 89.8% if it ends at the center, and 98.3% if it starts at the center. Increasing the length also increases the odds (3 times for each increase of 2 units). By changing the baseline path and refitting the model,

Table 4.5 Moving Audio Task Completion Model

Variable	b(SE)	CI
(Intercept)	-844.89 (195.19)****	-1226.66, -463.13
Number of replay	6138.06(60.23)****	6020.25, 6255.86
Number of selection	2042.68(74.57)****	1896.83, 2188.52
RC	-47.04(170.06)	-379.65, 285.57
LR	-95.53(138.69)	-366.79, 175.72
RL	-123.81(138.8)	-395.28, 147.66
CL	-584.84(169.83)***	-917, -252.67
CR	-485.39(169.89)**	-817.67, 153.11
LLR	-418.09(170.06)*	-750.71, -85.47
LLL	-607.33(169.80)***	-939.41, -275.24
RRR	-893.58(169.82)****	-1225.72, -561.44
RRL	-113.49(170.00)	-445.98, 219

p-value: ****<.0001 ***<.001 **<.01, *<.05

we learned that movements crossing or starting from the center are significantly more likely to be recognized than others, and there is no significant difference between these two paths.

4.5.7 Interaction with Moving Audio

Like with stationary audio dataset, we also looked at interaction data of how participants behaved during trials. 1640 (72%) moving audio stimulus were played only once; 516 (23%) were played twice; 112 (5%) were played three times or more. The fitted multilevel linear model includes subject as a random variable, $x^2(1)=557.10$, p<.0001, and length as the only significant fixed variable, b=-0.03, $x^2(1)=71.29$, p<.0001. The fitter coefficient suggests that longer length is associated with less playbacks.

2136 (94%) trials were confirmed without changes; 73 (3%) were changed once; 59 (3%) were changed twice or more before the final decision was made. We did not find any difference among locations.

Finally, we explored task completion time with a multilevel linear model that controls playback and selection change. Subject is a random variable, $x^2(1)=392.33$, $p<.0001$, and movement type is a significant fixed effect variable, $x^2(9)=66.6$, $p<.0001$. The fitted model is presented in Table 4.5. The baseline movement is LC, which moves from the left region to the center. The model suggests that comparing to the baseline, trials featuring audio movement RC, LR, RL, and RRL take about the same time to complete, whereas trials featuring audio movement CL, CR, LLR, LLL, and RRR takes significantly less time to complete.

4.6 Findings

Based on the results, we have come to five main findings.

4.6.1 Horizontal Stationary Audio Localization

The first goal of this research is to establish a baseline of what Web Audio API's spatialization features can achieve in the horizontal plane. Based on our data, we calculated the recognition rate for some common audio resolutions. When the audio resolution is 3, i.e., the half-circle horizontal audio space in front of a user is divided into three parts, the corresponding recognition rate is 90%. When the audio resolution is 4, the recognition rate is 72%. When the audio resolution is 5, the recognition rate is 52%.

We have also learned the most effective locations for each audio resolution from the data analysis. Since the audio space in our experiments is constructed as a half circle space surrounding the user, i.e. the right edge of right 6 corresponds with 0° and the left edge of left 6 corresponds with 180°, we can also calculate the specific positions based on these locations. As reported earlier, when the audio resolution is three, we found that left 5, the center, and right 5 are the most effective locations.

They correspond to approximately 159° (left 5), 90° (center), and 21° (right 5). Similarly, when the audio resolution is four, the most effective positions are at approximately 159° (left 5), 118° (left 2), 76° (right 1), and 21° (right 5). Finally, when the audio resolution is five, the most effective positions are at approximately 173° (left 6), 132° (left 3), 90° (center), 48° (right 3), and 7° (right 6).

4.6.2 Audio Space Layout

Prior work featuring audio spatialized in multiple locations on the horizontal plane often divides the audio space approximately in even sizes. For example, when evaluating the perception of concurrent speeches placed at different horizontal locations, Guerreiro et al. had three conditions: two speakers, three speakers, and four speakers [50]. The respective design has 180°, 90°, or 60° between speakers. In the experiment conducted by Lorho et al, they evaluated localization of five locations on the horizontal plane [77]. They adopted a layout where five locations were apart from the neighboring location by either 40° or 50°. This design is intuitive. However, it was never empirically evaluated. Our data provides some evidences of the validity of this design. Based on the audio positions reported earlier, when the audio resolution is three, there is a 69° span between the left, the center, and the right. When the audio resolution is four, the three interspaces span for 41°, 42°, and 55°. When the audio resolution is five, the interspace spans for either 41° or 42°.

We also observed interesting error patterns. Our data shows that users tend to make recognition errors biased towards the region where the audio is located in. When an audio is placed in the left regions, users tend to think the source of the audio is further left. When an audio is placed in the right regions, users tend to think the source of the audio is further right. One possible design implication is that, while the audio space is divided into even parts for a given audio resolution,

the targeted audio position is not necessarily in the center of the audio zone. Instead, designers should take into consideration of a user's bias based on the region. If a targeted audio zone in the left region, it should have more space on its left to anticipate more user errors in that direction.

4.6.3 Extreme Ends of the Audio Space

In their audio browser, Goose et al. used a half-circle audio space at first, but then they found that users could not track the audio's position as accurately at the extreme ends of the half-circle [49]. This prompted them to revise the design to a stage arc audio space. However, they did not present the exact range of the arc. Based on their graphical illustrations, the insensitive area seems to be around 30° at either end, i.e. the stage arc spans from 30° to 150°.

From our data analysis, we found that audio moving from left 6 to left 4 and from right 6 to right 4 were significantly less likely to be recognized than almost all other movements with the same length. This aligns with the observation by Goose et al. as left 4 and right 4 in our experiment correspond to 145° and 35° on a half-circle. This is a conservative estimate. We did not use left 5 and right 5 in the experiment. Therefore, we do not know how sensitive they could be.

4.6.4 Horizontal Audio Movement Recognition

We found that the recognition rates of ten movements are 79.8% on average. There is no significant difference among the movement types. However, a movement's length has significant effect on the recognition of its movement. Specifically, when an audio's length is increased by 2 units (i.e. 30°) in our experiment, the odds of its movement being recognized will increase 1.49 times. In addition, the direction of a movement also has significant effect. In general, movements towards right can be recognized better than movements towards left. For example, an audio moving within the left region with a length 2 and towards the right has a probability of 73.5% of being recognized

correctly. However, the probability would decrease to 69.2% if it moves towards the left. We do not have any participant feedback on why this is the case. Prior work has reported asymmetry recognition accuracy [77]. This might be a similar case and should be investigated further.

If we only care whether users can detect a movement's direction, the average recognition rate can achieve 93.2%. There is no significant difference between the two directions. Movements that stay in the left or the right region generally have lower direction recognition rates comparing to other movements that start at, end at, or cross the center. Movements crossing or starting at the center have the best direction recognition rates. Length also has a significant effect. Increasing the length by two units (i.e. 30°) results in its odds of being recognized increased by 3 times.

If using moving audio to provide feedback, designers could consider direction if the information is binary. This will ensure accurate recognition. Combining direction with the two most recognizable movements, i.e., crossing the center, and starting from the center, allows designers to communicate four different states. In either case, when possible the moving audio should employ a long length.

4.6.5 Lateral Positions and Central Locations

We learned based on stationary audio playback patterns that users are more likely to play an audio placed in the central area more than once. The differences are subtle, though significant, since they do not lead to vastly different overall recognition time. However, fewer playbacks do not mean higher accuracy, as the center location always has better recognition rates. This seems to suggest that recognizing audio placed in the left or the right region might incur less cognitive cost. Similar observations were made when analyzing moving audio data. As reported earlier, our participants spent significantly less time when a trial featured one of three audio movement types that stay

within either the left or the right region, though participants did not recognize these movement better or worse than others.

Putting both observations together, it seems to suggest that users would have an easier time to process audio positioned in or moving within a lateral region. We did not have any feedback from our participants that could explain this behavior. However, Guerreiro et al. have reported in their work that their participants performed better when two concurrent audio streams were presented at either the left or the right of the participant than the condition where an addition audio stream was presented in front of the participant [50]. Their participants explained that they could focus on a lateral source by shutting down the other ear. However, with the frontal audio stream, they have to listen actively with both ears. Though our experiment setup is very different from theirs, we think the same explanation might apply. When only one lateral region is concerned, the participant can quickly eliminate all locations from the other lateral side. One possible design implication is that designers should utilize lateral locations or movements given everything else is equal. However, we should remind that longer movement length can have a positive effect in recognizing the movement or the direction, and lead to less processing time. Therefore, designers need to make this decision with consideration of all factors.

4.7 Limitations and Future Work

One main limitation of this work is that we conducted the experiment using Google Chrome browser on a Mac laptop. Different browsers should implement Web Audio API according to the same standard. However, there might have small differences. Therefore, our findings might not hold for all.

The exploratory nature of our experiments is another limitation. We tried to design experiments that are flexible enough to not only verify findings reported for other technologies but also provide space to probe new questions. As a result, some findings are derived based on lower level data. Though we are confident with our rationale, they should be validated with confirmatory evaluations.

One logical sequence of this work is to evaluate the other two dimensions in the audio space, i.e., elevation and distance. Though many psychoacoustic and HCI studies have concluded that these two dimensions are less effective, driven by different objectives, HCI researchers could uncover insights useful for design.

The next step in our research is to develop spatial audio interface prototype based on what we have learned here. We focused on audio's spatial properties specifically in this work as it is a less explored area. For the research prototype, we would like to combine spatial audio with other proven audio design elements, such as pitches, timbre, and conduct user studies to establish good or bad design practices.

4.8 Summary

The progress of Web Audio API has brought exciting new design opportunities to the web. Effective spatial audio designs, however, require a thorough understanding of this new technology. In addition, complex spatial audio features can incorporate many different properties. How to compose an audio to aid positive usability improvement has not been explored sufficiently. Aiming to make progress in this domain, we have conducted experiments to evaluate the localization of stationary audio and the recognition of moving audio in the horizontal plane. In this chapter, we presented the results and findings from our experiment.

In summary, we found that users can achieve recognition rate of 90%, 72%, and 52% if the horizontal audio space is divided into 3, 4, and 5 parts respectively. When localizing a stationary audio, users tend to make errors biased towards the regions where the audio is located. This can influence the choice of effective target location within an audio zone. Users can recognize audio movements with 79.8% accuracy. There is no significant difference among the ten movement types we have evaluated, but the movement's length and direction do have significant impact to the recognition. If only the correct direction recognition is evaluated, users can detect the direction with 93.2% accuracy. Users can recognize the direction of movements that cross the center or starting from the center significantly better. A movement's direction has no influence to correct direction recognition, whereas its length does. We have also provided evidence that when dividing the audio space to multiple zones each zone should have even sizes. In addition, prior work has reported that sound located in the extreme ends of a half-circle audio space is harder to recognize. We verified this and clarified the range spans from at least 35° to 145°. Finally, we have evidence that users could interact with audio better if it is placed or moves within one lateral region.

This study provides us some sights on how to design spatial audio interface. We took these lessons into the development of our proof-of-concept screen reader prototype. In the next chapters, we will present our user study with blind users. Though the main goal of the user study is to understand blind users' behaviors and reactions to the idea of introducing spatial audio feedback into screen readers, we also ask their feedback on the prototype designs. We will report these findings in the rest of the dissertation.

Chapter 5 Exploring the Potential of Spatial Audio with Screen Reader Users: Study Design

Chapter Three has identified a few areas where new technologies can reduce the communication barriers between sighted and blind web users. Among them is using spatialized non-speech audio to convey layout information. We find this option promising as capable audio synthesis technologies, such as Web Audio API, are readily available. In this chapter, we present our planning of a user study that will further explore the potential of spatial audio features with screen reader users. This chapter provides the motivation, prototype development, and study designs. Chapter Six will present the findings.

5.1 Introduction

From the text analysis study reported in Chapter Three, we have learned that one problematic area for the communication between screen reader users and sighted users is conveying web page layout information. Sighted web users use spatial terms to describe web pages, even when they are aware that the information will be used by blind users. Compared to limited use of colors and shapes in the same context, it suggests that the use of spatial terms might be hard to avoid by sighted users who have no accessibility training. Additionally, though the use of spatial terms is pervasive, there are only a few frequently used spatial concepts when describing web page components. A possible design to mediate communication is to convey spatial concepts using non-visual modality.

From the Web Audio API evaluation study reported in Chapter Four, we have learned that Web Audio API is capable of conveying limited stationary positions in the horizontal audio plane. When the number of locations is kept low, users are able to localize these locations with reasonable accuracy. Moreover, moving audio can convey movement direction reliably.

Informed by both studies, we believe that adding spatialized non-speech audio feedback to screen readers has the potential to help mediate the communication of web page layout information between sighted and screen reader users. Using Web Audio API, an audio cue can be dynamically generated for a web element and spatialized according to its position on the web page. Though spatial audio synthesized via Web Audio API does not have a fine resolution, it may be sufficient considering there are only a few frequently used spatial concepts. If screen reader users could perceive the relative layout of web elements or the overall layout of a web page based on spatial audio feedback, such information may help them incorporate spatial concepts into their interaction with web pages, i.e., interpret spatial terms or use spatial terms in their own descriptions.

This chapter and the next chapter present a study that explores the potential of spatial audio feedback for screen reader users. We focus on evaluating whether or not screen readers can actually derive layout information based on spatial audio cues and gathering early feedback on spatial audio designs from them. In addition, we also try to validate our findings from Chapter Three with blind users regarding to their experiences with spatial terms and gain a more in-depth understanding of how they with spatial concepts using screen readers.

In this chapter, we report our spatial audio feedback design considerations and the prototype development. We also present the user study design.

5.2 Related Work

One area of work that is closely related to this study is how blind users acquire information about a web page's layout. Webpage designs inherited some modular layout designs from paper-based mediums. Columns and modular sections are commonly seen in newspaper [85]. Narrow columns allow readers to read faster and avoid losing their places when changing lines. Modular layouts are also flexible: editors can easily merge columns when necessary; replacing content in one modular section would not impact other part of the newspaper.

Francisco-Revilla et. al studied how sighted users and blind users interpret the layout of web pages [36,37]. In their study of 20 sighted users, they used a think-aloud protocol and asked their participants to explain how information was structured and identify layout groups on a given webpage [36]. The researchers handpicked 8 shopping websites and 8 news websites for the experiment. Within each type of websites, half of them was assessed as having simple layouts while the other half as having complex layouts. This assessment was based on the number of groups and subgroups presented visually. After the study, they coded the interpretations so that they could analyze the order and frequency of elements mentioned. Their main finding was a confirmation of early work that users follow a top to bottom, left to right sequence when browsing a webpage. They have also provided evidence that users perform a quick parse of the webpage before looking at the content. They had some observation on how certain elements are more useful in guiding the interpretation process, for example, the page title is parsed earlier, search box and navigation bars are also parsed earlier and mentioned often. However, these elements are normally located in the top or left section of a webpage. Their reporting did not provide enough evidence that these elements' prominence in webpage interpretation is the same as their main finding.

They followed up with a study of how blind users interpret web page layouts [37]. In this second study, they recruited 4 blind users (visual impairment conditions were not reported). To control each session's duration, they tested 12 out of the 16 websites used in the previous study. The 12 websites still composed of equal number of shopping and news websites, as well as equal number of simple and complex websites. They also had to modify the measures somewhat in order to accommodate the characteristics of this targeted population, for example, removing the question assessing the perceived webpage layout complexity between live websites and screenshots of websites. This study did not reveal any significant user interaction patterns. However, they did observe that their participants tended not to traverse the webpage completely when trying to build a mental model of the page. Specifically, they tended to start with the top, then the left column. But after this, they would jump to the end the webpage, continue parsing from the bottom, then right column, and stop. As a result, the middle section was not examined. This observation is interesting. There might be many potential reasons, including misleading experiment instructions. However, since the researchers did not report enough details, it was difficult to explore further. Furthermore, the study also had a small sample. Therefore, this interaction pattern should be considered as inconclusive implications.

Accessibility researchers have studied sighted and blind web users separately before on how they process web page content. For example, Borodin et al. have documented screen reader users' unique browsing strategies in [13]. Jay et al. have studied how sighted users browse web pages in order to learn what critical information is not conveyed to blind users [60]. But these works do not help answer the question of how blind users make sense of spatial terms with the navigation strategies available to them.

5.3 Prototype

Most screen readers do not provide any spatial audio feedback. VoiceOver included in Apple's macOS (previously Mac OS X, then OS X) has a positional audio feature. However, there is very little knowledge of how blind users feel or use this feature. The lack of knowledge in this domain prompts us to implement a proof-of-concept prototype and use an exploratory study to investigate this uncharted area with targeted users.

5.3.1 Design

Our spatial audio feedback designs draw inspirations from how people use sound when navigating in a physical environment. Research in psychology has shown that humans and other species use two main methods to maintain orientation during travel [41]. The first method is landmark-based navigation. With this method, travelers recognize landmarks, which could be perceived via vision, audio, or odor etc. Then, travelers can maintain their orientation on an external map or a cognitive map based on position information associated with the landmark. The second method is path integration. With this method, travelers first identify their starting points on an external map or a cognitive map. Then, they update their current positions on the map with the movements they have completed.

Landmark-based navigation is an essential skill for blind people when traveling. During Mobility and Orientation training, blind people learn techniques to recognize orientation clues and landmarks from the surrounding environment using all sensory channels available, including auditory [59]. In the domain of Mobility and Orientation training, clues refer to information that is temporary in nature, for example, the sound of people walking by, or the sound of someone operating a copy machine in the background. In contrast, landmarks are information that is

permanently fixed in the environment, for example, the sound of elevator doors opening or closing. Functionally, clues provide hints that help blind people to derive their orientation, whereas landmarks allow blind people to recognize their positions instantly. By utilizing both clues and landmarks, blind people can find out where they are and where they should go next effectively. To distinguish from these usages and avoid implying the purpose that could be achieved, we refer to our spatial audio feedback as spatial audio "cues" in this document.

Landmarks are already familiar terms to screen reader users. There are two interpretations of the term "landmarks". Firstly, they can refer to the feature of navigating by landmarks provided by some screen readers. In this case, landmarks refer to semantic HTML tags or ARIA roles, such as main, article, banner, etc. Another interpretation is more similar to how it is used in physical navigation. A screen reader user can memorize the linear position of a web element and use it as an anchor or a reference point during navigation. For example, if an unlabeled button is a few steps after a "contact us" link, a screen reader user can use the link as a landmark. When trying to reach the button, she can do a search of the words "contact us", move to the link, then tab a few times to find the unlabeled button. This use is, of course, based on textual information. Yesilada et. al have considered such landmarks as one kind of travel objects that is essential in completing a journey successful [122]. The other main travel object class, memory, works with landmarks, i.e., blind travelers use memory objects to detect landmarks.

We hypothesize that the practice of using clues and landmarks in physical navigation can be utilized on web pages by accompanying each web element with a stationary audio cue that is spatialized in the audio space corresponding to its position on the web page. When users localize the location of the audio cue in the audio space, they would also obtain some ideas about where the web element is located on the web page. Some audio cues that have fixed positions on a web

page might have the potential to serve as landmarks. For example, logos are often placed at the top left corner, and login buttons are often on the top right corner. The audio cues associated with these web elements could be used as web page audio landmarks. When feeling confused about the orientation, screen reader users could play a landmark audio cue. Based on where the sound comes from, they could derive their current positions.

Path integration is also an important navigation technique for blind people. Prior research has suggested that path integration could be an effective technique regardless of one's experience with visual impairments [146]. When browsing web pages, screen reader users often memorize the procedures required to reach a goal. This can be considered a form of path integration as such a procedure includes clearly defined a starting point and step counts. When users feel lost on web pages, they often refresh the page and restart the process from the starting point or a reliable landmark half way through the procedure [119].

Spatial audio may be used to enhance path integration techniques by conveying movement directions. With current screen reader features, users can only utilize procedures involving sequential linear steps, for example, "tabbing five times" or "the third button". If users could perceive the direction of movement, they would be able to include additional directional steps in the procedure. To support this feature, screen readers can play moving audio cues corresponding to the direction when the user navigates from one web element to another. As presented in Chapter Four, users can reliably recognize audio's direction of movement in the horizontal plane. Prior research has shown that pitch is an effective audio property to convey vertical positions [106]. Combined together, a moving audio can deliver both horizontal and vertical location changes.

5.3.2 Technology

We decided to develop a custom-built prototype for this study. Using a screen reader that is commercially available and is already being widely used by blind users has the advantage of improving the validity of the study. However, there are also a few problems with this approach.

Firstly, it would pose many technical challenges to add spatial audio features to an existing screen reader. Most of the screen readers available in the market are proprietary technologies. Their source code is not open to the developer community. JAWS provides a mechanism that allows developers to control the screen reader's behavior using scripts (which are application-specific files that tell JAWS how to function, such as what to speak and how to navigate under various situations, when using the targeted application. JAWS includes only scripts for popular software by default. To ensure JAWS operates properly with other less popular software, developers can produce their own scripts and make them available to their users). This mechanism is not sufficient to support the features required by this research, as it only supports reading content using the existing audio features and does not allow adding spatial audio to the output. Among the popular screen readers, NVDA is the only open-sourced project. However, screen readers are complex software programs that often take years for a group of dedicated developers to produce. Adding spatial audio features by hacking into a branch of the source code and compiling a version for this research are technically possible, but it would take much time and effort to produce. The scope of this research will not justify the resource invested in such a development process.

Secondly, using an existing screen reader might also allow participants' experience with that particular screen reader to influence their behaviors and performance. Some participants might have more experience with a particular screen reader than other participants. This familiarity might

be confounded with the effect of new feature experimented. In addition, screen readers have provided shortcuts and features that are designed to support navigating among linearly ordered information. The goal of adding spatial audio features is to understand what new approaches could be possible given the additional layout information. Therefore, if participants are already familiar with existing ways to complete certain tasks, they are likely to resort to these approaches.

We used JavaScript to develop the prototype. In Chapter Four, we have learned that Web Audio API, which is available via JavaScript library, has the capability to deliver spatial audio reasonable for the purpose of this research. We also used JavaScript to support keyboard-based interactions. Typically, a web page is rendered from three types of code files: HTML pages that provide the content, CSS files that provide styling information, and JavaScript files that handle interactive functionalities. These three types of files are connected by references in the HTML pages. By implementing the prototype completely in JavaScript, web pages used for the study can simply reference the JavaScript files and have all features available to them without any further coding.

With JavaScript, it also requires little effort to reuse the code. All major browsers have some mechanisms in place to allow adding third-party extensions to the browser that could interact with the web page currently loaded (called Extension in Google Chrome browser and Add-on in Mozilla Firefox). Just like a regular web page, such an extension comprises HTML pages, CSS files, and JavaScript files. It also requires a configuration file that declares various access privileges necessary for the extension's functionalities to the browser. The files need to be packaged in certain structures required by the specific browser. Once installed, the extension has access to a lot information about the web page currently loaded in the browser. The extension and the web page can also communicate with each other in limited ways. By implementing the prototype using JavaScript, if the user study shows that certain implementation is immediately useful for users, the

relevant JavaScript code could be easily extracted and reused in an extension. The extension can then be released to the public quickly.

In addition to Web Audio API, we also used Web Speech API for the screen reader's text to speech features. Web Speech API allows various configurations. Developers can select the specific voice profile, set the pitch, rate, or adjusting volume of a speech. Developers can also programmatically cancel, pause, or resume a speech. There are also various events associated with a speech object. By adding callback functions to these events, developers can manage how an interface responds to user interactions.

5.3.3 Basic Functionalities

The prototype will be used as a probe in the user study. Though it does not need to have all features of a fully functioning screen reader, it needs to support common interactions that are likely to be used during the user study by participants. Chapter 2 has provided a summary of how screen reader users browse the web. In summary, screen reader users often use fast navigation techniques to find desired content or acquire an impression of what a web page presents. Fast navigation allows users to navigate by a certain type of element, such as by heading, or by paragraph. However, when the page is inaccessible or if the users are afraid of missing important content from the web page, they would adopt an exhaustive approach by going through everything on the web page one piece of content at a time, such as instructing screen readers to read continuously from start to finish, or by using Tab keys to go through all web elements that could receive focus (note: text-only content would be skipped).

The study prototype was implemented to replicate such supports (Table 5.1 for all functionalities supported). However, to reduce the time needed for training during the study, only a subset of fast

Table 5.1 Basic Screen Reader Functionalities Implemented in the Prototype

Key	**Function**	**In other screen readers?**	**Note**
N	Move to the next element	No	
H	Move to the next heading at any level, i.e., the level of the heading is ignored	Yes	
X	Move to the next checkbox	Yes	
B	Move to the next button	Yes	
E	Move to the next "edit box"	Yes	"Edit box" is a term used by JAWS and it refers to fields that accept text, such as <input> with type "text" or <textarea>
SHIFT	Navigate backward	Yes	When holding SHIFT and pressing the other navigation keys, the screen reader would find the previous matching element instead of the next one.
HOME	Jump to the top of the web page	No	Screen readers use different keys, e.g., JAWS uses key combination of CONTROL and HOME.
END	Jump to the end of the web page	No	Screen readers use different keys, e.g., JAWS uses key combination of CONTROL and END.
CONTROL	Mute the screen reader	YES	
Left Window Key	Replay the current element	No	This is a common screen reader feature. But the mapped keys are different.

navigation functionalities provided by popular screen readers was implemented. They are by headings, by checkboxes, by text fields, and by buttons. Later when developing web pages for the study, only these web elements were used. The limited use of different kinds of web elements also reduces the complexities of the web page and the tasks that study participants need to perform. When a feature is replicated, the prototype adopted the same shortcut key when possible. For a few functionalities that use combined keys, the prototype simplifies it by using just one key.

For exhaustive navigation, however, we decided to implement a new navigation key. One of the ways to understand how users interact with a web page is to observe what navigation functionalities they invoke when exploring the web page. The existing techniques supported by popular screen readers leave much ambiguity. When using the Speak All function, the screen reader will read everything from the current position until being interrupted or reaching the end of the web page. Users receive the information passively. This makes it hard to detect what information makes sense to the user. The more active approach, namely tabbing through all focusable web elements, gives better understanding of how the user responds to each navigation step. However, this process skips all textual content, which can have negative impact on navigation as the textual content surrounding a focusable web element can help clarify its purpose. The new navigation feature is a combination of Speak All and Tab-based navigation. A new shortcut key, N (for “Next”), was assigned to support actively traversing everything on a web page one piece of information or element at a time. With this feature, users can go through the whole page and are guaranteed not to miss anything.

The prototype also supports a few other common screen reader functionalities. When adding the SHIFT key to one of the navigation keys, i.e., holding SHIFT when pressing the other key, the screen reader will go backward, i.e., moving to the previous applicable element instead of the next

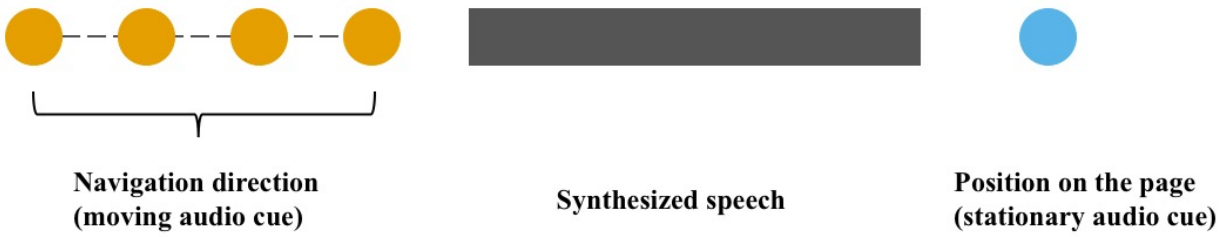
one. Users can use shortcut keys to jump to the top or the bottom of the web page. Users can also replay the current element in case they forget or do not hear it clearly the first time. When reaching the end of the web page (or the top of the web page if moving backward), the screen reader uses a short squeaky sound to indicate there is no more applicable content beyond this point.

One feature that is common to other screen readers, but not implemented in this prototype, is the ability to change speech rate. Normal sighted people listen to at about 300 words per minute. Experienced screen reader users can understand up to 500 words per minute [13]. The reason for not implementing this feature is similar to the motivation for developing a custom-built screen reader. As the participants' experience varies, they might have different expectations about how the spatial audio feedback should play. Not being able to adjust the speech rate helps ensure that all participants experience the same interface. It also helps reduce the training effort.

5.3.4 Spatial Audio Feedback and Additional Functionalities

The spatial audio features are implemented as a different module so that they can be turned on or off based on the user study conditions. The spatial audio features take a metaphor where the whole web page is mapped to a two-dimension audio space. The width of a web page corresponds to the horizontal plane whereas the height of the page corresponds to the vertical plane. The origin or the intersection of the two axes is located at the center of the web page. In addition to spatial audio, we also associated web elements at higher positions with higher pitches, as the localization of vertical positions is known to be unreliable.

The prototype produces two spatial audio cues. Therefore, the entire audio output can be divided into three sequential segments: the navigation direction audio cue, text to speech, and the positional audio cue. Figure 5.1 illustrates the composition of the audio output.



Navigation Direction Audio Cues

Figure 5.1 Screen Reader Prototype Audio Output

Navigation direction audio cues convey the direction of movement when a user navigates from one web element to another. Figure 5.2 shows three kinds of movements:

1. Horizontal position changes only
2. Vertical position changes only
3. Both horizontal and vertical changes

Horizontal position change is assessed by comparing the two web elements' left edges. If they differ more than a threshold, it is considered as having horizontal position change. The threshold is used to absorb minor alignment issues. In this prototype, we used 10 pixels as the threshold. The vertical position change is assessed by comparing the current web element's bottom value with the next web element's top value.

The prototype uses stereo panning to convey horizontal position changes and uses pitch changes to convey vertical position changes. Specifically, a navigation direction audio cue last 1.05 seconds. It comprises four sine wave pure tones. Each tone is 150-millisecond long and there is a 150-millisecond pause between two tones. If the navigation moves towards the right, the user will hear

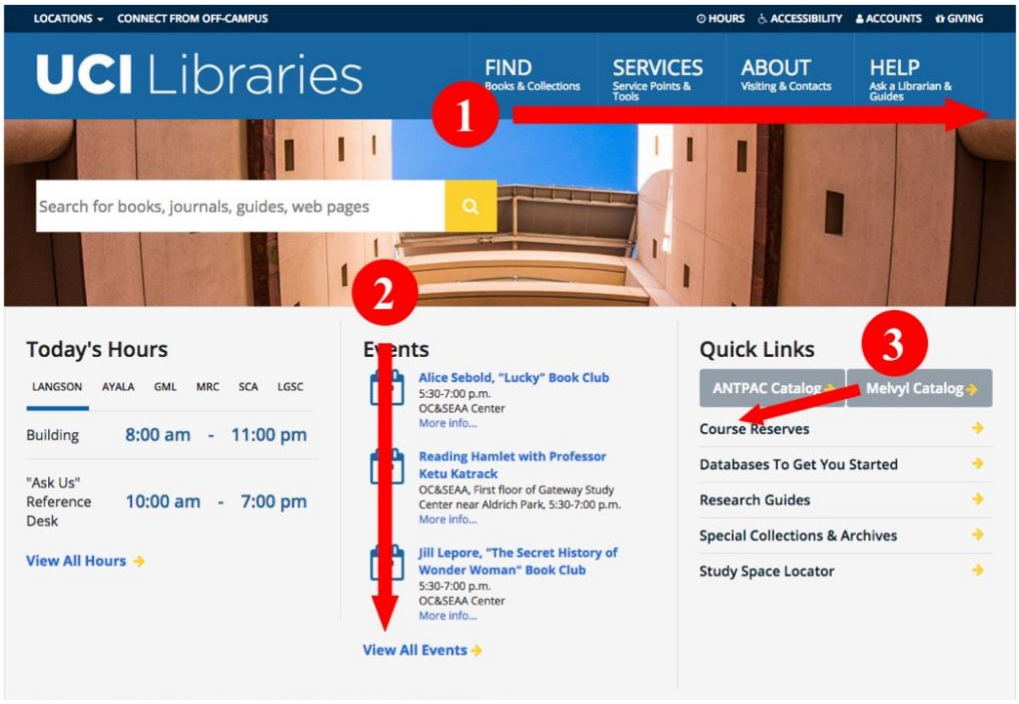


Figure 5.2 Navigation Directions

the first tone at the leftmost position and the last tone at the rightmost position. The middle two tones are positioned evenly between the first tone and last tone. If the navigation moves towards the left, the reverse order is used. If there is no horizontal position change, all four tones will be positioned in the middle of the horizontal plane.

If the navigation moves to a higher position, the first tone will have a lowest pitch, whereas the other three will have increasingly higher pitches. If the navigation moves to a lower position, the reverse order is used. The pitch effect is created by assigning different frequencies to the tones, i.e., lower frequency leads to lower pitched tone. The four values used are 200 Hz, 1000 Hz, 1800 Hz, and 2600 Hz. If there is no vertical position change, all four tones use the same frequency of

1400 Hz. If the navigation involves both horizontal changes and vertical changes, the audio cue features both horizontal and pitch changes.

It should be noted that the direction audio cues do not convey the distance of movement. For example, when navigating to a web element located on the right of the current web element, users hear the same audio cue regardless whether two web elements are next to each other or further apart. We used the full range of the horizontal positions or pitches to make sure users can recognize the change reliably.

Synthesized Speech

After the navigation direction audio cue, the prototype pauses for 300 million seconds before playing the synthesized speech output. The current web element's content or alternative content is provided to create a SpeechSynthesisUtterance object. The object uses default settings for speech rate, pitch, and volume, which correspond to normal speaking voice.

Positional Audio Cues

After the speech is over, there is another pause of 300 million seconds before the positional audio cue for the current web element starts. The audio cue is a sine wave pure tone that lasts 200 million seconds. The whole web page is mapped to a square audio space where the center of the web page is located at the audio space's origin (Figure 5.3). A web element's position on the web page is calculated by finding the center position of the rectangle area surrounding the web element. Then the web element's position in the audio space is calculated accordingly. For example, if a web element is to the left of the center and is higher on the page, it would have a negative horizontal value and positive vertical value.

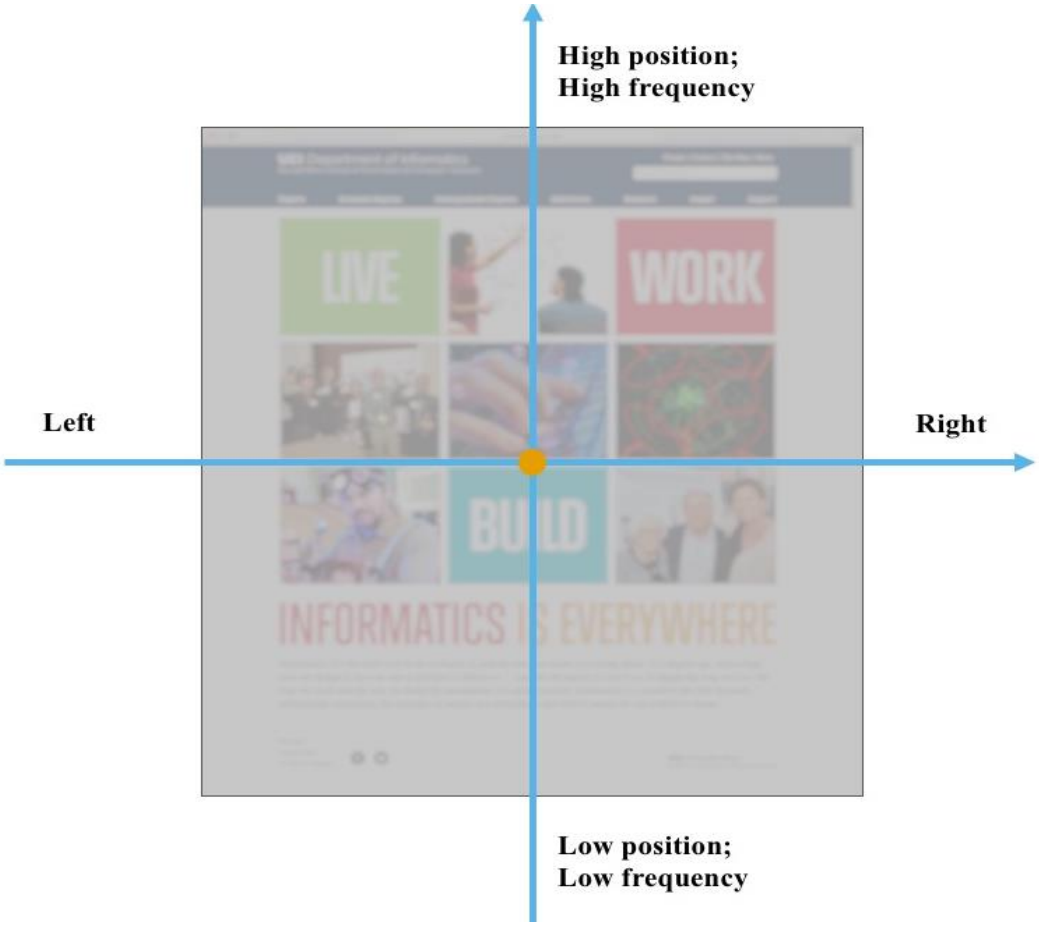


Figure 5.3 Mapping Web Pages to the Audio Space

Since elevation in an audio space is hard to localize, we also used pitches to help convey a web element's vertical position. The pitches used span from 220 Hz to 4220 Hz. If a web element is located at the lowest position possible on the page, i.e., the bottom edge, its audio cue will have a frequency of 220 Hz; if it is located at the highest position possible on the page, i.e., the top edge, its audio cue will have a frequency of 4220 Hz. However, since there are always some margins and paddings around a web page's content, positional audio cues will always have a frequency between 220 Hz and 4220 Hz.

There are a few main differences between direction audio cues and positional audio cues. Direction audio cues are expected to be easier to recognize, whereas positional audio cues are expected to be hard to recognize. Direction audio cues take longer to play, whereas positional audio cues are shorter. Direction audio cues convey the movement relative to the current and next web element. It is therefore expected to help path integration during the navigation. In contrast, positional audio cues convey the current web element's position relative to the whole web page. It offers more straightforward assistance in landmark-based navigation. However, if a user could tell the difference between the current positional audio cue and the previous one, which might be very subtle, the positional audio cues would also help with path integration.

Additional Screen Reader Features

To support the spatial audio feedback, we also provide some additional features (Table 5.2). As described earlier, pressing Left Windows key will replay the text or alternative text of the current web element when spatial audio features are disabled. If the spatial audio features are enabled, pressing Left Windows key will replay the speech and the positional audio cue afterwards. In addition, the Right Windows key can be used to replay the positional audio cue only. Since it is difficult for users to localize a tone when it is presented alone, users can combine SHIFT and Right Window key to replay the positional audio of the previous web element. In this way, a user can easily play the current positional audio and the previous positional audio consequently and recognize their differences.

To take advantage of the layout information that users might perceive, we also implemented directional navigation using four arrow keys. Users can press one of the arrow keys to move towards a direction. When there is no more web element in a direction, the screen reader will play

Table 5.2 Additional Screen Reader Features Implemented in the Prototype

Key	**Function**	**In other screen readers?**	**Note**
Right Window Key	Replay the current element's positional audio cue	No	
SHIFT + Right Window Key	Replay the previous element's positional audio cue	No	
Up Arrow	Move to the web element directly above the current web element	No	
Down Arrow	Move to the web element directly below the current web element	No	
Left Arrow	Move to the web element to the immediate left of the current web element	No	
Right Arrow	Move to the web element to the immediate right of the current web element	No	

the same squeaky sound to notify users. Because this feature requires web elements to align perfectly, we anticipate that it would only be useful when users recognize that web elements are arranged as a table or a grid.

5.4 Study Design

The main goal of the user study is to learn spatial audio design guidelines. With little existing understanding on how to design spatial audio feedback, we deliberately started with two most basic kinds of spatial audio cues. Though they seem trivial, there is still much to learn.

Most screen readers do not have any spatial audio feature. One exception is VoiceOver on Apple's Mac desktop and laptop computers. But it provides minimal documentation. There is no

information in the public domain about the effect and efficiency of applying such a feature in screen readers. Using moving audio cues in interfaces has received even less attention. There is no screen reader offering this feature. There is also no known HCI project, about screen reader or non-screen reader technologies, that has a focus on using synthesized moving audio to communicate information related to interface interaction.

There are more questions regarding screen reader user experience when using spatial audio features. For example, given limited stationary audio resolution, screen reader users can only learn a web element's approximate location on the web page. In this case, do they still benefit from such knowledge? And if so, how do they utilize the spatial information? What information conveyed via spatial audio cues is considered useful and what is not useful? Does the spatial information introduce new ways for screen reader users to learn web pages or do users integrate the spatial information into their existing navigation practices? How do screen reader users learn the additional spatial audio features?

These questions provide motivations for our user study. Our first goal is to hear participants' feedback on the two kinds of spatial audio cues provided in the prototype. In addition, we hope the prototype will inspire participants to provide broader feedback on using spatial audio in assistive technologies. Insights gained from the user study will help produce design revisions and clarify the research agenda for future research on this topic.

5.4.1 Exploratory and Qualitative

This user study is exploratory in nature since there is not sufficient existing research on the same topic that provides a clear framework and agenda. Given the lack of understanding, an exploratory study that utilizes qualitative research methods can make the greatest contribution at this stage, as

it would help gain a clear picture of how users approach the proposed technologies and understand what questions are meaningful to pursue further.

Using qualitative methods can also address a main challenge of conducting accessibility research. Lazar et al. has discussed the difficulty of recruiting people with disabilities for research activities [73]. In summary, a typical HCI experimental study requires 30~40 participants in order to create enough control groups and produce datasets that can be reliably analyzed statistically. However, such a sample size is often not possible for research targeting users with a particular kind of disability. Researchers have come up with a few strategies in response to this challenge. For example, an experiment can be designed to include only one treatment group and one control group, so that a small sample size could still lead to valid analysis. Researchers can also recruit participants without disabilities for the control group. However, this approach makes the assumption that these two groups, with otherwise similar demographics, are indeed equivalent. As Lazar et al. points out [73], this assumption is hard to measure. Another strategy is to conduct distributed research. This allows researchers to recruit participants without being constrained by locations. An obvious drawback of this approach is the lack of control. Researchers are not always certain of the user environment's configurations. Nonetheless, this option is suitable for survey studies, diary studies, and longitude studies based on log data. Finally, if it is appropriate for the topic, accessibility researchers can utilize case studies, which focus on getting rich data from a small number of participants over a long period of time. When planning the research, we had a sense by talking to our recruitment channels that we were likely to be able to have between 10 and 20 participants. It was clear that we could achieve more valid findings using qualitative research methods.

One-on-one interviews are the primary method in the user study. We used the prototype to demonstrate the spatial audio designs and as a probe to focus the conversation. Participants were encouraged to think with and beyond the prototype.

We also collected some quantitative data using questionnaires. However, the answers are mainly used to guide the interviews. This power of quantitative data is limited by concerns over the study duration. We tried to keep the study sessions within two hours since participants are likely to feel fatigue for longer sessions. In addition, most people with visual impairments utilize local public transportation catering to people with disabilities, which are less frequent. After adding the required transportation time, a two-hour study session is practically a half-day activity for the participant. An experiment design involving a control condition was considered. However, given the likely small number of participants, we can only use a within subject design and counter balance the conditions. After running a few pilot sessions, it was clear that a within subject design could lead to study sessions lasting four hours or longer.

5.4.2 Study Procedure

Each study session includes ten steps and takes about two hours:

Step 1: Introduction

At the beginning of a study session, a researcher introduces the study motivation and agenda to the participant. We make an effort to communicate that both positive and negative feedback are

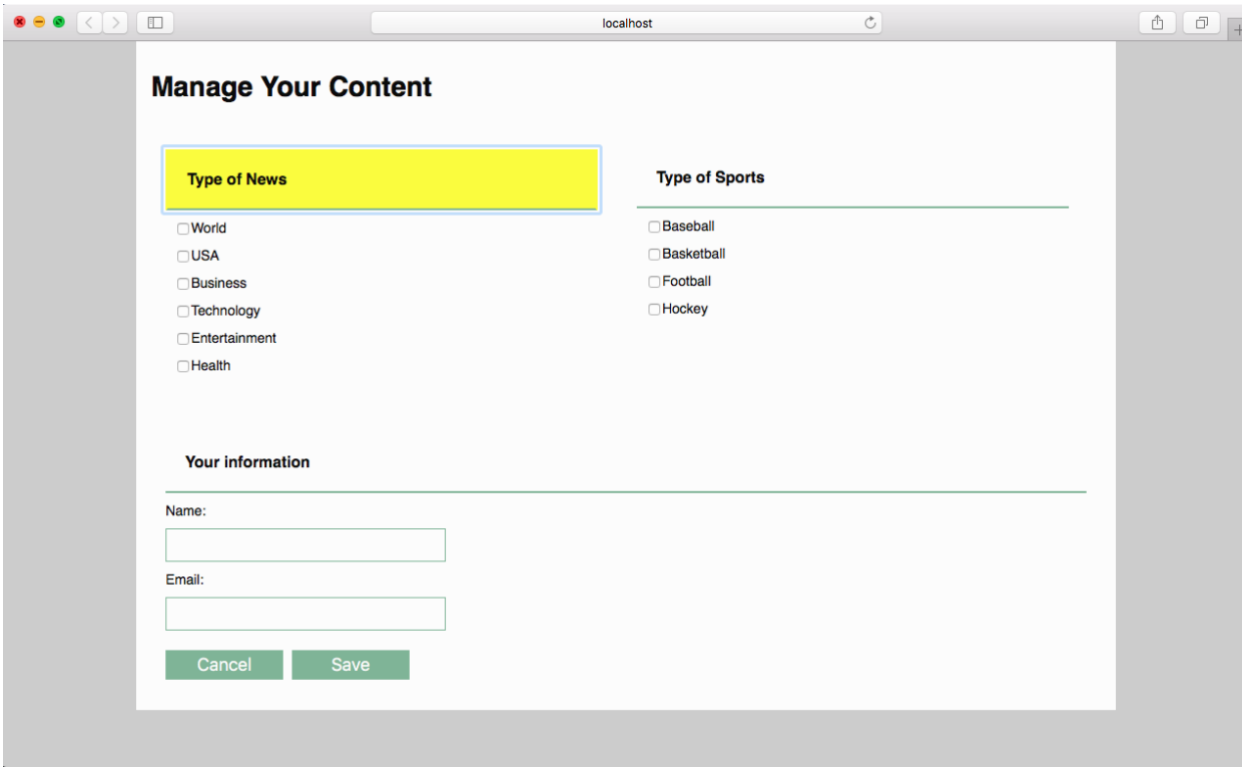


Figure 5.4 Screen Reader Prototype Training Page

equally useful and important for the research. The participant is encouraged to share anything coming to her mind. Once the participant gives the consent to proceed, the researcher conducts a survey that gathers demographic information, visual and hearing conditions, and experiences with computers.

Step 2: Screen Reader Training

Next, the researcher provides training on the screen reader prototype. At this step, the spatial audio features are disabled and the training only includes non-spatial audio features. The training is facilitated with a simple web page (Figure 5.4). The web page simulates a newsletter content setting page. It is composed of only web elements supported by the screen reader. The participant is guided to try each feature one by one. Afterwards, the participant is asked to explore the page

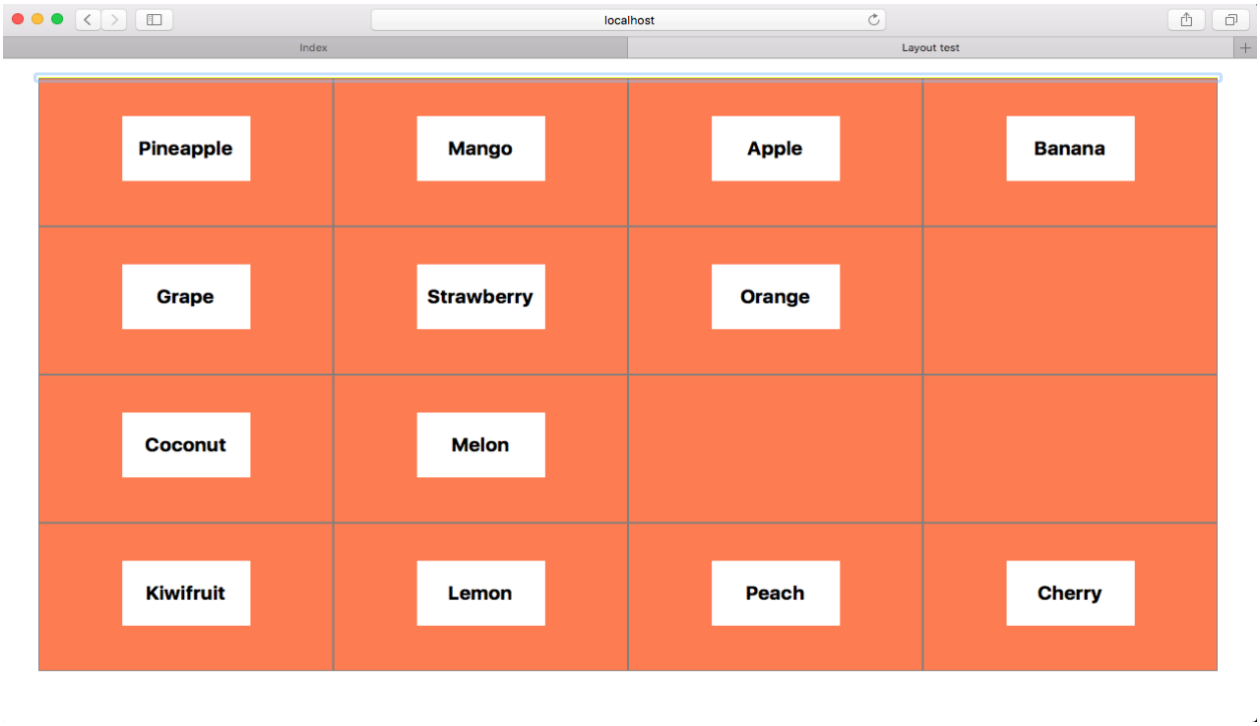


Figure 5.5 Web Page for Layout Test

freely and practice. This step lasts until the participant indicates that she is ready to move on to the next stage.

Step 3: Layout Testing

In this step, the researcher first provides a brief summary of the text analysis study reported in Chapter Three, which helps set the context of the following tests. The participant is encouraged to ask questions and share her experience of dealing with spatial terms during web browsing. Next, the researcher presents a simple web page that will be used to facilitate the layout testing (Figure 5.5). This web page features only 13 buttons, organized in a four-by-four grid (i.e., three cells are empty). Then the researcher asks five layout related questions about the web page. The five questions are derived from the text analysis study findings. They test the concepts of one web element next to or below another web element, grid layout, and corners. The participant is free to

use the computer during the process and actively explore the answers. After the participant answers a question, the researcher will also ask whether or not she can answer the question if using her regular screen readers and how.

Once the participant finishes all five questions, she is presented with a five-item questionnaire. The questionnaire surveys the participant's experience of using the screen reader prototype to complete the given tasks. The responses are collected using a 7-point Likert scale. The five items are:

Q1: This screen reader provides features sufficient for the tasks.

Q2: This screen reader gives too much information.

Q3: I felt stressed when using this screen reader.

Q4: I felt lost on the web page.

Q5: I have clear ideas about what is on the web page.

The goal of this step is manifold. It helps participants' learning by giving them time to practice. It also helps focus the interview on spatial concepts using regular screen readers. To make sure their answers were not limited by the prototype's features, we also asked participants how they would find the answers to the layout related questions if using a screen reader that they are mostly familiar with. The data collected from this interview provides a portrait of the current practice of how screen reader users deal with spatial concepts.

Step 4: Spatial Audio Feature Training

Next, the researcher provides training on the additional spatial audio features. First, the participant is guided to familiarize herself with the spatial audio used. This step is facilitated with a simple

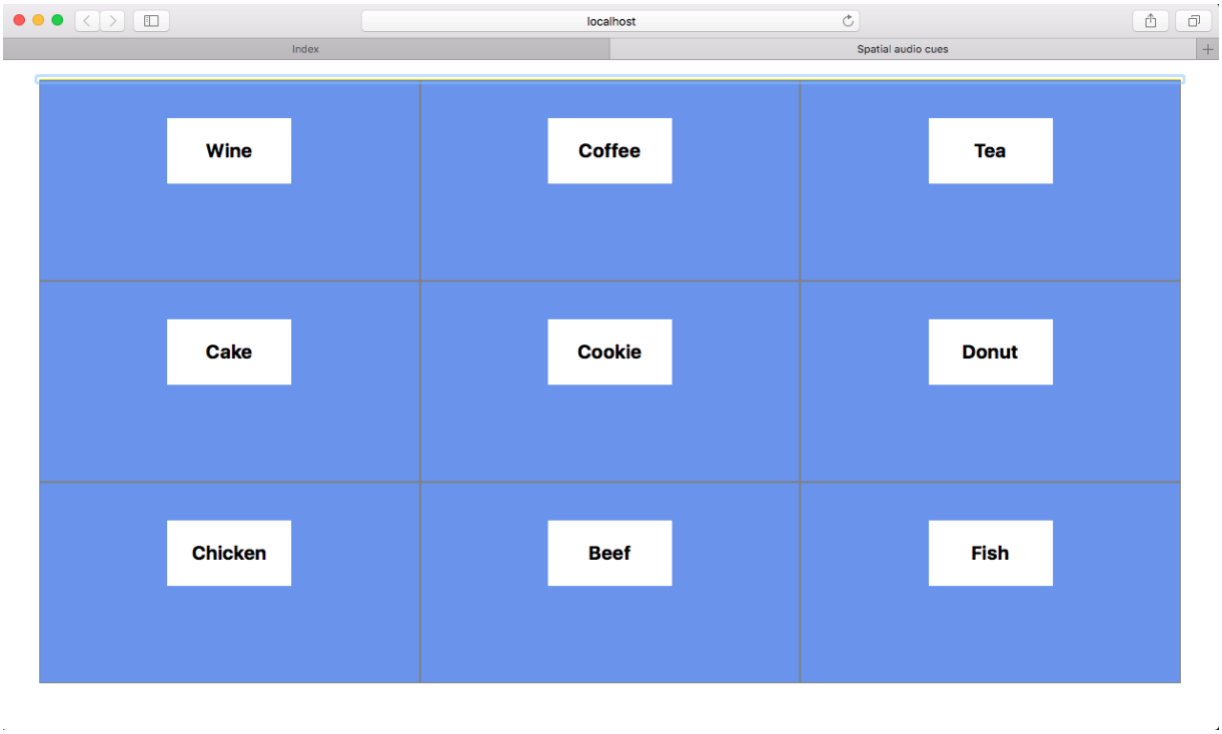


Figure 5.6 Web Page for Spatial Audio Training

web page comprising only nine buttons organized in a three-by-three grid (Figure 5.6). The researcher demonstrates the audio cues from the simpler ones to the more complex ones. Once the participant can recognize the audio cues with confidence, the researcher introduces the additional supporting screen reader features. The participant is encouraged to practice on this page, which is perfectly suitable for the directional navigation techniques due to its grid layout. Next, the participant is presented with the same training page as in Step 2 again, except she hears spatial audio feedback with every move and can use the additional features now. This page is more complex as its web elements do not align perfectly. The participant is asked to explore this page freely and practice new features. This step lasts until the participant indicates that she is ready to move on to the next stage.

Step 5: Layout Testing with Spatial Audio Features Enabled

In this step, the participant is presented with the same five layout questions from Step 3. This time, the participant is asked to find out the answers utilizing the spatial audio features. When necessary, the researcher probes her approach during the process. After completing the questions, the participant answers the same five-item questionnaire. The participant is instructed to respond based on her experience from this task, i.e., answering layout questions using the spatial audio enabled screen reader prototype.

Step 6: Break

The participant takes a break of about ten minutes.

Step 7: Mental Representation Task

After the break, the participant is presented with a web page that simulates an online food order service (Figure 5.7). The web page only features web elements supported by the screen reader. It includes horizontal layouts and vertical layouts. In this task, the participant is asked to explore the web page freely using the spatial audio enabled screen reader to gain a general idea of its content and organization. Then she is asked to illustrate the layout perceived visually using customized magnetic pieces and a whiteboard. The pieces represent different types of web elements and have different patterns on the front side. The patterns help the participant recall what a piece represents if needed. However, she can also simply ask the researcher to assist in identifying a piece's meaning. This prevents the task from becoming a test of one's memory. For the same reason, the participant is told that she does not need to memorize the content during the exploration phase as she can use the computer when working on the board to gather more detailed information. The researcher observes and asks questions during the process when there is any notable behavior. The

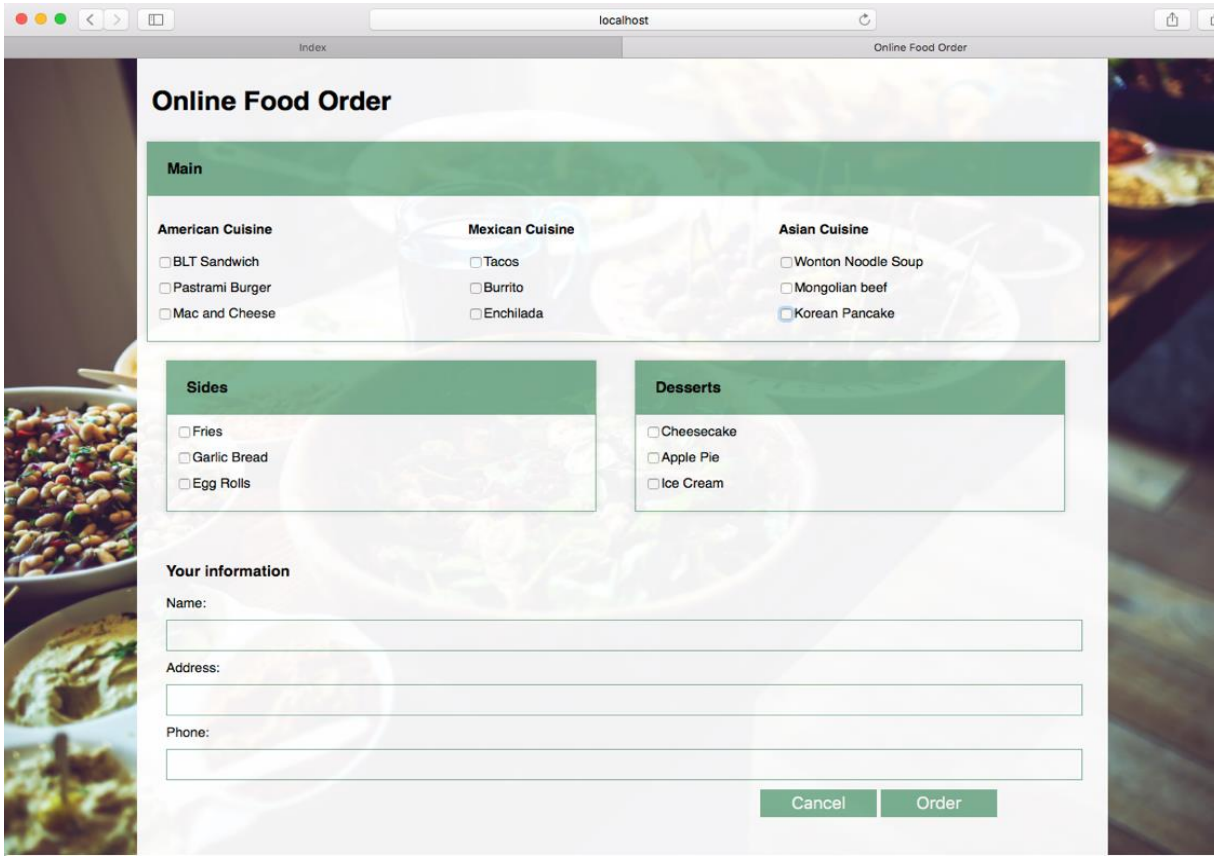


Figure 5.7 Web Page for Mental Representation Task

participant can ask for help if she cannot proceed due to some critical confusion. In this case, the researcher will ask what the confusion is, why it is critical to her process, and any other questions relevant before providing hints.

After the task, the participant answers the five-item questionnaire again. She is instructed to answer based on her experience from this task.

Step 8: Usability Questionnaire

Next, the participant completes a System Usability Scale (SUS) questionnaire [20]. SUS is a widely used usability scale in the field of Human Computer Interaction. It comprises ten questions.

Responders answer each question using a 5-point Likert scale. The calculation process generates a number score between zero and 100. The higher the score, the better the usability.

Step 9: Semi-structured Interview

Finally, the researcher conducts a semi-structured interview. First, the researcher goes through the five-item questionnaire responses and asks for explanations on notable answers, such as when a certain item differs among iterations. Then, the researcher asks the participant to share her feedback around four topics: screen reader spatial audio features, spatial audio cues themselves, challenges related to spatial concepts when browsing online resources, and potential other uses of spatial audio in assistive technology interfaces.

Step 10: Conclusion

The researcher thanks the participant and presents the incentive.

5.4.3 Apparatus and Study Environment

The screen reader prototype and all web pages used for the user study are hosted on an Apple Macbook Pro laptop (Model: MD313LL/A). The computer runs a local hosting server. The Google Chrome browser is used during the study.

Since most screen reader users are familiar with JAWS, which was developed for Windows operating systems, participants are presented with a standard, full-sized PC keyboard. The screen reader shortcut keys are mapped accordingly so that participants do not need to deal with Mac-specific keys.

In a typical user study session, the researcher and the participant sit face to face (Figure 5.8). The computer is placed in front of the researcher and the keyboard is placed in front of the participant.

Figure 5.8 Study Setup

Therefore, the participant cannot see the computer screen even if they have any residual vision. If a study session is conducted at the participant's preferred location (e.g., their homes) and there is space constraint, a similar set up, such as sitting on the side, is used to ensure that the participant has enough space to work but cannot see the computer screen.

The participant is provided a set of over-ear headphones. The headphones are wired to make sure the connection is good. They have a frequency response range of 20Hz to 20kHz.

When performing the mental presentation task, the whiteboard is placed next to the participant so that the participant can reach both the keyboard and the board comfortably.

5.4.4 Data Collection

We collected four types of data during the user study.

- Questionnaire data: in each study session, the participant answers the five-item questionnaire three times. The participant also completes one SUS questionnaire [20] at the end.

- Interview data: this includes the semi-structured interviews at the end of each user study session, as well as conversations during the session when we followed up with any interesting observations. We audio-recorded most of the user study sessions (after they gave verbal consents). The recordings are transcribed selectively.

- Video recording: we had a camera set up to capture the board when the participant is completing the mental representation task. The participant's face was not captured due to privacy concerns. The videos are used to review the process of creating the mental representation.

- Keystroke data: all keyboard interactions with web pages are logged using a script. The data is used to review what steps a participant takes to achieve a goal.

All data was stored in a secure computer with password. We removed all identifiable information from the data and notes. We used a random code generator to create a code for each participant. The codes were used in managing the dataset.

5.4.5 Data Analysis

Due to the design of the study, the application of statistical methods is limited. We mainly used descriptive statistic methods on the quantitative data (questionnaires, keystroke) to gain an accurate understanding of participants' usage patterns.

We employed an inductive and iterative approach to analyze the qualitative data. The focus was to learn their feedback on using spatial audio feedback in screen readers, spatial audio cue design, challenges related to spatial concepts, and potential of spatial audio in other interfaces. We used the open-coding method from the grounded theory [46]. A software tool, ATLAS.ti [147] was used to support the coding process. The initial codes were often associated with specific problems or comment. The subsequent iterations grouped codes with similar nature into themes such as "recognition problem," or "learning curve."

5.5 Recruitment

We recruited participants via two venues: UCI Disability Services Center and Braille Institute Anaheim Center at Orange County. We used passive recruitment at UCI Disability Services Center. The staff forwarded the IRB approved material to UCI students with visual impairments. Students interested in participating in the research then contacted the researchers directly. Braille Institute Anaheim Center at Orange County is a non-profit organization dedicated to training people with visual impairments in various independent living skills, including using computers with screen readers. They do not have a similar mailing list. Their staffs facilitated the process by presenting the study to their students first and passing the contact information to the researchers if permission was acquired. We also used snow balling recruitment through existing contacts and current participants.

We offered $30 cash incentive to participants who completed the study. Participants also had the choice of coming to the UCI campus or having the study conducted in a location convenient to them (mostly their homes). If they chose to come to the UCI campus, we also offered assistance in transportation.

5.6 Summary

This chapter provides details of the development and planning of our user study. The user study features a screen reader prototype with spatial audio features implemented using JavaScript. Specifically, the prototype uses stationary audio cues spatialized in a 2D audio space to convey the positions of web elements on a web page. The prototype also plays moving audio cues when users navigates from one web element to another. We designed an interview-based user study that aims to gather feedback from blind users on the spatial audio designs implemented and the broader implication of using spatial audio interfaces to support their use of technologies. We will report the user study results in the next chapter.

Chapter 6 Exploring the Potential of Spatial Audio with Screen Reader Users: Results

This chapter reports the results from the user study. Three general questions drive the interpretation of the data. First, does spatial audio feedback help screen reader users interpret spatial terms, such as those identified from the study reported in Chapter Three? Second, can screen reader users perceive a web page's overall layout based on spatial audio feedback? Third, how do screen reader users feel about the spatial audio feedback regarding our prototype specifically and in general? The first and second questions are explored around the tasks that participants performed during the study, whereas the last question draws from interview and observation data. We also discuss design implications for spatial audio interfaces.

6.1 Participant Information

20 participants took part in the user study. Below is a summary of their backgrounds:

- Eight participants were female and twelve participants were male.
- The average age was 42.5 (median: 40.5, SD: 14). The minimum and maximum ages were 24 and 67 respectively.
- 12 participants identified as Hispanic, five participants identified as Caucasian, one participant identified as Asian, one participant identified as Pacific Islander, and one participant identified as African American.

- Nine participants were congenitally blind, five participants were adventitiously blind, five participants were legally blind, and one participant was low vision.
- 13 participants reported normal hearing, five participants reported better than normal hearing, and two participants reported less than normal (but sufficient) hearing.
- The average self-rated computer experience was 5.35 (median: 5.5, SD: 1.46) on a 7-point scale (7 is the best).
- The average self-rated screen reader experience was 5.4 (median: 6, SD: 1.64) on a 7-point scale (7 is the best).
- The average years of using screen readers was 16.7 years (median: 16.5, SD: 10.63). The minimum was 1 year and the maximum was 36 years.
- 15 participants used computers to access the web primarily, five participants used their phones to access the web primarily.
- Microsoft Internet Explorer was the primary browser for 11 participants. The other nine participants listed Google Chrome, Firefox, or Safari as their primary browsers. Each of these three browsers was reported by three participants.
- JAWS was the primary screen reader used by 13 participants. VoiceOver was the primary screen reader for four participants. Three other screen readers were used by one participant each. They were Natural Reader, NVDA, and ZoomText (this is an assistive technology for people with low vision. It was used by one participant with low vision, whose secondary choice is JAWS).
- Seven participants were unemployed. Six participants were students. Four participants worked as Assistive Technology consultants. Three participants worked as administrators.

6.2 Prototype Usability Overview

To check whether or not the usability of the study prototype itself caused any problems to the participants, we conducted the System Usability Scale questionnaire (SUS) [20] after each participant completed the last activity using the spatial audio enabled prototype, i.e., the white board activity. SUS is widely used in the field of Human Computer Interaction to measure an interface's usability. Because it generates a numeric score in the range between zero and 100, it allows comparison between interfaces.

Based on responses from all 20 participants, our prototype scored 77.6. Bangor et al. have reported an empirical evaluation of SUS [5] in order to establish a usability benchmark. They gathered 2,324 usages of SUS from 206 studies over nearly a decade. The average score is 70.14 with a median of 75. Comparing our score to the benchmark, we can conclude that our prototype has decent, average usability. Therefore, we do not expect the prototype's usability to have unusual impact to participants' responses.

6.3 Spatial Term Interpretation

To understand whether or not spatial audio feedback helps screen reader users interpret common spatial terms, we look into participants' performance and comments from the five layout-related questions (Chapter 5.4.2, Step 3&5). We presented the questions to participants twice, first when they just finished the training of the prototype without spatial audio features, then again after they received the training for the spatial audio features. If they expressed difficulties in finding the answers, we would follow up and ask clarifying questions, such as how they would find the answers using other screen readers and what problems they experienced. We compare the two sets

Table 6.1 Layout Question Performance

	To the right	**Below**	**Grid**	**N-Row**	**Corners**
Without spatial audio feedback	Maybe	Maybe	No	No	No
With spatial audio feedback	Yes	Yes	Most	Most	Most

of data to gauge the effect of the spatial audio feedback (Table 6.1). We reproduce the questions along with the results below. The results provide portraits of participants' current practice and their interactions with the spatial audio features.

6.3.1 Without Spatial Audio Feedback

In summary, with regular screen readers, i.e., no spatial audio feedback, it is generally hard for participants to make any spatial assessment beyond the immediate context.

Right or below: Find the Web Element to the Right of / Below the Current Web Element

The problems of interpreting these two spatial terms are rooted in the linear nature of the mental models (more on this in 6.4.1) that users construct based on the screen reader's audio output. The challenge is that, though users are confident that the previous or the next element is close to the

current element, they cannot be sure of the spatial positions of these web elements. Technically, a web designer could customize tab orders of web elements on a web page, which would take precedence over the default left to right, top to bottom order that screen readers follow. However, none of our participants shared any experience of encountering web pages that employ unconventional ordering.

When asked about their strategies to find content to the right or below the current element, participants listed a few common answers. All these answers involve ways to find the end of a line. The most frequently mentioned strategy is to use arrow keys. This refers to screen reader's read commands. When reading text content, users can switch among reading by characters, by words, by lines (ends at the end of the line), and by sentences (ends at the end of the sentence, which could span less than one line or multiple lines). In JAWS, users can do so using arrow keys alone or combined arrow keys with other keys. If reading by lines, the screen reader will stop when reaching the end of the line. It can give clues of the order of the words. However, this feature is only available when reading text. It is not available during navigation. Participants with more experience are aware of this limitation, whereas the less experienced users did not seem to understand the difference.

Some screen reader features can indirectly convey some layout information. One example is JAWS Virtual Viewer feature. When invoked, this feature copies the content from the current window into a virtual buffer. Then, users can read the text using arrow keys. This feature provides a way for users to access text embedded in an interface that is otherwise inaccessible. However, this feature can only read the text content. Though not intended for conveying layout, if virtualized text is on the same line, a user can derive that the original web elements might be at the same

horizontal level too. However, this feature is limited at conveying layout, and the user will have to go back to the regular interface to interact with the web elements.

Finally, alternative output interfaces can also offer some hints about a web page's layout. P16 shared that she could find where a line breaks when using a braille display with the screen reader. Braille displays can output braille characters using a row of refreshable electro-mechanical pins. As the braille display has limited width, a line of content will require a few refreshes before reaching the end of a line. When the braille characters stop in the middle of the display, she knows that she has reached the end of the line on the web page.

Grid: Find the Dimensions of the Grid

The web page has 13 buttons organized in four rows. The first and the last row have four buttons, whereas the second row has three buttons and the third row has two buttons. We asked participants to determine the number of columns and rows on the web page.

Participants generally said there were few hints for them to perceive such a layout on their own. All participants were aware that screen readers would read the layout information of a table, i.e., number of rows and columns, as well as shortcut keys for navigating among cells. However, if a structure is not labeled as a table, they would not be able to learn the layout. When being pushed further, some participants said they might guess the layout of a grid based on their past experiences. For example, if there were 12 items (instead of 13 used in the study), they could imagine the items organized in three by four or four by three formation. However, this is a complete guess.

N-Row: Find the Three Buttons that are Organized in a Row on this Web Page

We asked participants to find which row had three buttons. Multiple items in a row is a spatial concept related to the grid. Because being able to perceive that there are multiple rows could lead to the recognition of a grid that is composed of these rows.

Participants said that they could learn about multiple web elements of the same type positioned next to each other, i.e., forming a group. They start with the first web element and continue navigating to the next item until encountering a different kind of web element. By counting, they could gather how many of such web elements are grouped together. But since there is no easy way for them to know when the line breaks, it is hard to know how many elements are in the same row. One strategy, raised by P20, is to analyze the similarities among the web elements and derive the possibility of them being in the same row.

Corners: What Buttons are Located at the Top Right / Bottom Left Corners?

Most participants knew the basic features of jumping to the start of a web page or the end of a web page. Since screen readers read from left to right, top to bottom by default, it is understandable that one would assume that the first item is located at the top left corner and the last item is located at the bottom right corner. However, some more experienced participants pointed out that the first or the last web element on a page was not necessarily at the top left or bottom right corner.

In contrast, the other two corners, i.e., top right and bottom left, are tricky to get to. One common answer is based on the false belief that the user can move to the end of the line during navigation using arrow keys, e.g., jumping to the start of the page and then moving to the right using the Right Arrow key. However, as mentioned earlier arrow key-based directional movement only works for reading text. It is not available for navigation.

Another common strategy is to move to the start or the end of the page and then tab forward or backward a few steps. Assuming the web page starts at the top left corner and ends at the bottom right corner, a few steps after or before them are approximately the top right and bottom left corners. This strategy, of course, relies much on guess work. P16 said she would try two or three steps since "*there is not much space between top left and top right*".

One participant also shared his experience of using frames to find corners before the use of frames fell out of fashion. At that time, web pages often employed multiple frames as containers for different sections of the web page. For example, there might be a header frame, a left menu frame, and a main content frame. According to him, screen readers used to use the position to identify each frame. Therefore, it was easy to find the "top right frame". However, the practice of using frames has largely stopped now.

6.3.2 With Spatial Audio Feedback

Some participants demonstrated insufficient understanding of the spatial audio features when performing the activity's second iteration. The activity followed the training session of spatial audio features. We started the activity when participants indicated that they were ready to move on. However, during the activities some participants did not seem to remember certain features or did not realize what a feature supported. This is probably due to a lack of goal-oriented, active learning during the training. In addition, some participants performed the activity using arrow keys. This behavior does not necessarily indicate problems with using the spatial audio features. Participants might have chosen to use arrow keys because they were familiar with similar features in existing screen readers. These factors should be taken into consideration if the results are referenced elsewhere.

Right: Find the Web Element to the Right of the Current Web Element

All participants answered this question correctly:

- Five of them simply used the Right Arrow key and did not seem to pay attention to the accompanying spatial audio cues.
- The other 15 participants found the answer by navigating to the next element. They made either explicit or implicit comments suggesting that they reached the conclusions based on the spatial audio cues.

Below: Find the Web Element Below the Current Web Element

All participants answered this question correctly:

- 13 participants found the answer using the Down Arrow key. They either did not pay attention to the accompanying spatial audio cues or could not be sure about the answer based on spatial audio cues and turned to the Down Arrow key to confirm.
- The other seven participants found the answer by navigating to the following elements and counting the element's position, i.e., by listening to the audio cue indicating line change and counting to the corresponding position.

Grid: Find the Dimensions of the Grid

Most participants answered this question correctly. Participants often navigated through all buttons and counted the buttons in each row. When they finished, they often had answers for both this question and the next question regarding the N-Row.

- Six participants found the answer using arrow keys, i.e., by moving towards a direction and counting until reaching the end.
- Ten participants found the answer by navigating forward and paying attention to audio cues that indicate changing lines. Some of them also tried using arrow keys. But their interaction with the web page showed that they could recognize the grid formation based on spatial audio cues without any additional information.
- Four participants could not find the answer.

N-Row: Find the Three Buttons that are Organized in a Row on this Web Page

Most participants answered this question correctly.

- Seven participants used arrow keys to find the answer.
- Eight participants found the answer based on spatial audio cues.
- The same four participants who could not recognize the grid formation also could not answer this question. One additional participant, who answered the grid question based on audio cues, could not find the answer to this question due to fatigue.

Corners: What Buttons are Located at the Top Right / Bottom Left Corners?

Most participants answered this question correctly.

- Ten participants moved to the corresponding cells using arrow keys.
- Five participants found the answers by navigating linearly on the web page and counting the position based on knowledge learned about the grid from earlier questions.
- The same five participants who could not answer the N-Row question also could not answer this question.

6.4 Overall Web Page Layout

The first task deals with interpreting common spatial terms in simplified scenarios. We used the second task to evaluate whether or not participants could perceive the web page's overall layout. Specifically, we asked participants to explore a provided web page freely and then illustrate their perceived layout of the web page using whiteboard and some custom-made tactile puzzle-like pieces (Chapter 5.4.2, Step 7). This activity is more complex and the web page used is more realistic.

As described in the previous chapter, we could only perform this task once due to time constraint. Therefore, there is no direct comparison of using a screen reader with- or without- spatial audio feedback. However, when conducting the first task, participants shared much information on how they dealt with web page layouts, which offers insights on how they perceive the overall layout. In short, they organize content on a web page in a linear manner. This is consistent with prior work [1,87].

6.4.1 Linear Mental Model Without Spatial Audio Feedback

Prior research has reported that screen reader users construct mental models of web pages that resemble lists. Murphy et. al conducted interviews of 30 blind and partially sighted computer users to understand their navigation strategies using screen readers and their perceptions of web page layouts [87]. They found that screen reader user's mental model of a web page was best described as a vertical list of points and links. The linear representation was attributed to the linear nature of screen reader output. Abidin et al. also investigated screen reader user's mental model of web pages [1]. Based on diagrammatic representations of perceived web page layout collected from ten screen reader users, they found that most of their users organized web page information in a single-

column format, i.e., a vertical list. Some experienced users adopted two-dimensional format to reflect the logical grouping of the content.

Our data confirms these findings. The linear nature of screen reader user's mental model is most evident when participants searched for the item to the right or below the current item. They often navigate to the next item and gave it as the answer. If the next item is considered to be to the right of the current item, it suggests that the user's mental model is a horizontal list. In contrast, if the next item is considered to be below the current item, it suggests that the user's mental model is a vertical list. In our dataset, both orientations were equally preferred. In either case, the participant's reasoning was often informed by their understanding of the technologies or other similar medium. For example, P14 explained that the next item could be to the right or lower, but not higher because screen readers read from left to right, top to bottom. P18 drew experience from how print media is designed. She explained that she was moving to the right because we read from left to right in daily life. She understood that the content would break at some point to the next line. But she said that there was no way to know when it did. For this reason, she could not know for sure what was below a given web element.

Whereas the horizontal list model seems to reflect web page layout more accurately, the vertical list model is somewhat being enforced by how screen readers function. P7 shared that he often used the Down Arrow key to explore a page when visiting it for the first time because pressing Down Arrow key would read the content on a web page line by line. P20 also mentioned the JAWS features that presented a particular element in a list and allowed the user to traverse the list using the Up Arrow and Down Arrow keys. This also conveniently avoids line break as each piece of information is a single line on the vertical list. The action of pressing the Down Arrow and the Up Arrow keys may have prompted users to organize all the information vertically in their mind. Even

though they know that the web page is not necessarily organized in such a manner, the vertical list models align with how they interact with the web page using screen readers. Therefore, it is an effective model to work with.

In most cases, participants were aware of the limitation of this linear mental model. They acknowledged that they were mainly speculating if one item was below or to the right of the current item when answering the respective questions. They emphasized that they did know the previous or next in the order read by screen readers, but not where these elements were on a web page, because the screen reader "*doesn't tell you*" (P14).

In absence of spatial layout information, some participants find alternative grouping schemes based on logical relationships. P2 is an inexperienced screen reader user. When asked to find the item to the right of the current item, she struggled to come up with an answer. When probed, she guessed that maybe the content (in this case, all fruit names) was organized alphabetically. P16 is a very experienced screen reader user. When exploring the page, she also considered the possibility that all fruit names were organized alphabetically. However, in her case, after going through all content on the page, she concluded that the content was not organized in such a way.

Another useful alternative grouping strategy is to derive relationships based on headings. This strategy has been reported by other researchers [13]. In summary, headings are often used on web pages to outline information hierarchy. For example, an article's title might use a heading tag whereas the content of the article uses a regular style. Another example is applying decreasing heading levels to an article's title and subtitle to indicate their relationship. Such usage naturally implies the logical structure among information. Screen reader developers have designed features

to take advantage of this design pattern. As a result, fast navigation using headings is one of the most effective navigation strategies that screen reader users use.

The use of this strategy is best demonstrated by P2 and P4. Both are inexperienced screen reader users. They had trouble interpreting the spatial audio feedback. Consequently, the results of their board activities only show how they perceived the web page based on headings (Figure 6.6, top two). They positioned headings either horizontally or vertically. Then they listed items after the heading in a subgroup closer to the heading. This practice is not new. However, this exemplifies heading-based navigation as a powerful technique that is easy to learn and effective in organizing information.

One interesting interpretation of headings comes from P15, who is an inexperienced screen reader user. When he explored the web page during the whiteboard activity, he was surprised to learn that the page title was heading level 1 (<h1>) and the category title right after it was heading level 3 (<h3>). He seemed to think that skipping heading level 2 (<h2>) was unexpected. It might just be insufficient knowledge about HTML since he was also surprised late that there was multiple heading level 4 (<h4>) elements.

6.4.2 Perceived Web Page Layout with Spatial Audio Feedback

Data collected from the whiteboard activity reflected more truly on participants' performance and interaction. By this point, participants had gone through repeated training of the screen reader prototype and had used the screen reader to resolve layout related questions multiple times. Though some still needed reminders occasionally, in general they had shown enough experience to understand strategies and generate meaningful preferences.

15 participants produced reasonably accurate visual representations of the web page used in the study on the whiteboard, including one participant who could only finish the core part of the web page due to time limitation. Two participants produced linearly oriented visual representations of the web page. Three participants had severe problems in figuring out the web page layout. At the end, they were asked to construct their perceived layout of the headings from the web page only.

On average, the participants spent 25 minutes and 36 seconds to construct the mental representation of the perceived web page layout using the whiteboard. The median time is 23 minutes and 7 seconds. The longest session was 41 minutes and 42 seconds, whereas the shortest session was 14 minutes and 50 seconds (Figure 6.1).

We analyzed the participants' keystroke logs to see if there was any obvious navigation pattern. All navigation features that are available in the prototype were used. On average, participants used 218 keystrokes to complete the board activity. The minimum is 59 keystrokes, the maximum is 564 keystrokes, and the median is 167. If we only count navigation keys, i.e., excluding keys used to control the speech or audio cues, participants used 160 keystrokes on average. The minimum is 59, the maximum is 338, and the median is 145.

When we looked into what keys were used (Figure 6.2), we noticed that among the 15 participants who finished the whiteboard activities, six of them did not use arrow keys at all during their sessions. Another six participants used primarily arrow keys (50% or more). The remaining three participants used a mix of arrow keys and other navigation keys. The presentations that participants

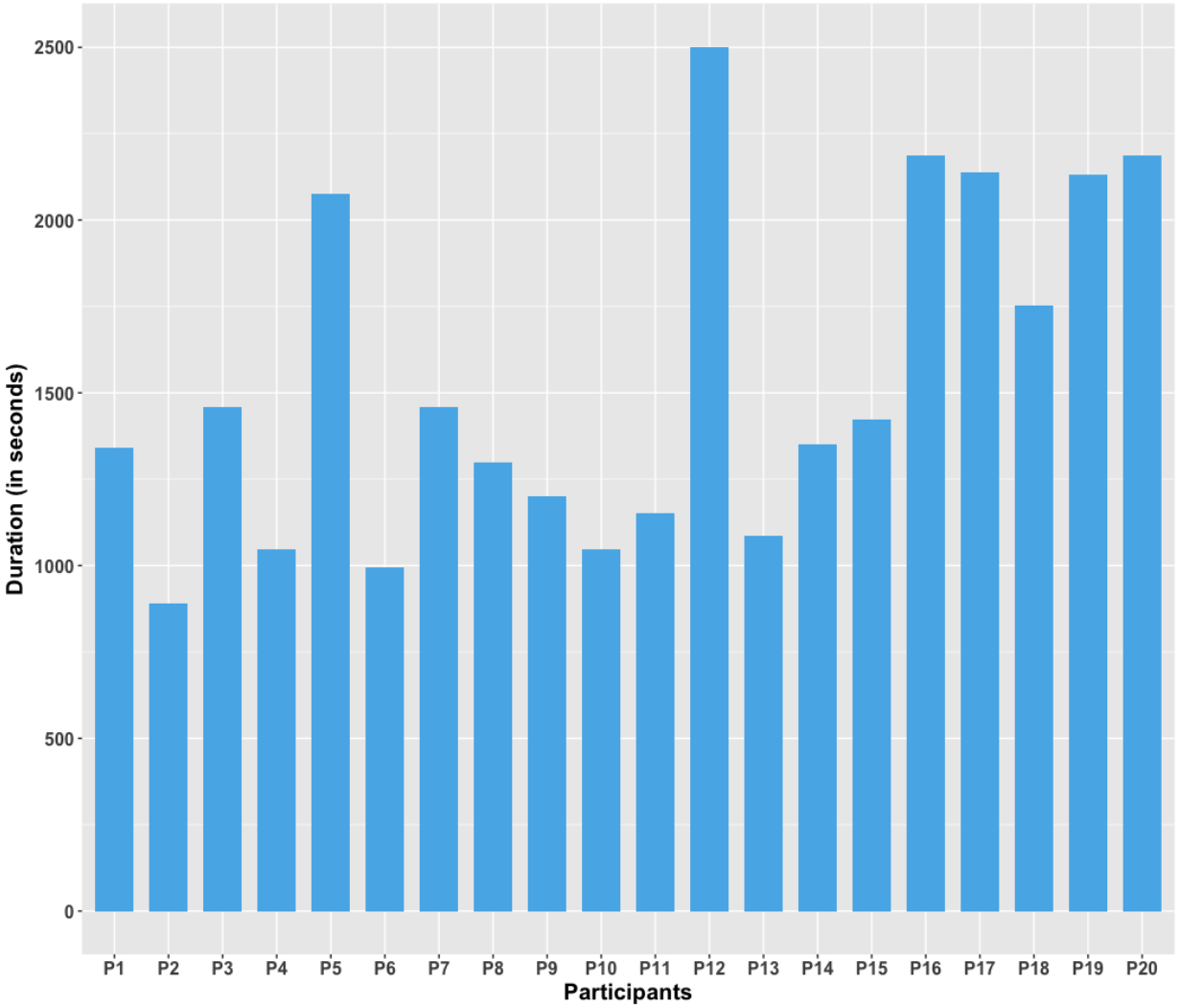


Figure 6.1 Time Spent Completing the Whiteboard Task by Participants

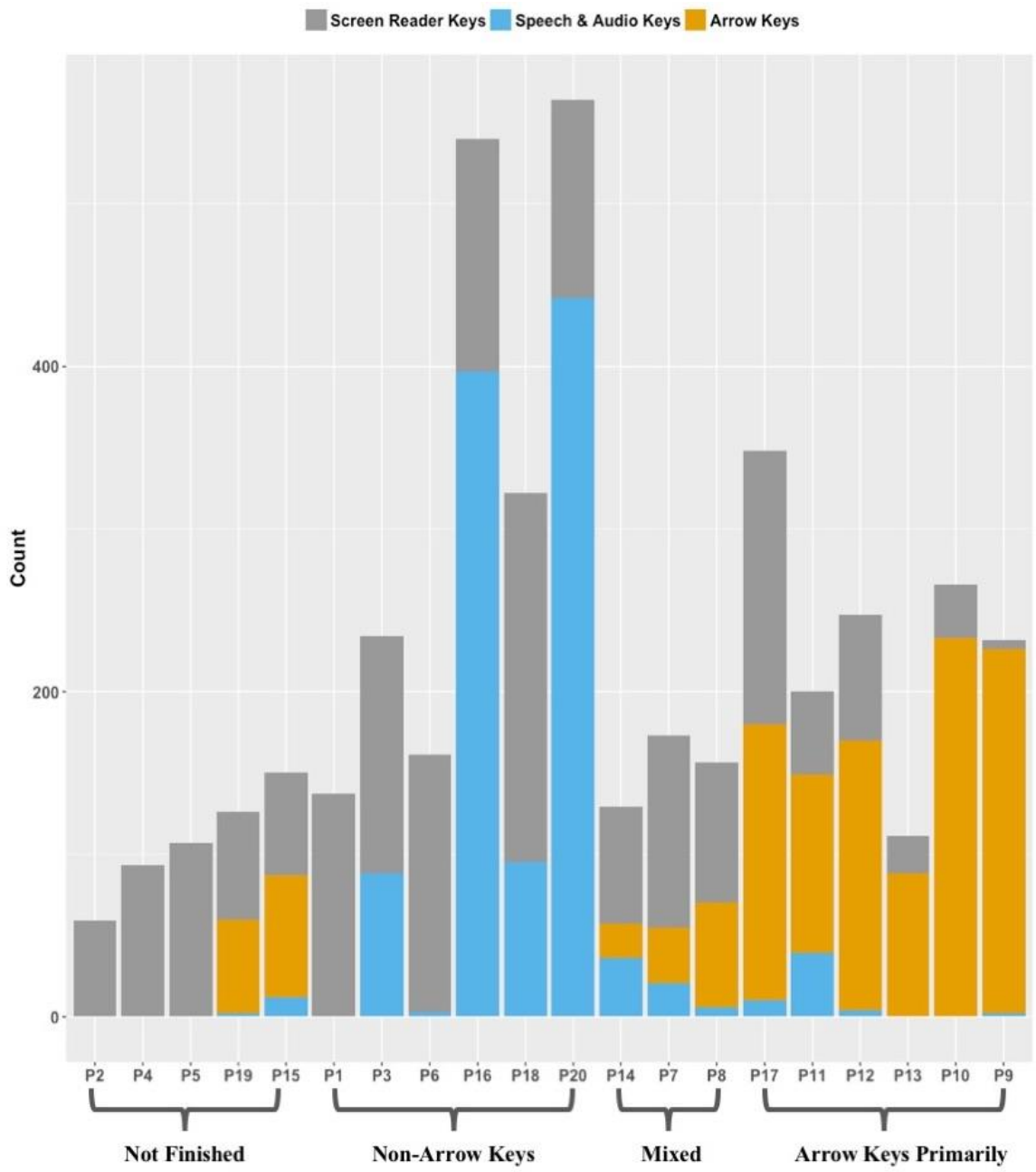


Figure 6.2 Keys Used by Participants

have produced on the whiteboard do not have clear differentiating visual attributes among these three groups (Figure 6.3-6.5). Moving by arrow keys can give the user a sense of direction even without spatial audio feedback. Therefore, we looked into these three groups separately and explore whether or not they employed different strategies to figure out the web page layouts.

No arrow key: for the six participants in this group, the use of heading navigation (using H to go to the next or previous heading) and exhaustive navigation (using N to go to next or previous web element) were the main navigation strategy. These keys account for 87% keystrokes used on average (minimum is 71% and maximum is 98%). Four participants used them more than 90% of the time. They frequently used the feature of replaying the spatial audio cue for the current web element. On average, this feature was used 166 times (minimum is 0, maximum is 426, and the median is 88). Though the user would always hear spatial audio cues when navigating from one element to the next, this particular feature is standalone and it is a clear indication of the use of spatial audio cues.

Based on the keystroke logs, P1 and P6, who differed from the other four participants in that they did not use the spatial audio feature mentioned earlier (P6 used the Control key to mute the audio a few times), explored the page using typical screen reader navigation strategies: going through all headings first, then going through other elements one by one. The perceived layouts represented on their finished whiteboards were fairly accurate. Since they did not use the spatial audio features to invoke additional spatial audio feedback, their sense of positions was mostly based on the spatial audio cues provided during the navigation. When asked, they both said that they mainly utilized the stationary audio cue at the end of the speech output. P3, P16, P18, and P20 used similar typical screen reader navigation techniques. In addition, they used the feature comparing single audio cues

from neighboring elements frequently when they were not sure about the element's relative position.

Primarily arrow key: for the six participants in this group, the use of heading navigation and exhaustive navigation only account for 19% of the overall keystrokes (minimum is 0% and maximum is 31%). In contrast to the group who did not use arrow keys, this group of participants did not use speech and audio related features very much. On average, they only used it 6 times (minimum is 0, maximum is 33, and the median is 0).

The typical strategy used can be described as active directional navigation. They started exploring the web page using fast or exhaustive navigation features. Then, once they had some rough ideas of the web page, they used arrow keys to derive the relative positions among neighboring elements. For example, they would move to the left, then to the right to figure out what neighboring web elements were; or they would move towards one direction until reaching the end. Though not apparent from the keystrokes used, they commented that they did pay attention to the audio cues. Their finished whiteboards showed some detailed spatial features, such as slight indentations of the checkboxes under a heading, which would not be possible without spatial audio feedback. However, their actions suggested that the audio cues were most likely used as confirmations of the expected movements.

This active navigation style worked for some, but not others. For example, P13 were able to figure out the position of each element and memorize the whole page layout. He finished the whiteboard quickly and produced a very good visual representation of the web page. In contrast, P12 and P17 appeared to have a hard time grasping the content's structures on the web page. During the whiteboard activities, they only found a linear vertical list of items and missed most elements

further right on the web page. Since web elements on this page are not perfectly aligned, moving using arrow keys sometimes skips content between elements or is blocked. This might have created much confusion when they tried to imagine the layouts. Ultimately, they finished their whiteboard activities, but only after the researcher pointed out what mistakes they had made.

Mixed keys: for the remaining three participants who used a mix of arrow keys and other navigation keys, the use of heading navigation and exhaustive navigation accounted for 41% of their total keystrokes (minimum is 33% and maximum is 45%). They used speech and audio features moderately. On average, they used these features 16 times (minimum is 0, maximum is 30, the median is 18). Their navigation strategies were similar to the group primarily using arrow keys. They started exploring the web page using typical screen reader fast navigation features. But they seemed to quickly recognize that the non-heading web elements were grouped under respective headings as vertical lists. For example, P8 tried using the Up Arrow and the Down Arrow keys to explore the text boxes under the heading "Your Information" first. Then she tried the same with check boxes under a few other food categories. She declared that she had a good general idea about the web page after just trying a few food categories. Later, when working on the board, she utilized the spatial audio cue replay feature when there were confusions.

P7 and P14 also used a similar navigation strategy that was not used by other participants. When working on the whiteboard, they would navigate by Checkboxes and pay attention to what they heard. When they noticed that they had moved to a different category of information, they would go backward to find the nearest heading. For example, the web page features one group of Mexican food followed by a group of Asian food. When the participant noticed that they had just heard a food item belonging to Asian food category, they would go backward and find the nearest heading from this point. There is no conceivable connection between this technique and the spatial audio

features. However, it does demonstrate how experienced screen reader users utilize various hints to help them understand a web page.

Unfinished: for the five participants who could not perceive the spatial layout of the web page and produce satisfactory visual representations (Figure 6.6), three only used navigation features available in normal screen readers, i.e., navigation based on a particular web element, exhaustive navigation, etc. Two participants used regular navigation keys and arrow keys evenly. They used the feature to replay current web element a few times but did not use the feature playing current audio cue at all. Three of them reported only a few years' experience using screen readers. Another participant reported 18 years' experience using screen readers. However, he explained that he counted from the time when he started using assistive technologies. He is also the oldest participant. The self-rated computer experience of these four participants were the lowest in the dataset. The fifth participant rated his experience with computer and screen reader higher and reported 15 years' using screen readers. However, we suspect his experience might be misrepresented since he was recruited from an introduction-level screen reader training class.

In summary, the board activities showed that our participants who were experienced screen readers could derive the main visual layout of the web page using the screen reader with spatial audio feedback regardless what strategies they used. However, there are some minor errors in the details, such as the precise location of a web element. There is no significant difference among the perceived layouts acquired using different navigation strategies.

Figure 6.3 Perceived Web Page Layout by Participants Who Primarily Used Non-Arrow Keys

Figure 6.4 Perceived Web Page Layout by Participants Who Primarily Used Arrow Keys

Figure 6.5 Perceived Web Page Layout by Participants Who Used Mixed Keys

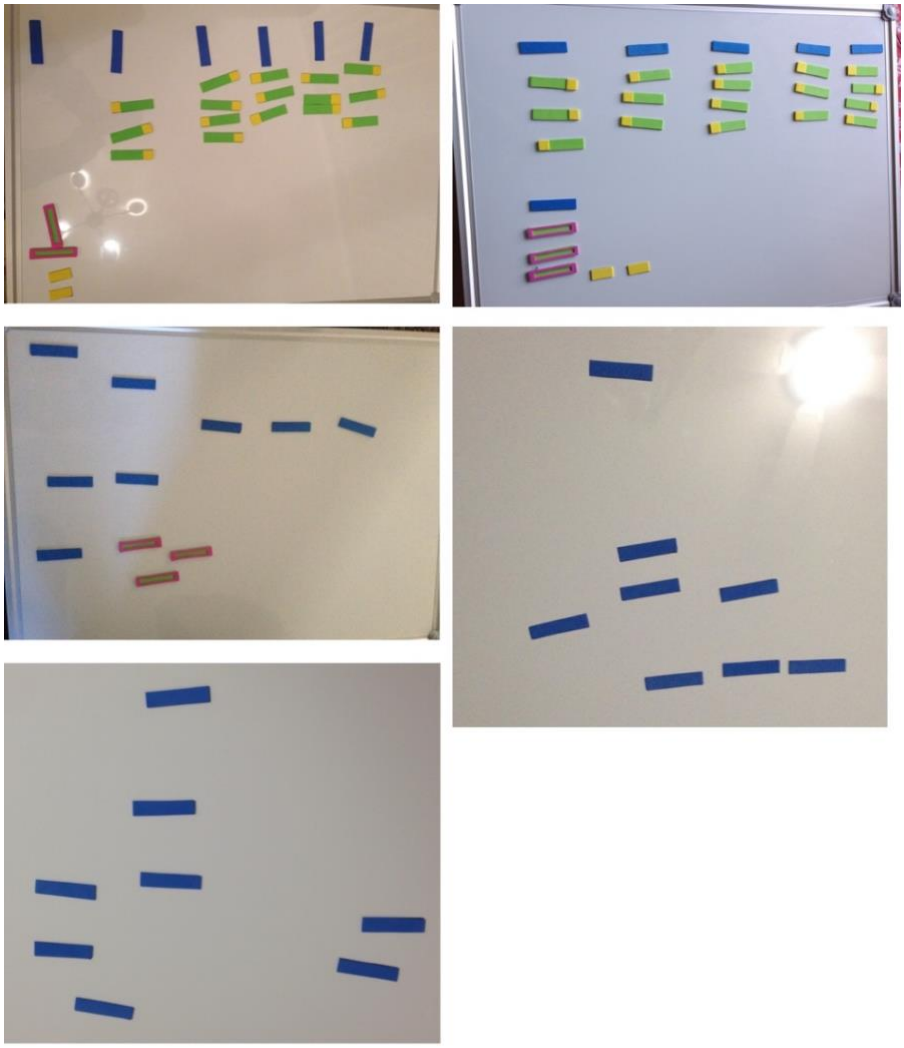


Figure 6.6 Perceived Web Page Layout by Participants Who Could Not Finish Successfully

6.5 Questionnaire Responses

In addition to SUS, we also took the same five-item questionnaire three times: once after the first round of layout task with basic screen reader, once after completing the layout task with spatial audio feedback, and one after completing the whiteboard activity. The questions are:

Q1: This screen reader provides features sufficient for the tasks.

Q2: This screen reader gives too much information.

Q3: I felt stressed when using this screen reader.

Q4: I felt lost on the web page.

Q5: I have clear ideas about what is on the web page.

The order of the three questionnaires was not counter-balanced. Therefore, we cannot perform meaningful statistical comparison among them. However, the questionnaire answers provide a narrative on how participants felt about the spatial audio features. We presented the answers of participants who finished the whiteboard activity and those of participants who did not produce satisfactory whiteboard representations separately.

For the 15 participants who produced good visual representations of the web page in the whiteboard activities, their answers showed that they were very comfortable using the new spatial audio features (Figure 6.7). When they were first introduced to the prototype and its basic screen reader features, they felt that the prototype was not sufficient for interpreting layout related questions (Q1). Once they learned the spatial audio feedback and its related features, they responded positively to the prototype's support of finding layout-related information and perceiving a web page's overall layout. Their answers to Q2 and Q3 remained about the same for

all three questionnaires, i.e., they did not consider the additional spatial audio feedback as overwhelming, they also did not have any significant changes on their perceived stress level. Q4 asked whether or not they felt lost on the web page. The answers suggested that they felt less lost once the spatial audio features were made available. Answers to Q5 also confirmed that they had better ideas of what was on the web page once the spatial audio features were made available.

For the five participants who did not produce satisfactory visual representations of the web page in the whiteboard activities, their struggles were reflected in their questionnaire answers (Figure 6.8). After learning the spatial audio features, they felt that the prototype supported the layout-related tasks a little better (Q1). However, the increase of their scores was much smaller and ambiguous compared to the other group. The participants did feel that the additional spatial audio feedback was a little more than they could handle (Q2). Maybe the additional information was overwhelming, they also felt more stressed once the spatial audio feedback was provided (Q3). In terms of task performance, they felt more lost when using the prototype with spatial audio features enabled (Q4) and they were less clear about what the web pages had (Q5).

6.6 Feedback on Spatial Audio Features

The qualitative data collected offers much insight into the potential of using spatial audio feedback in screen readers. Throughout the sessions, we encouraged participants to freely share any problems that they had or their reactions to the tasks. We did not require them to think-aloud as that might interrupt their workflow given all information was audio based. But if we observed certain interactions that were not self-explanatory, we asked participants to clarify when they reached a natural break. In the semi-structured interview, we also asked participants for their direct feedback regarding the spatial audio features, as well as their preferences and challenges

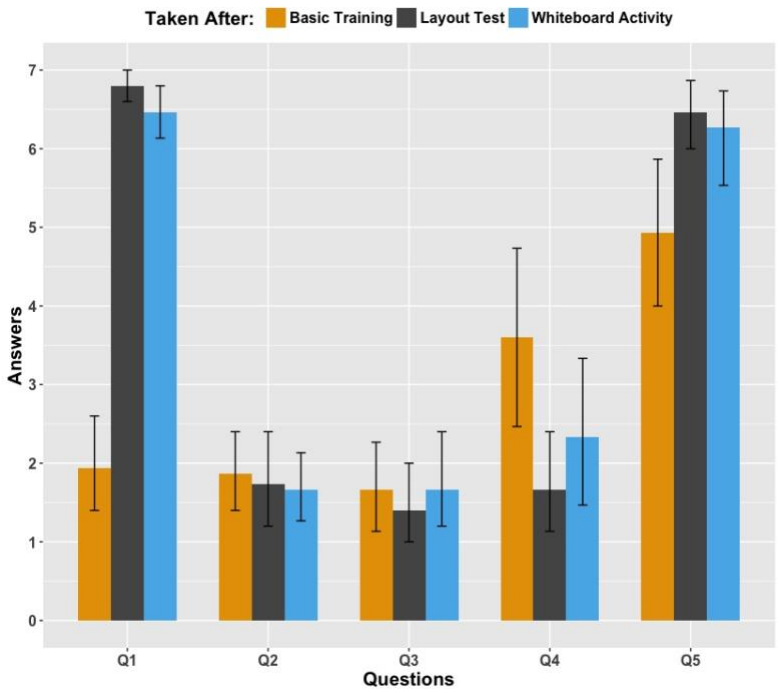


Figure 6.7 Five-Item Questionnaire Responses from Participants Who Completed the Whiteboard Task Successfully

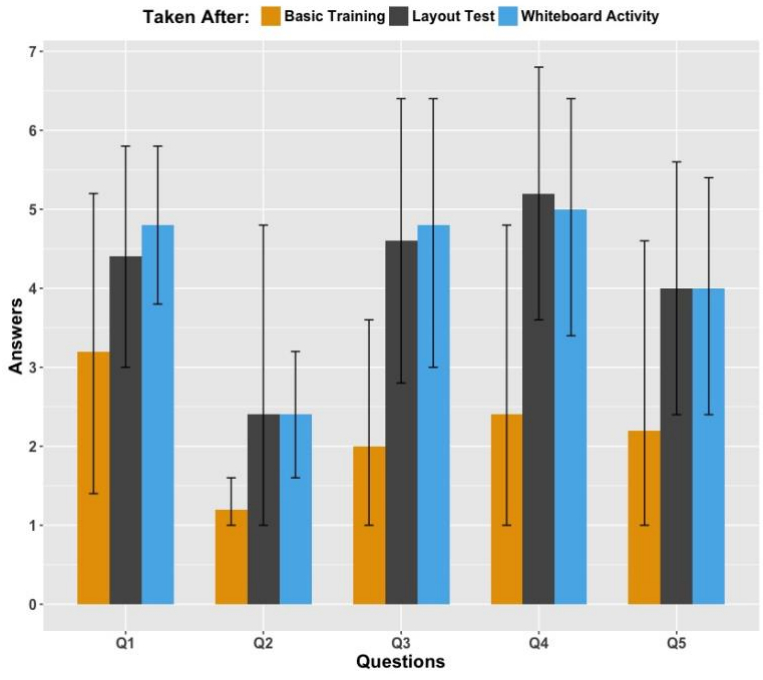


Figure 6.8 Five-Item Questionnaire Responses from Participants Who Could Not Complete the Whiteboard Task

experienced. Here we organize the feedback to four themes: general layout challenges, reaction and preference of spatial audio feedback, audio cue design, other possible applications.

6.6.1 General Layout Challenges

Participants shared some layout related challenges that they had experienced in their daily lives. Two main challenges identified concern interacting with sighted people and learning new web pages.

Interpreting spatial terms used in instructions provided by sighted people is a commonly shared frustration among the participants. This echoes what we have found from the text analysis study reported in Chapter Three. For example, P17 said:

> *If you are talking to a sighted person, they will ask you to go, you know, they would say, "the link you are looking for is on the top right". Well…, but where is the top right?*

> *They will say, well, it's in the middle or on the top right. I'll say that doesn't have any bearing to me, 'cause I don't know where that's at.*

This is not just a problem when talking to people who do not normally interact with people with visual impairments. Participants also shared stories when seeking help from Customer Support, who should have some training on interacting with customers with disabilities.

Even if a user has some vision, interpreting spatial terms is still problematic. P18 is legally blind. He said he could figure out the layout information if he had to. However, he also mentioned that it was not easy.

> *Because I have partial vision, sometimes I can manage. The issue is that, it's a lot strain on my part. And sometimes, depending on like what kinda colors or designs the programmers use, it might be hard for me to see it, like, 'cause it might be small print, the color might not have enough contrast for me. So it's pretty difficult, like, I really don't like doing that way.*

On the other hand, screen reader users also wish to be able to use spatial terms. This is useful when a screen reader user needs to direct a sighted people to a certain part of the web page, for example, when seeking help from Customer Support on inaccessible web elements or collaborating with sighted co-workers. Some screen readers, such as JAWS, track the navigation flow using invisible pointers. The invisible pointer can be different from the visible cursor or mouse pointer position on the screen. In addition, screen readers do not need to move the part of the web page that contains the currently focused element to the browser's viewport (the visible area of a web page), i.e., a screen reader user can read text from a web page that is not visible on the screen. This can cause confusion when a screen reader user tries to show something to a sighted person, as the object of interest might not be obvious to the sighted person. In this case, it would be useful if the screen reader user can perceive some layout information about the web element and describe it using spatial terms, which are natural for sighted users to interpret. P7 shared his experience:

I always have my colleague helping me out with pages… that I tell her, "the buttons are not talking. Can you tell me what're there." She's trying to figure out where I am on the screen. So, she has to look at the screen for a minute to see where my cursor is and where I am talking about, to be able to tell me "oh, you are on this side." So, if I can tell her right away where I am at…

The second challenge is rooted in the importance of layout when reading a document. When sighted people reads a web page, its layout implicitly suggests the relationships among content. One underlying reason is the principles of grouping (or Gestalt laws of grouping) [23]. For screen reader users, since the visual clue is not available, if the web designer does not present the same structural information in other perceivable forms, screen reader users are often left in a state of confusion or rely much on guesswork. This problem is best illustrated by an example from P18:

Basically, I don't know if things are grouped together. So, let's say, "important information" or whatever, and then I am listening the things that come right after it. Sometimes I won't be sure if that is within that group or they belong to another group.

Sometimes, visually they look together. So, if you're just using your vision, it's easy to tell what belongs to what; but sometimes they are kinda separated on the page for some reason, at least as far as like HTML code concerned. So, I will go like, "iPhone", and underneath that, it will say, like, give you a little summary, it'll give you like a "buy" link, and then I will go to the next thing, and it'll just tell me the next phone. Then, I will be like, "wait, where is the

price". And, so I have to keep scrolling down, and then the next column over will be the price. And then… that makes it pretty much impossible… unless I will just count: OK, this is three phones down; this is three prices down; that matches. But sometimes that does happen where, again, things just aren't group together sometimes, or VoiceOver doesn't read it one right after another for some reason. So that makes it… then from there, you basically start having to, like… what I'd have to do, is I'll have to keep reading the whole page, like I'll have to read it through until in my brain I sorta figure out the pattern, where the layout is. And then from there, I'll have to like go, basically, finish whatever I am doing.

In addition, participants believed that it was important to know a web page's layout in order to navigate effectively. Some participants said that they tried to construct a mental image of what they think the web page looked like when learning a new web page. Then they could use this mental image much like a "map" to guide their navigation. A few participants even went so far as to actually construct tangible models. P7, who is an assistive technology instructor, said he also gave students tactile models when training them to use touch-based devices, such as iPhone. These responses show similarities with the practice of using physical models of an environment to teach visually impaired people navigate the environment during Orientation & Mobility trainings.

6.6.2 Reactions and Preferences of Spatial Audio Feedback

Is Spatial Audio Feedback Useful?

In general, participants responded positively to the spatial audio feedback. When we followed up, participants' responses indicated that spatial audio feedback could address both layout-related challenges reported in the previous section.

First, receiving spatial audio feedback allows them to communicate more effectively with sighted people when spatial terms are involved. P20 shared his reactions:

> *It definitely adds important elements to your experience, specifically focusing on where buttons are physically. That way you can tell a sighted person, "hey, can you click on that unlabeled button on the bottom left corner of the screen on the web page?" 'cause I can finally give directions to a sighted person. 'cause, you know, sighted person, they don't know. So, the interaction between sighted person and visually impaired person, the way they explain it is very confusing. So, we have to explain to them. "Okay, can you go to the bottom… point your mouse literally at the bottom right corner, the last one at the bottom right corner of the screen." So that way, you know, 'cause maybe a label. That's the developer's fault. But that's where this program becomes so powerful, to me. You know. You know what I mean?*

P16 also shared an example:

> *Because sometimes people… they will have like a headshot or whatever graphic in the middle of the page. So being able to navigate from one… I don't know, not only from graphic to graphic, as you can currently do, but also know where those graphics are in relation to one or another, that'd be really helpful so that when people say, oh, find the picture of this thing, x picture in the middle of the page, and then from there move to the right to go to wherever you have to go, you could find the picture more easily. And if the picture is labelled with alt-text, then even better because you can find it and verify with the label and what the spatial looks like*

In both examples, participants implied that they did not need to know the detailed layout of a web page. They only referred to prominent positions or landmark web elements. This is similar to blind people locating clues and landmarks in physical environments for effective navigation.

Second, having the additional spatial audio feedback helped maintain orientation, i.e., "*pinpointing where exactly in the screen are you"* (P13). Participants stated that being able to create a mental image of the web page helped them stay oriented when browsing. P8 says:

> *When I do orientation and mobility with someone and learn a new route, if I can make a mental map of the route in my head, it helps me to do it the next time. I think mental mapping of web pages could be really important, especially with the complicated web page. It kinda give you an idea.*

Participants mentioned that knowing the placement of at least the main web elements could guide them in devising effective navigation plans. These views were not contingent upon a participant's vision condition since both late blind participants and congenital blind participants mentioned similar points. But they all considered themselves "visual people," referring to their abilities to visualize structures in their minds. P6 drew parallels between reading web pages and reading magazines. He believed that knowing the layout would help users find out what navigation path was relevant given the current position.

> *It might help if you actually knew what the page… let's say you lost your eyesight, but you have an idea what a web page looks like, so you have an understanding where they are. If not, you can relate to them is like, to say they look like magazines. If they remember what a magazine looks like, then they might be able to do the left, right, top, bottom.*
>
> *…*
>
> *I believe that's very important for blind people. because we all visualize things. If you visualize things on the screen, you can close your eyes and kinda see it, and kinda know where you are.*

Orientation can also be understood from a cognitive perspective. Being aware of the current web element's physical position on a web page can help users process information. P1 gave an example:

P1: I think so (it is useful), especially if you are gonna do the website pages like that. Because you don't know where you are at. It's very helpful. I will use it… you know.

Interviewer: why would that be useful? Why do you want to know where you are on the page?

P1: because it doesn't describe exactly what you are doing. It just reads. A regular screen reader just reads. And this's gonna help me indicate where I am reading.

Interviewer: but why do you want to know where you are reading?

P1: to understand the article I am reading. Because if it's towards the end, and it's just finishing touch to the article. Then I know. Versus, OK, it's just the middle of a new story, or, I know. It's more direct to what I'm doing. And it gives me a more visual picture of the whole thing. So… I would definitely use it.

Maintaining orientation on a web page can also help users deal with accessibility incidents. Inaccessible web elements can displace the screen reader focus to unexpected positions on the web page or reset the focus to the beginning of the page. Knowing the location of problematic web elements can help users avoid those areas or get back to the previous position after the accessibility incident. P3 says:

I think it'd help, especially if there're pictures on the web page. So, if you want to try to get away from looking at those pictures, because you know, there is no descriptive text. You can avoid them by knowing where they are on the page. So, it might help.

…

Might be interesting, you know how the Mac has a trackpad. If there is some way to interact with the trackpad so that if you know where the position of what you want is, you can tap on the trackpad based on where you think the thing you are looking for is.

It should be noted that the positive attitude towards spatial audio feedback is not fully accounted by these two reasons. Some were enthusiastic about the feature but could not articulate what exact benefits that they have gained. The benefit could be emotional or increased engagement in the long run. For example, P8 says:

(Screen readers) they've taken the position: as long as they (users) get to them (content) it's ok, they (users) don't have to know. I think it's kinda nice to have a feel for where things are on the page, just like sighted person would.

Her response seems to root at a sense of equal access to information. However, our study did not include appropriate metric to measure such possible qualities.

We also tried to understand why some participants did not think spatial audio feedback was useful. Only one experienced participant, P9, stated explicitly that the spatial audio feedback was not useful. He is congenitally blind and a lifelong screen reader user. He felt that the spatial audio was not useful to him because he could find out what was on a web page using existing screen reader features. During the study session, he showed a strong inclination to use techniques that he was familiar with, such as arrow key-based navigation. However, when asked how he dealt with spatial terms, he said that he simply ignored such information, as screen readers would not be able to convey them. P14 is also congenitally blind and a lifelong screen reader user. But P14 offered a rebuttal to P9's opinion. He says:

> *They are useful… I may not want them on all the time. But if somebody was going to describe the screen to me and tell me that I needed to be on a particular part of the screen, I need them. As a computer user with JAWS, I get around the screen pretty well, and I don't usually have to use spatial concepts. But if somebody is giving you coordinate related to spatial concepts, you have to have them. Then I'd need to have them. And with JAWS, if you move Down Arrow, that isn't gonna necessarily move you down in the row.*

Unsurprisingly, participants who did not finish the whiteboard task successfully tended to be less enthusiastic about the spatial audio feedback. However, they did not directly comment on the usefulness of spatial audio feedback. The main problem given was often information overflow, i.e., they expressed confusion associated with not being able to handle so much information at once.

Preferences Between Stationary Audio Cues and Moving Audio Cues

To understand how spatial audio feedback is used, if a participant made a layout related comment when completing an activity, when there was a chance, we would ask her to explain how she derived the spatial information. During the semi-structured interviews, we also asked participants directly their preferences between the stationary audio cues and moving audio cues.

Based on the information collected, participants strongly preferred stationary audio cues. The main reason was that they believed that stationary audio cues communicated richer information about web page layouts. P20 explained:

> *I think, again, the minute, as you could see here at the headings, the reason why I listen to it is because I am looking for the minute, the final result of that beep, which was actually more accurate in my opinion. Because when you're moving it, it's not as accurate, you just went from "beep, beep, beep, beep," but you don't actually know specifically where it's gonna end up. I mean, yeah, it's gonna be on your left or far right, or it could be minute. So, I will say accuracy is… the last beep is the act of the final accurate coordinate... corresponding point. So, I used it like, like a graph in my mind. Okay. So, you know, when you find the final point, that last beep tells me, okay, it's minutely to the left, or it's dead center, dead right. 'cause when you're moving around, like with anything, you are not gonna be accurate because it's moving too much. You know, it's … I get the idea of it and I think it's great that you know, okay, it's gonna be somewhere on the left, could be minute, could be far left. So that's why I go to for that last beep. That's why I determine my final…you*

know. I used the Shift Tab, Shift Right Window and Right Window key a lot.

That's very cool shortcut keys. I like those.

This is, of course, not a surprise. Stationary audio cues are designed to indicate both a web element's horizontal and vertical positions on the web page. In contrast, moving audio cues are designed to portrait only the direction of movement. What we were not sure about during the design process was whether or not participants would find stationary audio cues practical, as we anticipated that the spatial audio localization would be challenging. During the study sessions, participants indeed had many problems when trying to extract information encoded in stationary audio cues. In fact, when the researcher observed that a participant was struggling and followed up, the participant would often explain that she was having difficulty localizing a stationary audio cue. Some participants utilized the stationary audio cue comparison feature extensively in order to determine the subtle differences between two audio cues. (More on audio recognition in 6.6.3.)

Interestingly, we observed, despite the apparent challenge of localizing stationary audio cues, participants seemed to have strong interest in utilizing stationary audio cues even though some information could be easily derived based on moving audio cues. One possible explanation is simply a participant's curiosity. However, we observed it from multiple participants consistently. So, another possible explanation is that users are willing to invest more effort in techniques believed to be more useful in the long run.

Furthermore, we noticed that participants were very attentive to details. For example, many participants added indentation to checkbox after a heading when working on the whiteboard activity. Due to styling, the checkboxes do not always have the same horizontal positions in pixel values as their preceding headings when being rendered. The values would then result in small

differences in the stationary audio cues' horizontal positions. The difference was minor and we did not expect that it would be noticeable to participants. However, the fact that many participants added indentations showed that they did notice the differences and it had an impact on their perceived web page layouts.

Nonetheless, it should be noted that participants did not always recognize the differences correctly, i.e., they realized that there were differences, but were not able to tell what exactly the differences were. There were other ways to resolve the confusion. One can listen to the moving audio cues when moving from the heading to the first checkbox, which would be accompanied by an audio cue with only vertical changes. A participant can also move using arrow keys to confirm whether or not the web elements were aligned vertically. For participants who have residual vision and some knowledge of common visual layouts, they may also realize that it is odd to apply different styles to content sharing the same structure on the same web page. With all these options available, it is therefore interesting to see our participants adhered to one method. It adds to the previous observation that there is a certain "stickiness" associated with using the stationary audio cues to reach the conclusion.

Another reason for preferring stationary audio cues is their short duration. Some participants pointed out that, with just a single beep, stationary audio cues delivered information quickly and to the point. Some also said that they would use the moving audio cues more if they were shorter. This may refer to two aspects of moving audio cues' design. First, stationary audio cues are 200 million seconds long and moving audio cues are 1.05 seconds. So, listening to stationary audio cues is much faster. Second, with its short duration, it is practical for users to replay the current stationary audio cue, especially because our prototype allows users to replay only the audio cue.

In contrast, to hear the moving audio cue again, a user has to go back to the previous web element and redo the movement. This adds significantly more time.

A small number of participants indicated their preferences for moving audio cues. The main reason provided was the ease of recognition. They acknowledged that moving audio cues did not communicate as much information as stationary audio cues. But they felt that there was enough layout information conveyed for the purpose of completing activities during the study sessions, which suggests that their conclusions might be contextualized in the activities at hand.

Some participants viewed both audio cues as equally important and having a complementary nature. P6 says:

> *I think they have to go together. 'Cause they complement each other. ... the point of this is to know where you are on the page. the beep tells you where you are. The movement, the other one, tells you how you move in the page. So, you need a combination of both.*

Moving audio cues, which occur before synthesized speech output, inform where navigation is going; stationary audio cues, which happen after synthesized speech output, confirm the movement. Participants felt that moving audio cues provided hints or prepared them for the upcoming, more specific spatial information encoded in the stationary audio cues.

6.6.3 Audio Cue Design

Usability

We identified three main usability issues with the prototype's spatial audio feedback designs: spatial audio cues recognition, cognitive overload, and feature customizations.

Recognition issues were present for both horizontal and vertical positions. Though participants reported that horizontal changes in the audio space were easier to recognize, the visual representations produced using whiteboards showed that there were still some challenges. This is best exemplified by the first two headings on the web page. These two headings occupy the full width and are aligned to the left. The stationary audio cues representing them have horizontal values in the middle. However, some participants incorrectly perceived them to be on the far left, whereas some perceived them to be on the far right. This is a little unexpected. From the Web Audio API evaluation study reported in Chapter Four, we have learned that users could localize horizontal positions accurately if the resolution was low enough, e.g., 3 to 5 regions. In these two cases, there was only one item on the same row. We expected that participants would perceive the position perfectly.

We suspect that horizontal recognition errors may be attributed to lack of training. In the Web Audio API evaluation study, participants spent a lot time on recognition tests. They had the chance to hear all possible horizontal audio cues. In contrast, participants in this study had only limited time to practice. So, they might not be able to learn or remember the full range of the horizontal audio space. Therefore, when they perceived a slight horizontal position offset from the middle, they had problems putting the difference in perspective. To make matters worse, as reported by other research and presented in our Web Audio API evaluation dataset, users often exhibit a little

localization bias, i.e., an audio positioned in the middle might sound a bit to the left or to the right for different people. While this subtle bias is likely to be a part of how the human auditory system functions, users could learn from experience over time how an audio cue's horizontal position corresponds to the web element's position on a web page and how to correct consistent bias.

Participants reported that vertical position recognition based on the audio cue's pitch was difficult. Prior research has shown that pitch could be naturally associated with vertical position [106]. During training, participants did quickly grasp the idea of web elements higher on the web page being associated with higher pitches. The training page features web elements in three rows, which made it easy to distinguish among high, regular, and low pitches. However, during the whiteboard activities, participants often had to distinguish audio cues of web elements in close proximity. In such cases, the pitch differences were often very subtle. This led to great challenges. Nonetheless, it should be noted that most participants correctly identified the vertical order of web elements, i.e., vertical position recognition based on pitch is challenging, but not necessarily prone to errors.

We had concerns when designing the study that it would be difficult for regular users who do not have music training to recognize pitches. It motivated us to provide a feature where the user can quickly replay and compare the stationary audio cues of the previous and current web elements. Unfortunately, lots of participants forgot about this feature during the study. For those who used the feature, they reported that it was very helpful in determining two web elements' relative positions.

The issue of cognitive overload is more related to moving audio cues. Participants reported that they understood how to make sense of the moving audio cue's two audio attributes, i.e., horizontal

change indicated horizontal movement and pitch change indicated vertical movement. However, during the study session, participants had much trouble actually utilizing moving audio cues.

One specific problem is the difficulty of paying attention to both attributes at the same time. It was common for participants to choose one audio attribute, in most cases the more intuitive horizontal changes, and focus on that when listening to the moving audio cues. In the case of an audio cue for vertical movement, i.e., featuring only pitch changes, participants would often notice that there was no horizontal position change but made no comment about the pitch change. Even for participants who had positive feedback about the moving audio cues, they reported using the horizontal direction change as a signal for line changes instead of the pitch change. For example:

> *Interviewer: if you only can keep either the moving audio or the single beep, which one would you keep?*
>
> *P14: I would say the moving audio.*
>
> *Interviewer: OK. Then you won't really know where exactly you are. You only know the direction of moving.*
>
> *P14: Correct.*
>
> *Interviewer: Why do you think that's more useful to you?*
>
> *P14: Because when they start going this way, you know you are going back the other direction. they are gonna be below what you just did. I mean, they are*

not gonna be, you know what I am saying, they are not gonna be on the same row.

P6 commented that he was expecting two audio cues, one for horizontal movement and one for vertical movement. This might be a viable alternative design: rather than having two audio effects playing concurrently, splitting the information into two non-overlapping phases within the same audio cue. Furthermore, a few participants also shared that keeping track of how many steps they had moved based on moving audio cues added additional strain to their memories. They suggested that it would be preferred if the screen reader could assist in counting in some way, such as reading the numbers out loud.

Another problem was processing and making sense of the audio cues, e.g., "*sometimes it's hard to remember all what they mean* (P8)". There were a few cases where a participant had problems figuring out layouts when the moving audio cue provided sufficient information. When the researcher replayed the audio cue and asked participants what they had heard, they would identify the audio cue's attributes correctly. However, they could not make correct connections between the moving audio cue and its implication in terms of layout. Participants explained that there were too many things going on at the same time. One had to pay attention and really focus in order to understand all the information. P16 says:

You have to be in a place where you don't have anything else distracting you. You have to just strictly use your hearing for this and only this. You can't be on the phone. If someone is talking to you, you have to tell them to be quiet for you to listen.

…

> *As useful as it might be, I wouldn't be able to focus on three different sounds, the spatial sound, the screen reader sound, and the people trying to explain where to go. So that'd be somewhere where I probably would not use it. In addition to navigating outside. Because… again, how much stuff can you hold in your head at some point.*

The cognitive overload issues impacted all users. For participants who could not finish their whiteboard layout activities successfully, the negative effect was severe. They often appeared very stressed when working on the tasks. Since these participants had less experience or even still struggled with using screen readers, it is understandable that additional audio cues would be perceived as noises and only impede their abilities to make sense of the screen reader's regular speech output. Other experienced participants were more at ease with the study activities. However, a few of them who are longtime screen reader users still mentioned that dealing with all the audio cues was a challenge.

It should also be noted that negative feedback was mostly raised when talking about the whiteboard layout activity. Participants rarely gave similar comments when talking about the spatial term interpretation tasks. We speculate that the simpler web page used in the spatial term interpretation task might have made it easier for participants. During the tasks, participants only needed to plan their navigation a few steps from a fixed anchor location, such as the current element or the beginning of a certain row. It is conceivably easier to track the focus of the screen reader in these

tasks. In addition, the questions provide some hints that the web elements were organized neatly, which made it easier to predict where the screen reader's focus was moving to.

The third usability issue was the lack of customization of features. One popular suggestion is to speed up the moving audio cues. Research has reported that screen reader users often turn up the reading speed so that they can get the information most quickly. Some participants offered the same reason for why they wanted customization. Another popular suggestion was to allow turning off one or both spatial audio cues when layout information was not necessary. This way users can focus on content or data. When the situation arises in which understanding the spatial relationship is necessary, users can switch on the spatial audio feedback relevant to the task. P1 gave an example:

> *I probably won't need it 100% if I am reading, say, an email. And I don't really need to know where I am at. Then I probably won't use it. But if I am looking for something and I need the detail, yeah, it'd be more useful.*
>
> *…*
>
> *Say I am doing a Word document. Sometimes like, say I am gonna sign it and, you know, "thank you" or whatever, I wanna be directly knowing exactly where I am at. That'd help me real quick, versus me counting the spaces stuff. Yeah, it'd help me direct myself faster.*

Quality

We asked participants to comment on the audio cues themselves. Most participants found the audio cues pleasant enough for the purposes. We chose basic tones to keep the prototype simple and easily reproducible. However, we were worried that participants would find the tones boring and unnatural. To our surprises, most said they were happy with them, even compared to possible alternatives such as using musical instrument notes or natural sounds. P7, who is congenitally blind, said he heard that some people would get headache from hearing synthesized tones like our audio cues, though he was fine with them and explained that it might have to do with him growing up with synthesized audio.

Some participants complained about the different pitches used in the vertical moving audio cue. The moving audio cue includes four tones. When it moves up or down, the four tones have different frequencies, which resulted in different pitches. The frequency differences between two neighboring tones are the same for all three pairs. However, some participants perceived the middle two tones having similar frequencies. In contrast, the difference between the first one and the last one was more distinguishable. An obvious solution is to increasing the overall frequency span and consequently the interval between neighboring tones. However, we have chosen the maximum range where we felt the two extreme frequencies were still comfortable. Increasing the range could lead to irritating extreme tones and might create other problems.

There are others less prevalent issues. Some participants reported that it was harder to recognize a stationary audio cue's horizontal position when the audio cue has a very high or a very low frequency. P18 also commented that the stationary audio cues for web elements at the bottom of the page had lower pitch, which sounded similar to the moving audio cues. It was hard for him to

distinguish. However, the moving audio cue and stationary audio cue were designed to be separated by the synthesized speech output exactly for the purpose of them not interfering with each other. P18 also shared that the positional audio used in Mac's VoiceOver, which is a feature similar to our stationary audio cues, was much easier to register in his mind when he heard it.

6.6.4 Other Possible Applications

At the end of the interview, we asked participants what other situations where they think providing spatial audio feedback would be useful.

The most popular answer is formatting, especially when visual formatting plays an important part of how a document and its author is perceived. P7 suggests:

> *Something like a resume. Something like a more complicated report. Not just like a regular text document. But something would have headings, footers. To help you… if you navigate by heading, you wanna know heading to the left, to the right. So I think, formatting for people who need to do reports, resumes, and research papers, I think it could be helpful for them.*

Resume often takes advantage of layouts to emphasize key information or simply fitting more information on the page. Other participants also shared that they needed to get sighted people's help to check resume formatting to make sure it appeared professional. Being able to receive spatial audio feedback can enable them check some basic styling issues, such as column alignment, themselves.

Participants also suggested that spatial audio feedback could help convey serial information, i.e., data that is sequential or stepwise in nature. For example, P10 described:

> *When I am buying sheets, back to the sheets example, I am making, you know, three decisions at once in the example we gave. In buying food, I am only making one decision at once, yes or no fries, yes or no apple pie, whatever. So, in those situations I don't care what is on the screen. I just wanna know the data. But if I am making multiple decisions, then it'd be good to know, you know, pick one from this column, one from that column.*

He also suggested that train schedule timetable could be accompanied by audio feedback spatialized in the horizontal plane according to the station's order in the journey. Similar designs can apply for any instructions that include multiple steps. The main benefit of having such audio feedback is to remind the user about her current position in case she skips any information by mistake, in which case she would hear the audio feedback moving in a larger step than expected. When used with continuous data, an audio cue can be placed somewhere between a starting position and an ending position to indicate the percentage. One example is to convey the status of downloading a file. Currently, screen readers speak the percentage in words, e.g., "50% completed." Spatial audio cues can communicate the information faster and more intuitively. P7 says:

> *The screen readers, when you're downloading something, they give a percentage, you know, "10 percent," "30 percent," "50 percent." But I noticed, for example, NVDA uses beeps to represent the progress. And it just felts like it's much faster. So I think… versus depending on how fast you have*

your screen reader, just hearing sound and not having to focus on so much what it's saying might work. So it's kinda a way for you kinda still listen, but I think it's faster than waiting for the synthesizer to finish talking because, especially if you're listening to progress, it gets behind. By the time it says 50 percent, it's probably 70. The beeps seem to stay more in sync with what's happening on the screen.

A surprising answer is to provide spatial audio feedback for spreadsheet software, such as Microsoft Office Excel. It is surprising because information on a spreadsheet is in tabular formats naturally. It is perfectly organized for a user to navigate using arrow keys. Since arrow keys have directional nature inherently, users can derive spatial information when using them. In addition, users also receive speech output announcing the current row and column numbers. Therefore, spatial audio feedback will not provide any new information. When asked, participants explained that spatial audio cues can be more intuitive and serve as redundant feedback to remind and confirm the user actions. P18 explained:

it's hard to follow it mentally. It's hard to be, like, okay, I'm on column D row 6. Because you are having a constant count. You constantly have the sort of like keep track of where you are. It's just a lot of work. So, it's hard to keep up. But, having the sound, sorta like follow it, just kinda move around. So, you can just hear, like, okay, I am on the right side of things. It's kinda just want, like, an idea. Then I think that's pretty helpful.

Finally, spatial audio feedback may also find a place in assistive technology training. Some participants mentioned using alternative medium to assist learning. Both P3 and P11 shared that they would paint or sketch web pages when visiting them for the first time. P7, who is an assistive technology instructor, said he used mock-up iPhones with raised icons when teaching students using the iPhone. After trying the prototype, he commented that spatial audio feedback could serve similar purpose but require less cost and effort.

6.7 Implication for Design

The study results provide some implications on how spatial audio feedback designs can improve as well as other issues such as feature designs and training strategies.

First, the moving audio cues should be shorter. As reported earlier, users found the stationary audio cue's duration appropriate but the moving audio cue takes too long to complete. In addition, we learned that one main reason why participants had problems recognizing the current moving audio cues, which include both horizontal change and pitch change, was the cognitive load caused by paying attention to two audio elements at the same time. Therefore, instead of creating the longer moving effect, we may consider using distinctive short single tones to indicate movements. With one starting location and one ending location, there are only eight movement combinations. If using eight intuitive short sounds to represent these movements, it could help user achieve better recognition and more practical to play (or repeat).

Second, other pitch effects should be explored. Participants did agree that pitch changes were good metaphor for vertical position changes. However, recognizing our pitch changes created by modifying the tone's frequency is challenging. We used simple tones to keep the prototype simple and reproducible. For real product and more mature user studies, professional sound designers

should be consulted to create more effective sounds. The sound should remain short and simple. It can incorporate some pitch variation to still communicate the implication intuitively.

Third, the audio cues should be positioned according to the corresponding web element's position on the web page, not the viewport. In this study, this difference did not play a part because the web page is short and the whole page is presented in the current viewport. Considering the scenarios shared by participants where they wanted to acquire a web element's spatial position in order to give sighted people directions, if an audio cue communicates a web element's position relative to the viewport, the audio cue would not be able to enable the user to describe the accurate position of the web element to sighted peers. Of course, this raises another issue. If a web page is very long, how does it affect the sound effect conveying the audio cue's vertical position? Many web pages also adopt endless scrolling designs, in which the web page will automatically load more content when a user moves closer to the bottom of the page. Both cases pose challenges to spatial audio cue designs.

Fourth, instead of mapping to the exact position of a web element on the page, some offset should be considered to counter small indentation used for styling purposes. Web pages typically use white space to group content, but the layout will not align perfectly like a table does. Working out the minor differences is tedious and unnecessary for the purpose of using layout to assist web accessibility. We were not sure whether or not automatically remove the offset was a good idea since users might not notice the minor differences anyway. However, during the study participants spent a lot time and effort to capture these small differences. It would be better to guides users away from doing such laborious work. For example, if two web elements' horizontal positions on the web page differ by a few pixels but less than a preset threshold, their respective audio cues would have the same horizontal positions. When trying to picture the layout, a user would

recognize them as aligned in a "column." The offset can be preset to a fixed number, or screen readers can analyze the page and figure out an appropriate number that would not mix legit grouping by whitespace with neglectable spacing.

We have also observed that a user's proficiency with screen readers had great impact to the user experience. In this study, both inexperienced users and experienced users exhibited problems adopting the new features: while inexperienced users are still getting used to process existing standard screen reader speech output and might be overwhelmed by extra audio feedback, experienced users have formed their own habit of using screen readers and the spatial audio feedback might not fit into their routines easily. Other participants also expressed that they did not need to know spatial information when they perform normal reading.

One design implication in response is that screen readers should provide shortcut keys that allow users to quickly enable or disable the spatial audio feedback. In addition, since it is not obvious what level of experience with screen readers a user has, the spatial audio feedback should be defaulted to OFF and allow users to configure the feature to their own preferences on their own terms. Screen readers can also allow users to pre-configure combinations of settings that could assist particular tasks that a user has in mind. Once the user encounters such a task, she can easily switch to the suitable group settings and enjoy the rich spatial audio feedback immediately. Upon completion, the user can quickly switch back to the normal speech mode.

However, if the feature is defaulted to OFF, what motivations can we provide to encourage the adoption? We have heard examples from participants where they experienced communication barriers with sighted people due to spatial terms. Participants also proposed specific scenarios where spatial audio feedback could be useful. Screen readers can create lightweight addon tools

or preset modules for specific tasks. For example, one tool could focus on using spatial audio feedback to guide a user learning a new web page's layout. The screen reader can prompt the user when she visits a web page for the first page. Another useful tool or module could focus on stepwise instruction interpretation. For example, the screen reader can help parse instructions. If the instructions include spatial terms, the screen reader can prompt the user to turn on spatial audio feedback.

Another potential feature is to enable comparison of two non-neighboring audio cues. The current design only provides stationary audio cues for the current and the previous web elements. During navigation, users only hear the movement from the current web element to the next web element. However, participants mentioned frequently using landmark locations as references when trying to understand the current position on the web page or progress, for example, the center of the page (derive whether or not they are on the top part or bottom part), large images. Therefore, it may be useful to designate a few prominent landmarks on the web page, and users can configure to hear the spatial audio cues relative to a specific landmark. It would be more interesting if the user can also set temporary landmarks. For example, a user can set a heading to be a temporary landmark. Then when she navigates on web elements following the heading, she could derive how these web elements organized in relation to the heading. An alternative design of this feature is to play audio cues using the current web element as a reference, i.e., taking an egocentric approach. This design is more similar to how human uses landmarks to navigate in physical environment. For example, knowing the position of a landmark building, one can derive her own position on the map depending on the angles and distance of the landmark building. Such features can provide future research opportunities to evaluate effective landmark-based navigation techniques when screen reader users navigate on web pages.

We also believe that adequate training will be critical to initial user experience. In this study, participants only spent less than an hour learning the new spatial audio features before performing various tasks. Some comments and behaviors suggest that they were still unsure about the features or had not reached sufficient proficiency to deal with complex tasks. In future studies, more training should be included and, if possible, a longitudinal study design should be adopted. When training a new user, screen readers should provide goal-oriented exercises situated in practical spatial tasks. Only after mastering all these smaller tasks will users further the training to more complex tasks. We also heard participants who completed tasks successfully explaining how their knowledges of technologies or other comparable interfaces help them understand the tasks in hand. A training program can include similar materials to facilitate effective learning. For example, it can start with reviewing spatial concepts and related operations using magazines or newspaper. When later learning the spatial audio features, instructors can draw connections and use analogy to help users understand new features.

There are also some observations on screen reader user behaviors in general. We learned that screen reader users draw clues from surrounding context to help them reconstruct content structures or relationships. Since web pages with poor accessibilities often provide ambiguous information, having access to additional information is often critical. Therefore, it might be useful for screen readers to convey more information to users, even if the information is not expected to be beneficial, such as visual information. Though blind users cannot utilize this information directly, they may extract implicit clues that could help them understand ambiguous information.

Finally, the user study also raises some interesting questions. During the spatial term interpretation task, participants tended to utilize spatial audio feedback when dealing with horizontal spatial relationships and use arrow keys when dealing with vertical spatial relationships. Due to our small

sample size and the exploratory study designs, we cannot conclude that the type of spatial concept has an influence to a user's choice of navigation. However, it is an intriguing topic that should be examined in future studies. Participants who completed tasks successfully also often refer themselves as "visual person." The connection between being able to visualize information and abilities to use spatial audio feedback should be investigated further.

6.8 Summary

In this chapter, we summarized the results from the user study of 20 blind screen reader users. The participants reported that dealing with spatial terms was a practical challenge that has an impact in their interaction with sighted people. However, with screen readers currently available in the market, they can do little about it. After some short trainings, participants demonstrated that they were able to interpret common spatial terms when spatial audio feedback was provided. In addition, most of them can also successfully acquired the approximate layout of a web page based on spatial audio feedback. In the semi-structured interviews, they shared positive feedback on the idea of using spatial audio to convey layout information, including the designs supported by the prototype and other possible applications. However, they did report that the recognition of the audio cues imposed cognitive workload that cannot be ignored. We also discussed the negative impact of spatial audio output to inexperienced screen reader users. Finally, we proposed improvements, new designs, and future study topics based on the study results.

Chapter 7 Discussion and Conclusion

The overarching question that motivates this research is whether or not adding spatial audio feedback to screen readers can improve web accessibility. Web accessibility is a broad term and it can mean many different things. In this chapter, we take what we have learned from the work reported in previous chapters and discuss what impact spatial audio feedback makes in the context of the broader web accessibility. Then, we discuss the limitations of this study, topics, as well as questions that should be further investigated in future work.

7.1 Using Spatial Audio Feedback to Assist Communication with Sighted People

In Chapter Three, we identified one specific web accessibility problem, i.e., the communication of spatial terms between sighted users and screen reader users. The root of the problem is that screen readers convey little to no spatial layout information to their users. We hypothesized that this issue would be reduced if screen readers find alternative means to convey spatial concepts. This inspired a design that uses simple spatial audio feedback to inform users where they are on a web page and the direction of movement during navigation.

One goal of the study reported in Chapter Five and Chapter Six is to evaluate whether or not providing spatial audio feedback did allow screen reader users to make sense of spatial terms. The exploratory nature of the study does not allow us to make categorical conclusions. However, we believe that the answer is positive based on three observations.

First, participants were not able to initially answer questions concerning the spatial terms. Their explanations confirmed that the inability to answer the questions was the result of how typical screen readers function. Even in a few cases where experienced users pointed out viable alternative strategies determining the spatial relationship among web elements, they did not consider those options convenient or worth the effort.

After a quick introduction to the spatial audio features, participants quickly picked up on how to integrate the spatial audio feedback into their navigation strategies. Most were able to answer the same questions when presented with them again, though some of the participants used arrow keys and we were not sure how much they utilized the spatial audio feedback. For a few participants who still experienced difficulty answering the questions, it was often the case that they could recognize the spatial concepts implied by the audio cues but could not properly visualize web elements and utilize the additional layout information. These participants all have less experience with screen readers. In the later interviews, some participants explained that having to process all the information at once was a major challenge for them since they were still getting used to the standard screen reader output. It seems that more training and practice would improve their abilities to utilize spatial audio feedback over time.

Second, participants were able to perceive web page layouts via spatial audio feedback. Recognizing a web page's overall layout is more complex than interpreting simple spatial terms. Nonetheless, most participants successfully produced visual representation of their perceived web page layouts on the whiteboard. Even though some used arrow keys heavily and arrow keys have an inherent spatial nature, their models still exhibited subtle spatial details, such as indentation, that could only be learned from the spatial audio feedback. The keystroke data also showed that some participants were able to use spatial audio features strategically, such as for acquiring the

general idea using spatial audio feedback and using arrow keys to confirm relative spatial relationships.

Third, we directly asked participants to rate the screen reader prototype without- and later with- the spatial audio feedback. Based on the responses, participants believed that spatial audio feedback enabled them to complete layout-related tasks whereas typical screen readers were insufficient in that regard. In addition, participants experienced with screen readers did not think that the additional spatial audio created cognitive challenges (though other data suggested that it did incur some cognitive workload). Participants with less screen reader experience provided similar answers. However, their responses were not as positive.

Based on these three observations, we believe that providing spatial audio feedback is a promising approach to supporting screen reader users working with layout-related tasks. However, participants almost unanimously preferred being able to disable audio feedback when not needed. In general, participants did not believe that spatial audio feedback was necessary when they just needed to learn the web content. We identified two scenarios in which spatial audio feedback would be appreciated. The descriptions below could help screen reader developers understand the context and design suitable interactions.

The first scenario is when screen reader users need to follow instructions with spatial terms embedded. This scenario is more likely when the communication is asynchronous, as otherwise screen reader users could easily ask for clarifications. Since typical verbal communication is synchronous, this scenario more likely involves written instructions. Possible examples could be training manuals or email responses from customer service.

The second scenario is when screen reader users produce instructions. This may cover a wide range of cases, i.e., written, verbal, asynchronous, or synchronous. For example, it could be part of a job and the instructions will be provided for the general audience. Since the majority of the population is sighted, being able to use spatial terms can make instructions more effective. A staff member with visual impairment can utilize spatial audio feedback to recognize spatial relationships and use the corresponding terms in her writing. Another possible case is when a screen reader user needs to report an inaccessible web page to a web master. If the web master does not have adequate accessibility training and has difficulty determining the problematic area, the screen reader user can learn the approximate position of the inaccessible web element and guide the web master to it.

7.2 Using Spatial Audio Feedback to Manage Inaccessible Web Pages

Another goal of the study is to gather more general feedback from screen reader users on how they think the spatial audio could be useful. We heard comments on four cases in which being able to perceive layouts via spatial audio feedback would positively impact user experience.

The first case is to use the additional layout information to help understand web pages with poor accessibility. When web elements on a web page are not properly positioned, screen readers would in turn read them in disarray. The obvious mismatch among web elements will throw the users into a state of confusion. To recognize and correct the problem, users have to search for clues elsewhere. For example, users might notice patterns among the disorganized information and then they can reorganize the information in the intended order. The strategy's success is, however, case by case since the same kind of clues do not always present. Furthermore, it may be challenging to use this method as it incurs additional cognitive and memory demand.

With spatial audio feedback, users could acquire some sense of the web page's layout. The spatial information could be invaluable when dealing with incorrectly ordered web elements. Visual clues, such as white space and grouping, play important roles when sighted people process information visually. If the web page is usable by sighted people but not via screen readers, it is likely that its layout includes information implicitly clarifying the structure. Navigating web elements with the spatial audio feedback could help users notice patterns in the information's organization. For more advanced users, they could also compare the audio cues to recognize grouping information, e.g., some steps are bigger than others, which implies within-group spacing and between-group spacing. Since layout information is always present, over time screen reader users could even build more sophisticated skills for decoding layout information.

The second case is enabling a user to take into consideration a web element's position on the page when assessing the web element's role on the web page. Knowing that the web element is at the beginning or at the end of the page has different implications. If the web element is located at a more prominent position on the web page, it may also suggest the web element's influential role in the interaction. In contrast, if the web element is among the list of items on the left menu or in the footer, its role would be less central to the web page.

The third case is to help recover from the aftermath of an accessibility incident. Today's web pages are still full of inaccessible web elements. They can confuse screen readers and displace the reading focus to unexpected locations. In turn, users often have trouble figuring out what has happened and where they are now. A coping strategy is to reload the page, which would set the focus back to the beginning of the web page, then retrace back to the previous location on the web page. However, remembering the sequence of steps taken the last time can be challenging since short-term memory has limited capacity.

Spatial audio feedback can play a role in this process by providing additional clues for users to remember problematic area and web elements. The user can maintain an impression of where she is on the web page during navigation. After an accessibility incident, she can use the spatial memory as well as other existing clues to navigate back to the general area. For frequently used web pages, the user can also memorize the inaccessible web element's location on the web page so that she can avoid it when visiting again in the future. In this scenario, spatial audio feedback does not enable actions that users could not perform previously, i.e., users can still recall the previous location by memorizing the neighboring web element's content. However, spatial audio feedback is an additional clue that users may pick up.

The fourth case concerns the general positive feeling of gaining access to information that has been off limit to screen reader users previously. Layout information is readily available to sighted users, but elusive, at the best, for screen reader users. Screen reader developers have designed effective interactions to allow browsing web pages without learning layout information. However, one side effect of this mechanism is the unequal access to information between sighted and blind users. Having a way to perceive the web page layout provides the opportunity to bridge the divide. In some cases, it helps resolve tangible problems, such as the communication barrier discussed in the previous section; in other cases, it simply gives screen reader users the comfort of knowing they can find out what sighted people are referring to if necessary.

7.3 Using Spatial Audio Feedback to Promote Learning

We have also seen the potential for employing spatial audio feedback in blind users' learning. In Mobility and Orientation training, one technique for familiarizing a blind trainee with a physical environment, such as a building, is to present her with a physical model of it. By playing with the

model, the blind trainee can learn the spatial structure and available paths. Essentially, this helps the blind trainee build a mental model of the physical environment that could be revisited when actual navigation occurs.

A similar process happens when a blind user learns a new web page or interface. The user has to get a sense of what is on the web page or interface before she can perform tasks. Our participant's story of using a tactile iPhone model to help students learn how to use iPhone is a good example. In addition, new interfaces may adopt innovative designs that are drastically different from the old ones. It can be confusing if the blind user holds assumptions based on old interfaces. Certain operations might not make sense if the user does not know how the interface is organized.

The study reported in Chapter Six showed that providing simple spatial audio feedback during navigation allows screen reader users to perceive web page layouts. Therefore, it is possible for blind users to learn the layout of a 2D interface or map with web-based virtual models using spatial audio feedback-enabled screen readers. Compared to creating physical models, creating web-based models is less expensive and takes less effort. It only requires some basic HTML web development skills or being able to use visual coding tools. Accessibility professionals can create web-based models quickly when planning for assistive technology trainings or Mobility and Orientation trainings. Furthermore, the web-based model can be easily replicated and reused. An evolving model can be used to suit the learning goals throughout the training. Each student can even receive a customized model based on her learning pace, without incurring much more cost than using the same model.

The same idea can also apply to home appliance manuals. More and more appliances use flat, touchscreen control panels nowadays. Without any non-visual identifiable features on the buttons,

it makes it difficult for blind people to learn how to use them. A common practice is to ask sighted family members or friends to learn all the functionalities. Then, blind users can place tactile dots with different shapes and sizes on the buttons to help them recognize the associated functionalities. Getting help from others can be troublesome, as it always requires scheduling, not to mention that some interfaces can be complicated and the sighted helpers have to learn themselves first. Manufacturers can improve this process by providing accessible web-based models and manuals to support blind user self-learning. When learning to use the appliance, a blind user can locate the web-based model of the control panel and navigate the virtual interface one button at a time. She can discover how all the buttons are laid out on the control panel based on spatial audio feedback. Each button should be annotated with its functionalities so that the screen reader can also announce the information during the exploration. Manufacturers could even create step-by-step, interactive web-based lessons to guide learning. For manufacturers who try to improve their products' accessibilities, they can even track the user's interaction data with web-based models to gain insights on problematic designs.

7.4 Limitations

The most relevant findings of this research come from the final study. Therefore, the final study's limitation also overshadows the entire project. The main limitation of the final study is rooted in its exploratory nature. We have reported signs of the positive impact of adding spatial audio feedback to screen readers. However, our main contribution is the identification of various opportunities for accessibility improvements utilizing spatial audio feedback. Whether or not these designs could indeed improve blind users' experience of completing the respective tasks remains to be evaluated with studies designed for particular purposes.

Another limitation concerns how the spatial audio is delivered. Our studies used stereo headphones to deliver spatial audio cues. This is common in the study of spatial audio interfaces. It also aligns with how blind users use screen readers in public settings. However, perceiving audio via headphones is different from perceiving audio via speakers. Therefore, our findings cannot be extended to scenarios where blind users use screen readers with speakers.

Finally, the prototype used in the final study features arrow key-based navigations. This was designed with the intention of seeing if participants could utilize the layout information learned through spatial audio cues and derive more effective directional navigation strategies. During the study, we indeed observed participants using a combination of arrow keys and spatial audio cues to assess the web page layout. The two features appear to have a relationship where one indicates to the user the planned movement and the other confirms to the user what movement has happened. However, since either feature could serve to inform or confirm the layout, the specific role of spatial audio feedback is obscured.

7.5 Future Work

This research helped us gain a better understanding of how spatial audio feedback could be used by screen reader users. The topics listed below are areas that we believe future studies should explore or address:

- Collaboration tools between sighted and blind users: we have seen that spatial audio feedback is more useful when spatial terms are concerned, which is common when communicating with sighted people. Therefore, tools can be designed specifically to mediate the interaction between sighted and blind users.

- Revised spatial audio cues: we used synthesized tones that include both horizontal position changes and pitch changes in this study. We have learned that users had problems perceiving both at the same time. In future work, we should explore and evaluate using distinct complex sound clips. The sound can still incorporate pitch variation to take advantage of the "height-pitch" metaphor. However, the top priority of the sound design should be incorporating clear characteristics to enable easy recognition.
- Longitudinal studies of using spatial audio features: one challenge of our user study was to train participants using the prototype in a short period of time. Though the features are straightforward to learn, it is difficult to see how the new features can be organically integrated into user interaction with screen readers. Our prototype does not provide all screen reader features. Therefore, it is not suitable for longitudinal studies. However, given the positive feedback we have received, the next study can invest more resources in developing the spatial audio features in a fully functional screen reader and conducting longitudinal study to evaluate usage in more realistic settings.

7.6 Summary

In this chapter, we bring the conversation back to the original question: can spatial audio help improve web accessibility for blind web users? We drew answers from the studies reported in this dissertation and considered them in the broader context of web accessibility. We believe that spatial audio can bring positive impact to screen reader users' experiences. Specifically, spatial audio can help users make sense of spatial terms and perceive overall web page layout. Additionally, spatial audio can also enable web-based auditory modeling tools to assist blind users' learning of new interfaces.

Bibliography

1. A. H. Z. Abidin, Hong Xie, and Kok Wai Wong. 2012. Blind users' mental model of web page using touch screen augmented with audio feedback. In *2012 International Conference on Computer Information Science (ICCIS)*, 1046–1051. https://doi.org/10.1109/ICCISci.2012.6297180

2. Hend S. Al-Khalifa. 2010. Exploring the Accessibility of Saudi Arabia e-Government Websites: A Preliminary Results. In *Proceedings of the 4th International Conference on Theory and Practice of Electronic Governance* (ICEGOV '10), 274–278. https://doi.org/10.1145/1930321.1930378

3. Najd A. Al-Mouh, Atheer S. Al-Khalifa, and Hend S. Al-Khalifa. 2014. A First Look into MOOCs Accessibility. In *Computers Helping People with Special Needs* (Lecture Notes in Computer Science), 145–152. https://doi.org/10.1007/978-3-319-08596-8_22

4. Javier De Andrés, Pedro Lorca, and Ana B. Martínez. 2010. Factors influencing web accessibility of big listed firms: an international study. *Online Information Review* 34, 1: 75–97. https://doi.org/10.1108/14684521011024137

5. Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction* 24, 6: 574–594. https://doi.org/10.1080/10447310802205776

6. M. K. Baowaly and M. Bhuiyan. 2012. Accessibility analysis and evaluation of Bangladesh government websites. In *2012 International Conference on Informatics, Electronics Vision (ICIEV)*, 46–51. https://doi.org/10.1109/ICIEV.2012.6317487

7. D. W. Batteau. 1967. The Role of the Pinna in Human Localization. *Proceedings of the Royal Society of London B: Biological Sciences* 168, 1011: 158–180. https://doi.org/10.1098/rspb.1967.0058

8. Durand R. Begault and Tom Erbe. 1993. Multichannel Spatial Auditory Display for Speech Communications. Retrieved December 7, 2015 from http://www.aes.org/e-lib/browse.cfm?elib=6525

9. Tim Berners-Lee. The original proposal of the WWW. Retrieved December 26, 2017 from https://www.w3.org/History/1989/proposal.html

10. Jeffrey P. Bigham, Ryan S. Kaminsky, Richard E. Ladner, Oscar M. Danielsson, and Gordon L. Hempton. 2006. WebInSight:: Making Web Images Accessible. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility* (Assets '06), 181–188. https://doi.org/10.1145/1168987.1169018

11. Jeffrey P. Bigham, Richard E. Ladner, and Yevgen Borodin. 2011. The Design of Human-powered Access Technology. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '11), 3–10. https://doi.org/10.1145/2049536.2049540

12. Meera M. Blattner, Denise A. Sumikawa, and Robert M. Greenberg. 1989. Earcons and Icons: Their Structure and Common Design Principles. *Human–Computer Interaction* 4, 1: 11–44. https://doi.org/10.1207/s15327051hci0401_1

13. Yevgen Borodin, Jeffrey P. Bigham, Glenn Dausch, and I. V. Ramakrishnan. 2010. More Than Meets the Eye: A Survey of Screen-reader Browsing Strategies. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)* (W4A '10), 13:1–13:10. https://doi.org/10.1145/1805986.1806005

14. Yevgen Borodin, Shinya Kawanaka, Hironobu Takagi, Masatomo Kobayashi, Daisuke Sato, and Chieko Asakawa. 2010. Social Accessibility. *No Code Required: Giving Users Tools to Transform the Web*: 347.

15. Erin Brady and Jeffrey P. Bigham. 2014. How Companies Engage Customers Around Accessibility on Social Media. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility* (ASSETS '14), 51–58. https://doi.org/10.1145/2661334.2661355

16. Giorgio Brajnik, Yeliz Yesilada, and Simon Harper. 2011. The Expertise Effect on Web Accessibility Evaluation Methods. *Human–Computer Interaction* 26, 3: 246–283. https://doi.org/10.1080/07370024.2011.601670

17. Judy Brewer. 2004. Web Accessibility Highlights and Trends. In *Proceedings of the 2004 International Cross-disciplinary Workshop on Web Accessibility (W4A)* (W4A '04), 51–55. https://doi.org/10.1145/990657.990667

18. Stephen Brewster, Joanna Lumsden, Marek Bell, Malcolm Hall, and Stuart Tasker. 2003. Multimodal "Eyes-free" Interaction Techniques for Wearable Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '03), 473–480. https://doi.org/10.1145/642611.642694

19. Andrei Broder. 2002. A Taxonomy of Web Search. *SIGIR Forum* 36, 2: 3–10. https://doi.org/10.1145/792550.792552

20. John Brooke. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194: 4–7.

21. Douglas S. Brungart and William M. Rabinowitz. 1999. Auditory localization of nearby sources. Head-related transfer functions. *The Journal of the Acoustical Society of America* 106, 3: 1465–1479. https://doi.org/10.1121/1.427180

22. Michele A. Burton, Erin Brady, Robin Brewer, Callie Neylan, Jeffrey P. Bigham, and Amy Hurst. 2012. Crowdsourcing Subjective Fashion Advice Using VizWiz: Challenges and

Opportunities. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '12), 135–142. https://doi.org/10.1145/2384916.2384941

23. D. Chang and K. V. Nesbitt. 2006. Identifying Commonly-Used Gestalt Principles as a Design Framework for Multi-Sensory Displays. In *2006 IEEE International Conference on Systems, Man and Cybernetics*, 2452–2457. https://doi.org/10.1109/ICSMC.2006.385231

24. Michael Cohen and Lester F. Ludwig. 1991. Computer-supported Cooperative Work and Groupware. Part 2Multidimensional audio window management. *International Journal of Man-Machine Studies* 34, 3: 319–336. https://doi.org/10.1016/0020-7373(91)90023-Z

25. Gina Conti-Ramsden and Miguel Perez-Pereira. 1999. Conversational Interactions Between Mothers and Their Infants Who Are Congenitally Blind, Have Low Vision, or Are Sighted. *Journal of Visual Impairment & Blindness* 93, 11: 691.

26. Vivienne Conway. 2011. Website accessibility in Western Australian public libraries. *The Australian Library Journal* 60, 2: 103–112. https://doi.org/10.1080/00049670.2011.10722582

27. Vivienne Conway, Justin Brown, Scott Hollier, and Cam Nicholl. 2012. Website Accessibility: a Comparative Analysis of Australian National and State/Territory Library Websites. *The Australian Library Journal* 61, 3: 170–188. https://doi.org/10.1080/00049670.2012.10736059

28. Andrew Crossan and Stephen Brewster. 2008. Multimodal Trajectory Playback for Teaching Shape Information and Trajectories to Visually Impaired Computer Users. *ACM Trans. Access. Comput.* 1, 2: 12:1–12:34. https://doi.org/10.1145/1408760.1408766

29. Laurent Demany and Catherine Semal. 2008. The Role of Memory in Auditory Perception. In *Auditory Perception of Sound Sources*, William A. Yost, Arthur N. Popper and Richard R. Fay (eds.). Springer US, 77–113. https://doi.org/10.1007/978-0-387-71305-2_4

30. Jan B. F. Van Erp, Hendrik A. H. C. Van Veen, Chris Jansen, and Trevor Dobbins. 2005. Waypoint Navigation with a Vibrotactile Waist Belt. *ACM Trans. Appl. Percept.* 2, 2: 106–117. https://doi.org/10.1145/1060581.1060585

31. Jody Condit Fagan and Bryan Fagan. 2004. An accessibility study of state legislative Web sites. *Government Information Quarterly* 21, 1: 65–85. https://doi.org/10.1016/j.giq.2003.12.010

32. Steve Faulkner and The Paciello Group. ARIA in HTML. Retrieved August 1, 2017 from https://www.w3.org/TR/html-aria/

33. W. E. Feddersen, T. T. Sandel, D. C. Teas, and L. A. Jeffress. 1957. Localization of High-Frequency Tones. *The Journal of the Acoustical Society of America* 29, 9: 988–991. https://doi.org/10.1121/1.1909356

34. Patricia J. Flowers and Chao-hui Wang. 2002. Matching Verbal Description to Music Excerpt: The Use of Language by Blind and Sighted Children. *Journal of Research in Music Education* 50, 3: 202–214. https://doi.org/10.2307/3345798

35. Susannah Fox. 2013. 51% of U.S. Adults Bank Online. *Pew Research Center: Internet, Science & Tech*. Retrieved December 26, 2017 from http://www.pewinternet.org/2013/08/07/51-of-u-s-adults-bank-online/

36. Luis Francisco-Revilla and Jeff Crow. 2009. Interpreting the Layout of Web Pages. In *Proceedings of the 20th ACM Conference on Hypertext and Hypermedia* (HT '09), 157–166. https://doi.org/10.1145/1557914.1557943

37. Luis Francisco-Revilla and Jeff Crow. 2010. Interpretation of Web Page Layouts by Blind Users. In *Proceedings of the 10th Annual Joint Conference on Digital Libraries* (JCDL '10), 173–176. https://doi.org/10.1145/1816123.1816148

38. Euan Freeman, Graham Wilson, Dong-Bach Vo, Alex Ng, Ioannis Politis, and Stephen Brewster. 2017. The Handbook of Multimodal-Multisensor Interfaces. In Sharon Oviatt, Björn Schuller, Philip R. Cohen, Daniel Sonntag, Gerasimos Potamianos and Antonio Krüger (eds.). Association for Computing Machinery and Morgan & Claypool, New York, NY, USA, 277–317. https://doi.org/10.1145/3015783.3015792

39. Andre P. Freire, Cibele M. Russo, and Renata P. M. Fortes. 2008. A Survey on the Accessibility Awareness of People Involved in Web Development Projects in Brazil. In *Proceedings of the 2008 International Cross-disciplinary Conference on Web Accessibility (W4A)* (W4A '08), 87–96. https://doi.org/10.1145/1368044.1368064

40. José L. Fuertes, Ricardo González, Emmanuelle Gutiérrez, and Loïc Martínez. 2009. Hera-FFX: A Firefox Add-on for Semi-automatic Web Accessibility Evaluation. In *Proceedings of the 2009 International Cross-Disciplinary Conference on Web Accessibililty (W4A)* (W4A '09), 26–35. https://doi.org/10.1145/1535654.1535661

41. Charles R. Gallistel. 1990. *The organization of learning*. The MIT Press, Cambridge, MA, US.

42. William W. Gaver. 1989. The SonicFinder: An Interface That Uses Auditory Icons. *Hum.-Comput. Interact.* 4, 1: 67–94. https://doi.org/10.1207/s15327051hci0401_3

43. Greg Gay and Cindy Qi Li. 2010. AChecker: Open, Interactive, Customizable, Web Accessibility Checking. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)* (W4A '10), 23:1–23:2. https://doi.org/10.1145/1805986.1806019

44. Michele Geronazzo, Alberto Bedin, Luca Brayda, Claudio Campus, and Federico Avanzini. 2016. Interactive spatial sonification for non-visual exploration of virtual maps. *International Journal of Human-Computer Studies* 85: 4–15. https://doi.org/10.1016/j.ijhcs.2015.08.004

45. Teresa D. Gilbertson and Colin H. C. Machin. 2012. Guidelines, Icons and Marketable Skills: An Accessibility Evaluation of 100 Web Development Company Homepages. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility* (W4A '12), 17:1–17:4. https://doi.org/10.1145/2207016.2207024

46. Barney G. Glaser and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine.

47. Ramiro Gonçalves, José Martins, Jorge Pereira, Manuel Au-Yong Oliveira, and João José P. Ferreira. 2012. Enterprise Web Accessibility Levels Amongst the Forbes 250: Where Art Thou O Virtuous Leader? *Journal of Business Ethics* 113, 2: 363–375. https://doi.org/10.1007/s10551-012-1309-3

48. Morten Goodwin, Deniz Susar, Annika Nietzio, Mikael Snaprud, and Christian S. Jensen. 2011. Global Web Accessibility Analysis of National Government Portals and Ministry Web Sites. *Journal of Information Technology & Politics* 8, 1: 41–67. https://doi.org/10.1080/19331681.2010.508011

49. Stuart Goose and Carsten Möller. 1999. A 3D Audio Only Interactive Web Browser: Using Spatialization to Convey Hypermedia Document Structure. In *Proceedings of the Seventh ACM International Conference on Multimedia (Part 1)* (MULTIMEDIA '99), 363–371. https://doi.org/10.1145/319463.319649

50. João Guerreiro and Daniel Gonçalves. 2014. Text-to-speeches: Evaluating the Perception of Concurrent Speech by Blind People. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility* (ASSETS '14), 169–176. https://doi.org/10.1145/2661334.2661367

51. Richard Guy and Khai Truong. 2012. CrossingGuard: Exploring Information Content in Navigation Aids for Visually Impaired Pedestrians. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), 405–414. https://doi.org/10.1145/2207676.2207733

52. Vicki L. Hanson and John T. Richards. 2013. Progress on Website Accessibility? *ACM Trans. Web* 7, 1: 2:1–2:30. https://doi.org/10.1145/2435215.2435217

53. Kotaro Hara, Vicki Le, and Jon Froehlich. 2013. Combining Crowdsourcing and Google Street View to Identify Street-level Accessibility Problems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 631–640. https://doi.org/10.1145/2470654.2470744

54. Ira J. Hirsh. 1971. Masking of Speech and Auditory Localization. *Audiology* 10, 2: 110–114. https://doi.org/10.3109/00206097109072548

55. W3C Web Accessibility Initiative (WAI). Introduction to Web Accessibility | Web Accessibility Initiative (WAI) | W3C. *W3C Web Accessibility Initiative (WAI)*. Retrieved December 28, 2017 from https://www.w3.org/WAI/intro/accessibility.php

56. W3C Web Accessibility Initiative (WAI). Web Accessibility Initiative (WAI) - home page. *W3C Web Accessibility Initiative (WAI)*. Retrieved December 28, 2017 from https://www.w3.org/WAI/

57. W3C Web Accessibility Initiative (WAI). Target Corporation - A Cautionary Tale of Inaccessibility | Web Accessibility Initiative (WAI) | W3C. *W3C Web Accessibility Initiative (WAI)*. Retrieved January 12, 2018 from https://www.w3.org/WAI/bcase/target-case-study

58. Emily Jackson-Sanborn, Kerri Odess-Harnish, and Nikki Warren. 2002. Web site accessibility: a study of six genres. *Library Hi Tech* 20, 3: 308–317. https://doi.org/10.1108/07378830210444504

59. William H. Jacobson. 1993. *The Art and Science of Teaching Orientation and Mobility to Persons with Visual Impairments*. AFB Press.

60. Caroline Jay, Robert Stevens, Mashhuda Glencross, Alan Chalmers, and Cathy Yang. 2007. How people use presentation to search for a link: expanding the understanding of accessibility on the Web. *Universal Access in the Information Society* 6, 3: 307–320. https://doi.org/10.1007/s10209-007-0089-5

61. Shaun K. Kane, Jessie A. Shulman, Timothy J. Shockley, and Richard E. Ladner. 2007. A Web Accessibility Report Card for Top International University Web Sites. In *Proceedings of the 2007 International Cross-disciplinary Conference on Web Accessibility (W4A)* (W4A '07), 148–156. https://doi.org/10.1145/1243441.1243472

62. Joyce Karreman. 2004. *Use and Effect of Declarative Information in User Instructions*. Rodopi.

63. Kellar Melanie, Watters Carolyn, and Shepherd Michael. 2007. A field study characterizing Web-based information-seeking tasks. *Journal of the American Society for Information Science and Technology* 58, 7: 999–1018. https://doi.org/10.1002/asi.20590

64. Nigel J. Kemp and Derek R. Rutter. 1986. Social Interaction in Blind People: An Experimental Analysis. *Human Relations* 39, 3: 195–210. https://doi.org/10.1177/001872678603900302

65. David E. Kieras and Susan Bovair. 1984. The Role of a Mental Model in Learning to Operate a Device. *Cognitive Science* 8, 3: 255–273. https://doi.org/10.1207/s15516709cog0803_3

66. Minoru Kobayashi and Chris Schmandt. 1997. Dynamic Soundscape: Mapping Time to Space for Audio Browsing. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (CHI '97), 194–201. https://doi.org/10.1145/258549.258702

67. Sreekar Krishna, Dirk Colbry, John Black, Vineeth Balasubramanian, and Sethuraman Panchanathan. 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. Retrieved April 24, 2018 from https://hal.inria.fr/inria-00325432

68. Joanne Kuzma. 2009. Regulatory Compliance and Web Accessibility of UK Parliament Sites. *Journal of Information, Law & Technology (JILT)* 2: 1–15.

69. J. Lazar, B. Wentz, C. Akeley, M. Almuhim, S. Barmoy, P. Beavan, C. Beck, A. Blair, A. Bortz, B. Bradley, M. Carter, D. Crouch, G. Dehmer, M. Gorman, C. Gregory, E. Lanier, A. McIntee, R. Nelson, D. Ritgert, R. Rogers, S. Rosenwald, S. Sullivan, J. Wells, C. Willis, K. Wingo-Jones, and T. Yatto. 2012. Equal Access to Information? Evaluating the Accessibility of Public Library Web Sites in the State of Maryland. In *Designing Inclusive Systems*. Springer, London, 185–194. https://doi.org/10.1007/978-1-4471-2867-0_19

70. Jonathan Lazar, P. Beavan, J. Brown, D. Coffey, B. Nolf, R. Poole, R. Turk, V. Waith, T. Wall, and K. Weber. 2010. Investigating the accessibility of state government web sites in Maryland. *Designing inclusive interactions*: 69–78.

71. Jonathan Lazar, Patricia Beere, Kisha-Dawn Greenidge, and Yogesh Nagappa. 2003. Web accessibility in the Mid-Atlantic United States: a study of 50 homepages. *Universal Access in the Information Society* 2, 4: 331–341. https://doi.org/10.1007/s10209-003-0060-z

72. Jonathan Lazar, Alfreda Dudley-Sponaugle, and Kisha-Dawn Greenidge. 2004. Improving web accessibility: a study of webmaster perceptions. *Computers in Human Behavior* 20, 2: 269–288. https://doi.org/10.1016/j.chb.2003.10.018

73. Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research Methods in Human-Computer Interaction*. Morgan Kaufmann.

74. Jonathan Lazar, Brian Wentz, Abdulelah Almalhem, Alexander Catinella, Catalin Antonescu, Yeveniy Aynbinder, Michael Bands, Edward Bastress, Brandon Chan, Brian Chelden, Darin Feustel, Nabin Gautam, Whitney Gregg, Michael Heppding, Cory Householder, Alex Libby, Corey Melton, Jack Olgren, Loren Palestino, Morgan Ricks, Scott Rinebold, and Matthew Seidel. 2013. A longitudinal study of state government homepage accessibility in Maryland and the role of web page templates for improving accessibility. *Government Information Quarterly* 30, 3: 289–299. https://doi.org/10.1016/j.giq.2013.03.003

75. Sooyeon Lee, Chien Wen Yuan, Benjamin V. Hanrahan, Mary Beth Rosson, and John M. Carroll. 2017. Reaching Out: Investigating Different Modalities to Help People with Visual Impairments Acquire Items. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '17), 389–390. https://doi.org/10.1145/3132525.3134817

76. Rui Lopes, Karel Van Isacker, and Luís Carriço. 2010. Redefining Assumptions: Accessibility and Its Stakeholders. In *Computers Helping People with Special Needs* (Lecture Notes in Computer Science), 561–568. https://doi.org/10.1007/978-3-642-14097-6_90

77. Gaëtan Lorho, Juha Marila, and Jarmo Hiipakka. 2001. Feasibility of multiple non-speech sounds presentation using headphones. In *Proceedings of ICAD '01*, 32–37.

78. Jennifer Mankoff, Holly Fait, and Tu Tran. 2005. Is Your Web Page Accessible?: A Comparative Study of Methods for Assessing Web Page Accessibility for the Blind. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '05), 41–50. https://doi.org/10.1145/1054972.1054979

79. Ana B. Martínez, Javier De Andrés, and Julita García. 2014. Determinants of the Web accessibility of European banks. *Information Processing & Management* 50, 1: 69–86. https://doi.org/10.1016/j.ipm.2013.08.001

80. Hans van der Meij, Peter Blijleven, and Leanne Jansen. 2003. What makes up a procedure. *Content & Complexity. Information design in technical communication*: 129–186.

81. Stephen W. Mereu and Rick Kazman. 1996. Audio Enhanced 3D Interfaces for Visually Impaired Users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '96), 72–78. https://doi.org/10.1145/238386.238406

82. A. W. Mills. 1958. On the Minimum Audible Angle. *The Journal of the Acoustical Society of America* 30, 4: 237–246. https://doi.org/10.1121/1.1909553

83. A William Mills. 1972. Auditory localization (Binaural acoustic field sampling, head movement and echo effect in auditory localization of sound sources position, distance and orientation). *Foundations of modern auditory theory.* 2: 303–348.

84. Durgaprasad Misra, Alka Mishra, Sunil Babbar, and Sunita Singh. 2017. Web Accessibility Assessment of Government Web Solutions: A Case Study in Digital India. In *Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance* (ICEGOV '17), 26–34. https://doi.org/10.1145/3047273.3047323

85. Daryl R. Moen. 2000. *Newspaper Layout & Design: A Team Approach*. Iowa State University Press.

86. Brian C. J. Moore. 2012. *An Introduction to the Psychology of Hearing*. BRILL.

87. Emma Murphy, Ravi Kuber, Graham McAllister, Philip Strain, and Wai Yu. 2007. An empirical investigation into the difficulties experienced by visually impaired Internet users. *Universal Access in the Information Society* 7, 1–2: 79–91. https://doi.org/10.1007/s10209-007-0098-4

88. Mala D. Naraine and Peter H. Lindsay. 2011. Social inclusion of employees who are blind or low vision. *Disability & Society* 26, 4: 389–403. https://doi.org/10.1080/09687599.2011.567790

89. Jakob Nielsen. 2000. Designing web usability: the practice of simplicity New Riders Publishing. *Indianapolis, Indiana*.

90. Donald A. Norman. 1983. Some observations on mental models. *Mental models* 7, 112: 7–14.

91. 2014. World Wide Web Timeline. *Pew Research Center: Internet, Science & Tech*. Retrieved December 26, 2017 from http://www.pewinternet.org/2014/03/11/world-wide-web-timeline/

92. Bonnie O'Day. 1999. Employment Barriers for People with Visual Impairments. *Journal of Visual Impairment & Blindness* 93, 10: 627.

93. Uran Oh and Leah Findlater. 2014. Design of and Subjective Response to On-body Input for People with Visual Impairments. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility* (ASSETS '14), 115–122. https://doi.org/10.1145/2661334.2661376

94. M. Ohuchi, Y. Iwaya, and Y. Suzuki. 2006. Cognitive-map forming of the blind in virtual sound environment. Retrieved August 25, 2016 from https://smartech.gatech.edu/handle/1853/50493

95. Abiodun Olalere and Jonathan Lazar. 2011. Accessibility of U.S. federal government home pages: Section 508 compliance and site accessibility statements. *Government Information Quarterly* 28, 3: 303–309. https://doi.org/10.1016/j.giq.2011.02.002

96. Government of Ontario. 2014. Accessibility for Ontarians with Disabilities Act, 2005. *Ontario.ca*. Retrieved January 9, 2018 from https://www.ontario.ca/laws/view

97. Joanne Oud. 2012. How Well Do Ontario Library Web Sites Meet New Accessibility Requirements? *Partnership: The Canadian Journal of Library and Information Practice and Research* 7, 1. https://doi.org/10.21083/partnership.v7i1.1613

98. M. Perez-Pereira and J. Castro. 1992. Pragmatic functions of blind and sighted children's language: a twin case study. *First Language* 12, 34: 17–37. https://doi.org/10.1177/014272379201203402

99. D. R. Perrott and A. D. Musicant. 1977. Minimum auditory movement angle: Binaural localization of moving sound sources. *The Journal of the Acoustical Society of America* 62, 6: 1463–1466. https://doi.org/10.1121/1.381675

100. Antti Pirhonen, Stephen Brewster, and Christopher Holguin. 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* (CHI '02), 291–298. https://doi.org/10.1145/503376.503428

101. Beryl Plimmer, Peter Reid, Rachel Blagojevic, Andrew Crossan, and Stephen Brewster. 2011. Signing on the Tactile Line: A Multimodal System for Teaching Handwriting to Blind Children. *ACM Trans. Comput.-Hum. Interact.* 18, 3: 17:1–17:29. https://doi.org/10.1145/1993060.1993067

102. Christopher Power, André Freire, Helen Petrie, and David Swallow. 2012. Guidelines Are Only Half of the Story: Accessibility Problems Encountered by Blind Users on the Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), 433–442. https://doi.org/10.1145/2207676.2207736

103. Cynthia Putnam, Kathryn Wozniak, Mary Jo Zefeldt, Jinghui Cheng, Morgan Caputo, and Carl Duffield. 2012. How Do Professionals Who Create Computing Technologies Consider Accessibility? In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '12), 87–94. https://doi.org/10.1145/2384916.2384932

104. Pei-Luen Patrick Rau, Lianhui Zhou, Na Sun, and Runting Zhong. 2016. Evaluation of web accessibility in China: changes from 2009 to 2013. *Universal Access in the Information Society* 15, 2: 297–303. https://doi.org/10.1007/s10209-014-0385-9

105. Lord Rayleigh. 1907. XII. On our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 13, 74: 214–232.

106. Dimitrios Rigas and James Alty. 2005. The rising pitch metaphor: an empirical study. *International Journal of Human-Computer Studies* 62, 1: 1–20. https://doi.org/10.1016/j.ijhcs.2004.06.004

107. Ravic Ringlaben, Marty Bray, and Abbot Packard. 2013. Accessibility of American University Special Education Departments' Web sites. *Universal Access in the Information Society* 13, 2: 249–254. https://doi.org/10.1007/s10209-013-0302-7

108. Nitin Sawhney and Chris Schmandt. 2000. Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments. *ACM Trans. Comput.-Hum. Interact.* 7, 3: 353–383. https://doi.org/10.1145/355324.355327

109. David Sloan and Sarah Horton. 2014. Global Considerations in Creating an Organizational Web Accessibility Policy. In *Proceedings of the 11th Web for All Conference* (W4A '14), 16:1–16:4. https://doi.org/10.1145/2596695.2596709

110. Edward E. Smith and Lorraine Goodman. 1984. Understanding Written Instructions: The Role of an Explanatory Schema. *Cognition and Instruction* 1, 4: 359–396. https://doi.org/10.1207/s1532690xci0104_1

111. Jaka Sodnik, Christina Dicke, Sašo Tomažič, and Mark Billinghurst. 2008. A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies* 66, 5: 318–332. https://doi.org/10.1016/j.ijhcs.2007.11.001

112. Hironobu Takagi, Shinya Kawanaka, Masatomo Kobayashi, Takashi Itoh, and Chieko Asakawa. 2008. Social Accessibility: Achieving Accessibility Through Collaborative Metadata Authoring. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility* (Assets '08), 193–200. https://doi.org/10.1145/1414471.1414507

113. Titus J. J. Tang and Wai Ho Li. 2014. An Assistive EyeWear Prototype That Interactively Converts 3D Object Locations into Spatial Audio. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers* (ISWC '14), 119–126. https://doi.org/10.1145/2634317.2634318

114. Hitoshi Terai, Hitomi Saito, Yuka Egusa, Masao Takaku, Makiko Miwa, and Noriko Kando. 2008. Differences Between Informational and Transactional Tasks in Information Seeking on the Web. In *Proceedings of the Second International Symposium on Information Interaction in Context* (IIiX '08), 152–159. https://doi.org/10.1145/1414694.1414728

115. Silvanus P Thompson. 1882. LI. On the function of the two ears in the perception of space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 13, 83: 406–416.

116. Shari Trewin, Brian Cragun, Cal Swart, Jonathan Brezin, and John Richards. 2010. Accessibility Challenges and Tool Features: An IBM Web Developer Perspective. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)* (W4A '10), 32:1–32:10. https://doi.org/10.1145/1805986.1806029

117. Nicole Ummelen. 1997. *Procedural and Declarative Information in Software Manuals: Effects on Information Use, Task Performance and Knowledge*. Rodopi.

118. Yolanda Vazquez-Alvarez and Stephen Brewster. 2009. Investigating Background & Foreground Interactions Using Spatial Audio Cues. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '09), 3823–3828. https://doi.org/10.1145/1520340.1520578

119. Markel Vigo and Simon Harper. 2013. Coping tactics employed by visually disabled users on the web. *International Journal of Human-Computer Studies* 71, 11: 1013–1025. https://doi.org/10.1016/j.ijhcs.2013.08.002

120. Elizabeth M. Wenzel, Marianne Arruda, Doris J. Kistler, and Frederic L. Wightman. 1993. Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America* 94, 1: 111–123. https://doi.org/10.1121/1.407089

121. Yeliz Yesilada, Giorgio Brajnik, and Simon Harper. 2009. How Much Does Expertise Matter?: A Barrier Walkthrough Study with Experts and Non-experts. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility* (Assets '09), 203–210. https://doi.org/10.1145/1639642.1639678

122. Yeliz Yesilada, Robert Stevens, and Carole Goble. 2003. A Foundation for Tool Based Mobility Support for Visually Impaired Web Users. In *Proceedings of the 12th International Conference on World Wide Web* (WWW '03), 422–430. https://doi.org/10.1145/775152.775212

123. William A. Yost. 1974. Discriminations of interaural phase differences. *The Journal of the Acoustical Society of America* 55, 6: 1299–1303. https://doi.org/10.1121/1.1914701

124. William A. Yost and Raymond H. Dye Jr. 1988. Discrimination of interaural differences of level as a function of frequency. *The Journal of the Acoustical Society of America* 83, 5: 1846–1851. https://doi.org/10.1121/1.396520

125. Pavel Zahorik, Douglas S. Brungart, and Adelbert W. Bronkhorst. 2005. Auditory Distance Perception in Humans: A Summary of Past and Present Research. *Acta Acustica united with Acustica* 91, 3: 409–420.

126. Xiaoming Zeng. 2004. Evaluation and Enhancement of Web Content Accessibility for Persons with Disabilities. Retrieved August 10, 2015 from http://d-scholarship.pitt.edu/7311/

127. H. Zhao, C. Plaisant, B. Schneiderman, and R. Duraiswami. 2004. Sonification of geo-referenced data for auditory information seeking: Design principle and pilot study. Retrieved August 28, 2016 from https://smartech.gatech.edu/handle/1853/50918

128. 2015. Americans' Internet Access: 2000-2015. *Pew Research Center: Internet, Science & Tech*. Retrieved December 26, 2017 from http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/

129. Convention on the Rights of Persons with Disabilities. Retrieved December 28, 2017 from https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/article-9-accessibility.html

130. Section-508-Of-The-Rehabilitation-Act | Section508.gov. Retrieved December 28, 2017 from https://www.section508.gov/section-508-of-the-rehabilitation-act

131. Disability Discrimination Act 1992. Retrieved January 9, 2018 from https://www.legislation.gov.au/Series/C2004A04426

132. Web Content Accessibility Guidelines 1.0. Retrieved December 28, 2017 from https://www.w3.org/TR/WCAG10/

133. Web Content Accessibility Guidelines (WCAG) 2.0. Retrieved December 28, 2017 from https://www.w3.org/TR/WCAG20/

134. WAVE Web Accessibility Tool. Retrieved December 29, 2017 from https://wave.webaim.org/

135. WHO | International Classification of Diseases. *WHO*. Retrieved July 19, 2017 from http://www.who.int/classifications/icd/en/

136. WHO | Visual impairment and blindness. *WHO*. Retrieved June 11, 2017 from http://www.who.int/mediacentre/factsheets/fs282/en/

137. Key Definitions of Statistical Terms - American Foundation for the Blind. Retrieved July 19, 2017 from http://www.afb.org/info/blindness-statistics/key-definitions-of-statistical-terms/25

138. Results of the 2016 GOV.UK assistive technology survey | Accessibility. Retrieved June 3, 2017 from https://accessibility.blog.gov.uk/2016/11/01/results-of-the-2016-gov-uk-assistive-technology-survey/

139. WebAIM: Screen Reader User Survey #6 Results. Retrieved June 3, 2017 from http://webaim.org/projects/screenreadersurvey6/

140. JAWS Screen Reader - Best in Class. Retrieved July 23, 2017 from http://www.freedomscientific.com/Products/Blindness/JAWS

141. NV Access. *NV Access*. Retrieved July 23, 2017 from https://www.nvaccess.org/

142. Vision Accessibility - Mac. *Apple*. Retrieved July 24, 2017 from https://www.apple.com/accessibility/mac/vision/

143. HTML5. Retrieved March 23, 2017 from https://www.w3.org/TR/html5/

144. Web Audio API. Retrieved August 29, 2016 from https://www.w3.org/TR/webaudio/

145. Omnitone: Spatial audio on the web. *Google Open Source Blog*. Retrieved March 23, 2017 from https://opensource.googleblog.com/2016/07/omnitone-spatial-audio-on-web.html

146. Navigating without Vision: Basic and Applied Research : Optometry and Vision Science. *LWW*. Retrieved June 18, 2017 from http://journals.lww.com/optvissci/Fulltext/2001/05000/Navigating_without_Vision__Basic_and_Applied.11.aspx

147. ATLAS.ti: The Qualitative Data Analysis & Research Software. *atlas.ti*. Retrieved April 26, 2018 from https://atlasti.com/