Towards a User Interface for Audio-Haptic Exploration of Internet Graphics by People who are Blind and Partially Sighted

Sabrina Marie Knappe

Department of Electrical & Computer Engineering

McGill University

Montréal, Québec, Canada

December 15, 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of

Master of Science

Abstract

Graphical media make up a large part of the Internet content, yet are largely inaccessible to people who are blind or partially sighted. This makes many websites less appealing to these users and is a large gap in current web accessibility. This thesis has three components, all contributing to the development of a system allowing users who are blind and visually impaired to interact with images on the Internet. The first is a survey of blind individuals in Canada about their habits and needs with regards to interacting with images on the Internet, and qualitative analysis of subsequent interviews to better understand their perspectives. The second is preliminary work on overall system design based on initial communication with users, and iterative design enhancements of the user interface of the system. The third is an experiment gauging the efficacy of a system that provides spatialized audio representations of photographs. The wide diversity in user abilities and desires were challenging aspects of system design, and ultimately we found that while audio spatialization is a suitable solution for some users, it will not fulfill the needs of every user.

Abrégé

Les médias graphiques constituent une grande partie des contenus Internet, mais sont en grande partie inaccessibles aux personnes aveugles ou malvoyantes. Ceci rend de nombreux sites Web moins attrayants pour ces utilisateurs et constitue une lacune importante dans l'accessibilité actuelle du Web. Cette thèse comporte trois éléments, qui contribuent tous au développement d'un système permettant aux utilisateurs aveugles et malvoyants d'interagir avec les images sur Internet. Le premier est une enquête auprès de personnes aveugles au Canada et à l'étranger sur leurs habitudes et leurs besoins en matière d'interaction avec des images sur Internet, ainsi qu'une analyse qualitative des entretiens ultérieurs pour mieux comprendre leurs points de vue. Le deuxième est un travail préliminaire sur la conception globale du système basé sur la communication initiale avec les utilisateurs, et des améliorations itératives de l'interface utilisateur du système. Le troisième est une expérience visant à évaluer l'efficacité d'un système qui fournit des représentations audio spatialisées de photographies. La grande diversité des capacités et des souhaits des utilisateurs a été un défi pour la conception du système, et nous avons

finalement constaté que la spatialisation audio est une solution adaptée à certains utilisateurs, mais elle ne répondra pas aux besoins de tous.

Acknowledgements

A brief thanks to all who supported me in this endeavour. My parents put up with me during some of my hardest moments and encouraged me when I thought things were hopeless. My grandparents provided me with a warm place to come home to, and sharp wits to bounce ideas off of. My sisters, though geographically distant, were always a phone call away to inspire me with their joy and creativity. My church community has been a rock in difficult times and a celebrant in times of jubilation. The project of my education would not have been possible without all of these wonderful people, and I am very thankful to have them in my life.

I would also like to express my gratitude for my academic colleagues, who worked tirelessly alongside me on this project. To Jeremy, who always pushes me forward and challenges me to do better, Cyan, whose reliability is matched only by her graciousness, and Juliette, my partner in crime. You have all been instrumental in getting me to the finish line. Thank you.

Contents

List of Figures

List of Tables

List of Acronyms

AI	artificial intelligence.
API	application programming interface.
ARIA	Accessible Rich Internet Applications.
BPSP	blind and partially sighted people.
CCB	Canadian Council of the Blind.
CCTV	closed-circuit television.
CNIB	Canadian National Institute for the Blind.
DOF	degrees-of-freedom.
DOM	Domain Object Model.
GUI	graphical user interface.
HCI	human-computer interaction.
IDE	integrated development environment.
IMAGE	Internet Multimodal Access to Graphical Exploration.
ML	machine learning.

NFB	National Federation of the Blind.
OCR	optical character recognition.
RAAMM	Regroupement des aveugles et amblyopes du Montréal métropolitain.
REB	Research Ethics Board.
SNS	social networking services.
TTS	text to speech.
UCD	user centred design.
UI	user interface.
UX	user experience.
VR	virtual reality.
WCAG	Web Content Accessibility Guidelines.

Student Contributions

This thesis contributes to the existing body of work in the following ways. First, it provides an overview and analysis of Internet use by blind and partially sighted individuals in Canada, using information obtained through a survey of users. Second, it outlines the iterative design of user interfaces deployed to enable blind and partially sighted users to access Internet graphics through haptic and audio media. Third, a user study evaluating the efficacy of audio-haptic and audio interpretations in conveying spatial information in images is described and its results are presented and discussed.

Chapter 1

Introduction

The Internet is an increasingly visual medium [1]. Even websites that were previously considered accessible, such as Twitter.com, are becoming more and more image heavy [2]. The increased use of images in posts is implicitly encouraged, because posts with images on social media see more interaction than their text-only counterparts [3]. A side effect of this change is the reduced accessibility of websites to blind and partially sighted people (BPSP) who often choose to avoid images altogether rather than try to make sense of incomplete and inaccurate information [4].

The most common way to make images accessible is alternative text captions (alt-text) which can be added by the individual posting an image at the time of the image upload. Users rarely take the time to write these captions, however, so auto-captioning services that use machine learning to identify image contents are frequently used to supply missing captions

[5]. Some screen readers, such as JAWS, have built in functionality for providing image captions [6]. Unfortunately, these computer-generated captions are often inaccurate and incomplete. Furthermore, they do not allow BPSP to interact with an image the way sighted users do. When a sighted user encounters an image on the web, they can quickly glance at it to understand its general contents and then choose to examine the image more closely. For BPSP, there is only a single description which doesn't allow for further interaction. Prior work aiming to make graphics accessible and interactive to BPSP has had several limitations: It was either largely theoretical and not implemented, implemented exclusively for touchscreens, or targeted a specific subset of graphical media, such as graphs and charts.

The Internet Multimodal Access to Graphical Exploration (IMAGE) project is the Shared Reality Lab's effort to provide exploratory access to visual media on the Internet from a laptop or desktop computer. It consists of a Chromium browser extension that allows users to select an image on a web page, and sends that graphic to a server that interprets it and returns a text, an audio and an audio-haptic rendering [7]. Its high-level architecture affords a significant amount of flexibility, since the components that deal with the extraction of information from the image (processors), and transform that information into media consumable by BPSP (handlers) are modular and open source. Developers and designers are free to create their own processors and handlers to suit their particular needs. For example, a company wishing to sonify product images on their website might make a preprocessor that identifies images hosted on the website and presents users with a sound clip of the product

colours represented by pitch. The results of all the renderings are delivered to users in a "renderings window" that allows them to interactively navigate between sections of an audio file.

The open-ended nature of such a system presents a challenge when considering how to allow users to access the renderings through a web browser. Like the underlying architecture, the interface needs to be flexible, allowing many different types of rendering to be presented and explored. Designing the interface and user experience is further complicated by the variety in needs and abilities of the target user base. A person is considered legally blind if they "have a visual acuity of 20/200" even with corrective lenses, or if they have a highly restricted field of vision (less than twenty degrees across) [8]. In practice, this means that some individuals who are legally blind may still be able to read printed type or see images if magnified. At the other end of the spectrum, there are those with no vision, who cannot make any use of visual representations. Comorbidities such as hearing loss can further complicate the needs of a user.

This thesis consists of work that was a part of the IMAGE project, from the design to evaluation phase. The rest of this thesis proceeds as follows: First, literature relevant to the work is reviewed. Second, the results of a survey and subsequent follow-up interviews are analyzed and discussed. Next, the iterative development of user interfaces for the IMAGE system are summarized. Lastly, the manuscript of a user study evaluating the efficacy of the system on the spatialization of photographs is presented, followed by a discussion of the

body of work and the conclusions and consequences for future work on the project.

Chapter 2

Background

There have been many efforts made to increase accessibility of images for BPSP, ranging from crowdsourcing efforts for captions, to sonification of the colours contained in images [9]. To date, there is no widely adopted solution by the BPSP community. This chapter will outline the psychological realities that make representing images for BPSP challenging and explore how other researchers have applied this knowledge to the creation of interfaces and image representations for BPSP.

2.1 Mental models of visual information

How do BPSP understand visual media? This is often dependent on when an individual lost their vision. For those who are congenitally blind, visual stimuli never make their way to the visual cortex. Instead, this area is developed according to input from the auditory

system [10]. The congenitally blind have no point of reference for visual stimuli, and so may struggle with visual concepts such as colour and perspective, which are obvious to sighted individuals [11] [12] [13]. Conversely, those who lost sight later in life benefit from both development in their visual cortex and the enhanced auditory and tactile acuity found in the congenitally blind [14] [15].

Raised-line drawings are a very close haptic analog to visual drawings, and are used to study how BPSP understand visual methods of representing objects. There have been efforts to improve the ability of congenitally blind individuals to produce and recognize line drawings of common objects, but this requires a significant degree of learning [16]. A study did show, however, that raised-line drawings of faces of various emotions were correctly identified by congenitally blind adults [17]. Basic shapes such as circles, squares, and triangles are also easily identified [18]. More complex raised line drawings which precisely copy visual images are unlikely to be interpretable to congenitally blind BPSP [19, 20]. Systems aiming to represent visual media need to take into account these perceptual limitations if they want to be useful to the congenitally blind.

For BPSP who lost vision later in life, and even those who lost vision as children, there is far more understanding of visual concepts [21]. These individuals may be able to both identify and create raised-line drawings in the same way as sighted people. For some tasks, those who go blind later in life have an advantage over both the early blind and the sighted. Heller found that those who went blind later in life could identify pictures in either raised-

line or embossed format much more quickly than early blind and sighted participants [22]. A wider variety of audio and haptic representations may be useful to these individuals as a result of the mental models they acquired early in life.

2.2 Interfaces for the Blind and Partially Sighted

To understand the challenges involved in designing an interface for BPSP it is crucial to understand how BPSP browse the Internet, and more generally, use computers. Screen readers have been the primary avenue of access to computers for BPSP since the 1980s, when IBM created the first screen reader for use with DOS [23]. A screen reader, at its core, is an interface that allows the user to go line by line through the Domain Object Model (DOM) structure of an HTML page and hear the name of the part of a hierarchy as well as the text held within. For example, if a user were to tab onto a first-level header (H1), the screen reader would first read out "header level one" and then read out the header text. Screen readers are often supplemented by a refreshable Braille display, which displays the selected text in Braille form. BPSP using refreshable Braille displays may choose to use them in conjunction with or instead of the speech output of the screen reader, although the screen reader is generally what communicates the text information to the display. These technologies developed when computers were almost entirely text-based interfaces, and over time have had to be supplemented by additional technologies such as Accessible Rich Internet Applications (ARIA) to deal with more dynamic web content [24]. ARIA provides accessible

widgets to developers such as modals and disclosures, as well as the ability to name regions and landmarks on web pages, even in DOM structures that do not provide a native field for alt-text.

This method of presentation of Internet content is linear in nature, and the user can only inspect a single element at a time. This is very different from the way that sighted people explore a website. Sighted users "scan" web pages. This applies to both the text content and the page overall. Whereas a screen reader reads out a web page sequentially, sighted users skip around a lot, both within blocks of text and around the entire web page [25]. While screen readers do have built-in functionalities to skip to headers and other potential points of interest on a web page, they do not give BPSP the awareness of peripheral content that sighted users have [26].

This limitation has spawned several efforts to augment or replace screen reader technologies. These were especially common in the eighties and nineties, before screen readers had crystallised as the primary method for human-computer interaction (HCI) for BPSP. Edwards' 1987 thesis proposed an interface composed of adjacent blocked regions on a computer screen that when passed over by a mouse would sound a tone representing a certain button [27]. While the concept seemed feasible, it was not widely adopted, potentially because of the difficulties users had with the mouse. Traditional screen readers have the benefit of not relying whatsoever on mice, which are difficult for BPSP to use because they provide no haptic feedback as to where objects are located on a screen; they

are forced to operate in a void. After screen readers solidified as the dominant technology used by BPSP to access computers, researchers began investigating the usability of various interfaces when accessed through a screen reader. Kurniawan, Sutcliffe, Blenkhorn, and Shin explored the usability of Windows computers when used with LookOut, a screen reader offered by Microsoft at the time [28]. They found that users have great difficulty using a system when it does not adhere to their previously established mental model, which may include capabilities of the screen reader they habitually use.

Some researchers attempted to create interfaces that would be accessible to both BPSP and sighted users according to principles of universal design. Savidis and Stephanidis created a user interface management system to help developers make interfaces that would work for BPSP and the sighted [29]. Their strategy was never widely adopted, although some of their concepts showed promise. Their picture exploration application created a hierachical model of a photograph which BPSP could explore. The collaboration module allowed blind users to work remotely with sighted users, facilitated by the translation of an abstract hierarchy (similar to a DOM) into either sighted or blind metaphors. Both applications were well received. Bouraoui created widgets that can be incorporated into an integrated development environment (IDE) [30]. Preliminary use by a blind developer revealed that such an approach was practical and a potentially fruitful avenue for continued exploration. Emery et al. used a universal design philosophy in their development of a multimodal feedback system to assist users in a drag-and-drop task on a computer [31]. They found that auditory feedback

enhanced the performance of older adults, even those who were sighted. Sjostrom reviewed several additional researchers' attempts at creating haptic systems as alternatives to the traditional screen-reader paradigm [32]. Approaches included force feedback to haptically render graphical elements, software for haptic mice, and a mouse with pins allowing BPSP to scan images on their computers. Leuthold, Bargas-Avila, and Opwis built an enhanced text user interface to augment the usual graphical user interface (GUI) of web pages by increasing the labeling and navigation features available [33]. They devised a set of guidelines based on HCI principles that they used to adhere to user-centred design methodologies in their development process. As a result, their interface was well received by users.

Many researchers have created frameworks and guidelines for the development and evaluation of user interfaces for BPSP. Morley's 1999 thesis presented methodology for the design and evaluation of non-visual interfaces [34]. She proposed a variety tasks for other researchers to use in evaluating non-visual interfaces with users. Casali emphasized the importance of choosing interfaces that match the capabilities of target users [35]. Fukuda, Saito, Takagi, and Asakawa proposed the metrics of listenability and navigability as two potential new metrics for web usability [36]. Navigability is the extent to which BPSP can understand and navigate the structure of a web page, based on the time it takes to reach target elements, whether headings or skip links [37] to main content are present on a page, what percentage of links on a page are accessible, whether HTML form elements are appropriately used, and whether tables are used as actual tables or to define a page layout.

Listenability has to do with the use of text on a web page, and has four components. The first is whether alt-text has been used appropriately, the second is whether there is any text that is repeating due to the way alt attributes are used, and last is whether the spelling of words or character spacing of letters within words has been modified for stylistic reasons that may interfere with a screen reader. These two metrics are a useful rule of thumb when designing an streamlined interface for BPSP to use on the web. Alonso, Fuertes, Gonzalez, and Martinez produced two papers on interfaces for BPSP. The first explained the connections between the user requirements of BPSP, HCI principles, and a set of guidelines they produced [38]. The second presented a toolkit for creating interfaces based on their guidelines [39].

Non-standard interfaces have also been considered for a variety of applications to help BPSP. Tzovaras et al. designed a virtual reality (VR) system with the goal of allowing BPSP to interact with virtual objects in 3D space [40]. They used a CyberGrasp glove as their haptic device and created several test environments that allowed users to manipulate and touch objects in virtual space. The initial feedback was positive. Tanaka and Parkinson attempted to create a multimodal interface to help BPSP who are audio producers work with audio waveforms in computer programs [41]. They developed a custom haptic device that would allow users to feel the waveforms as they scanned them with a knob. Users found it useful as it integrated into their everyday workflows.

The literature on interfaces for BPSP shows that typical visual interface strategies must

be adapted or replaced for a non-sighted user group. This can be done through completely novel interface strategies, although these can be difficult for users to learn, or it can be done through proper use of existing web conventions. Interfaces that use multiple modalities can help substitute or supplement the presentation of information.

2.3 Image Representations for Blind and Partially Sighted People

Alt-text, as previously mentioned, is the default method of representing images on the Internet for BPSP. Representations that go beyond a simple verbal description of an image are called "rich representations" and may combine sound, haptic, and even olfactory stimuli in order to represent visual media [42].

Morris, Johnson, Bennett, and Cutrell provide a taxonomy of the design space for rich representations of images in their 2018 paper [42]. The taxonomy consists of five "categories" that describe choices made when designing a non-visual image representation. The first is "interactivity." A representation is either "passive" or "active"—can the user determine what representation they receive in real time? The second is "stability". Is the representation produced "static" or is it "evolving". Thirdly, "representation" refers to the media used to convey the image to the user. These can include sonic as well as haptic methods. Fourth, "structure" refers to the existence of an overarching form that dictates

the organization of a representation. Image representations are either "structured" or "unstructured". Finally, "personalization" is whether or not representations are customizable- representations are either "generic" or "personalized". This taxonomy is useful for characterizing existing projects that represent images for BPSP.

Previous attempts to represent images in a non-visual manner generally fall into one of two categories: audio or haptic. Because haptics are largely insufficient to communicate the amount of information in a piece of visual media, they are often combined with some kind of audio feedback, and thus would be considered "audio-haptic" representations.

2.3.1 Audio Representations

The most basic audio representation of an image is a verbal description of its contents. There is a good deal of literature on how to best describe images. Girgis' Masters' thesis reviewed several authors' attempts at providing captions for images and used them to formulate guiding questions for describing images to BPSP [43]. These questions can be roughly summarized as "who", "where", and "what". Users wanted to know who was in an image, where the image was taken, what the subjects (people, animals, etc.) of the image were doing, and what the emotional valence of the image subjects was. Some form of captioning is almost always required to give BPSP a complete understanding of an image, so although these approaches are incomplete in terms of interactivity, they are still useful.

Some researchers have simply attempted to provide more information about visual

media to BPSP in text form. Wu, Wieland, Farivar, and Schiller used artificial intelligence (AI) to improve the automatic alt-text available to BPSP on social networking services (SNS) [44]. Low et al. created a Twitter extension that uses a combination of optical character recognition (OCR), finding matching images that do have captions, and crowdsourcing captions to provide better descriptions to users [45]. Choi et al. used machine learning (ML) techniques to extract the raw data from line graphs, bar charts, and pie charts, then rendered them in HTML table format that BPSP could navigate with their screen reader [46]. While increasing the amount of information available to BPSP is a good start, these text-based approaches are not an adequate substitute for visualization, as they do not communicate multidimensional information in a concise manner the way visualizations do.

When captions or text-based descriptions are insufficient to convey the information present in an image, sonification techniques may be used. These techniques involve representing certain data in an image as sounds, creating a sort of sonic language for visual characteristics. One common approach is the sonification of numeric data which replaces visual representations such as line graphs, bar charts, and other methods of numerical data visualization with sound [47]. HighCharts, a data visualization API commonly used on the web includes a sonification function that turns either single points or lines into tones whose pitch indicates the value at a given point [48]. Grond and Hermann explored the possibility of sonifying mathematical functions using vowel sounds [49]. The vowel sounds indicated a

variety of commonly seen graph shapes, and the changes between two vowels indicated inflection points in the graph. This technique had mixed results, with users often not being able to distinguish between different curves.

Another common approach is to communicate colour and shape information about a picture by mapping the aspects of colour to sound characteristics. Cavaco et al. made a system that transformed the colour properties of images into sound [50]. Pitch was mapped to colour, saturation to timbre, and value to perceived loudness. The pixels of the image would then be played back to users who could reconstruct an idea of the colours and shapes present. Users had some difficulty distinguishing colours with neighbouring frequencies, but it is possible this could be remedied with additional training. Toff and Mignote took a similar approach, but presented the colour contents based on regions that users could explore via a touch-screen [51]. Ferwerda and Kwok also used colour sonification techniques, but exclusively for Matlab via a function that displays data points with a colour scale [52].

When dealing with images that convey qualitative or affective information, soundscapes may be the correct sonification approach. Winters, Joshi, Cutrell, and Morris combined multiple approaches to create sonic representations of Twitter posts containing images. They used soundscaping, earcons, background music, text to speech, and sonification, each representing a different aspect of a post. The demo of their system was promising, but they never deployed it or tested it with BPSP. Rector et al. created audio interpretations that could be used by BPSP in a museum setting to enjoy art [53]. This system was proxemic

and played audio representing different aspects of an image depending on how close an individual was to the image. First, an individual would hear background music that represented the overall aesthetic qualities of the painting. As they approached, they would hear sonification representing the colours present, sound effects forming a soundscape of the objects and subjects present in the painting, and finally a verbal description of the painting. This strategy was very well received by users. Both of these projects took on the difficult task of sonifying images that have multiple layers of meaning and may interest people for different reasons.

When there is an option for a visual display, audio is often sidelined, since it has a much narrower bandwidth in terms of the information it can communicate [54]. The problem of presenting images to BPSP motivates researchers, designers, and engineers to step away from the visual frame and consider how information might be communicated in creative and novel ways. However, the limited bandwidth of auditory information often prompts the use of additional modalities, such as haptics, to increase the amount of information that can be conveyed [55].

2.3.2 Haptic and Audio-Haptic Representations

The earliest attempts to represent images for BPSP through a haptic medium were largely analog. These include raised-line drawings, 3D printed maps and paintings, and specially produced maps [56]. All of these methods have been substantially refined, and are widely

accepted by the BPSP community. Unfortunately, these analog representations are costly and time-consuming to produce, making them inaccessible to many BPSP. Digital representations of visual media have the potential to be more economical, and quickly provide BPSP with an idea of the contents of an image. There are currently four main haptic and audio-haptic approaches to rendering images:

Force Feedback

Force feedback devices consist of an end-effector which is held in the hand of the user and can exert force back against the user through a combination of motors and joints. These devices can be used to simulate surfaces or textures that the users feels or runs up against by moving the end effector. Zhang, Duerstock, and Wachs used a Force Dimension® Omega 6 force feedback device to render histological images for BPSP [57]. Force feedback can also be useful in rendering data, such as vectors, lines, and other constructs usually graphed visually. The PHANToM force feedback device is a popular choice for these renderings, and was used by multiple researchers [55] [58]. Yu, Reid, and Brewster made use of the Logitech Wingman force feedback mouse to enable BPSP to interact with charts and graphs [59]. Jeong also used a force feedback mouse to allow blind users to touch images and Braille cells found on a computer [60].

Pin Arrays

Pin arrays are similar to a regular braille array commonly used by members of the blind community. These consist of a tablet whose surface is covered with cells of pins that can be raised or lowered to create contours, braille, or textures. The Orbit Graphiti® is an example of one such device available commercially [61]. Researchers have investigated how graphics can be rendered for BPSP with these devices. O'Modhrain et al. reviewed several other efforts using pin arrays, along with other haptic methods of rendering visual media and concluded that pin arrays are the best approach for rendering images graphically for exploration by hand [62]. Rastogi, Pawluk, and Ketchum used a mouse mounted with a small eight-pin display to render graphics [63]. Their system allowed users to "zoom" into various areas of an image with a press of a button, an attempt at solving the resolution problem common to pin arrays that O'Modhrain et al. identified.

Vibrotactile Displays

Vibrotactile motors are likely the most widely used type of haptic motors, and are commonly found in cellular phones and video game controllers. The vibrations of these small motors at different amplitudes and frequencies are used to either simulate natural phenomena or communicate information [64]. Wacker et al. built a vest that uses an array of vibrotactile motors to signal the proximity of objects to the wearer [65]. The closer the object to the wearer, the stronger the motors would vibrate. Zhao et al. used vibrotactile motors placed

on the hand in conjunction with a tablet interface to enable graphical exploration [66]. Users were given directional and progress cues so that they could explore line graphs. Overall, the device enabled participants to fairly accurately follow shapes and graphs. Palani et al. used vibrotactile feedback within a tablet for a similar task and determined the best arrangement and strength for vibrotactile cells for line detection, line discrimination, and orientation discrimination tasks [67].

Touch Screens and Tablets

Finally, some researchers have attempted to convey information about visual media using touch screens. These have the advantage of already being widely distributed among the BPSP community [68]. Tantribeau's 1992 thesis is an example of some early work on this approach. He used a tablet and two styluses that users could move around the tablet to hear the pixels at their locations in audio form [69]. Zhong et al. used crowdsourcing to tag images and then built a system that enabled users to explore images with their hands on their phones' touchscreens [70]. When a user passed an object on the phone with their finger, it would vibrate and a description of the object would be read out.

The primary obstacle to the distribution of audio-haptic representations to BPSP is the prohibitive cost of the specialized hardware required [62]. Commercially available pin arrays can cost over ten thousand dollars, about a sixth of the median annual income in Canada [71]. The second obstacle is the lack of widely available software for displaying graphics through

a haptic device. The vast majority of the projects discussed above were never deployed for public consumption, so even if BPSP had access to a haptic device, there would be no support for its use. The haptic system that is the most used in the blind community is VoiceOver's object detection capabilities on Apple's iOS devices [68]. Users can pass their fingers over an image, and VoiceOver will read out what is present in the image at the location of their finger [72]. This approach is functional, but limited, and BPSP would benefit from a more diverse set of options of haptic access to graphics [62].

Chapter 3

Understanding the wants and needs of users

The first task when deciding how to construct a system that would allow BPSP to explore Internet graphics was to talk with members of the blind and partially sighted community to understand their habits, needs, and desires. Due to COVID restrictions at the time, we were unable to meet with them in person and observe their workflows, and instead had to rely on the habits users reported to us. These can be compared against literature on computer habits of BPSP to help contextualize our results.

We started with a handful of informal conversations with BPSP to better understand the state of efforts to make graphics more accessible and the attitude of the community towards them. Several of these were held with David Brun, our primary contact in the BPSP

community, while two others were held in a group setting with community organizations. It was important to us to make early connections with users to help set the direction of the IMAGE project. These discussions helped us identify which questions to ask in our survey and semi-structured interviews.

3.1 A Survey of Blind and Partially Sighted People

After gaining a preliminary understanding of the potential design space and community attitudes towards assistive technologies used with computers, we distributed a survey to the blind community through the Canadian Council of the Blind (CCB), a community organization that has chapters across Canada and seeks "to promote the well being of those who are blind or have low vision" [73]. The survey was distributed through their email newsletter, with both French and English language options available.

3.1.1 Methods

The survey was developed by the author through consultation with various team members as to what they felt was necessary to know, as well as investigation of literature to find common metrics taken when surveying the BPSP community. The questionnaire was hosted on Microsoft Forms because of its accessibility to BPSP and its compliance with McGill University regulations regarding data storage. McGill's Research Ethics Board (REB) approved the administration of the questionnaire under REB #21-04-010.

Participants were asked basic demographic questions relating to age and type of vision loss, as well as questions about their weekly Internet habits, the devices they use to access the web, and what they find most difficult about accessing web graphics. In order to maximize accessibility and minimize complexity, all questions were some form of multiple choice, with the exception of a handful of short answer questions. The survey was tested for screen reader compatibility by team members, including David Brun, before being distributed.

3.1.2 Results

Quantitative Results

There were 63 survey respondents, 59 for the English language option and 4 for the French language option. One respondent's data was removed because they were not located in Canada and the survey was targeted towards Canadian residents. Of the 63 respondents, 36 identified as blind, 21 identified as visually impaired, 2 identified as deaf-blind, and 1 identified as blind and hearing disabled (see Figure 3.1). Two participants did not provide any information about their identified disability. Less than half of our respondents (26) were under 50 years in age. The largest age bracket of respondents was those aged 65 to 75 years. Additional details on the age distribution of our participants can be seen in Figure 3.2. Forty-eight percent of respondents were Braille-literate. Most respondents said that they use a web browser daily to interact with websites that contain graphics. Forty-five percent of respondents said they do this many times a day.

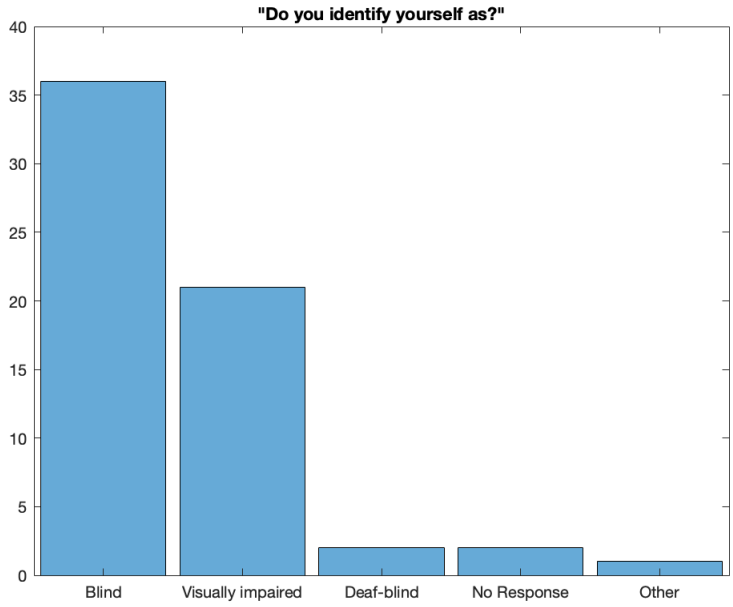


Figure 3.1: Disability identification of survey respondents.

Over half of respondents said they use an Apple iPad, but less than a quarter make frequent use of the device. Smart speakers were frequently used by about half of respondents, but wearables such as a smartwatch were only frequently used by approximately a quarter of respondents. Detailed information on the frequency of device use can be found in Figure 3.3. For the 11 respondents who said they used another type of device than was listed in the survey, 4 mentioned Braille displays, 2 mentioned Victor audio players, 1 said they used an Apple Watch, 1 used a Linux desktop computer, and 1 mentioned using a closed-circuit television (CCTV), a device used to magnify screens. One respondent said they take out borrowed or public devices (such as those available through accessibility services in a library) but did not specify which devices they use. Ninety-six percent of respondents did not have

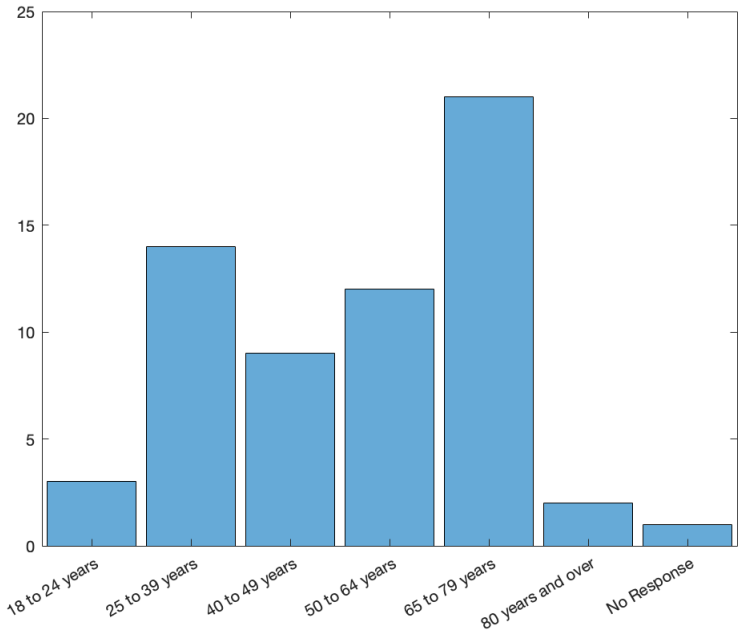


Figure 3.2: Age distribution of survey respondents.

easy or affordable access to a 3D printer.

Respondents were nearly unanimous that access to graphics is a problem for their community. All but one of the respondents felt that they missed out on information contained in visual media on the Internet at least some of the time. Respondents were fairly mixed on whether or not accessibility features enhanced their understanding of web graphics, however. Approximately one third felt that accessibility features were most often helpful, a quarter felt they sometimes enhanced their understanding, and approximately another quarter felt they either rarely or never helped. This question was rather open, and whether participants were including the availability of accessibility features in their answer is unclear. For example, it is possible a respondent might have chosen “rarely” as their

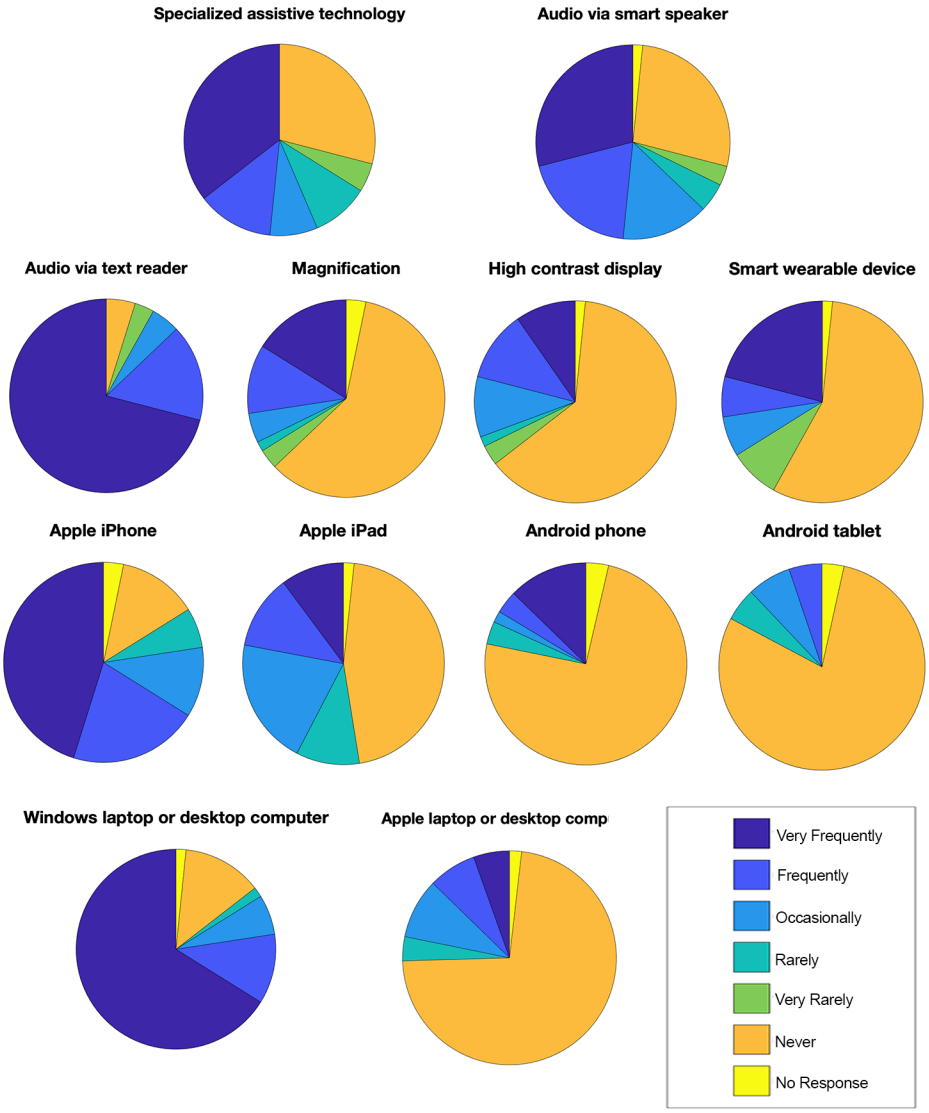


Figure 3.3: Frequency of use by device.

response because they rarely encounter proper accessibility features on the web.

Qualitative Results

We asked users to describe a specific instance in the last month when they encountered a web graphic they found frustrating or confusing. Forty-five of the 62 respondents answered the question. Violations of accessibility guidelines were the most commonly cited problem, with 12 respondents mentioning some sort of violation. Half of these were due to difficulties with contrast and colour, problems that are common for partially sighted users. The other half related to missing captions, improper use or labeling of images, or moving image contents. All of these problems are a result of website designers, programmers, and users violating Web Content Accessibility Guidelines (WCAG). Visual renderings of numerical data were the second most frequently mentioned problem, with graphs, charts, and tables being mentioned by 10 respondents. Some respondents mentioned that not being able to access these graphics affected their ability to do their work properly, highlighting a very serious consequence of image inaccessibility.

Annoyance, frustration, and the feeling of being left out were common sentiments. One respondent said they "hate hearing Image Image Image," referring to the feedback a screen reader provides when there is an image present but no description given. Another user said they "needed sighted assistance" because they could not find what was likely an improperly labeled submit button. Two users mentioned that they did not like that they were missing out on information that others were privy to.

3.1.3 Discussion

Our respondents were significantly more Braille literate than the BPSP community on average with 48% of respondents saying they read Braille. The common wisdom in the community is that 10% of BPSP are Braille literate, but estimates based on surveys of Braille literacy in the United States and elsewhere range anywhere from 2% to 25% [74]. Sheffield et al. suggest that differing definitions of the words "blindness" and "literacy" and inconsistency in populations surveyed may be the causes of the wide range of estimates [74]. The discrepancy in Braille literacy rates between our sample group and the community at large is important to keep in mind because of the differences in cognition between BPSP who are Braille literate and those who are not. The age distribution of our respondents skews significantly older than the age distribution of the general population, but it is in line with the age distribution of the blind community itself since visual disabilities are comorbid with age [75]. This has consequences for the development of assistive technology, given the declines in cognitive, tactile, and auditory acuity with age [76]. There were not enough francophone respondents to the survey to determine whether there are any disparities in experience between francophone and anglophone BPSP in Canada. The large difference in number of respondents by language is likely a consequence of the primarily anglophone constituency of the CCB, which is headquartered in Ontario.

Overall, the habits of BPSP in our respondent group aligned with known community

trends. WebAIM conducts a regular survey of desktop screen reader users. Their survey has consistently shown that VoiceOver, NVDA, and JAWS are the most popular screen readers [77] . Like in our sample, they also found that JAWS and NVDA are significantly more popular than VoiceOver. This is likely related to the preference for Windows computers, although it is difficult to know if the relationship between these trends is causal, and if so, in which direction. Our results were also consistent with the common wisdom regarding the ubiquity of the Apple iPhone as the preferred mobile device of the BPSP community, with 66% of respondents saying they used an iPhone frequently or very frequently. The Apple iPad appears to be a fairly popular secondary device, since it is used by the majority of respondents, but not at the same frequency as the iPhone or a Windows desktop computer.

The data surrounding device usage also aligned with respondents' self-identified disabilities in expected ways. Twenty percent of visually impaired respondents were Braille literate, compared to 66% of respondents who identified as blind. Assistive technology to increase contrast of magnified images were much more popular among those identifying as visually impaired. Over 90% of those who identified as blind never used magnification, whereas 70% of those who identified as visually impaired used magnification either frequently or very frequently.These results reflect the tendency of those with little to no vision to identify as blind, while those who still have usable vision are more likely to identify as visually impaired. All deaf-blind respondents were Braille literate.

The results of our survey confirm long-held wisdoms about the preferred technologies of

the BPSP community and reaffirm the need for better access to graphical media online. The prevalence of use of specialized devices suggests that BPSP would be open to using novel devices if they improved their access to graphics, as the majority of respondents indicated that they feel they are missing out due to lack of access to graphical media. Providing blueprints for use with 3D printers, an option discussed by the team, was found to be ulikely to improve accessibility due to lack of easy access to such devices by members of the community. The differences in device usage between those who identified as blind and those who identified as visually impaired indicate that the ideal solutions for access to graphics for those two groups are likely to differ.

3.2 Semi-structured Interviews

3.2.1 Methods

Respondents were asked if they would be open to participating in a follow-up interview as part of the initial survey. 54 of the 62 respondents agreed to participate and of these, 8 were interviewed. The interviewees were chosen to get a reasonable demographic spread across age and gender. Our primary objectives were to ascertain the daily Internet browsing habits of users, as well as what types of images were most interesting to users, and what they wanted to know about the images that interest them. The interviews were conducted in a semi-structured manner using either telecommunications software such as Zoom or over the

phone, depending on user preference. A script of questions was used, but participants were allowed to direct the conversation if they had topics in particular they wanted to discuss. The questions participants were asked are shown in Table 3.1.

Participants were also sent audio samples created by our prototype system to review in advance of the interviews so that they could provide us with feedback on our strategy. The audio samples consisted of one image rendered using the semantic segmentation technique, and two images rendered with different variations of the object detection sonification technique. Object detection refers to a machine learning technique that identifies objects in an image and provides labels, bounding boxes, and centroids for these objects. Semantic segmentation provides labels and coordinates forming a contour around regions of an image. The object detection images were sonified by associating objects with a "popping" noise, whose pitch or volume was varied to indicate size or height within the image. In order to associate the "pops" with their respective objects, the label for the object would be read out by text to speech (TTS) software. The semantic segmentation samples used a buzzing noise to trace the contours of the regions identified by the semantic segmentation module. Interviewees were asked their opinions on the audio samples, specifically which they enjoyed the most, which they found the most and least intuitive, which was the clearest, and which they would want to use in their everyday lives.

Topic	Question
Nature of Disability	Describe your level of visual impairment.
	Are you congenitally blind? If not, around what age did your vision become impaired?
Current Behavior	How do you interact with electronics and the Internet during an average week?
	Which device do you use most frequently?
	What website or application do you use most frequently?
	Do you use audio on websites?
	How do you access and interact with audio content on the web?
	Has website audio ever interfered with your ability to navigate the web? If so, please describe a time when it happened and how you dealt with it.
Attributed Emotions	What feelings do you associate with your interactions with graphical media?
	What could make these interactions more positive?
	What is the most positive interaction with graphical media you've had?
Desired Outcomes	What do you do when you encounter a web graphic? When you encounter graphical media, what do you want to know about it?
	Do you ever ask sighted friends or family for additional context? Why?
	Describe a scenario in the past month in which you had a frustrating encounter with graphical media.
	Imagine that you have been given basic information about the image you just described. What else would you like to know about the image?
Brainstorming	Imagine your ideal experience interacting with graphical media. What would it be like?

Table 3.1: Semi-structured Interview Questions

3.2.2 Results

Eight survey respondents were interviewed (2 female, 6 male). Of these participants, 1 identified as visually impaired, 6 identified as blind, and 1 identified as blind with a hearing disability. One interviewee was between the ages of 18-24, two were between the ages of 25 and 39, two were between the ages of 40 to 49, one was between the ages of 50 to 64, one was between the ages of 65 and 79, and one was 80 years or older.

Current Behaviour

The interviewees used a variety of strategies when interacting with graphical media on the Internet. The participant who identified as visually impaired used a combination of a screen reader and zoom technology to navigate the web. They encountered difficulties in using both the Microsoft narrator and Microsoft zoom software simultaneously.

Many pain points for interviewees were attributable to violations of guidelines for accessible web development. These include improperly labeled buttons, images without provided captions, parts of web pages that move around, and image links. Lack of captions on images was such a frequent problem for users that some completely ignored any images they encountered, since they so often found them to be a waste of time.

Some users reported using their cell phones for image recognition. Both Android and iPhone devices had some capabilities to inform users of what was in an image, although the iPhone was used by more of the interviewees.

Attributed Emotions

Interviewees described being sad, disappointed, frustrated and annoyed when they encounter images on the Internet. One interviewee could not think of a positive experience they had with graphics on the Internet. For congenitally blind interviewees in particular, there is mystery and confusion associated with imagery. One user explained that they are aware that images allow the sighted to take in complex information very quickly, and expressed a desire to have a similar experience.

When our interviewees asked a sighted friend or family member about what was present in an image, they were usually asking for a quick and to the point description of the relevant details. The advantage of having a human describe the image to them is that they could make judgements about what aspects of the image were relevant, and also answer questions about details of an image the interviewee might be interested in. The disadvantage was the lack of independence and the additional time needed to get help. Interviewees saw sighted help as something to be used only when necessary.

Desired Outcomes

Users were very interested in the people in images. Many of their encounters with visual media were in an SNS context, where the emotions, actions, and demographics of the people in an image would be very important. Colour was also mentioned several times something users would like to know. In some cases, knowing the colour of an object had clear utility.

For example, one interviewee wanted to know what colour the jerseys were in a photograph of a hockey team, since it indicates whether they are playing a home or away game. In other cases, interviewees wanted to know the colours of objects "just because" or as an indication of aesthetic quality.

There was an overall desire to have access to as much information as possible. Interviewees wanted to have everything at their fingertips, and be able to personally review it to decide its relevance to them. However, they also wanted the ability to "preview" the contents of an image before spending a lot of time sifting through large amounts of information. There is a tension between these two requirements, as creating an "at-a-glance" overview inherently involves some curation on the part of the person designing the overview.

Audio Files

The reaction to the audio prototypes of object detection and semantic segmentation sonifications we presented was positive. One user said that they felt they were sitting in front of an image and looking at it the way a sighted person would. They felt this "normalized" the experience of exposure to images across users. There was a preference for the object detection rendering strategy over the semantic segmentation rendering strategy among interviewees. This preference is likely a consequence of the difficulty interviewees had in understanding what the semantic segmentation strategy was presenting. Many were

unable to identify what the audio files were meant to convey. Most recognized that the sound was moving, but they did not understand what the movement was supposed to mean. It was not uncommon for interviewees to mistake the moving audio to mean that an object itself was moving in the image.

For object detection, most interviewees correctly understood that the objects were being placed in space according to their location in the image. There was frustration with the speed at which some "pops" were played, especially when there was a group of many objects with the same label. In these cases, interviewees felt that they could not pick out the individual objects from the group.

Some users were concerned about the aesthetic value of the renderings. They felt that the popping and buzzing noises used were not pleasant to listen to and that this would cause frequent consumption of renderings in this format to become annoying. Interviewees emphasized that different solutions will work for different people. Even if they themselves did not find the audio samples compelling, they recognized that it would likely be useful for some portion of BPSP.

3.2.3 Discussion

Our conversations provided insights about the needs and desires of the BPSP community. The first is that users want to have what might be described as a "blind version of a sighted experience" with images on the web. They not only want access to the same information

sighted users have, such as colour, position, emotion, etc., they also want it to be delivered in a form that is similar to the way sighted users experience it. This means having the capacity for both at-a-glance absorption of the information contained in an image, and the ability to explore portions that interest them in additional depth. The second is that their computer use can be subdivided like that of sighted users into "work" and "recreation" purposes, and that their use of desktop and laptop computers is often more for work than recreation. This has implications for what types of images should be targeted by a system for desktop or laptop use for maximum impact.

Our discussions also informed the creation of a taxonomy of image types for the team to consider tackling, a matrix of user needs, and a set of personas derived from the taxonomy and matrix. Personas are fictional characters used by UX researchers and designers to help development teams keep end users in mind while developing. The taxonomy of image types (see Figure 3.4) is a non-exhaustive list of image types that BPSP encounter on the Internet and a short description of the information these images contain that would be important to convey to users. Comic strips, online shopping, charts and graphics, art and aesthetics, personal photographs, photographs on news sites, memes, screenshots, gifs, and educational diagrams were all identified as image types that could be rendered in an audio or audio-haptic format for users.

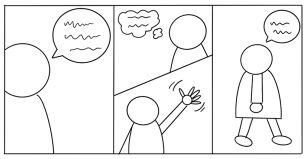
Maps were also frequently mentioned by users but were not included in the taxonomy for two reasons. First, the maps described by users, although visual in nature, were not

represented on the web in a typical image format. Instead, they were usually a specialized container that called an application programming interface (API) such as Google Maps to render contents. Secondly, we wanted to avoid users confusing the IMAGE system with a system for real-time navigation. Navigating spaces remains a challenge for BPSP, but was outside of the scope of our project. As the IMAGE system developed, a specialized preprocessor and handler were introduced for Google Maps so that users could hear audio renderings of their contents. This is done through a button in the HTML hierarchy and exists outside of the primary interface for the system.

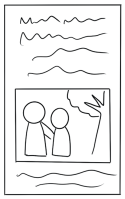
The plot of user needs (see Figure 3.5) describes how the nature of a user's disability determines the tools they might find useful. The plot has two axes: The x axis represents the degree of hearing loss of an individual. The y axis represents the degree of vision loss of an individual. Because our system targets those who are blind or partially sighted, the y axis does not include sighted individuals. We plotted five potential user types on the axes and described the types of solutions they might find helpful.

Our "typical" or "standard" user has no hearing loss and no usable vision. This user requires an audio or audio-haptic rendering of web images. They likely use a screen reader, and may use a Braille display as well.

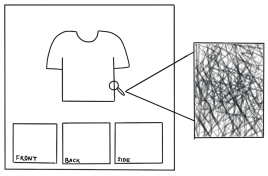
For those with usable vision, like the individual we interviewed who identified as visually impaired, magnification software and high contrast screens may be useful instead of or in addition to the audio feedback. Such an individual may even prefer aids to vision



a. Comic Strips
Paneled cartoons are common on the Internet. Knowing both the contents of speech bubbles in the actions taking place in the comic is important. It would also be nice to differentiate between types of speech bubble.

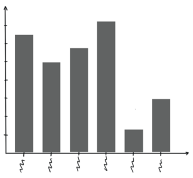


f. Photographs.in News Articles
These show specific events or lend credibility to the events communicated through text. The identities and demographics are important to these images. Celebrities or political figures are often featured.

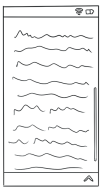


b. Online Shopping
When buying clothing and other items online, descriptions are often insufficient. Communicating colour and texture (perhaps with a zooming feature) would be useful,

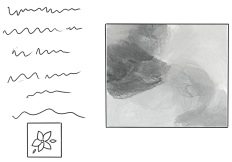
g. Memes
Memes are similar to comics, but being able to identify the format being used is very important to explain the joke. Common characters and distortions made to an image also have particular meanings.



c. Charts and Graphs
Being able to know the quantities being represented as well as general trends, maxima and minima, would be useful for users. The labels on the axes are also very important.



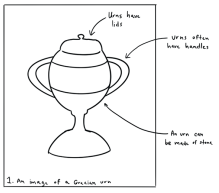
h. Screenshots
People often screenshot information to share it instead of copying and pasting or sending links. This is especially common on social media. A screenshot often contains both text, logos, and pictures.



d. Art and Aesthetics
Images of art may be presented on their own or as a way to emphasize the emotional aspects of a textual work. The affective information in a piece of art needs to be communicated to users.

i. Gifs
Gifs are in an image format but feature movement. Sometimes gifs are used to accentuate an aspect of a static image. The difference between frames causing the movement is important to communicate.

e. Personal Photographs
Users come across photographs taken by friends and family frequently on social media. They want to know the genders, ages, emotional states, and actions of the people in the photographs. Location, time of day, and what the individuals are wearing were also requested.



j. Educational Diagrams
Educational figures such as those found on Wikipedia may detail important attributes for prototypical examples of an object. The important features that make an object part of a category must be communicated to users.

Figure 3.4: A taxonomy of image types produced from conversations with potential users.

instead of methods involving sensory substitution when interpreting image contents. Similarly, individuals who have some usable vision but no hearing may not benefit from our system at all.

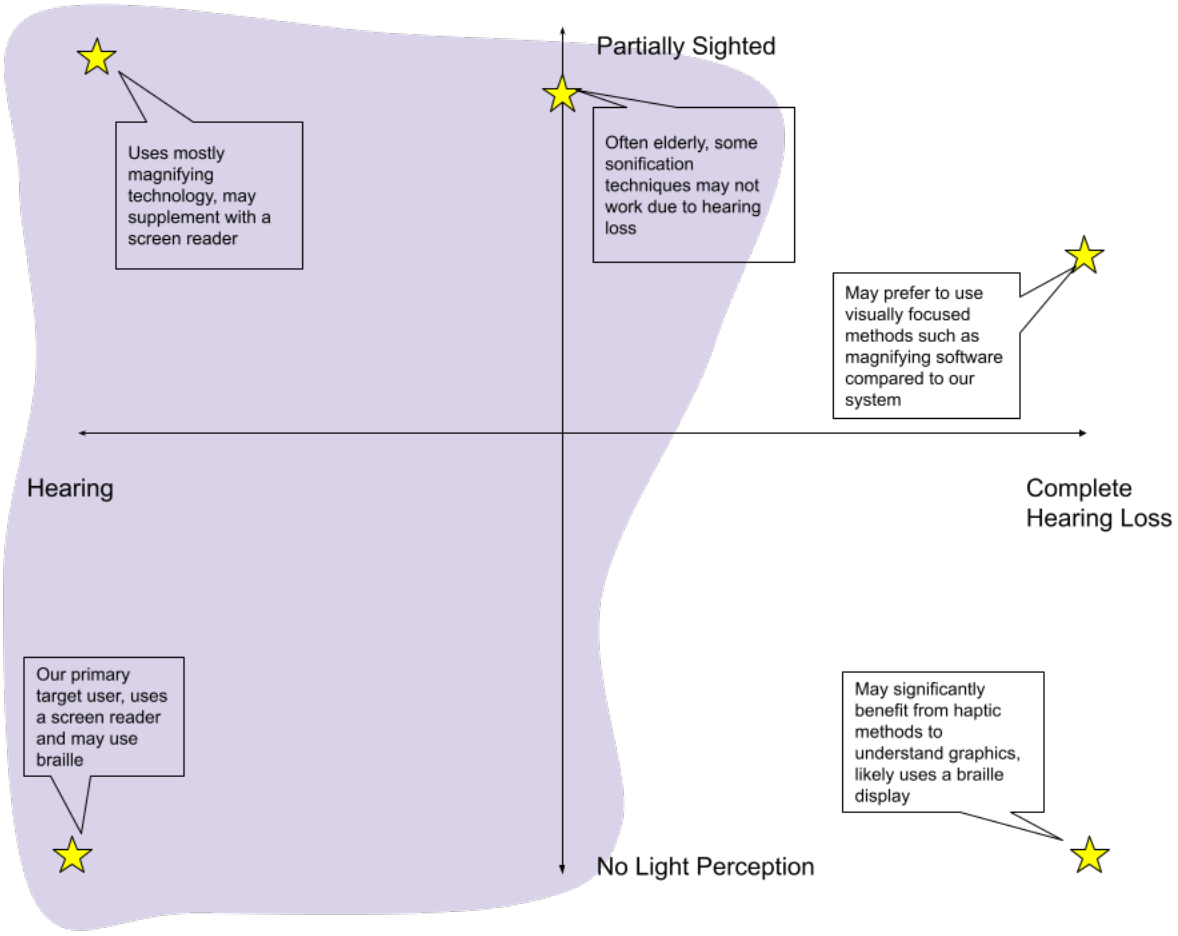


Figure 3.5: A plot of user needs based on nature of disability. The shaded area is the group of users initially targetted by our system.

Because of the age demographics of our user group, there are many individuals who consider themselves visually impaired and also have some degree of hearing loss [78]. These individuals cannot take advantage of every sonification technique, since they may not be able to hear some pitches. They may also have less hearing in one ear than another, making it difficult for them to use audio spatialization.

For the deaf-blind who have little to no usable hearing and little to no usable vision, haptic methods are essential for communicating information. These users may benefit from specialized haptic devices along with descriptions provided through a Braille reader. We combined our plot of user needs with our taxonomy of images types and our understanding of user habits gained from the semi-structured interviews to develop personas to represent potential users (see Figure 3.6).

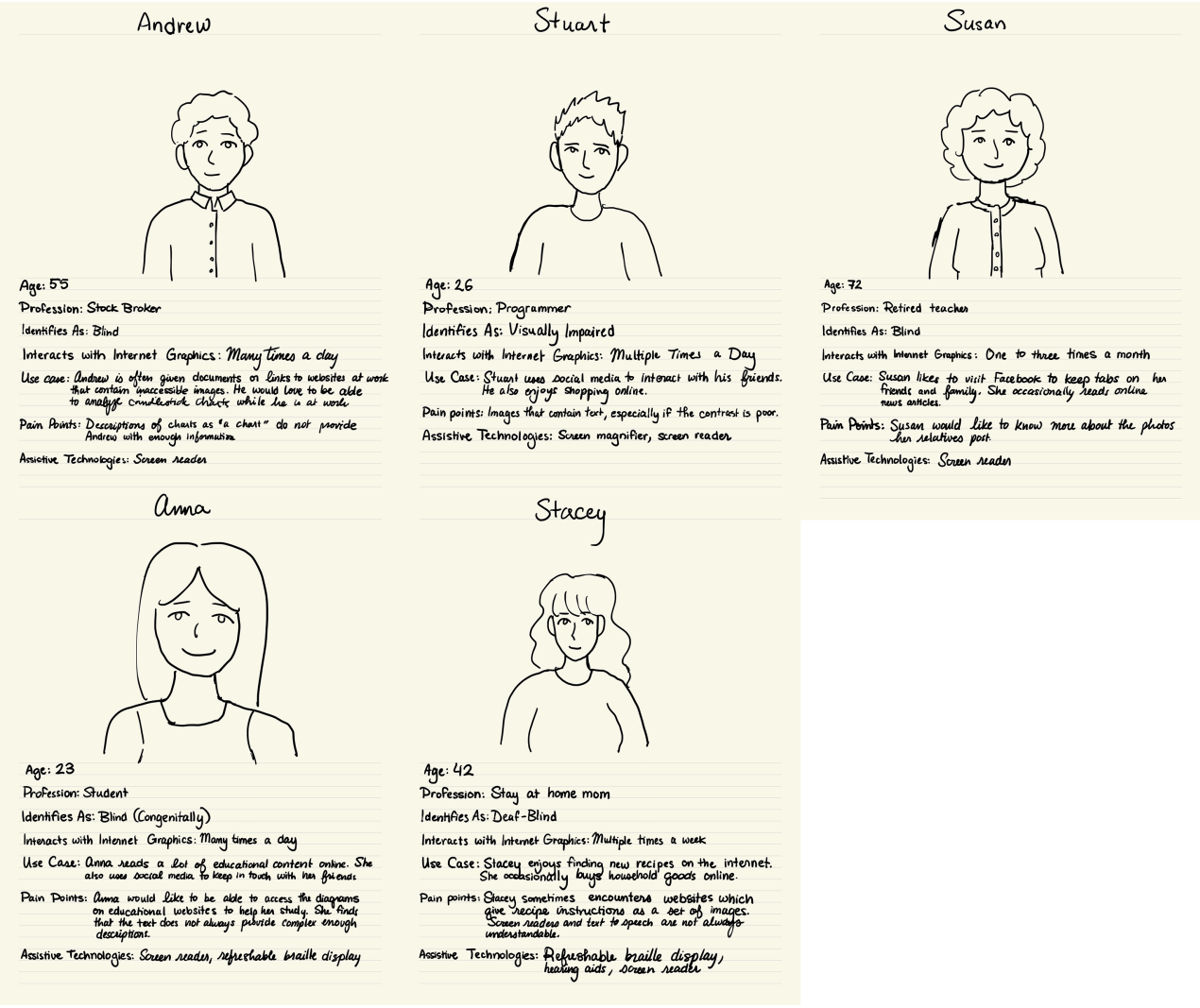


Figure 3.6: User personas for the IMAGE project

Chapter 4

Interface Design of the IMAGE System

After we had an understanding of what users wanted to know about images and how they interacted with them on a daily basis, a bare-bones version of the system was constructed to allow for iterative testing with users. Although the system architecture would allow for many types of renderings to be delivered to users, the handlers deliver two types of renderings by default. The first is a text-only transcription explaining the contents of an image, to be used by a screen reader if for some reason a user does not have the ability to use an audio rendering. The second is an audio segment consisting of both an object detection and semantic segmentation portion rendered using audio spatialization techniques. Audio spatialization was chosen by the team as the default rendering style based on the positive

feedback from users in our semi-structured interviews, and the guidance of our team audio expert.

The default audio rendering was arrived at through informal participatory design between the IMAGE team and BPSP in our pilot testing group. Initially, semantic segmentation and object detection were returned to the user as two separate renderings, but these were combined since they both communicate spatial information to the user. A rendering begins with the semantic segmentation portion, where regions are named and then a buzzing noise traces around the contour of that region, indicating the space it occupies. After all regions have been mapped in this fashion, object detection lists individual objects and provides a spatialized "pop" or "click" to describe the location of each object. If there are several objects of the same kind in one region, they are introduced together and the pops are played in quick succession.

It was decided that the IMAGE system would be available to users through a Google Chrome extension. This extension would be compatible with any Chromium browser, making it accessible to most users. Firefox is the only major browser that does not use chromium, and it makes up less than 10% of the market share of Internet browsers. Browser extensions are also relatively easy to develop, and protect user privacy.

Unfortunately, the benefits of browser extensions come with significant limitations to UI options. Because the IMAGE content is delivered to users through the browser, they will not only be interacting with it through their screen readers, but will expect behaviour that

is in line with what they usually experience in a web browser. Previous research found that users react poorly to systems that interfere with their screen readers or do not behave as expected [28, 79]. Our own semi-structured interviews found that systems that did not "play well" with screen readers were a major pain point for users. In light of this, we decided to stick strictly to using interface elements that could be created with a combination of HTML, CSS, and JavaScript.

4.1 The Settings Page

4.1.1 Design

The settings page was identified early on as an important feature for the IMAGE extension, as it would help accommodate the needs of a broad user group. The settings were also crucial at that point in development of a beta version of the system deployed early on in the project to developers both within and without the team. The system was released early to allow other teams to produce handlers for the system [7]. Individuals in and outside of the team required access to special options so that they could troubleshoot for development purposes.

A combination of user needs and available hardware dictates what renderings should be delivered to a user. For example, a user may have stereo headphones, but if they also have moderate hearing loss, they may need a non-spatialized rendering (see Table 4.1). We initially thought that a combination of devices recognized as being in use by the computer

	Stereo Headphones	Pin array and Stereo Headphones	Haply 2diy and Stereo Headphones
No Hearing Loss	Deliver spatialized renderings to the user.	Deliver spatialized rendering with text and line drawing through pin array.	Deliver spatialized rendering and 2diy rendering.
Mild or Moderate Hearing Loss	Deliver non-spatialized renderings, or allow selection of TTS voice.	Deliver non-spatialized rendering with text and line drawing through pin array.	Deliver non-spatialized rendering and 2diy rendering.
Severe or Profound Hearing Loss	Deliver text used by TTS.	Deliver text rendering through pin array along with line drawing.	Deliver text rendering and 2diy rendering.

Table 4.1: Hearing loss and hardware-based design objectives.

and disability-identification given by the user could be used to load one of a selection of presets of settings in the Chrome extension that would determine which renderings a user receives. Settings for Chrome extensions are located in an "options" page, however, and due to limitations of this page, hardware could not be identified automatically and would have to be manually enabled by users.

Our first design (see Figure 4.1) allowed users to select between three audio modes, which would have delivered to the user different renderings. It also allowed users to set a minimum volume, since both researchers on our team and community stakeholders informed us that the volume of some of the sonifications in our spatialized renderings was often too low, and due to the dynamic range of the audio, simply increasing the volume would make the rest too loud. This design also included the ability to choose between different text to speech options, because an interviewee had mentioned that as a result of their hearing disability,

they find certain text to speech voices more difficult to understand than others.

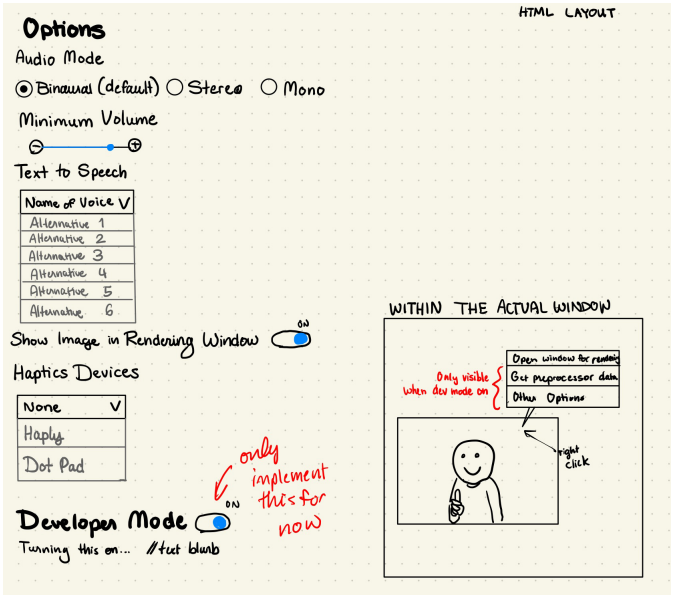


Figure 4.1: Sketch of the settings menu

The audio-related settings were included to accommodate users with mild to moderate hearing loss who might want to use our system but struggle with the default rendering provided. An option for showing the image that had been clicked in the renderings window was also included as a result of our conversations with the user community. The individual we interviewed who identified as visually impaired mentioned that they often use magnifying software while listening to alt-text to piece together the image contents. Because the renderings window opens as a pop-up box on top of the page where the image being rendered was selected, we thought it would be helpful to provide a copy of the image in the renderings window for that subset of users.

A drop-down menu with options for haptic devices to use with the IMAGE system was

included in the sketch work for the settings page. Three options were planned for, either no haptic device, a Haply 2diy, or a Dot pad, a pin array made by the Dot Incorporation. The developer mode toggle would enable two debug options for the IMAGE extension in the context menu, getting preprocessor data and getting request data. The sketches of the settings page (see Figure 4.1) were discussed by the IMAGE team to determine feasibility, and evolved into a wireframe.

4.1.2 Wireframe

The wireframe of the settings page made several changes to the first sketch (see Figure 4.2). First, as a result of audio and backend development teams' feedback regarding feasibility, the options for modifying the audio renderings were removed. In order to accommodate those unable to use the default spatialized audio rendering, an option to enable a second rendering that displays basic alt-text was included. This alt-text would also be accessible to any deaf-blind users.

An option for custom servers was introduced at the request of the development team so that they could test with secondary test servers. This option was not hidden under developer options because we wanted to make it easy for users to find it if they were part of another organization using the IMAGE system for specialized renderings on their own server. The haptic options were moved to be hidden under the developer mode, because of the prototypical state of the haptic renderings. The drop-down menu for choosing which

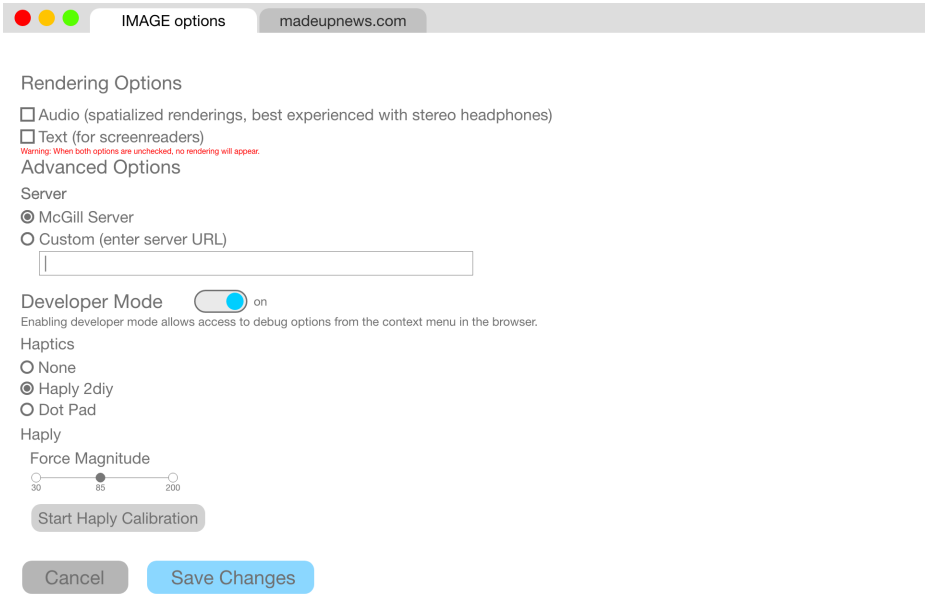


Figure 4.2: A wireframe of the updated settings page.

haptic device was swapped for three radio buttons, because although the drop-down reduced visual clutter, it was more tedious to use with a screen reader. An option for setting the force magnitude of the Haply 2diy was added at the request of the haptic development team, as well as a button to start the haptic calibration menu. Finally. we added a *save* button after investigating the capabilities of Chrome extensions and realizing the need for manually saving choices.

4.1.3 Implemented Version

When implementing the settings page, the breadth of options was once again reduced as a consequence of development priorities (see Figure 4.3). The force magnitude and calibration

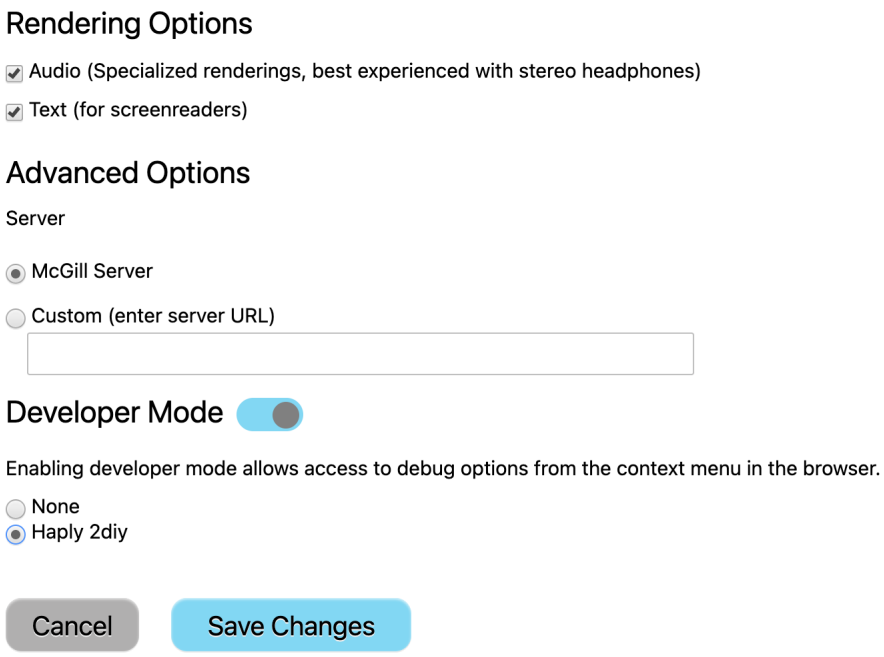


Figure 4.3: The currently deployed version of the settings page.

functions for the Haply 2diy were removed, as was the option for the Dot Pad, since there was no deployed functionality for pin arrays. A pop-up box confirming that options had been saved was added so that users would be aware of the system status.

4.2 The Renderings Window

The renderings window is the centrepiece of the IMAGE system. It takes all of the renderings obtained by handlers and presents them to the user. Renderings are delivered with timestamps, which can be used to skip to specific points in the audio file. This ability can be used to make navigation within an audio clip possible. Navigation is important

because it gives users agency to explore portions of an image that they find interesting. To allow for several renderings to be delivered without cluttering the window, each one is contained within a disclosure button that can be clicked to reveal or to hide its contents. The buttons are labeled with the rendering title so that BPSP can quickly hear which renderings are available before choosing one to access in depth.

4.2.1 Iteration 1

In the first version of the renderings window, a drop-down list was used to let users select various parts of an audio rendering (see Figure 4.4). A *play/stop* button was included below the drop-down to start or stop the portion of audio the user had selected. The default selection of the option was to play the entire rendering. Opening the drop-down would reveal each slice of the rendering listed in the order it is presented in the full audio sample. The drop-down menu approach adheres to the WCAG regarding operability, and does not have any functions that interfere with a screen reader. Pilot testing with a small group of BPSP found that all users could make use of the UI with their screen readers and were able to understand the function of the drop-down menu in selecting portions of the audio segment to play.

Unfortunately, this version of the renderings window had two significant problems. The first was the amount of clicks it took to play any particular part of a rendering. To select a portion of a rendering, the user had to first tab onto the drop-down menu, press *enter* to

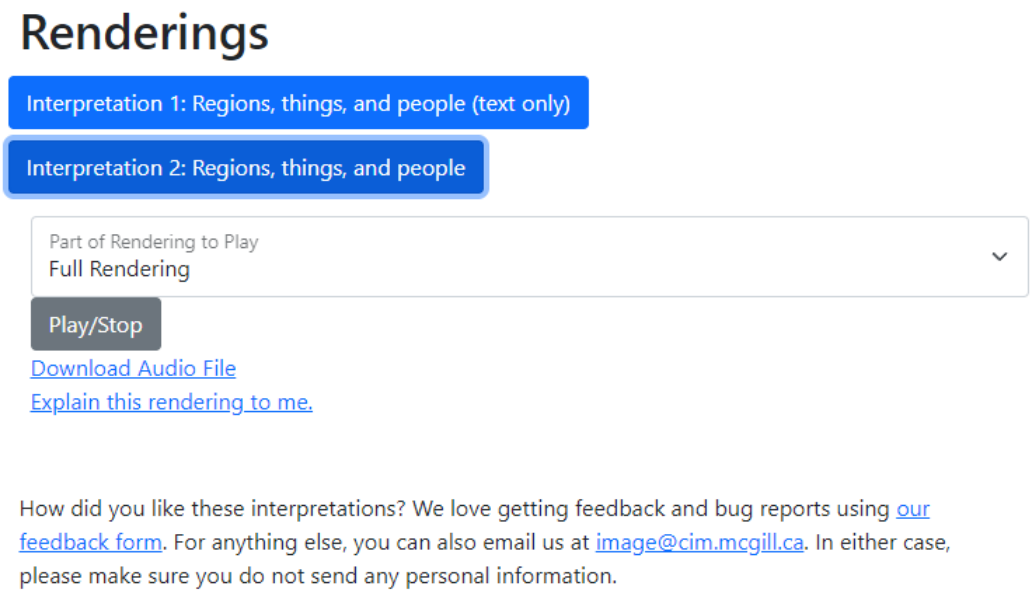


Figure 4.4: The first version of the renderings window.

open it, press the right arrow key as many times as it took to reach the rendering, press enter to select the rendering, then press the tab button (or its equivalent) until reaching the play button, and finally press the play button to hear the rendering. Effectively, the number of keyboard presses would be $5 + n$ where n is the number of the option desired in the drop-down menu.

The second major problem was that the the workflow for navigating through the drop-down menu did not give the impression of "exploration". There was no way for users to quickly toggle between segments the way a sighted person can choose to direct their gaze to an area of a photograph. Our second iteration was an attempt to fix both of these problems.

4.2.2 Iteration 2

Our second concept for an IMAGE renderings page relied on existing interfaces that deal with audio. Because the primary method of delivering renderings is through audio files divided based on timestamp, we wanted to use an audio player to allow users to scan the contents of the audio or jump between segments with the arrow keys. The audio player in Apple's Music application on the Mac operating system works in this manner, and it is very effective. However, we could not modify the functionality of the default HTML5 audio DOM, because it is not accessible. For IMAGE's project website we used Plyr, an HTML5 compatible audio player that is accessible to screen readers (see Figure 4.5). We decided to modify Plyr for the renderings window.

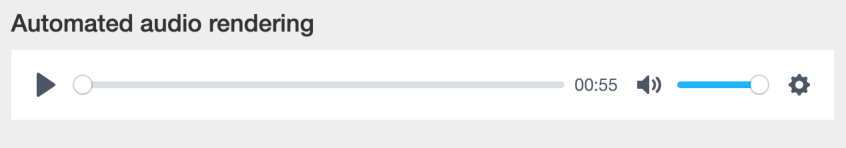


Figure 4.5: The plyr audio player.

We added timestamps within the audio player that users could navigate between by pressing the arrow keys while inside the timeline division. The rest of the functions of the audio player were left intact. While this solution appeared cleaner, it too had serious flaws. The most serious of these was that the audio timestamp visible on the right side of the player was read out by some screen readers whenever the audio played. This made it nearly impossible to understand the audio files. Additionally, in order to use the arrow keys to

move between segments in the player, users had to navigate into the timeline section of the player with their screen reader. When using some screen readers this first required exiting the division that contains the play button and then navigating into the division that contains the timeline. This was annoying for users because they could not choose between segments and pause and play at the same time.

4.2.3 Iteration 3

After some testing, it was understood that using any version of a slider to enable scanning of an audio sample would be incompatible with screen reader use. As a result, we were forced to return to our first iteration involving the drop down menu as a starting point. The core problems with this UI still needed to be resolved. After some experimentation, the UI shown in Figure 4.6 was developed. This version of the renderings window has each portion of the audio file, as well as the full rendering, able to be triggered by pressing a button labeled with the title of the segment that would be triggered. Each button is contained within an H1. Screen readers are equipped with keyboard shortcuts for headers, and pressing these shortcuts takes the user to the next object with said header. Placing the button objects within an H1 makes it extremely quick to navigate to the next portion of the audio sample. Pressing a button in this version now plays or pauses the audio segment instead of playing and stopping it. This means that if a user accidentally double clicks a button they will not have to restart the entire audio segment. Clicking on a different button while an audio

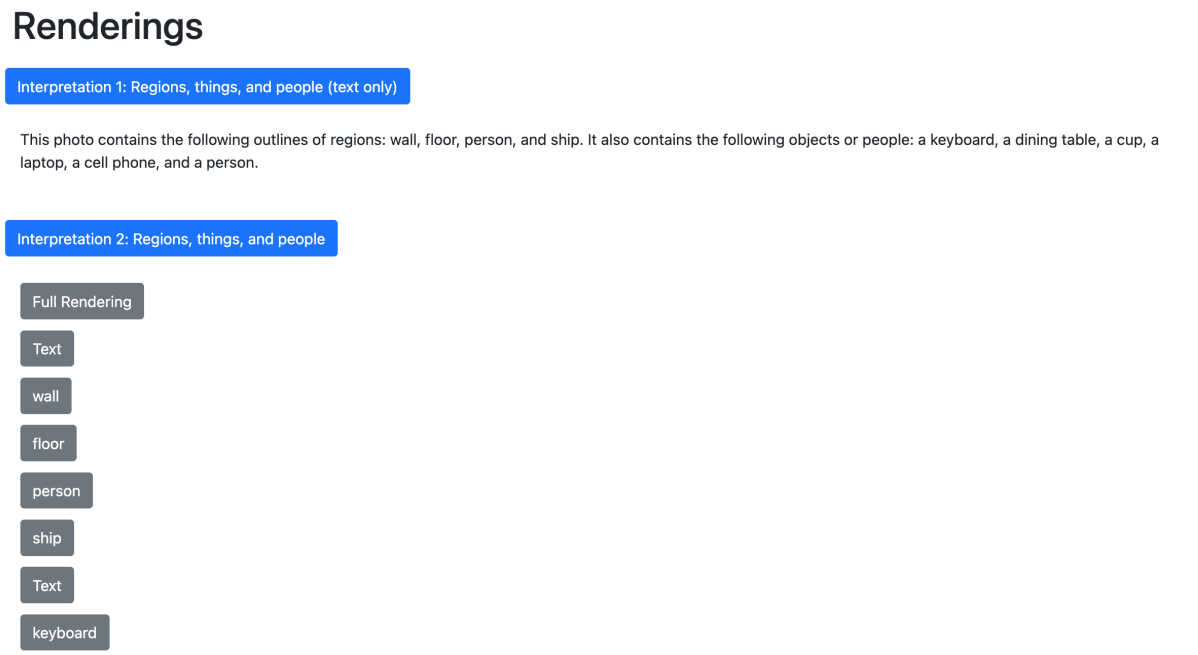


Figure 4.6: The currently deployed version of the renderings window.

segment is already playing will stop the first audio segment.

While visually unimpressive, this UI solves the key problems with the original dropdown UI. First, it reduces the number of button clicks required to reach the desired audio segment to $1 + n$ where n is the position of the target segment in the list of available segments. Secondly, the capacity to quickly move between segments is significantly closer to the way sighted users can change focus within an image. This is done while still giving the user a choice about whether or not to trigger the audio to play. Cursory feedback from our pilot testing group showed heavy preference for this UI design over the initial drop-down design. A pilot tester said they felt that they were able to get a good idea of what was contained in an image using the system. The biggest weaknesses of this interface are that it does not

easily translate to audio-haptic examples, and that the user still hears an object label twice when using it, once when they tab onto the button and once as part of the audio sample when they click on it.

4.3 Future Work

As the IMAGE system evolves, the amount of information it can communicate will increase. Currently, team members are working on preprocessors and handlers that will enable the detection and rendering of more granular aspects of an image, such as the clothing and emotions of people present in an image, and whether a person is a celebrity or not. optical character recognition (OCR) is also being added, which will allow text in images to be read out as part of a rendering. OCR, clothing, and emotion information are not useful without being properly tied to the object that contains them. An emotion must be tied to the person expressing it, and text content recognized by OCR must be tied to the object it is displayed on. This will be implemented using bounding boxes of detected objects to create a tree hierarchy for a particular rendering. For this functionality to be fully enabled withinthe IMAGE system, it must be determined how users will be able to explore these renderings in a way that effectively communicates which attributes are contained by which objects.

A potential evolution to the current interface that would allow users to explore these hierarchical renderings would be to add the attributes of objects, such as colour, emotion, etc., as a set of second-level-header-contained buttons under their respective H1-contained

button. This would carry over the principle benefits of the header-button system from Iteration 3 of the UI: the rapidity of navigation between options and the ease of use with a screen reader. This strategy would suffer from the same problems the current UI does, however, such as the redundancy of labels being heard in the samples themselves and as part of the screen reader navigation of the UI.

Other UI possibilities would violate the WCAG, but might be considered acceptable by users depending on the benefits provided. Removing the labels on the buttons triggering the audio samples on contact is an option, but it does not give the user any warning that they are going to trigger a potentially lengthy audio segment. Additional development and testing would be necessary to determine whether this would be acceptable to users, since the renderings window may be perceived as a separate application. A final possibility currently being considered is to move the object labels from the audio segments to the button labels themselves and simply play the spatialized "pop" when a button is pressed. This may be the best way to reduce auditory redundancy and continue to comply with the WCAG.

Chapter 5

Understanding System Efficacy

5.1 The Craft Study

The goal of the Craft Study was to understand how efficacious our system was, independent of any user interface choices. That is, in both the audio and audio-haptic methods of delivery, would users to be able to interpret the spatialized information they were given?

5.1.1 Methods

Participants were recruited through the Regroupement des aveugles et amblyopes du Montréal métropolitain (RAAMM), a French-language community organization for BPSP in the Montreal area. The study consisted of three segments: the first was a pre-experiment questionnaire, where participants were asked basic demographic questions,

such as their age, gender, and the nature of their vision loss. The last part of the study was a post-questionnaire, where users were asked to rate the enjoyability and utility of the system using Likert scales, as well as give general feedback to the researchers. The experiment was conducted in between the two questionnaires. McGill's REB approved the administration of the experiment under REB#21-10-031-03.

Participants were exposed to eight image renderings each: four rendered in the audio modality only, and four rendered through audio and haptic modalities. Images hosted on actual web pages were chosen to reflect the experience a user might have when selecting an image to explore. The pictures were chosen in pairs with similar composition and content, so that users could get an audio only and audio-haptic of each type of image. The chosen pairs were two photographs with people as the principle subjects, two photographs of outdoor scenes with some animals, two photographs of outdoor scenes containing buildings, people, and objects, and two photographs of indoor scenes containing only objects. For each participant, the audio only type of rendering was randomly assigned to one member of each image pair, with the other member of the pair getting the audio-haptic rendering. The order in which participants were exposed to each of the eight image renderings was also randomized.

The audio renderings were made using the deployed version of the IMAGE pipeline to create audio samples from images found on the web. Our chosen images were cropped so that they were all in a 2:3 aspect ratio, then uploaded to a web page, where the IMAGE extension

was used to obtain JSON data from the preprocessors using developer options. There was a chance that the centroid coordinates in the JSON files would not match those in the MP3 files obtained through the extension because a separate request is sent to obtain each of these files, allowing for inconsistencies in ML behaviour. To ensure consistency, the JSON data procured through the developer options was passed through the handlers manually using a script. The audio files received from the handler were trimmed to normalize the amount of data in each sample to avoid differences in cognitive load between samples. Each sample was modified to contain a single contour in the semantic segmentation style, and four objects conveyed using the object detection method. Since the RAAMM recruited Francophone or bilingual participants, and the IMAGE system does not currently have French language renderings implemented, the audio samples were further modified so that all the TTS segments were replaced with French-language translations. Translations were vetted by French-speaking team members.

We used a Haply 2diy, a 2 DOF force feedback device, for the audio-haptic renderings. It was chosen as a target for use with the IMAGE System because of its inexpensive price point (approximately CAD$300) and its familiarity to our development team. The 2diy consists of two single-jointed, motorized arms that connect together with a central knob, forming a diamond shape (see Figure 5.1). The knob is held by the user and moved around the 2diy's base, shown in grey. The 2diy is used with a physics simulation environment which can be programmed to mimic textures, surfaces, and barriers. The effects are simulated through

forces exerted by the arms against the user's hand. These forces can also move a user's hand around the surface, or to particular points within the haptic simulation. This form of interaction is called guidance.

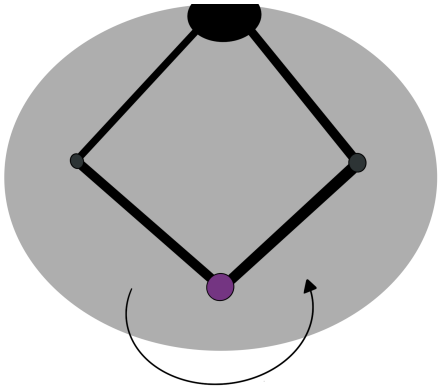


Figure 5.1: An illustration of the Haply 2diy force feedback device.

The haptic simulations used in the experiment consisted of four circles of radius 1, centred at the centroids found in the JSON data (shown in black in Figure 5.2). When a participant made contact with a circle using the end-effector, it would play the name of the object and then the associated spatialized "pop" that localizes the object. When two circles were touched at the same time, both audio tracks would be triggered so that the participant would understand that they were in contact with two objects. The circles were solids, meaning they could be intersected by the end effector (the red circle in Figure 5.2) as a result of forces exerted by the motors. There were also solid "walls" around the simulation environment forming a 6 unit by 9 unit rectangle matching the aspect ratios of both the images and the velcro board. These walls acted as boundaries, keeping the user inside the limits of the

Figure 5.2: An example of the haptic simulation environment.

photograph. No guidance was used so that the effects of basic kinesthetic feedback could be isolated.

The visuals of the simulation were hidden from participants to avoid changing the experience between participants with some and no vision, but they were visible to the researchers so that they could observe user strategies for finding objects within the image.

Participants were exposed to the renderings one at a time, and allowed to move around the haptic simulation or listen to the audio rendering as many times as they wanted. When they felt ready, they were encouraged to place tokens in the locations they had heard or felt them on a $6x9$ inch board covered in velcro located to their right, as seen in Figure 5.3. If they had been exposed to the audio-only rendering, they were also asked to use Wikki Stix,

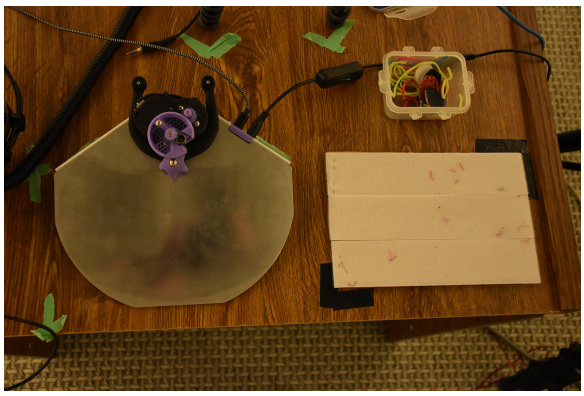


Figure 5.3: The experimental setup for the craft study

yarn pieces covered in wax that allow them to hold their shape, to designate the outline of the contour they heard. Participants were allowed to intersperse exposure to a rendering with placement of a specific token if they wanted to place objects one by one.

We hypothesized the following:

- (H1) Users would be able to more accurately place tokens in the audio-haptic condition compared to the audio-only condition.
- (H2) Users would be able to more quickly place tokens in the audio-haptic condition compared to the audio-only condition.

In order to test H2, we planned to time participants during their placement of tokens. A pilot of the study revealed that we would not be able to separate out the token placement and rendering exploration portions of the task because of the high cognitive load the task required. As a result, time was informally kept using the video and audio recordings of the sessions, but participants were not explicitly timed.

5.1.2 Results

Pre-Test Questionnaire

Eight participants were tested (3 female, 5 male). Only two participants were under the age of 50, making our sample significantly older than the general population, but fairly consistent with the age distribution in the BPSP community. Only one of our participants had any light perception, with the rest having no vision. Three were congenitally blind, three lost vision growing up, and two lost their sight in adulthood. Seven of our eight participants used a computer daily, and one participant had stopped using computers since retirement. All of our participants were Braille literate, and six of the seven who use computers daily use a Braille reader while using the computer. The most popular screen reader among our participants was JAWS, followed closely by NVDA. VoiceOver and Microsoft Narrator were used by one participant each.

Task Results

The accuracy of a participant's recreation of an image using tokens was measured by taking the Euclidean distance between the centre of the token placed for an object and the corresponding centroid generated by the IMAGE system. The x and y distances from the centroid were normalized to a unit square by dividing according to the image dimensions before calculating the Euclidean distance. The mean Euclidean distance of tokens from their respective centroids was 0.237. The standard deviation was 0.133, and the standard

Image	Audio		Audio-Haptic	
	Mean Distance	Standard Deviation	Mean Distance	Standard Deviation
Indoor1	0.195	0.089	0.183	0.074
Indoor 2	0.212	0.119	0.326	0.200
Outdoor 1	0.322	0.130	0.256	0.115
Outdoor 2	0.282	0.112	0.212	0.065
Mixed 1	0.292	0.197	0.238	0.105
Mixed 2	0.176	0.108	0.204	0.171
People 1	0.285	0.145	0.166	0.083
People 2	0.218	0.076	0.206	0.105

Table 5.1: Distances on tasks across participants.

error was 0.009. Table 5.1 shows the mean Euclidean distances by image, and Table 5.2 shows the mean Euclidean distances by participant. The distances broken out into separate x and y measures are available in Appendix A.1. The mean of these separated measures was .147, with a standard deviation of .116, and a standard error of .005.

The data were not normally distributed and had some gaps, so a Wilcoxon Paired-Rank Test was used to determine the significance of results. Across participants, audio tokens were an average of 0.264 units away from the centroid location, while the audio-haptic tokens were an average of 0.219 away from the centroid location. The results of the Wilcoxon Paired-Rank Test showed $p > 0.05$, which does not provide evidence that placements made under the audio-haptic condition were more accurate (H1). Token distances were significantly more accurate in the x dimension compared to the y dimension for both the audio and audio-haptic conditions ($p = 0.01$).

Participants sometimes missed objects altogether (see Figure 5.3). This was most

Participant	Mean Token Distance	Standard Deviation
P1	0.213	0.089
P5	0.190	0.122
P6	0.233	0.103
P7	0.275	0.180
P8	0.284	0.128
P9	0.262	0.130
P11	0.189	0.083
P15	0.262	0.138

Table 5.2: Average euclidean distance of tokens by participant.

Participant	Number of missed objects
P1	2
P5	2
P6	0
P7	1
P8	5
P9	1
P11	2
P15	0

Table 5.3: Number of missed objects by participant.

common in the audio-haptic renderings, where 9 objects were missed, but 4 objects were missed in the audio renderings as well. Only two participants missed no objects across all renderings. This is a significant drawback to the audio-haptic rendering method. Participants also found it very difficult to represent the contours rendered by the semantic segmentation sonification strategy using the Wikki Stix. Most participants chose to fill an area with the Wikki Stix, rather than use them to draw a line representing the contour. As a result, we were unable to collect meaningful quantitative data about the accuracy of participants' understanding of semantic segmentation.

Post-Test Questionnaire

	Mean	Median	Mode
Audio-Only Rating	3.28	3	3,4
Audio-Haptic Rating	3	3	4
Enjoyment of System	3.28	4	4
Utility of System	2.57	4	4

Table 5.4: Participant ratings of system out of five.

After completing the main body of the experiment, participants were asked to rate aspects of the system on a scale from 1 to 5, with 1 being the worst and 5 being the best. Participants rated the audio-only form of the system, the audio-haptic form of the system, and their perceived utility and enjoyment of the system as a whole. The averages given by participants with respect to the system are given in Table 5.4.

Participants were also asked to provide any additional thoughts they had about the

system. P1 and P15 both said they would prefer to have "realistic" sounds indicating the contents of an image and their location. P5 and P9 said they had a difficult time using the haptic device, especially trying to find objects. P11 said they really enjoyed using the system, and found the haptic rendering intuitive because it reminded them of echolocation techniques.

5.1.3 Discussion

Participants' mean distances from the centroids did not cluster around any particular point. There was a fairly even distribution of distances between .18 and .3 units out of 1 unit, with four of eight participants falling above and four of eight participants falling below the mean ($\mu = 0.238$, $\sigma = 0.035$) These results are not consistent with a normal distribution. The distance of even the most inaccurate participant was well below .5 units out of 1, meaning all participants were placing their tokens within the general vicinity of the centroid locations. We observed that participants tended to properly reproduce the patterns of centroids, even if the patterns were not in the correct location. The distances from the centroids were mostly a result of transpositions of the entire pattern as opposed to misplacing particular tokens. Figure 5.4 shows an example of a participant's tokens overlaid with the centroid locations (shown by black dots). The transposition of the pattern is clear. This suggests that although participants may not be objectively accurate with their placements, they are extracting a gestalt understanding of the images from the renderings.

The only statistically significant finding was the larger distance of tokens from centroids in the y direction compared to the x direction. This difference was found in both the audio ($p = 0.01$) and audio-haptic ($p = 0.01$) conditions. This is unsurprising in the audio case, as hearing discrimination is known to be worse along the y axis [80]. It somewhat unusual for the audio-haptic case, where researchers have show higher acuity in the forwards-backwards direction, which we had mapped to the y axis [81]. The mixing of audio and haptic media may be reducing the positive effect of the haptic system. Another potential cause is the slanted base of the Haply 2diy, which results in movements that are not strictly forwards-backwards.

Use of the audio-haptic system did result in smaller distances between the tokens and their respective centroids than the audio system. The mean x distances were very similar in the audio and audio-haptic cases (0.117 and 0.112, respectively). However, the audio tokens were on average 0.200 units away from the true centroid location along the y axis, whereas the haptic tokens were only 0.156 units away from the centroid along the y axis. This was mostly due to improved discrimination of vertical location. Unfortunately, none of these findings could be shown to have statistical significance. A continuation of the experiment with additional participants may be useful in establishing whether these trends are meaningful or not.

The major disadvantage of the audio-haptic system was that it resulted in significantly more missed objects than the audio system. Finding objects in the haptic simulation was very time consuming, taking over twice as long to place as objects rendered via audio only

and disproving our hypothesis that the audio-haptic system would have a time advantage (H2). More than one participant opted to continue on to a different rendering when they were unable to find one or more objects with the haptic device after what they considered to be a reasonable amount of time. This is probably a consequence of the lack of guidance on the haptic prototype, a conscious choice made because of the oscillation problems common to the Haply 2diy when guidance is used in a haptic simulation. Future prototypes will have the advantage of using the newest version of the Haply 2diy, which provides a significantly smoother experience. These prototypes will be more adequately able to test the effects of guidance on participant perception. The missed objects in the audio renderings were likely a result of multiple "pop" effects played in quick succession after the number of objects was listed. Across all participants, 4 audio objects were missed and three of these were from the same rendering, further supporting this theory. We had slowed the popping sequences down considerably within the IMAGE system after feedback from initial audio samples, but these results indicate that further slowing it may ameliorate users' abilities to interpret renderings.

We suspect that the age of participants played a significant role in their performance, and had an impact on the distribution of distances. Figure 5.5 shows the mean distance of a participant's tokens according to age bracket. In spite of the two outliers who scored .02 units better than the highest scoring of the rest of the participants, there is a strong positive correlation (0.628) between age bracket and mean distance score. This is supported by observations of participants. The older participants struggled with dexterity and spatial

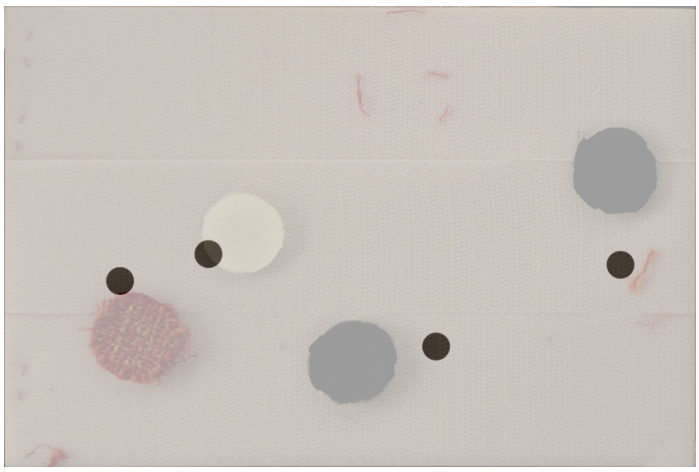


Figure 5.4: P5's token placements for image Indoor1 overlaid with centroid locations.

awareness more than participants in lower age brackets. This was most noticeable with the oldest participant who struggled to grasp the end effector of the haptic device with a single hand and mostly used a second hand to hold the "elbow" joint of the haptic device, making it difficult to operate. The impact of age on performance is also consistent with the literature. Stevens found that tactile acuity declines with age whether or not a participant is sighted [76]. It is also worth noting that the two participants whose scores were not linearly related to their age had very similar histories with regards to their vision loss. Both reported having been born with limited vision and having lost the rest of it in childhood. Their performance may be a variant of "the advantage of the late blind" as described by Morton A. Heller [22], where some development of the visual systems of the brain combined with experience in life as a blind individual results in better task performance. Further studies targeting this sub-demographic of BPSP would help determine whether or not there

is a connection.

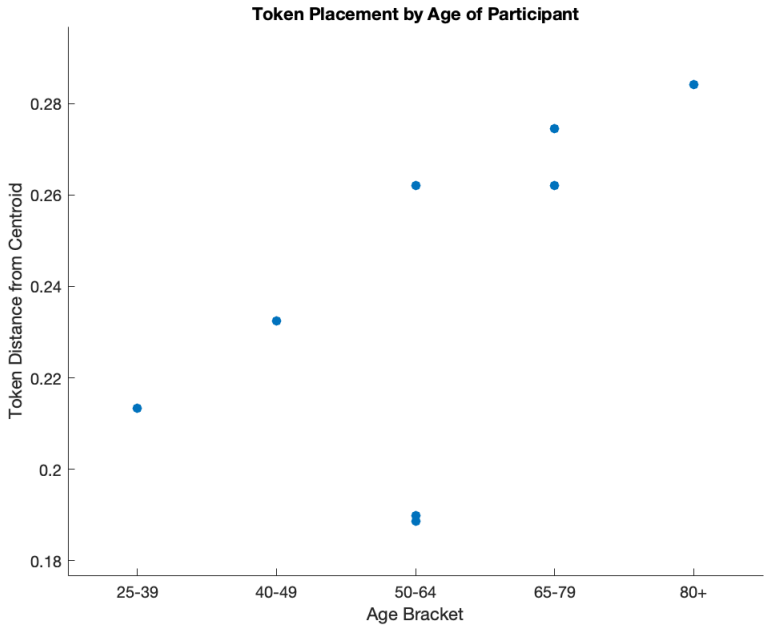


Figure 5.5: Distance of Tokens from Centroids by Age

There were noticeable patterns in the strategies used by participants when exploring the audio-haptic renderings. Most participants started by finding the limits of the simulation and tracing around those. From there, they would choose one of two methods. Some participants would make repetitive and expanding circular motions with the end effector, creating a radar-like pattern until they hit the object. As previously mentioned P11 remarked that using the system reminded them of using echolocation techniques in combination with a cane to find objects for navigation purposes in their daily life. BPSP use a cane by sweeping arcs in front of them with it as they move forward, hitting any objects in their path. Echolocation is a technique some BPSP use to navigate and consists of making sounds and using the

echoes that they hear to make conclusions about what is around them, similar to sonar [82]. Both of these techniques gradually give BPSP information about their surroundings within a particular radius, which is similar to the expanding circular motions made by some participants, including P11 with the 2diy's end effector. The other method used by participants was to draw either horizontal or vertical lines with the end effector across the simulation area, moving down or to the right in small increments after each line. This approach was often better at finding objects in the centre of the simulation space, which were often missed by participants using the sonar approach.

The most striking qualitative results were the widely disparate reactions to the system. Some users really enjoyed the audio rendering, but found the audio-haptic rendering frustrating, while other users found the audio rendering pointless, and thought the audio-haptic rendering was exciting. There are several potential explanations for this. First, the aforementioned reduced dexterity with age, which made it difficult for some participants to grasp the haptic device. This understandably impacted their enjoyment of the system. The time it took for participants to find objects in the audio-haptic rendering was by far the strongest predictor of their rating of the haptic system. The longer it took a participant to find all the objects, the more likely they were to get annoyed and give up on that rendering altogether. These participants tended to enjoy the audio condition more, since they found it easier to complete the task with this version of the system.

For approximately a third of the participants (3 out of 8), there was significant enthusiasm

towards the haptic device. This had an overlap with faster performance on the task in the audio-haptic renderings, but not necessarily with higher accuracy in token placement. One of these participants in particular (P6) showed a remarkable ability to recall where objects were located within the haptic simulation. Whereas other participants needed to retrace their steps multiple times within the haptic simulation to find an object they had hit, P6 was able to reliably find objects they had hit even only a single time. In spite of this, P6 scored in the middle of the pack in terms of token placement, suggesting a disconnect between exact accuracy in reproducing object locations and the general strength of the kinesthetic sense. This would be consistent with literature that indicates proprioceptive capacity and accuracy are not necessarily linked [81]. P6 also had the strongest positive reactions towards the system, rating it a 5 out of 5 for both utility and enjoyment.

Regardless of the remarks made during the session, when asked to sit down and rate the system, most participants gave it roughly the same rating (see Figure 5.4). The audio version of the system scored slightly higher than the audio-haptic version, but both scored between a three and a four. When explaining their scores, participants expressed that while they could see the potential in such a system, they did not feel like it was currently well developed enough to be something they would use in their day-to-day lives. Participants found them novel and interesting, but they did not understand what exactly these rendering strategies would be used for. This is potentially a reflection of the mismatch between what participants are interested in knowing about photographs and what information is given to them by our

rendering strategy. In our preliminary interviews, participants were not as interested in the exact locations of the objects in a photograph as they were in the actions, emotions, and physical characteristics of the objects and subjects. Our rendering strategy focuses on the location of objects and does not give much insight into the things participants said they cared about most. This could be resolved by improving detection of detail attributes by our ML, and is currently being worked on by members of the IMAGE team.

Several alternative rendering strategies were suggested by participants. P6 found that volume was not a very helpful indicator of object size, and found themselves instinctively assuming that it had to do with distance. This would be a sensible conclusion, since in the every day experience of BPSP, quieter sounds are usually further away. The rendering strategies suggested were related to the specifics of a participant's vision loss. The three participants with no vision from birth indicated that the technique were not helpful because they had such a limited understanding of photographs. One of these participants explained that when they were encountering an object in the audio-haptic rendering, they were imagining it as if they were exploring a top-down perspective of the room. This is consistent with P11's understanding of the end effector as a type of cane they were moving in front of them to encounter objects. The congenitally blind segment of participants also suggested the use of "realistic sounds" to indicate what is present in an image. For example, if there is a stream in a landscape, they would like to hear the noise of running water. This approach is similar to the proxemic system developed by Rector et al [53]. P15

described our current system as “a sighted approach to a blind problem” and emphasized the need for alternative approaches for the congenitally blind.

Chapter 6

Conclusion

The are many challenges in developing a flexible system that allows the diverse BPSP community to explore web images. Depending on the capabilities of a user, and the time of onset of their vision loss, their mental models of visual information will differ. Prior work to translate images to audio or audio-haptic representations focused primarily on specific use cases such as line graphs, and often did not give users a sense of exploration when interacting with graphic content on the web. This thesis examined and organized the capabilities and preferences of BPSP that make designing for them so challenging, designed a UI to account for these capabilities and preferences, and evaluated the efficacy of the IMAGE system's renderings when used by BPSP.

The survey and interviews with BPSP established a desire within the community for increased access to graphics on the web. We confirmed that the majority of our users use

Microsoft computers and Apple smartphones, and that Braille literacy is most common among those identifying as blind (as opposed to those who identify as visually impaired). We learned that although many BPSP currently ignore images they encounter, that does not mean they are not interested in their contents. BPSP want as much information as possible about images to pick and choose from, but they also want a summary of the information so that they do not waste their time. Finally, we discovered that system requirements would differ widely between users according to the nature of their disability.

The UI of the IMAGE system was designed using the findings of our survey and interviews. The settings page was created with developers and end users in mind, and has options targeted towards both groups. The renderings window went through multiple design iterations in order to find a UI that would adhere to WCAG standards while giving users a sense of exploration and agency.

As the IMAGE project evolves, the system will see expanded capabilities. The addition of the OCR preprocessor and handler, improved object recognition, and more sophisticated captioning will require corresponding advances to the system UI. The settings will similarly need to be enhanced in order to accommodate a wider breadth of renderings, as well as language options. As of December 2022, a French TTS module has been added to the IMAGE code base. Once fully integrated, IMAGE will also be accessible to the French speaking population.

Our experiences with users' reactions to the object detection and semantic

segmentation techniques presented indicate that while spatialization may be useful in representing graphics whose primary contents are spatial data, a different technique may be preferable for representing photographs. This theme was present both in the interviews with participants and in the results of the Craft Study. However, the spatialization combined with haptic feedback could be highly beneficial to the subset of the blind community with above-average proprioceptive abilities. The audio-haptic system gives these users an understanding of the locations of objects in an image and allows them to piece together a mental model of the composition of the picture they are exploring. The majority of participants were able to extract a gestalt understanding of the images, even using audio-only renderings.

In the future we would like to compare participant enjoyment of an approach relying more heavily on earcons and ambient sounds, to the spatialized audio described in this thesis. Using the sounds that would be present in a scene captured by a photograph as a rendering of the image has precedent, not only for generated representations, but also as a way to help BPSP take and share pictures [53, 83]. This approach was requested by participants both in the initial surveys and interviews, and as feedback during the Craft Study. The soundscape approach was initially avoided because of the technical difficulty of creating a sound library that corresponds properly with ML tagging capabilities, but the consistent support of such a system is a strong indicator that it should be pursued. We still have not arrived at a ubiquitous and universal method of conveying graphical web content to BPSP,

but the IMAGE system gives BPSP capabilities with regards to Internet photographs that they did not have before, and the breadth of offerings will expand as the system continues to develop.

Bibliography

[1] "The History of Online Photo Sharing: Part 1," Sep. 2015. [Online]. Available: https://twirpz.wordpress.com/2015/09/26/the-history-of-online-photo-sharing-part-1/ [Accessed: 2022-11-06]

[2] M. R. Morris, A. Zolyomi, C. Yao, S. Bahram, J. P. Bigham, and S. K. Kane, ""With most of it being pictures now, I rarely use it": Understanding Twitter's Evolving Accessibility to Blind Users," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: Association for Computing Machinery, May 2016, pp. 5506–5516.

[3] Y. Li and Y. Xie, "Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement," *Journal of Marketing Research*, vol. 57, no. 1, pp. 1–19, Feb. 2020.

[4] V. Voykinska, S. Azenkot, S. Wu, and G. Leshed, "How Blind People Interact with Visual Content on Social Networking Services," in *Proceedings of the 19th ACM*

Conference on Computer-Supported Cooperative Work & Social Computing, ser. CSCW '16. New York, NY, USA: Association for Computing Machinery, Feb. 2016, pp. 1584–1595.

[5] C. Gleason, P. Carrington, C. Cassidy, M. R. Morris, K. M. Kitani, and J. P. Bigham, ""It's almost like they're trying to hide it": How User-Provided Image Descriptions Have Failed to Make Twitter Accessible | The World Wide Web Conference," in *The World Wide Web Conference*, ser. WWW '19. New York, NY, USA: Association for Computing Machinery, May 2019, pp. 549–559.

[6] E. Whitaker, "Picture Smart In JAWS: Independently Selecting Your Artwork," Apr. 2020. [Online]. Available: https://blog.freedomscientific.com/picture-smart-in-jaws-independently-selecting-your-artwork/ [Accessed: 2022-11-06]

[7] J. Regimbal, J. R. Blum, and J. R. Cooperstock, "IMAGE: a deployment framework for creating multimodal experiences of web graphics," in *Proceedings of the 19th International Web for All Conference*, ser. W4A '22. New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 1–5.

[8] P. S. C. of Canada, "Guide for Assessing Persons with Disabilities - How to determine and implement assessment accommodations - Vision disabilities," Aug. 2007, last Modified: 2007-08-15. [Online]. Available: https://www.canada.ca/en/public-service-commission/services/public-service-hiring-guides/guide-assessing-persons-disabilities/

guide-assessing-persons-disabilities/guide-assessing-persons-disabilities-determine-implement-assessment-accommodations-vision-disabilities.html [Accessed: 2022-11-04]

[9] J. P. Bigham, C. Jayant, A. Miller, B. White, and T. Yeh, "VizWiz::LocateIt - enabling blind people to locate objects in their environment," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, Jun. 2010, pp. 65–72.

[10] P. Vetter, Bola, L. Reich, M. Bennett, L. Muckli, and A. Amedi, "Decoding Natural Sounds in Early "Visual" Cortex of Congenitally Blind Individuals," *Current Biology*, vol. 30, no. 15, pp. 3039–3044.e2, Aug. 2020.

[11] M. Szubielska, E. Niestorowicz, and B. Marek, "Drawing without eyesight. Evidence from congenitally blind learners," *Roczniki Psychologiczne*, vol. 19, no. 4, pp. 681–700, 2016, 681.

[12] ——, "The Relevance of Object Size to the Recognizability of Drawings by Individuals with Congenital Blindness," *Journal of Visual Impairment & Blindness*, vol. 113, no. 3, pp. 295–310, May 2019.

[13] G. Miletic, "Perspective Taking: Knowledge of Level 1 and Level 2 Rules by Congenitally Blind, Low Vision, and Sighted Children," *Journal of Visual Impairment & Blindness*, vol. 89, no. 6, pp. 514–523, Nov. 1995.

[14] D. Goldreich and I. M. Kanics, "Tactile Acuity is Enhanced in Blindness," *Journal of Neuroscience*, vol. 23, no. 8, pp. 3439–3445, Apr. 2003.

[15] C. J. Sabourin, Y. Merrikhi, and S. G. Lomber, "Do blind people hear better?" *Trends in Cognitive Sciences*, vol. 26, no. 11, pp. 999–1012, Nov. 2022.

[16] M. A. Heller, J. A. Calcaterra, L. A. Tyler, and L. L. Burson, "Production and Interpretation of Perspective Drawings by Blind and Sighted People," *Perception*, vol. 25, no. 3, pp. 321–334, Mar. 1996.

[17] D. Picard, C. Jouffrais, and S. Lebaz, "Haptic Recognition of Emotions in Raised-Line Drawings by Congenitally Blind and Sighted Adults," *IEEE Transactions on Haptics*, vol. 4, no. 1, pp. 67–71, Jan. 2011, conference Name: IEEE Transactions on Haptics.

[18] A. Theurel, S. Frileux, Y. Hatwell, and E. Gentaz, "The Haptic Recognition of Geometrical Shapes in Congenitally Blind and Blindfolded Adolescents: Is There a Haptic Prototype Effect?" *PLOS ONE*, vol. 7, no. 6, p. e40251, Jun. 2012.

[19] C. Willings, "Guidelines and Standards for Tactile Graphics." [Online]. Available: https://www.brailleauthority.org/tg/ [Accessed: 2023-02-22]

[20] "Creating Tactual Graphics for Students who are Blind or Visually Impaired." [Online]. Available: https://www.teachingvisuallyimpaired.com/tactile-graphics-guidelines.html [Accessed: 2023-02-22]

[21] J. M. Kennedy, "How the Blind Draw," *Scientific American*, vol. 276, no. 1, pp. 76–81, 1997.

[22] M. A. Heller, "Picture and Pattern Perception in the Sighted and the Blind: The Advantage of the Late Blind," *Perception*, vol. 18, no. 3, pp. 379–389, Jun. 1989.

[23] Knowbility, "A Brief History of Screen Readers." [Online]. Available: https://knowbility.org/blog/2021/a-brief-history-of-screen-readers [Accessed: 2022-10-24]

[24] Y. Borodin, J. P. Bigham, G. Dausch, and I. V. Ramakrishnan, "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)*, ser. W4A '10. New York, NY, USA: Association for Computing Machinery, Apr. 2010, pp. 1–10.

[25] J. Nielsen and K. Pernice, *Eyetracking web usability*. Berkeley, California: New Riders, 2010.

[26] A. Raj and R. Rosenholtz, "What your design looks like to peripheral vision," in *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, ser. APGV '10. New York, NY, USA: Association for Computing Machinery, Jul. 2010, pp. 89–92.

[27] A. Edwards, "Soundtrack: An Auditory Interface for Blind Users," *Human-Computer Interaction*, vol. 4, no. 1, pp. 45–66, Mar. 1989.

[28] S. H. Kurniawan, A. G. Sutcliffe, P. L. Blenkhorn, and J.-E. Shin, "Investigating the usability of a screen reader and mental models of blind users in the Windows environment," *International Journal of Rehabilitation Research*, vol. 26, no. 2, pp. 145–147, Jun. 2003.

[29] A. Savidis and C. Stephanidis, "The HOMER UIMS for dual user interface development: Fusing visual and non-visual interactions," *Interacting with Computers*, vol. 11, no. 2, pp. 173–209, Dec. 1998.

[30] A. Bouraoui, "Component based development of non-visual applications using braille-speech widgets," in *2007 IEEE/ACS International Conference on Computer Systems and Applications*, May 2007, pp. 84–91.

[31] V. K. Emery, P. J. Edwards, J. A. Jacko, K. P. Moloney, L. Barnard, T. Kongnakorn, F. Sainfort, and I. U. Scott, "Toward achieving universal usability for older adults through multimodal feedback," in *Proceedings of the 2003 conference on Universal usability*, ser. CUU '03. New York, NY, USA: Association for Computing Machinery, Jun. 2002, pp. 46–53.

[32] C. Sjöström, "Non-Visual Haptic Interaction Design - Guidelines and Applications," Doctoral Thesis (compilation), Certec, Lund University, 2002.

[33] S. Leuthold, J. A. Bargas-Avila, and K. Opwis, "Beyond web content accessibility guidelines: Design of enhanced text user interfaces for blind internet users,"

International Journal of Human-Computer Studies, vol. 66, no. 4, pp. 257–270, Apr. 2008.

[34] S. Morley, "The design and evaluation of non-visual information systems for blind users," Ph.D. dissertation, University of Hertfordshire, 1999.

[35] S. P. Casali, "A physical skills based strategy for choosing an appropriate interface method," in *Extra-ordinary human-computer interaction: interfaces for users with disabilities*. USA: Cambridge University Press, Dec. 1995, pp. 315–341.

[36] K. Fukuda, S. Saito, H. Takagi, and C. Asakawa, "Proposing new metrics to evaluate web usability for the blind," in *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '05. New York, NY, USA: Association for Computing Machinery, Apr. 2005, pp. 1387–1390.

[37] "WebAIM: Skip Navigation Links." [Online]. Available: https://webaim.org/techniques/skipnav/ [Accessed: 2022-12-10]

[38] F. Alonso, J. L. Fuertes, L. González, and L. Martínez, "A Framework for Blind User Interfacing," in *Computers Helping People with Special Needs*, ser. Lecture Notes in Computer Science, K. Miesenberger, J. Klaus, W. L. Zagler, and A. I. Karshmer, Eds. Berlin, Heidelberg: Springer, 2006, pp. 1031–1038.

[39] ——, "User-Interface Modelling for Blind Users," ser. Lecture Notes in Computer Science, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds. Berlin, Heidelberg: Springer, 2008, pp. 789–796.

[40] D. Tzovaras, G. Nikolakis, G. Fergadis, S. Malasiotis, and M. Stavrakis, "Design and implementation of virtual environments training of the visually impaired," in *Proceedings of the fifth international ACM conference on Assistive technologies*, ser. Assets '02. New York, NY, USA: Association for Computing Machinery, Jul. 2002, pp. 41–48.

[41] A. Tanaka and A. Parkinson, "Haptic Wave: A Cross-Modal Interface for Visually Impaired Audio Producers," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: Association for Computing Machinery, May 2016, pp. 2150–2161.

[42] M. R. Morris, J. Johnson, C. L. Bennett, and E. Cutrell, "Rich Representations of Visual Content for Screen Reader Users," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, Apr. 2018, pp. 1–11.

[43] R. Girgis, "Deep Learning to Assist Visually Impaired Individuals with Visual Exploration," M.E., McGill University (Canada).

[44] S. Wu, J. Wieland, O. Farivar, and J. Schiller, "Automatic Alt-text: Computer-generated Image Descriptions for Blind Users on a Social Network Service," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ser. CSCW '17. New York, NY, USA: Association for Computing Machinery, Feb. 2017, pp. 1180–1192.

[45] C. Low, E. McCamey, C. Gleason, P. Carrington, J. P. Bigham, and A. Pavel, "Twitter A11y: A Browser Extension to Describe Images," in *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, ser. ASSETS '19. New York, NY, USA: Association for Computing Machinery, Oct. 2019, pp. 551–553.

[46] J. Choi, S. Jung, D. G. Park, J. Choo, and N. Elmqvist, "Visualizing for the Non-Visual: Enabling the Visually Impaired to Use Visualization," *Computer Graphics Forum*, vol. 38, no. 3, pp. 249–260, 2019.

[47] D. Worrall, M. Bylstra, S. Barrass, and R. Dean, "SONIPY: THE DESIGN OF AN EXTENDABLE SOFTWARE FRAMEWORK FOR SONIFICATION RESEARCH AND AUDITORY DISPLAY," in *International Conference on Auditory Display (ICAD 2008), Montréal, Canada*, Jan. 2007.

[48] "Sonification | Highcharts," 2022. [Online]. Available: https://highcharts.com/docs/accessibility/sonification [Accessed: 2022-11-14]

[49] F. Grond and T. Hermann, "Singing function," *Journal on Multimodal User Interfaces*, vol. 5, no. 3, pp. 87–95, May 2012.

[50] S. Cavaco, J. T. Henriques, M. Mengucci, N. Correia, and F. Medeiros, "Color Sonification for the Visually Impaired," *Procedia Technology*, vol. 9, pp. 1048–1057, Jan. 2013.

[51] O. K. Toffa and M. Mignotte, "A Hierarchical Visual Feature-Based Approach For Image Sonification," *IEEE Transactions on Multimedia*, vol. 23, pp. 706–715, 2021.

[52] J. A. Ferwerda and V. T.-H. Kwok, "Multimodal MATLAB: data visualization for the blind," 2006. [Online]. Available: https://www.cis.rit.edu/people/faculty/ferwerda/publications/2006/ferwerda_mist_poster_v12.pdf

[53] K. Rector, K. Salmon, D. Thornton, N. Joshi, and M. R. Morris, "Eyes-Free Art: Exploring Proxemic Audio Interfaces For Blind and Low Vision Art Engagement," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 93:1–93:21, Sep. 2017.

[54] A. Supper, "Lobbying for the ear, listening with the whole body: the (anti-)visual culture of sonification," *Sound Studies*, vol. 2, no. 1, pp. 69–80, Jan. 2016.

[55] J. Fritz and K. Barner, "Design of a haptic data visualization system for people with visual impairments," *IEEE Transactions on Rehabilitation Engineering*, vol. 7, no. 3, pp. 372–384, Sep. 1999.

[56] D. T. Pawluk, R. J. Adams, and R. Kitada, "Designing Haptic Assistive Technology for Individuals Who Are Blind or Visually Impaired," *IEEE Transactions on Haptics*, vol. 8, no. 3, pp. 258–278, Jul. 2015, conference Name: IEEE Transactions on Haptics.

[57] T. Zhang, B. S. Duerstock, and J. P. Wachs, "Multimodal Perception of Histological Images for Persons Who Are Blind or Visually Impaired," *ACM Transactions on Accessible Computing*, vol. 9, no. 3, pp. 7:1–7:27, Jan. 2017.

[58] S. Brewster, "Visualization tools for blind people using multiple modalities," *Disability and Rehabilitation*, vol. 24, no. 11-12, pp. 613–621, Jan. 2002.

[59] W. Yu, D. Reid, and S. Brewster, "Web-based Multimodal Graphs for Visually Impaired People," in *Universal Access and Assistive Technology*, S. Keates, P. Langdon, P. J. Clarkson, and P. Robinson, Eds. London: Springer, 2002, pp. 97–108.

[60] W. Jeong, "Force feedback textual and graphic displays for the blind," *Proceedings of the American Society for Information Science and Technology*, vol. 43, no. 1, pp. 1–11, 2006.

[61] Chavda, Niraj, "Orbit Research Introduces the Graphiti® Interactive Tactile Graphic Display." [Online]. Available: http://www.orbitresearch.com/orbit-research-introduces-the-graphiti-interactive-tactile-graphic-display/ [Accessed: 2022-11-15]

[62] S. O'Modhrain, N. A. Giudice, J. A. Gardner, and G. E. Legge, "Designing Media for Visually-Impaired Users of Refreshable Touch Displays: Possibilities and Pitfalls," *IEEE Transactions on Haptics*, vol. 8, no. 3, pp. 248–257, Jul. 2015.

[63] R. Rastogi, T. V. D. Pawluk, and J. Ketchum, "Intuitive Tactile Zooming for Graphics Accessed by Individuals Who are Blind and Visually Impaired," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 4, pp. 655–663, Jul. 2013, conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering.

[64] H. C. Stronks, D. J. Parker, J. Walker, P. Lieby, and N. Barnes, "The Feasibility of Coin Motors for Use in a Vibrotactile Display for the Blind," *Artificial Organs*, vol. 39, no. 6, pp. 480–491, 2015.

[65] P. Wacker, C. Wacharamanotham, D. Spelmezan, J. Thar, D. A. Sánchez, R. Bohne, and J. Borchers, "VibroVision: An On-Body Tactile Image Guide for the Blind," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '16. New York, NY, USA: Association for Computing Machinery, May 2016, pp. 3788–3791.

[66] K. Zhao, M. Serrano, B. Oriola, and C. Jouffrais, "VibHand: On-Hand Vibrotactile Interface Enhancing Non-Visual Exploration of Digital Graphics," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. ISS, pp. 207:1–207:19, Nov. 2020.

[67] H. P. Palani, J. L. Tennison, G. B. Giudice, and N. A. Giudice, "Touchscreen-Based Haptic Information Access for Assisting Blind and Visually-Impaired Users: Perceptual Parameters and Design Guidelines," in *Advances in Usability, User Experience and Assistive Technology*, ser. Advances in Intelligent Systems and Computing, T. Z. Ahram and C. Falcão, Eds. Springer International Publishing, 2019, pp. 837–847.

[68] "Mobile Phones and Tablets." [Online]. Available: https://www.cnib.ca/en/mobile-phones-and-tablets [Accessed: 2022-11-15]

[69] P. Tantribeau and H. C. Lee, "A stereo auditory image-perception aid for the blind," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 4, 1992, pp. 1580–1581.

[70] Y. Zhong, W. S. Lasecki, E. Brady, and J. P. Bigham, "RegionSpeak: Quick Comprehensive Spatial Descriptions of Complex Images for Blind Users," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: Association for Computing Machinery, Apr. 2015, pp. 2353–2362.

[71] S. C. Government of Canada, "The Daily — Canadian Income Survey, 2020," Mar. 2022, last Modified: 2022-03-23. [Online]. Available: https://www150.statcan.gc.ca/n1/daily-quotidien/220323/dq220323a-eng.htm [Accessed: 2022-11-16]

[72] "Use VoiceOver for images and videos on iPhone." [Online]. Available: https://support.apple.com/en-ca/guide/iphone/iph37e6b3844/ios [Accessed: 2022-11-15]

[73] "About CCB – Canadian Council of the Blind." [Online]. Available: https://ccbnational.net/shaggy/about-ccb/ [Accessed: 2022-11-07]

[74] R. M. Sheffield, F. M. D'Andrea, V. Morash, and S. Chatfield, "How Many Braille Readers? Policy, Politics, and Perception," *Journal of Visual Impairment & Blindness*, vol. 116, no. 1, pp. 14–25, Jan. 2022.

[75] "Blindness Statistics | National Federation of the Blind." [Online]. Available: https://nfb.org/resources/blindness-statistics [Accessed: 2022-11-16]

[76] J. C. Stevens, E. Foulke, and M. Q. Patterson, "Tactile acuity, aging, and braille reading in long-term blindness," *Journal of Experimental Psychology: Applied*, vol. 2, pp. 91–106, 1996.

[77] "WebAIM: Screen Reader User Survey #9 Results." [Online]. Available: https://webaim.org/projects/screenreadersurvey9/ [Accessed: 2022-11-15]

[78] J. Besser, M. Stropahl, E. Urry, and S. Launer, "Comorbidities of hearing loss and the implications of multimorbidity for audiological care," *Hearing Research*, vol. 369, pp. 3–14, Nov. 2018.

[79] J. Lazar, A. Allen, J. Kleinman, and C. Malarkey, "What Frustrates Screen Reader Users on the Web: A Study of 100 Blind Users," *International Journal of Human-Computer Interaction*, vol. 22, no. 3, pp. 247–269, May 2007.

[80] M. A. Akeroyd, "An Overview of the Major Phenomena of the Localization of Sound Sources by Normal-Hearing, Hearing-Impaired, and Aided Listeners," *Trends in Hearing*, vol. 18, Oct. 2014.

[81] E. T. Wilson, J. Wong, and P. L. Gribble, "Mapping Proprioception across a 2D Horizontal Workspace," *PLOS ONE*, vol. 5, no. 7, p. e11851, Jul. 2010.

[82] A. J. Kolarik, S. Cirstea, S. Pardhan, and B. C. J. Moore, "A summary of research investigating echolocation abilities of blind and sighted humans," *Hearing Research*, vol. 310, pp. 60–68, Apr. 2014.

[83] S. Harada, D. Sato, D. W. Adams, S. Kurniawan, H. Takagi, and C. Asakawa, "Accessible photo album: enhancing the photo sharing experience for people with visual impairment," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: Association for Computing Machinery, Apr. 2013, pp. 2127–2136.

Appendix A

Tables

P1								
Image	Obj1x	Obj1y	Obj2x	Obj2y	Obj3x	Obj3y	Obj4x	Obj4y
indoor1	0.022	0.081	0.079	0.156	0.057	0.091	0.048	0.104
indoor2	0.189	0.136	0.137	0.273	0.195	0.076	0.013	0.254
outdoor1	0.238	0.297			0.029	0.269		
outdoor2	0.197	0.347	0.201	0.046	0.153	0.1	0.109	0.240
mixed1	0.089	0.127	0.145	0.029	0.187	0.065	0.043	0.261
mixed2	0.041	0.030	0.156	0.132	0.110	0.124	0.024	0.196
people1	0.119	0.052	0.081	0.182	0.132	0.001	0.053	0.204
people2	0.201	0.128	0.327	0.191	0.005	0.179	0.205	0.313
P11								

Image	Obj1x	Obj1y	Obj2x	Obj2y	Obj3x	Obj3y	Obj4x	Obj4y
indoor1	0.030	0.186	0.125	0.150	0.059	0.046	0.163	0.254
indoor2	0.003	0.055	0.003	0.120	0.140	0.227	0.131	0.230
outdoor1	0.184	0.312	0.085	0.103	0.075	0.136	0.131	0.080
outdoor2	0.056	0.158	0.109	0.127	0.059	0.213	0.032	0.195
mixed1	0.034	0.154			0.102	0.078	0.123	0.221
mixed2	0.147	0.204	0.033	0.046	0.212	0.119	0.067	0.078
people1	0.119	0.052	0.075	0.258			0.059	0.032
people2	0.122	0.246	0.023	0.156	0.174	0.014	0.256	0.173
P15								
Image	Obj1x	Obj1y	Obj2x	Obj2y	Obj3x	Obj3y	Obj4x	Obj4y
indoor1	0.088	0.210	0.221	0.198	0.021	0.073	0.175	0.197
indoor2	0.189	0.649	0.485	0.270	0.197	0.082	0.015	0.260
outdoor1	0.060	0.483	0.161	0.202	0.067	0.137	0.039	0.074
outdoor2	0.210	0.002	0.199	0.136	0.265	0.001	0.254	0.168
mixed1	0.049	0.163	0.197	0.079	0.186	0.238	0.261	0.246
mixed2	0.053	0.106	0.051	0.069	0.139	0.050	0.019	0.014
people1	0.024	0.211	0.122	0.020	0.053	0.071	0.038	0.228
people2	0.067	0.187	0.138	0.227	0.079	0.108	0.141	0.079

P5								
Image	Obj1x	Obj1y	Obj2x	Obj2y	Obj3x	Obj3y	Obj4x	Obj4y
indoor1	0.026	0.123	0.051	0.042	0.121	0.029	0.009	0.203
indoor2	0.155	0.335	0.057	0.342	0.019	0.025	0.051	0.070
outdoor1	0.050	0.327	0.033	0.403	0.105	0.425	0.223	0.422
outdoor2	0.183	0.064	0.097	0.006	0.045	0.209	0.048	0.221
mixed1	0.034	0.154	0.187	0.038	0.102	0.078	0.123	0.221
mixed2	0.057	0.009	0.104	0.047	0.048	0.039		
people1	0.026	0.073	0.040	0.204	0.004	0.085		
people2	0.189	0.004	0.030	0.041	0.153	0.093	0.202	0.054
P6								
Image	Obj1x	Obj1y	Obj2x	Obj2y	Obj3x	Obj3y	Obj4x	Obj4y
indoor1	0.100	0.168	0.087	0.105	0.007	0.098	0.027	0.292
indoor2	0.125	0.311	0.033	0.273	0.079	0.032	0.081	0.115
outdoor1	0.084	0.129	0.081	0.313	0.079	0.407	0.123	0.428
outdoor2	0.198	0.176	0.148	0.108	0.073	0.233	0.104	0.162
mixed1	0.071	0.309	0.210	0.014	0.256	0.077	0.141	0.130
mixed2	0.010	0.040	0.077	0.113	0.149	0.169	0.087	0.339
people1	0.047	0.346	0.002	0.081	0.062	0.112	0.137	0.032

people2	0.209	0.197	0.134	0.266	0.005	0.378	0.068	0.252
P7								
Image	Obj1x	Obj1y	Obj2x	Obj2y	Obj3x	Obj3y	Obj4x	Obj4y
indoor1	0.018	0.207	0.157	0.033	0.163	0.005	0.223	0.005
indoor2	0.061	0.290	0.113	0.180	0.007	0.29	0.003	0.148
outdoor1	0.028	0.168	0.345	0.251	0.191	0.110	0.053	0.127
outdoor2	0.238	0.060	0.133	0.101	0.105	0.14	0.010	0.133
mixed1	0.028	0.686	0.254	0.621	0.354	0.692	0.076	0.387
mixed2	0.042	0.040	0.114	0.263	0.077	0.145	0.218	0.337
people1	0.079	0.410	0.083	0.420			0.073	0.376
people2	0.087	0.142	0.027	0.070	0.081	0.036	0.012	0.135
P8								
Image	Obj1x	Obj1y	Obj2x	Obj2y	Obj3x	Obj3y	Obj4x	Obj4y
indoor1	0.084	0.090	0.065	0.342	0.161	0.096	0.119	0.193
indoor2								
outdoor1	0.028	0.168	0.345	0.251	0.191	0.110	0.053	0.127
outdoor2	0.183	0.347	0.237	0.316	0.148	0.399	0.121	0.381
mixed1	0.103	0.327	0.426	0.046			0.050	0.072
mixed2								

people1	0.061	0.152	0.406	0.287				
people2			0.113	0.136	0.019	0.249		
P9								
Image	Obj1x	Obj1y	Obj2x	Obj2y	Obj3x	Obj3y	Obj4x	Obj4y
indoor1	0.118	0.084	0.205	0.219	0.189	0.145	0.033	0.319
indoor2	0.129	0.368	0.089	0.375	0.196	0.106	0.045	0.127
outdoor1	0.088	0.131	0.391	0.112	0.337	0.104	0.303	0.110
outdoor2	0.128	0.221	0.046	0.026	0.063	0.320	0.283	0.154
mixed1	0.039	0.210			0.028	0.164	0.010	0.212
mixed2	0.125	0.057	0.023	0.619	0.121	0.606	0.186	0.195
people1	0.09	0.025	0.129	0.204	0.175	0.077	0.127	0.239
people2	0.067	0.191	0.137	0.230	0.077	0.108	0.14	0.070

Table A.1: x and y distances from centroid, by participant. Black squares indicate missed objects or incomplete tasks.