VIRTUAL WORLDS BEYOND SIGHT: DESIGNING AND EVALUATING AN AUDIO-HAPTIC SYSTEM FOR NON-VISUAL VR EXPLORATION

by

Aayush Shrestha

Submitted in partial fulfillment of the requirements
for the degree of Master of Computer Science

at

Dalhousie University
Halifax, Nova Scotia
August 2024

Dedicated to my parents, my ***Mamu and Papa****.*
I find solace in your unconditional love and selfless support. Your sacrifices are the foundation of my achievements. Thank you for everything!

Table of Contents

List of Tables

List of Figures

Abstract

Virtual Reality (VR), predominantly focusing on visuospatial renderings in its contemporary approach, has created a conservative narrative, making VR solely analogous to a mediated visual experience. While accessibility is included in the developmental phase of commercial VR applications, it is often considered an add-on, resulting in sub-par virtual experiences that often exclude visually impaired users. This research addresses these limitations by designing a hapto-acoustic VR system that leverages spatial audio and haptic feedback for sensory substitution of visual dominance in VR. A large-scale urban Virtual Environment (VE) was created using the Unity Game Engine, incorporating a physical cane prototype coupled with a virtual cane for interaction and an omnidirectional slide mill for navigation. A user study with 20 normally sighted participants evaluated and compared the system's effectiveness in texture differentiation and navigation tasks under two conditions: with visual cues and exclusively through audio-haptic feedback. The study results indicated that even with minimal training and limited prior VR experience, participants could navigate the environment effectively in non-visual conditions, though at the cost of increased cognitive load and error rates compared to visual conditions. The evaluation highlights the necessity for improved feedback mechanisms and suggests further validation with visually impaired users. The overall research contributes to the development of accessible VR systems through a novel white cane prototype, realistic spatial audio effects and a comprehensive evaluation demonstrating the system's potential in aiding non-visual navigation in a complex, large-scale VE while also engendering empathetic literacy among sighted users.

List of Abbreviations and Symbols Used

6DoF Six Degrees of Freedom.

AMD Age-Related Macular Degeneration.

ANOVA Analysis of Variance.

ASIC Application-Specific Integrated Circuit.

DFTs Discrete Fourier Transforms.

DHH Deaf and Hard of Hearing.

GE Game Engine.

GO Game Object.

HWD Head Worn Display.

IMU Inertial Measuring Unit.

IR Infra Red.

MoCap Motion Capture.

NASA-TLX NASA Task Load IndeX.

NLP No Light Perception.

NPC Non-Playable Character.

NV Non-Visual.

O&M Orientation and Mobility.

OOB Out-of-the-Box.

PWM	Pulse Width Modulation.
RQ	Research Question.
RTPCs	Real-Time Parameter Controls.
SFX	Sound Effects.
SUS	System Usability Scale.
TTR	Time to Reach.
UI	User Interface.
VE	Virtual Environment.
VI	Visual Impairment.
VR	Virtual Reality.
WAV	Waveform Audio File.
WV	With-Visual.

Acknowledgements

> "The most difficult thing in life is the decision to act, the rest is merely tenacity."
>
> – Amelia Earhart

This thesis, once a distant dream, has now come to fruition. I have learned that to realize your dream, one must carve it into existence. With heartfelt gratitude, I thank everyone who has been part of this incredible journey.

I would like to acknowledge the support of my supervisor - Dr. Joseph Malloch, who always guided me, supported me, praised me and, unbeknownst to himself, has changed my life in ways words can't entail. Thank you, Joe! I will always be grateful for your trust in starting my graduate journey, and I will surely miss our weekly one-on-ones and casual cabin conversations.

Special thanks to my thesis committee members, Dr. Derek Reilly and Dr. Mayra Barrera Machuca, for taking the time to read my thesis. I also extend my gratitude to Dr. Lizbeth Escobedo for graciously agreeing to chair my thesis defense.

I am deeply grateful to the GEM lab members, my "GEM Fam," for being a constant source of learning and inspiration, in every way possible - from casual water cooler chit-chats to formal meetings. Special mentions to Matthew Peachey and Nirmal Adhikari: Matt, for patiently listening to my rants and mentoring me when things made no sense; and Nirmal, for reminding me not to be too hard on myself and for the countless ping-pong breaks. Thank you both for always encouraging me.

Finally, I thank my parents in Nepal, who have always been there to support me and back me despite the distance and the time difference. Your ever-present love and encouragement always pushed me to strive for the best.

I conclude by acknowledging the Almighty for having made this academic journey full of growth and serendipity.

Chapter 1

Introduction

Virtual Reality (VR), often touted as a democratizing way to access new worlds and experiences, has traditionally been associated with visual phenomena and for most, it simulates spatial information through stereoscopic rendering presented through a Head Worn Display (HWD) [64]. While the potential of VR as a technology is immense in various use cases, most of the applications primarily rely on realistic visual feedback to provide an immersive experience to sighted people for whom visual modality is most dominant [141]. Thus, this has resulted in a very conservative narrative of VR, making it analogous to any mediated abstraction of reality that is only perceived 'visually.' The downside to this take on VR is that users who are visually impaired aren't included in the target demographics, segregating them from being able to use VR applications. One of the key principles of human-computer interaction is 'User-centered design design,' where usability concerns are incorporated from the start to make informed design decisions throughout the design process. Unfortunately, in practice, fixes for usability concerns perceived as affecting small populations (such as people with visual disabilities) are often tackled at the end of the software design life-cycle, resulting in sub-par user experiences [64]. Thus far, accessibility has not been a primary consideration during the development of commercially available VR systems for a large population of visually challenged users.

Although it is patent that the visual bandwidth supersedes other human senses [62], the immersiveness of any VR experience is not entirely defined by how virtual spaces are realistically depicted. This necessitates that other sensory feedback mechanisms, such as audio and haptics, be thoroughly leveraged. Thus, if the dominance of visuals in VR can be extricated through sensory substitution [38], it can be used as a capable assistive technology for visually impaired users. Currently, researchers have begun introducing accessibility to VR content for people who are blind or have low vision [141]. Their approach focuses on achieving this through the creation of soundscapes in the virtual world [38] as well as through telepresence techniques [88], which propose multimodal interactions with VR representations.

Unlike other screen-based interfaces, VR is inherently immersive as it requires the users to move and orient for exploring and navigating a VE. These explorations employ redirection of VR content or natural walking in physical spaces that define the scale of the VE [108] [141]. Moreover, to enable interactions with virtual objects, haptic feedback through wearable [119][123][111], or hand-held interfaces [94][67][34] mapped to a virtual representation are widely used. However, the issue for research that has been conducted in the field for visually impaired users is the disparity between the scale and complexity of VEs along with the lack of interfaces that resemble the standard white cane, which is commonly used by the visually impaired in real life. Thus, the gap in the literature suggests the need for a cane-like interface and the availability of a large-scale VE that can be explored in a safe setting for the visually impaired.

Based on these requirements, I wanted to create a large-scale urban environment that simulates a cityscape where a standard white cane mapped to a virtual counterpart acts as the interaction medium. The idea was to create a hapto-acoustic VR system to enable these requirements and devise a user study involving sighted participants in non-visual conditions to conduct a preliminary evaluation of the system's overall effectiveness. Important research questions outlined below helped model the user study that focused on assessing the system's overall efficacy, usability and cognitive impact in terms of participants being able to interact and navigate in the VE, along with insights into their experiential feedback and reported empathetic gains.

1. To what extent can users differentiate between virtual textures in non-visual conditions using the system?

2. How effectively can users navigate an unfamiliar virtual environment using only audio cues and haptic feedback, without any visual information?

3. In the absence of visual information, does navigating the virtual environment reduce cognitive load for individuals accustomed to visual cues, or does their inherent reliance on these cues increase their overall cognitive load?

The VE that hosts the complex urban infrastructure is designed and rendered in the Unity Game Engine (GE), which acts as the central facilitator for different interconnected components to work in tandem. The audio design and haptic rendering are supported through separate sound and haptic engines. The system utilizes a physical cane prototype coupled to a virtual cane through a motion capture system, while the navigation is made possible through an omnidirectional slide mill. The study involved a comprehensive evaluation of the system through a series of tasks designed to test the navigation and textural identification capabilities of participants under conditions with visuals and exclusively through audio-haptic feedback. Participants included 20 normally sighted individuals who were made devoid of visuals through the use of a blindfold and performed tasks such as road crossings and scavenger hunts conducted to assess the effectiveness of audio and haptic feedback. Quantitative data were collected on task completion times, error rates, and collision incidences, complemented by qualitative feedback through participant interviews and post-study questionnaires.

The study found that participants, despite having had limited training on the system and minimal prior VR experience, could effectively navigate the VE using only audio and haptic feedback. However, their performance in non-visual conditions was less efficient compared to when visual information was available. Participants exhibited a natural preference for auditory feedback over haptic feedback, although combined modalities provided a balanced outcome. Without visual cues, participants took longer to complete tasks and had higher rates of errors, such as vehicular collisions, indicating a cautious approach and increased cognitive load. The absence of visual information significantly heightened cognitive and physical demands, requiring greater mental effort to compensate for the lack of visual bandwidth. Interviews confirmed these findings, with participants describing heightened mental strain and difficulty maintaining orientation and direction. While the system is usable, the results highlight the innate reliance on visual cues, the need for improved feedback mechanisms and the inclusion of visually impaired participants for further validation.

The contribution of this research entailed in this thesis are as follows:

1. A handheld white cane prototype that renders tactile sensations of various virtual textures commonly found in a real-world cityscape.
2. A simulation of a large-scale virtual environment renders physically realistic audio effects modelled on the spatial properties of various urban elements, including attenuation, diffraction, occlusion, and binaural spatialization.

3. A comprehensive evaluation that demonstrates that sighted participants in non-visual conditions with no prior specialized training and limited VR experience were able to integrate multi-modal feedback to navigate and explore an unfamiliar VE.
4. A thorough assessment demonstrating behavioural changes and emotional impact in sighted participants concerning empathy towards non-sighted counterparts.

The thesis is organized as follows: The first chapter - *Literature Review* reviews the existing research on VR systems for visually impaired users, various exploration techniques and related work on haptics and audio rendering, identifying key gaps and challenges. *Methodology* outlines the phasewise ideation and iterative implementation of various components that make up the VR system. *User Study and Evaluation* details the rationale of the study design, participant recruitment strategies, data collection methods and analysis structure. The *Results* chapter presents the qualitative and quantitative findings from the study that highlight key behavioural patterns and experiential results to answer the proposed research questions. The *Discussion* chapter interprets the results in the context of the research questions, incorporates strengths and weaknesses of the current work and distills significant alterations to be implemented as future work. *Conclusion* summarizes the key aspects of the research and the essence of the thesis through its study and findings. Lastly, the appendices provide more details on the study design artifacts, as well as various apparatuses used for data collection.

Chapter 2

Literature Review

Virtual Reality (VR) has evolved significantly from entertainment and gaming to become a transformative tool in diverse fields such as education, healthcare, and assistive technology. This literature review focuses on VR's implications and applications for the visually impaired community. As technology advances, the integration of VR for visually impaired users aims not only to enhance accessibility but also to provide immersive, sensory-rich experiences that can significantly improve their quality of life. The visually impaired often face unique challenges in navigation, spatial awareness, and interaction with their environment. While beneficial, traditional methods and tools have limitations that VR technology can potentially overcome. By simulating real-world conditions through haptic and auditory feedback, VR can offer more dynamic and interactive experiences beyond what is possible with conventional aids.

This chapter reviews existing research on the use of VR for visually impaired individuals, emphasizing technological methodologies and user experiences. It explores how haptic feedback—conveying sensations through touch—and spatial audio—providing directional and environmental sound cues—are utilized to create immersive VE. These modalities are crucial in helping visually impaired users navigate and interact with VR spaces effectively. Moreover, the review is structured to provide a comprehensive analysis of the current state of VR technologies for the visually impaired, highlighting significant findings and practical applications. Additionally, it aims to identify gaps in the literature and propose areas where further research is needed.

2.1 Literature Review Process

I conducted a two-phase literature review; in the first phase, I did a wide-breadth brute-forced search for relevant literature over the past ten years. For the second phase, I followed a snowballing approach to pick up the literature that shared relevance to my planned research.

2.1.1 First Phase: Brute-Forced Wide-Breadth Review

For this phase, I conducted a broad search across multiple platforms, including the ACM Digital Library [1], Springer [8], IEEE Xplore [4], Google Scholar [3], and the Google Search Engine [2], to identify relevant research articles. I used terms like VR for the visually impaired,' soundscape and haptic-based navigation in VR,' spatial audio in VR for the visually impaired,' and navigation in VR for the blind,' among others, to search these platforms.

The procedure entailed examining all research paper titles from the past decade. Titles that suggested relevance led to a review of their abstracts. If the abstract confirmed relevance, the reference was cataloged with a rationale for its relevance. Papers that did not demonstrate relevance were disregarded. After this initial scanning, papers showing potential relevance underwent a more detailed review to confirm their significance. The focus of this literature review phase was on several key conferences:

1. ACM Conference on Human Factors in Computing Systems (CHI),
2. ACM Symposium on User Interface Software and Technology (UIST),
3. IEEE International Symposium on Mixed and Augmented Reality (ISMAR),
4. ACM International Conference on Tangible, Embedded and Embodied Interaction (TEI),
5. IEEE Conference on Virtual Reality and 3D User Interfaces (IEEEVR),
6. ACM Symposium on Virtual Reality Software and Technology (VRST),
7. International Conference on PErvasive Technologies Related to Assistive Environments (PETRA).

However, other relevant conferences like the *IEEE Haptics Symposium, IEEE Transactions on Haptics*, journals, and book chapters were also explored, which showed relevance to the research.

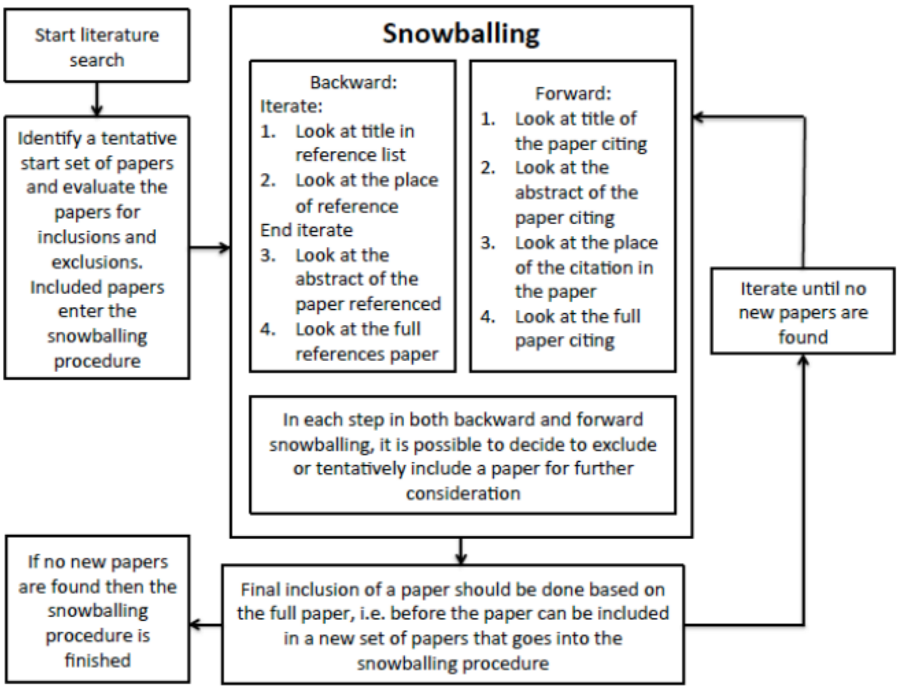


Figure 2.1: Wohlin's Snowballing Procedure for Literature Review
Image reused from [134]

2.1.2 Second Phase: Snowballing Based Review

For the second phase of the literature review, I followed Wohlin's backward snowballing procedure [134] as illustrated in Figure 2.1. By evaluating their relevance to my research, I identified 63 articles and book chapters from the first phase. I then went back and forth through the citations until no new papers were found. Ultimately, I catalogued each paper under the topic that served to create the basis of my literature review, each covering a specific facet of relevant literature such as haptics, soundscape, accessibility, empathy, assistive technology, etc. The rationale for this step was to manage the procured literature better and form a structured approach to writing the review with easy referencing.

2.2 Demographics Related Terminologies

2.2.1 Definition of Blind or Visually Impaired

According to the WHO (World Health Organization) World Report on Blindness [135], globally, Visual Impairment (VI) affects at least 2.2 billion people. The term "VI" encompasses a range of conditions characterized by a reduced visual function that cannot be fully corrected by conventional glasses, contact lenses, or surgical interventions [20]. According to WHO, VI can be classified into several categories based on visual acuity and the field of vision. “Low vision” refers to visual acuity less than 6/18 but better than 3/60 or a corresponding visual field loss of less than 20 degrees in the better eye with the best possible correction. On the other hand, “blindness” is defined by visual acuity that is less than 3/60 or a visual field loss of less than 10 degrees in the better eye [32]. These definitions help categorize the severity and type of vision impairment, ranging from mild to moderate to severe and profound, where the individual may have some light perception but no usable vision [10].

Users’ task-specific requirements and abilities of (physically) visually impaired users are oftentimes different from actually sighted users [131]. However, to follow a consistent and precise focus in our research, as outlined by Kreimeier et al. [63], we follow the definition of the term visually impaired as users with impaired vision not only because of physical impairment but also because them being in a non-visual virtual environment with an overload or unavailability of the visual modality.

2.2.2 Simulation of Visual Impairment

Simulating VI in VR presents a unique set of challenges and considerations. One critical decision in this process is whether to involve non-sighted participants or to simulate VI in sighted individuals. Current literature is ambiguous regarding the simulation of VI in sighted individuals and the correct approach to simulation. One argument might be that while providing authentic insights, there can be variability in individual experiences of blindness, while, on the contrary, the use of sighted users might allow for a more controlled and uniform baseline, facilitating a clearer analysis of a VR system’s efficacy. VI, by its inherent nature, is something that is purely subjective and evidently diverse. There are numerous VIs ranging from blurred or partial vision to complete darkness. Although only 15% of people with eye disorders have total blindness - those with a complete lack of light perception, documented as No Light Perception (NLP), the rest of the population with VI have some level of vision [73]. Research in simulating VI bifurcates in this regard as to simulating the extreme end of the spectrum (total blindness with complete darkness) or varied forms of specific visual disorders (blurred vision, restricted peripheral vision, etc). The former is something that can be accurately simulated while at the cost of portraying an exaggerated form of VI that the majority of the population does not suffer from, and the latter is something that might fall short in realistically conveying the condition bounded by the constraints of technical feasibility.

Head Worn Displays (HWDs) are a common method for simulating VIs such as glaucoma [94], cataracts [68], diabetic retinopathy [137], and Age-Related Macular Degeneration (AMD) [93]. While it might be able to correctly simulate one instance of a particular VI, as quoted in the study by Maher and Haegele [75], VI is too diverse and complex to be authentically replicated; no two people have exactly the same eye condition, and each is like a snowflake. While it is not uncommon to use HWDs to simulate VI in VR, the use of 'blindfolds' is a popular method of stimulating total blindness. Massiceti et al. [77] sought to investigate spatial navigation by using two different methods of locomotion in a virtual environment in blindfolded sighted participants. Similarly, Fialho et al., in their study of audio-based spatial navigation [42], used sighted people who were blindfolded for wayfinding tasks in the VE. Researchers even employed complete darkness to eliminate any residual light perception that might occur with less effective blindfolds [118]. Research indicates that simulating complete darkness to simulate total blindness can significantly affect participants' reliance on non-visual cues, thus providing valuable insights into the design of assistive technologies for the visually impaired [70]. However, some argue that sighted individuals wearing blindfolds are not a suitable proxy for participants living with VI and may act as an oversight or omission rather than as a conscious design decision [16].

2.2.3 Scope of 'VR'

The physical attributes of real objects and environments are perceived through various human senses in the form of vision, audition, touch, gestation and olfaction. When special software and hardware can simulate these modalities to give the user an impression of actually being in a different environment or seeing, touching, smelling, hearing or tasting an object that doesn't really exist, that becomes a mediated abstraction of reality, which is 'virtual.' However, current technical limitations of commercially available hardware and software can only represent a virtual object or environment using only three sensory modalities: visuals, audio and haptics. Thus, this research will focus on only these facets that contribute to the meaning of VR. The most well-known definition of VR is by Milgram and Kishino, who describe it as an environment "in which the participant-observer is totally immersed in, and able to interact with, a completely synthetic world." [79]. The basic interaction between humans and machines is characterized by two interlocking control loops: one on the human's and one on the computer's side [114]. Both are equipped with (mechanical or biological) sensors and actuators so that the computer simulation can give the user the visual, audio or haptic impression of seeing, hearing or touching virtual content and thus can be adopted to all sensory modalities [63].

In the context of this research, the scope of VR is a culmination of unimodal concepts of soundscape [44] and haptification [139], providing spatial and semantic information through audio and haptic feedback, respectively, for interactive exploration in a walkable VR experience. These sensory modalities are innately supplemented by the visual rendering of the mediated virtual space but are sometimes made absent for experimental integrity.

2.2.4 Orientation and Mobility Training

Orientation and Mobility (O&M) are two critical skills affected by VIs. The functional loss of these skills has significant repercussions on almost every aspect of life [49]. Orientation is defined as the "knowledge of one's distance and direction relative to things observed or remembered in the surroundings and keeping track of these spatial relationships as they can change during locomotion" [33]. Mobility is defined as "the act of moving through space in a safe and efficient manner" [40]. As a result of functional losses in these two skills, people with VI experience difficulties navigating dynamic and unfamiliar areas [110].

Wei et al.[59] describe O&M training as a critical educational and rehabilitative process to assist visually impaired individuals in navigating safely and independently. The training harnesses non-visual senses such as hearing, smell, and touch, enabling trainees to interpret their surroundings effectively. According to Virgili and Rubin [127], utilizing mobility aids is a core aspect of O&M training as the white cane teaches obstacle detection, curb locating, and route negotiation techniques. The Texas Deafblind Project [87] underscores the adaptability of O&M training to the trainee's needs, including those with additional disabilities such as deaf-blindness.

O&M training, conducted by certified specialists, occurs in diverse environments to mirror real-life situations, enhancing practical applicability. The training incorporates spatial awareness and protective techniques, including navigating using the upper protective arm technique to detect head-level obstacles and the lower protective technique for ground-level hazards [59]. Moreover, the training extends to real-world navigation across various settings, fostering critical thinking and problem-solving skills essential for making safe travel decisions [127]. This comprehensive approach not only boosts mobility but also enhances confidence, fostering greater independence and improving the overall quality of life for visually impaired individuals [87].

However, O&M training has two impediments that hinder effective, rapid, and safe learning: first, current O&M assessment and rehabilitation techniques are not standardized with regard to best practices and second, O&M training is not free of risk, whereby people with VI are exposed to potentially serious harm [93]. These are largely due to difficulties in establishing common guidelines for performance assessment of existing techniques, and trainees may sometimes experience accidental falls, undesired contact with people and objects, or trauma related to indoor or outdoor mobility [30].

2.3 VR Applications for the Visually Impaired

2.3.1 Skills Training

As discussed in section 2.2.4, VI doesn't only lead to loss of sight but also comes with mobility challenges that can prove to be a grave impediment to one's quality of life. Furthermore, navigating through tight spaces in an indoor environment and developing a sense of tactile maturity to identify real-life objects correctly add to the burden of daily life activities. Current literature that utilizes VR for the VI focuses on leveraging the power of simulation-based teaching techniques in these domains. The 'Skill Training' aspect of VR can be summarized as exploratory research on alternate methodologies to teach O&M, using a white cane in indoor settings, and object recognition through hand-held manipulation.

To circumvent the risks of O&M training that require physical presence in outdoor settings, VR simulations provide specific advantages for the training. It provides a controlled environment where users can perform specific tasks to measure their performance

[61], and the environment constitutes an ideal platform for performance assessment of different O&M techniques and interventions [69]. Thus, introductory VR sessions can be an avenue to familiarize oneself with new techniques in a completely safe space before performing real training sessions [70]. Ricci et al. [94] experimented with simulating glaucoma (peripheral vision loss) in a controlled study that constituted tasks primarily focusing on outdoor mobility obstacle avoidance and transportation system use. Moreover, Orly Lahav's study [69] focused on testing VR controller-aided orientation techniques for people with VI to acquire indoor and constrained spatial information.

Moreover, the feasibility study [123] by Tzovaras et al. that used a Cyber Glove to train users in understanding the shape and geometry of virtual objects along with object manipulation highlights the use of VR for training novice visually impaired individuals to develop tactile sense of daily objects. Furthermore, white cane simulations for training O&M skills in indoor spaces through proxy object collision [108] [141] is another example of skills training in VR for novice cane users.

2.3.2 Empathetic Literacy

According to Singer [107], empathy refers to the ways and extent to which an individual can develop knowledge and understanding of the physical and social world from the perspective of an embodied other. VR is substantially used for engendering a sense of empathy towards people with disabilities as it is considered an effective tool for abled body individuals to get a firsthand experience of what it feels like to be in the shoes of their differently-abled counterparts. Simulating VI to sighted individuals is one practice that is possible through the use of VR, where a spectrum of various VIs are simulated, ranging from glaucoma[94], cataracts [68], and diabetic retinopathy [137], AMD [93] to complete blindness. By simulating the visual limitations experienced by blind individuals, VR enables sighted users to engage in embodied cognition, where they see and feel the limitations and frustrations of VIs.

According to Herrera et al., [51], VR simulations that mimic visual impairments provide users with a firsthand experience of the challenges faced by the blind, fostering a deeper understanding and emotional connection. This immersive engagement is crucial for perspective-taking [51], a cognitive process that fosters empathy by allowing individuals to understand and share the feelings of others. Krösl et al., in their cataract simulation study [68], specifically highlighted how they wanted the relatives of cataract patients to benefit from the simulation. Additionally, studies by Schutte et al. [102] and Ahn et al. [13] reveal that these immersive experiences improve empathetic responses and promote positive long-term attitudes and behaviours toward individuals with visual impairments.

2.3.3 Accessibility Tools

Beyond navigation and education that are directly applicable to the VI community, indirect applications of VR that are instrumental in enhancing daily living skills and promoting independent living constitute VR-aided accessibility channels to test out inclusive assistive technologies and create social spaces.

In the study by Zhang et al., [140], a VR system was developed to assess the architectural design of aged living spaces to assist designers in understanding design guidelines to accommodate blind spots and challenging areas often undetected in the interior design review process. Moreover, virtual replicas of houses can easily be rendered in VR, which people can interact with and learn to navigate, reducing the learning curve and increasing confidence.

Moreover, VR technology is also being leveraged to develop and refine assistive devices for the demographics, as it provides a robust preliminary basis for researchers to test and optimize the design and functionality of these devices before real-world deployment. The research by Ricci et al., [93] to test out their 'Electronic Travel Aid' (ETA) prototype with users through VI simulations is an example of such a use case. In their study, VR provided a platform to assess whether the ETA feedback system could promptly convey environmental information to novice users to engage in safe and efficient travel in a realistic and safe experimental setting.

Additionally, VR facilitates the design of inclusive social spaces and nurtures interactions and shared experiences between visually impaired individuals and their sighted peers. The study about the inclusive design of public interactive displays (PIDs) by Yao et al. [137] is an example of how VR can be used for the empathetic modelling of social spaces, PIDs in this case, without the hassle of physical prototyping to test out the concept. Aan et al. [53] created a multiplayer game in VR for players with VI to engage in a collaborative gaming environment where all participants, regardless of their visual abilities, could engage equally. These kinds of applications foster a sense of communal belonging and not only improve social engagement for visually impaired people but also challenge societal perceptions of disability by showcasing their capabilities and adaptability.

2.4 Virtual Environment Exploration: Scale and Interface

Although the terms VR and VE are often used interchangeably, there are distinct nuances as VR is considered a subset of VEs but more immersive in nature. This research uses the term 'VE' as a three-dimensional digitally created virtual environment accessed through HWDs that enables interaction within the space and with different virtual entities.

The flexibility of a VE allows to represent only the part of the world that is considered relevant for the final user: the proper choice of the information, representation and rendering included in the virtual world can strongly simplify the perception and interpretation efforts required to the users [41]. However, organizing information, designing techniques to interact with them and configuring the size of the environment determine the overall immersiveness and efficacy of the VE.

2.4.1 Scale of Virtual Environments

The classification of scales of VEs that is referred to in this research comes from the works of Kreimeier et al. [63]. According to them, the scale of a VE is interpreted as the size of the virtual content and not as the user's input space that classifies them in *Small Scale, Medium Scale, Large Scale.* Moreover, according to the hapto-acoustic interaction metaphors of Felice et al., [41], they compare the scale to levels defined by the parametric classification of technical implementation of an interaction possibility. These parameters include the exploration interface, common Hardware used for interaction, the absolute or relative positioning of the user in the 3D space and the overall virtual content scaled to fit within the confines of the tracked physical space or an inflated area with no space limitations.

Small Scale Virtual Environment

In a Small Scale VE, the user interaction is defined by their engagement with virtual objects confined within their hand reach. Thus, the interaction is considered to be 'intimate'. Moreover, this scale employs absolute positioning of the user's avatar or hand such that their spatial relation with the virtual objects is fixed. Similar to the VEs explained in the works of Roberts et al. [95] and Wau et al. [138], the content of the VE is scaled to fit the available space and uses grounded force feedback devices like *Phantom series, Data Gloves* to provide haptic feedback to the user, enhancing the tangible sensation while working in a very limited working space.

Medium Scale Virtual Environment

The medium-scale VE expands the interactive space for users to explore larger VEs by physical walking. However, the space is within a 'tracked' walkable area, often a room and the VE is adjusted to fit within the room's boundaries. Here, similar to smaller-scale VEs, the interaction still uses absolute positioning, meaning the virtual objects maintain a consistent spatial relationship relative to the user. The content of a VE of this scale implements only sections of a larger environment that are displayed in real size. Due to the degree of freedom of movement, grounded force-feedback implementations are replaced with non-grounded haptic feedback modalities through controllers.

This size of virtual environments provides a balance between the polar ends of the VE scale spectrum (small and large scale). In contrast to smaller-scale VEs, the interaction is more natural and permits the use of whole-body motion while representing virtual content on a real-life scale without the complexity and cost associated with the technical aspects of larger-scale VEs. As a result, Medium Scale VEs are frequently highlighted in the literature as a practical and popular choice, evidenced by their prevalent classification in the reviewed literature summarized in Table 2.1.

Large Scale Virtual Environment

Similar to the implementations by Ricci et al. [93] and Armougum et al. [17], Large-scale VEs liberate users from spatial constraints, offering a virtual environment that can extend beyond the physical boundaries of their immediate surroundings. The interaction of the users is managed via relative positioning as if they are avatars in it, and the movement is controlled by joysticks and game controllers that provide passive haptic information. This scale is suitable for exploring extensive virtual spaces like buildings or urban environments without physical space limitations. However, controller-guided exploration can often be less immersive, which can be replaced by devices that provide a walk-in-place approach that mimics actual walking *(covered in Section 2.4.2)*.

2.4.2 Exploration Interface

In the context of this research, the exploration of VE is considered an amalgamation of the user's ability to locomote and interact with the different virtual objects through an interface. The term 'exploration' is directly correlated with 'user tasks' as referenced from the works of Gabbard [45] and Esposito [39] that organizes them as *Navigation and Object Selection/Manipulation*. Based on this classification, the locomotion interfaces comprise user tasks for navigation, while the interaction interfaces comprise user tasks for selection, manipulation, modification or querying any virtual object. The different Locomotion and Interaction Interfaces are described as follows, and the literature using them is illustrated in Table 2.1.

Locomotion Interface

Navigation refers to how the users move within a VE; the interface or medium through which it is made possible is a locomotion interface. Felice et al. [41] explain that navigation has two main components: travel and wayfinding, where the former is the physical component and the latter is the cognitive component. Moreover, they explain these components through metaphors, which are interpreted along the terms of interfaces described below.

Physical movement-based Locomotion

A key aspect of the (mostly egocentric) exploration of VEs is the movement of the users' perspective; for example, when exploring an audio-haptic VE that is an approximation of a real place, the users certainly want to be able to 'walk' around in it [63]. Locomotion in VE is enabled through physical movements, such as walking, which uses the body's motion to drive a user's movements. It can be classified as follows:

Table 2.1: Overview of Virtual Environment Scales and Interfaces in Selected Publications

Publication	**Scale of Virtual Environment**	**Locomotion Interface**	**Interaction Interface**
Kreimeier et al. [64]	Medium	Sidemill	
Zhao et al. [141]	Medium	Physical walk	Physical cane based mappings
Kreimeier et al. [62]	Medium	Slide mill	Controller based cane mappings
Siu et al. [108]	Medium	Physical Walk	Physical cane based mappings
Ricci et al. [94]	Medium	Controller	Controller based hand mappings
Lahav et al. [69]	Medium	Controller	
Krösl et al. [67]	Medium	Controller	Controller based hand mappings
Armougum et al. [17]	Large	Controller	
Ricci et al. [93]	Large	Controller	
Fialho et al. [42]	Medium		
Roberts et al. [95]	Small	Grounded force-feedback	
Wai et al. [138]	Small	Grounded force-feedback	
Thevin et al. [120]	Medium		
Tzovaras et al. [122]	Medium		
Tatsumi et al. [119]			Data-Glove based hand mappings

- Physical Walking
- Walking in Place
- Devices Simulating Walking

These movements are based on the relative and absolute positing of users' avatars in the VE and can be either tracked through sensors within a space or through devices-aided meditation.

'Physical Walking' is one of the most direct forms of travelling in a VE, where a definite physical space has set boundaries, and the user's walking movements are tracked and mapped through sensors or tracking systems. The works of Zhao et al. [141] and Siu et al. [108] describe the use of HTC Vive trackers [54] that are mounted on top of the user's head and track the movement within a specified room layout. Although it is the most natural and immersive technique of VE locomotion as it provides vestibular cues, the spatial limitation often makes it less feasible. Thus, an alternative to this is 'Walking in Place,' where users move their feet by taking strides similar to natural walking while always remaining in the same place. It does not have spatial limitations and still allows users to drive the movement with their body, but it does not provide vestibular cues and gives a lower sense of presence in the VE [41]. Based on walking in place, devices such as omnidirectional slide mills simulate walking based on the user's foot movement while standing atop the device. Studies conducted by Kreimeier et al. [64] [62] make use of these slide mills to simulate walking. However, the downside of these devices is their expensive technical aspect and inability to provide the perception of natural walking [41].

Controller-based locomotion

Controller-based locomotion is highly preferred for explorations of large-scale VEs where the spatial requirement is highly constrained or the need for technical implementations is limited. This locomotion interface uses devices ranging from keyboards to joysticks and VR controllers, where the navigation is done through the relative positioning of the avatar in the virtual landscape. Moreover, controller-based locomotion can enable the exploration of medium and large-scale VEs.

Armougum et al. [17], in their experimental design, used VR controllers for the locomotion of users to study cognitive load during VR experiences, whereas Krösl et al. [67] used the same VR controllers for a user study that enabled navigation of users through

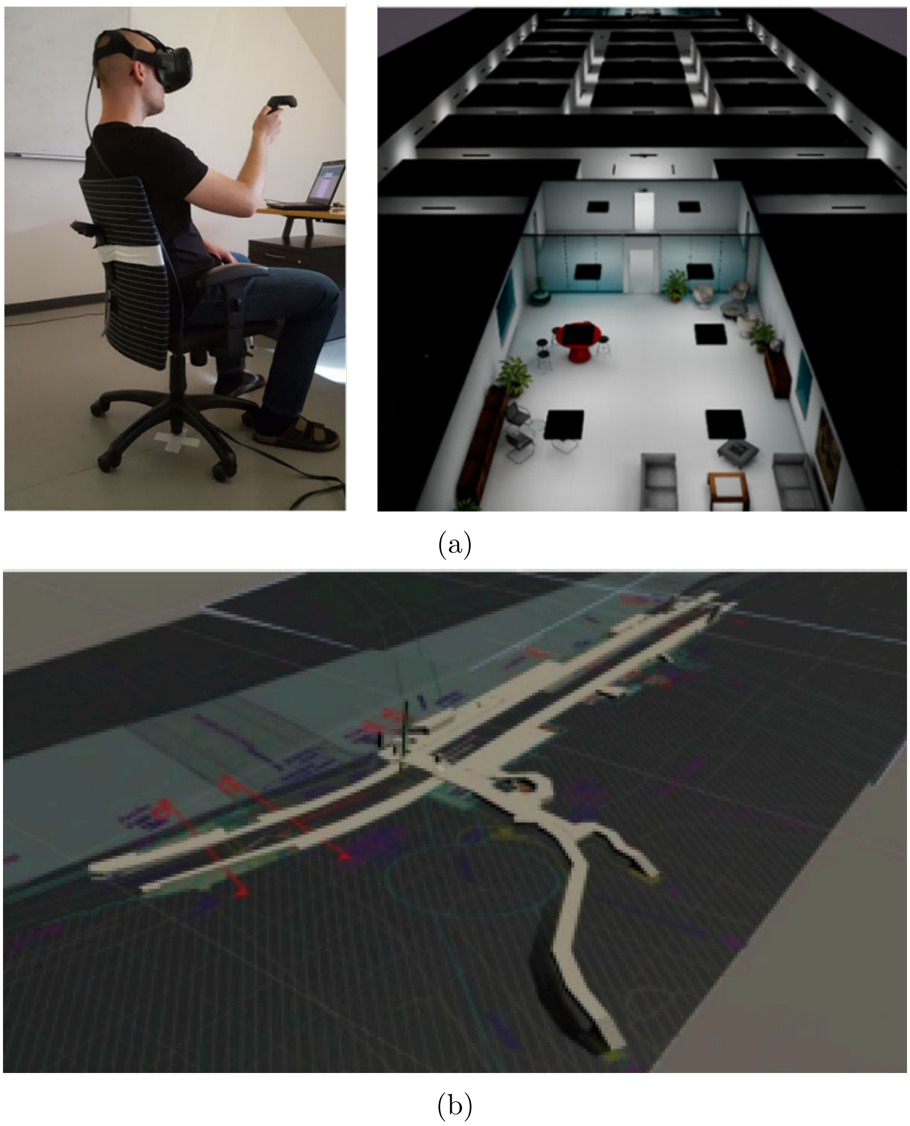


Figure 2.2: Controller-based locomotion. (a) Medium Scale VE (Building setting) [67]. (b) Large Scale VE (Metro station setting) [17]

escape-routes in a building. The former is an example of a large-scale VE, whereas the latter is an example of a medium-scale VE, as illustrated in Figure 2.2.

Interaction Interface

Interaction in VEs is acquiring knowledge of a virtual object's properties by interacting with it through translation, rotation, and scaling for a particular intent. The interaction process can be exocentric and egocentric, wherein the former, a user in a VE, interacts through a god's eye view while, in the latter, acts from inside the world and interacts with a subjective point of view [41]. The object's properties become more discernable with the increase of modality in the interaction as it adds more feedback loops: shape, geometry, material and weight with sight and haptics and acoustic properties with hearing.

Egocentric interactions allow a subjective viewpoint to the user through a first-person setting, while exocentric interaction doesn't constrain a user to a local area through a god's eye viewpoint. However, in the context of this research - in non-visual settings, relevant literature falls short in relation to exocentric interactions, and it is plausible that much work needs to be done to systematically observe the behaviour of blind or eye-busy users engaged in articulated multimodal tasks through this form of interaction [41]. Thus, the following interaction interfaces are all egocentric in nature and are feasible implementations even in non-visual settings.

Virtual Hand

One of the interactions in VEs is through tracked virtual cursors, where the cursor resembles a 3D model of a human hand. This virtual hand is tracked through a 3 DoF input device through Data/Cyber Glove or controller-based mappings as illustrated in Figure 2.3. These mappings directly map a user's hand motion to the virtual hand motion by calculating its 3D position and orientation in the VE [41]. This interface is considered highly intuitive and natural, as it allows real-life interactions similar to everyday life. However, the spatial requirement is the biggest drawback due to the limited range of the virtual hand, which means distant objects can only be interacted with after the virtual displacement of the user.

In Data/Cyber Glove mappings, the virtual hand is decomposed in a Tracked Hand, directly coupled to the user's real hand data [41]. Tatsumi et al. [119] used Data Glove-based mappings for the virtual hand in their research to assess their technique of haptic feedback for white cane usage. The feedback loop for this method is in the form of haptics, which is non-grounded/wearable in nature. Moreover, the interactions follow a 'Proxy Configuration' [41] to enhance realism where if the virtual hand intersects a virtual object, the tracked hand is constrained to not let it pass through the geometry.

In controller-based mappings, the virtual hand is coupled to the controller, which is responsible for tracking the orientation and rotation along with feedback in the form of haptics. This feedback loop for controllers can be grounded (Phantom) or non-grounded

(VR controllers) in nature. Moreover, more recent controllers that come with Meta Quest Pro and Quest 3 have enhanced mechanisms to track finger level trackings, such as grip and release, through sensors and based on the physical hand's fingers on the controller buttons. Controller-based interactions are widely used in research due to their technical simplicity and feasibility, which are evident in the works of Kreimeier et al. [62], Ricci et al. [93] and Krösl et al. [67].

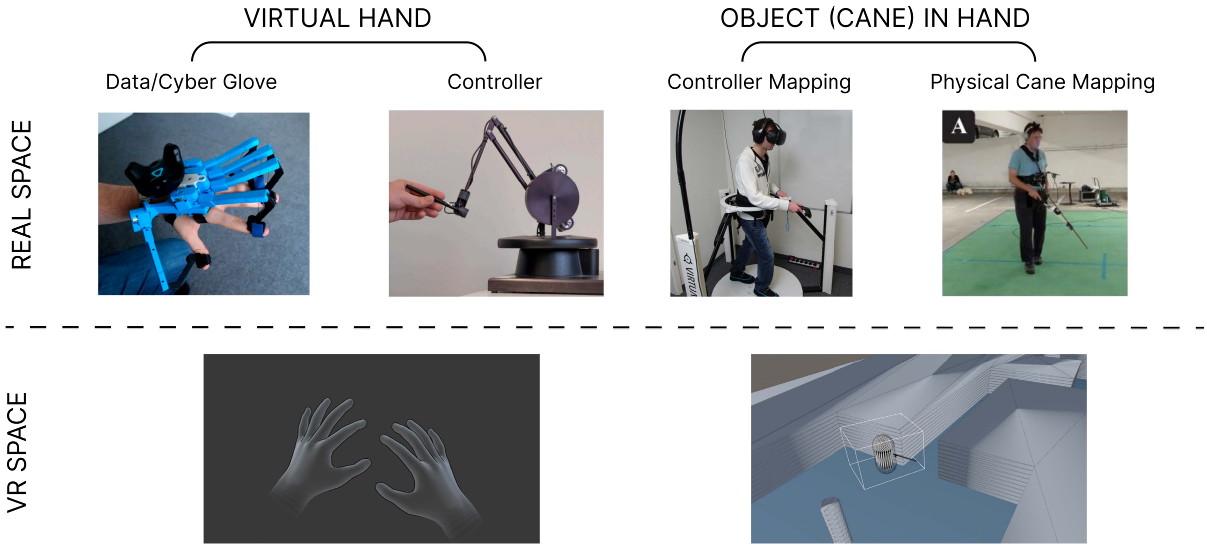


Figure 2.3: Different interaction Interfaces and their virtual counterparts. *Images reused from (in order of appearance) [41], [41], [62], [108], [124], [62].*

Object in Hand

In non-visual settings, 'Objects in hand' refers to using the 'White Cane' in VEs, where it is used as an interface for interaction involving kinesthetic and tactile feedback. 'Kinesthetic forces' allow users to interact with virtual objects to understand their shape, while 'tactile feedback' conveys information about an object's surface properties [108]. The virtual cane is tracked either through mappings based on controllers or an actual physical cane. When tracked through (VR) controllers, the orientation and position of the controller determine the virtual cane's spatial properties in the VE, whereas when tracked through a physical cane, additional sensors such as HTC Vive trackers or advanced motion capture system such as Qualysis [90] are used. Microsoft Research's Canetroller [141] is an example of physical cane-based mapping, whereas the use of VR controllers in the study conducted by Kreimeier et al. for their BlinWalkVR experiment [64] is an example of controller-based mappings. Similar to the concept of ‘Proxy Configuration’ in virtual Hand mappings discussed above, physical cane mappings can be effective in simulating virtual object collision through the use of brakes and actuators [108], which provides a highly immersive medium. On the contrary, they can be technically tedious to implement as compared to their controller-based counterparts and for usage that requires ‘quick and easy’ tracking.

2.5 Current Scenario of VR for the Visually Impaired Demographics

Accessibility in VR for non-sighted users is a usability concern that traditional VR systems have been trying to patch in order to establish a more inclusive experience. However, considerations at a later stage of the system design life cycle have proven to produce subpar fixes to these usability concerns. Thus, the design process should inherently promote key inclusions to make the VR experience more accessible for non-sighted users as it helps even sighted users during situational disabilities [81], analogous to how a rising tide raises all boats [38]. Fostering empathy for a disability through the use of VR is an effective means of communicating a perspective of another individual’s situation; however, challenges arising from misrepresentations of blindness simulations, lack of personal experience, and presence of misconceptions have led to fostering causal empathy and not long-term perspective shifts.

2.5.1 Inclusive Design

Inclusive design in VR for people with VI is inadequate in terms of two key areas of concern: the scale of a VE and the exploration interface/hardware. Current research that does make use of large-scale VE in their implementation lacks in terms of interaction interfaces that resemble a ‘standard white cane’ while research which does make use of the cane is either part of a wearable contraption or doesn’t appear to be humane in design. Moreover, current research with such cane-like interaction interfaces lacks the scale of their VE. Thus, the gap right now is to create an inclusive approach that makes use of large-scale VEs with exploration interfaces similar to real-world modalities.

Prior research in the field that simulated white cane used grounded haptic systems that did not allow users to move and explore the space physically [108]. Research by Tatsumi et al. [119] and Tzovaras et al. [123] make use of a hand-worn exoskeleton that simulated contact forces from holding a cane. These implementations lacked the use of any physical cane, and although they could simulate the textural and geometrical properties of virtual surfaces interacting through the virtual cane through haptics to a certain extent, the force feedback necessary to mimic the presence of the virtual ground/floor would be missing as there would be no mechanism to restrict hand movement in such regards. This is presented by Zhou et al. in their research of handheld sticks for virtual geometry detection as they highlight the need for design considerations around aspects such as *Continuation, Synchronization and Extended contact.* This issue was solved by Zhao et al., who used a wearable cane (Canetroller) [141] that used brakes to act as proxy surface boundaries, limiting the range of movement of the worn cane. This was further refined by Sui et al. [108] by introducing kinesthetic haptic feedback in multiple axes when holding the cane in different styles and supporting the techniques necessary to navigate more complex spaces successfully. The issue which still persists is that wearable cane implementations are not only bulky by design but lack the resemblance that visually impaired users are used to. Thus, if a hardware medium that closely resembles a standard white cane is used in explorative research of VEs, it would be less strenuous for them to adapt and use and enrich the inclusivity aspect as well.

Dolins et al., in their study [37] of spatial navigation, state that there is a trade-off between the naturalness of the (virtual) environment and experimental control in such experiments. This is evident in most prior research that employs different scales of VEs. For research that uses medium-scale VEs [67][108][141], the exploration interface is dictated through the tracked physical area of the experiment. For research that uses large-scale VEs [94][17], the exploration interface is limited to spatial navigation through only auditory cues. The problem with the former is that only a limited area of a large VR setting is available to be interacted with at any given time, and for the latter, the use of spatial audio through immersive can't provide a cognitive sense of objects' geometry or placement in the space [108] and is often temporary rather than persistent which needs to repeat if it is to remain salient, adding to the complexity [60].

2.5.2 Accessibility Conundrums

One of the key principles of Human-Computer Interaction is User-Centered Design, where usability concerns are incorporated from the start to make informed design decisions throughout the design process. Unfortunately, in practice, fixes for usability concerns perceived as affecting small populations (such as people with disabilities) are often tackled at the end of the software design lifecycle, resulting in sub-par user experiences [81]. Accessibility has thus far not been a primary consideration during the development of mainstream VR systems. Mammot et al. introduce the concept of situational disability [81] where a normal user can become disabled at certain instances (e.g. When the hands are occupied holding groceries and unable to do any other task) thus emphasizing the concept of considering accessibility as a core part of a VR system's design process. Although contemporary commercial VR systems seem to have a deficit in incorporating accessible mediums for a large diaspora of disabled users, in this literature review, the accessibility concerns are bluntly aimed at the visually impaired population.

Visual Dependency

VR experiences typically rely on visual stimuli, making them inherently challenging for individuals with VI. The immersive nature of contemporary VR applications relies on visual graphics, spatial awareness, and visual cues, which can create significant barriers for those who are blind or have low vision. Lack of visual information limits their ability to perceive and interact with the virtual environment. VR environments can be disorienting, especially without visual cues for navigation and orientation. Visually impaired individuals may struggle to understand the layout, identify objects, or navigate within the virtual space. Developing effective auditory or haptic navigation aids and spatial awareness techniques is essential to provide a sense of direction and orientation [133].

Lack of Non-visual Feedback

VR typically lacks alternative modalities or non-visual feedback to compensate for the absence of visual cues. While some VR applications may include audio cues, haptic feedback, or speech synthesis, their implementation is often limited, inconsistent, or insufficiently utilized. Providing meaningful and comprehensive non-visual feedback becomes crucial to conveying information and enabling participation for visually impaired users [57][64]. One primary reason is that, unlike visual cues, feedback from audio and haptics is impermanent and intermittent, due to which it needs to be repetitive to remain prominent[60].

Key Areas of Accessibility Concerns in VR

Existing literature points out overlapping themes of concern for accessibility in VR for visually impaired users that revolve around standards of content, interaction modality, and device usability. The papers used in this review that highlight these areas are congruently grouped up in the following Table 3.

Table 2.2: Literature Sharing Congruent Areas of Accessibility Concerns in VR

Publication	**Content Accessibility**	**Interaction Accessibility**	**Device Accessibility**
Mamott et al. [81]	✓	✓	✓
Jain et al. [56]	✓		
Philips et al. [12]	✓	✓	✓
Elmqvist et al. [38]	✓	✓	
Williams et al. [133]			✓
Zhao et al. [141]	✓		
Kreimeier et al. [64]		✓	

Content Accessibility

Smith and Nayar [109] point out the issue of content accessibility by stating that "accessibility alone is not enough". They compare blind-accessible video games to a wheelchair ramp-enabled library that is filled only with digests of the original books and argue about the term equivalent accessibility — allowing people with disabilities to experience the world as others do, "with a similar sense of control (intention) and efficiency as sighted players can" [109].

Interaction Accessibility

VR has the potential to create a "level playing field," providing spaces in which all users may be equal in their capabilities. Due to this very reason, it is necessary that VR systems and their representations account for end-user diversity within their design to be comfortable for and usable by a large audience. Thus, for the interaction in VR to be inclusive, two main areas of focus comprise accessible inputs and user interfaces [81]. The input in VR systems is typically through hand-held and one-size-fits-all controllers that assume users have an articulate range of motion. Unlike interaction with stationary or mobile devices, this implicit dictation of users' input actions offers limited to no support for direct input apart from 3D controller motions in mid-air. Thus, there is an opportunity to design VR interfaces to support direct 3D input as well as input through alternative modalities, such as voice and gaze [81].

Device Accessibility

VR hardware typically makes many assumptions about users' abilities, which can lead to accessibility problems [81] more specifically when the challenges are due to motor skills [12]. The physical interfaces of VR systems, such as controllers and head-mounted displays, may not be designed with accessibility in mind and often rely on visual feedback or complex button configurations, making them challenging to use for individuals with VI. Incremental changes in the design of this hardware that aim to reduce the number of individual components and introduce more flexible and ergonomic types of equipment prioritizing non-visual interaction methods are necessary to ensure equal access to VR experiences [133][81].

2.5.3 VR and Empathetic Literacy: Hits and Miss

VR is often revered as an effective means of promoting empathy as the system enables the simulation of an actual experience for people to better understand the perils and perspectives of another individual's situation. For simulations that focus on users experiencing a disability, the primary agenda is to foster some kind of empathetic concern and desire to help accommodate people belonging to the group. However, current literature suggests that such kind of disability simulation comes short in one way or another, misleading realities of a disability which can contribute to paternalistic discrimination [106]. Moreover, empathy that is fostered through these simulations is found to arouse compassionate feelings but not necessarily encourage users to imagine other peoples' perspectives [76].

In terms of disability simulation concerning vision, no ample research has been conducted that highlights the significant implications of post-simulation empathy. However, the present literature on studies concerning VR blindness simulation presents the following challenges.

Lack of Personal Experience

VR experiences aimed at fostering empathy for blind individuals require sighted users to understand and embody the perspective of blindness. However, sighted individuals naturally lack personal experience and understanding of the daily challenges faced by blind people. This knowledge gap can make it difficult to develop accurate and meaningful VR simulations that effectively convey the actual blind experience. Studies [76][108] have shown that following the simulations, empathetic concern (warmth) toward disabled people increased with certain boundary conditions to this effect; however, attitudes about interacting did not improve. Moreover, in some conditions, VR was no more effective at increasing empathy than less technologically advanced empathy interventions such as reading about others and imagining their experiences [76].

Misrepresenting Visual Impairment

Creating a realistic representation of VI in VR poses a challenge. Blindness is not merely the absence of vision; it encompasses a range of VI with different degrees and types of vision loss. Designing visual effects or simulations that accurately depict various types of blindness is a complex task that requires careful consideration and can also mislead people about blindness because it highlights the initial trauma of becoming blind rather than the realities of being blind [106].

Stereotypes and Misconceptions

When designing VR experiences to foster empathy, it is crucial to avoid reinforcing stereotypes or perpetuating misconceptions about blindness [133]. Misrepresenting blind individuals as helpless or dependent can undermine the goal of promoting empathy and understanding. Redmond et al. [108] found that with a number of participants in disability simulations, most of the participants felt more confused, embarrassed, helpless, and more vulnerable to becoming disabled themselves compared to baseline, which in itself contradicts the idea of empathy generation.

2.6 Hapto-Acoustic Design Space for Visually Impaired Users in VR

The visual sensory bandwidth exceeds the bandwidth of haptics and hearing and therefore enables faster and more interactive types of locomotion [64]; however, related work in VR for users with VI and/or low vision has mostly examined navigation techniques in indoor spaces with less focus on video spatial renderings and more on spatialized audio and texture mimicry through vibrotactile feedback mechanisms. Niklas Elmqvist in his article [38] terms this approach as sensory substitution which in the case of non-sighted users implies visualization only through the two practical options - touch or sound. Moreover, the article talks about shifting the idea of visual representation to spatial representations through the creation of a soundscape where the user exhibits navigational expertise through a mental map using echolocation and kinesthetic feedback from a cane tip and ambient environmental sounds in a simulated environment [38].

Microsoft Research's Canetroller [141] is an example that demonstrated how rendering virtual objects haptically, including simulating materials' properties and textures, could enable users who were completely blind to successfully navigate and understand virtual scenes when paired with a novel haptic controller that mimicked the interaction of a white cane [81]. A successor to this study [108] employed a multiple degrees-of-freedom cane enabling users to adapt the controller to their preferred techniques and grip along with a three-axis brake mechanism to emulate large shaped virtual objects. Additionally, as three-dimensional rendering, the main channel of information used in current virtual reality, is unavailable to people who are blind, new haptic rendering technologies capable of detecting and displaying surface roughness in Telepresence [88] can also be explored as a means to convey virtual 3D objects to people who cannot see [81].

Jain et al. [56] in their research on sound accessibility in VR for Deaf and Hard of Hearing (DHH) users, suggest a morphological search-based design space of input devices and information visualization. The design space includes both syntactic and semantic dimensions for sounds, visuals, and haptics. However, since the review concentrates on visually impaired users, vocabulary encompassing sound and haptics has been explained as follows.

2.6.1 Hapto-Acoustic Vocabulary

Sound Vocabulary

The taxonomy of sound vocabulary covers two dimensions: source and intent. As illustrated in Table 2.3, the “source” is the origin of the sound (e.g., from a character or an object), which is further categorized depending upon the sound emanating from an actual object/character in the VR world or from those playing in the background [56]. The “intent,” on the other hand, is the impact of a sound on the users’ experience that increases realism and/or conveys critical information.

Table 2.3: Jain et al. [56] Sound Source Categories and Descriptions

Sound Source	**Category**	**Description**
Speech	Localized	Spatially positioned speech of a character
	Non-localized	Ambient speech of a narrator
Objects	Inanimate	Sounds from non-living objects
	Animate	Non-speech sounds from living beings
Interaction	-	Sounds from the interaction between multiple objects
Ambient	Point	Spatialized ambient sounds that belong to the VR world
	Surrounding	Non-spatialized ambient sounds
System	Notification	Sounds of critical alerts of specific events
	Music	Background music

Haptics Vocabulary

As per Jain et al. [56], the taxonomy of the haptics vocabulary is articulated as having three feedback dimensions depending upon the delivery, location, and qualitative elements. The first dimension is the haptic form factor, which describes the device on which the haptic feedback is delivered. It encompasses a range of delivery mechanisms, from traditional VR controllers to non-conventional and emerging haptics-based commodity devices. The second dimension is the location of the body on which the haptic feedback is delivered, called the haptic feedback location, which includes prominent localized areas of palms, arms, torso, head-mounted, and legs. Finally, the haptics elements constitute three primary quality factors of a realistic haptic experience, namely (i) Intensity (amplitude/strength), (ii) Timbre (sharpness/pitch), and (iii) Rhythm (beats/interval between feedbacks).

Table 2.4: Jain et al. [56] Categories of Sound Intents

Sound Intent for	**Description**
Conveying critical information	All sounds that are critical for progression in an app
Increasing realism	Ambient or object sounds that increase immersion
Rhythm or movement	Sounds that correlate to or enhance a particular user/object/actional movement
Generating an affective state	Emotional sounds with varying level intonations
Aesthetics or decoration	Non-critical sounds that increase the beauty
Non-critical interaction	Interaction sounds that are not critical to game progression

2.6.2 Related Work on Haptic Rendering

Haptic manipulation in non-visual settings aims at understanding objects rather than moving them around, changing their position and orientation. Applications are mainly based on one-point interaction devices that are more stable and effective if compared to whole hand armatures that are still not very effective in contexts where haptic is the only feedback channel [41].

The implementation of grounded and non-grounded haptic feedback is primarily through commercial voice-coil actuators, the size of which depends on the contraption used. Besides actuators, in-built force feedback mechanisms in modern-day VR controllers are another technology that implements haptic feedback. Moreover, the medium for haptic feedback through different modalities in the form of wearable, handheld and foot-based devices has also been discussed. It has been found that in VR, haptic feedback is primarily used to convey object geometry or mimic textures.

Haptic Technologies and Modalities

Wee et al., through their systematic literature review [130], classify haptic devices used in VR into five categories, namely *Handhelds, Wearables, Encountered types, Physical props and Mid-air.* These classifications constitute the interaction modality of the devices with the body; however, for the scope of VR literature that is primarily focused on visually impaired users, encountered type and mid-air haptic devices were found to be non-relevant due to their novel feedback mechanism that is spatially inaccessible to users with VI. Moreover, it was found that some physical prop-based haptic devices in this regard could be accommodated under handheld haptic devices, and few had specific use cases that were primarily feet-based. Thus, in this section, the haptic devices in VR for visually impaired users can be classified into three primary categories as follows.

Wearable Haptic Devices

Wearable haptic interfaces include attachments that can typically be worn on the user's fingers, wrists, hands, or anywhere on the body [130]. These are generally designed to be compact and comfortable in order not to impede the movement of users [86] and provide unobtrusive and precise tactile feedback.

Spagnoletti et al. [111] introduced a wearable cutaneous device that could provide the sensation of making/breaking contact with virtual objects (texture such as brick felt and aluminum) by providing variable pressure and texture stimuli through a moving platform and a vibrotactile motor. Tatsumi et al. [119] used force feedback to enhance interaction

Table 2.5: Literature on haptics rendering in VR using different modalities in non-visual conditions

Publication	**Modality**	**Application**	**Hardware**
Turchet et al. [121]	Feet-based	Textural Simulation *(solid and granular textures)*	Prop-based (sandals)
Giordano et al. [48]	Feet-based	Textural Simulation *(wood, ceramic, marble, gravel - ∅ 16-20 mm)*	Prop-based (shoes)
Zhou et al. [142]	Hand-held	Object Geometry *(3D shape tapping and contour tracing)*	Prop-based (Stick)
Culbertson et al. [34]	Hand-held	Textural Simulation *(Vinyl, Denim, Wood, Plastic, Paper, Cardboard)*	Prop-based (Tablet Stylus)
Siu et al. [108]	Wearable	Textural Simulation *(Concrete, Metal, Tile, Carpet)* and Object Geometry	Custom navigation cane controller
Spagnoletti et al. [111]	Wearable	Textural Simulation *(felt, aluminium, brick)*	Cutaneous device
Tatsumi et al. [119]	Wearable	Textural Simulation *(Two types of studded pavement blocks)* and Object Geometry	Data Glove
Tzovaras et al. [123]	Wearable	Object Geometry *(Recognition of shape and proportion understanding of virtual objects)*	Data Glove

by enabling users to recognize virtual space geometry. Their study focused on creating a haptic VR system for cane walk simulations, where feedback forces are provided through a data glove worn on the hand. Similarly, the use of data gloves in the study by Tzovaras et al. [123] as the primary source of haptic feedback for the identification of virtual objects is another example of wearable haptic devices.

Handheld Haptic Devices

Handheld haptic interfaces are novel devices that are typically held and generally resemble a controller and may also include add-ons that enhance the haptic feedback of default controllers [130]. With a simple form factor that embraces the 'grab and go' design philosophy and affordable manufacturing cost, handheld controllers are the default haptic interface accompanying commercial virtual reality hardware [130] apart from VR controllers.

Culbertson et al. [34] developed a system that captures and recreates tool-mediated texture contact vibrations. The system uses a sensorized handheld tool to record three-dimensional tool acceleration, tool position, and contact force over time. This data is then used to render realistic virtual textures on a tablet using a stylus augmented with small voice coil actuators. In another study [108], researchers from Microsoft created a physical cane-mediated haptic device. The cane was held in hand, similar to an actual white cane, but was equipped with various braking mechanisms and voice coil actuators that provided the sensation of different virtual textures along with proxy collisions to convey the sense of different object shapes and sizes laid out on the virtual ground.

Foot-Based Haptic Devices

Similar to certain handheld haptic devices that use physical objects like a cane to deliver haptic feedback, physical props to deliver sensations are not uncommon and are known as 'passive haptics' [130]. With the use of props, complex hardware is not needed, which enables the delivery of realistic haptic feedback at a low cost [105]. Building upon the 'prop-based passive haptics,' certain devices that provide feedback through the feet are classified as foot-based haptic devices, analogous to their handheld counterparts.

These devices provide tactile feedback through the feet and are primarily used to simulate different ground textures and surfaces, as discussed in section 2.6.2. Turchet et al. [121] developed a system that simulates the auditory and haptic sensations of walking on various surfaces using sandals equipped with pressure sensors and actuators. The sensors detect interaction forces during walking and control several physically based synthesis algorithms, which drive both the auditory and haptic feedback. A similar experiment by Giordano et al. [48] used four sets of vibrotactile actuators throughout the sole of shoes that mimicked virtual textures, which had to be identified by users in blind-folded conditions to assess the kinesthetic ability of participants in non-visual settings.

Application of Haptic Feedback in VR

Unlike in visual conditions, where haptic feedback acts as an add-on to make auditory or visual stimuli more realistic, in non-visual settings, the applications of the haptic channel are dominant and act as one of the primary sensory mediums for kinesthetic realism. An overview of related literature depicts that on a high level, the usage scenario of haptics in VR in situations devoid of visuals can be categorized into two primary applications: to convey object geometry and render virtual textures.

Conveying object geometry

One of the primary applications of haptic feedback in VR is the accurate representation of object geometry and surface properties of virtual objects. These geometrical properties are assessed through parameters such as weight distribution, stiffness, deformability, etc., and are conveyed through haptic devices to simulate normal and tangential force feedback.

'Dynamic redirection' is a concept that has been used to increase the versatility of a handheld prop, i.e., by modifying the prop's position and orientation in the virtual world to provide more convincing haptic feedback [116]. Yang et al. explain that a VR grabbing tool that provides haptic feedback for precise virtual object manipulation increased its grabbing range through dynamic redirection [136]. Zhou and Popescu, in their experiment [142], proposed a method based on the same redirection principle that synchronizes real and virtual contacts to provide convincing haptic feedback for various virtual objects using a physical handheld stick mapped to its virtual counterpart. This method supports both tapping and extended contact, enabling users to trace the contours of virtual objects and recognize their complex shapes. Tatsumi et al., in their experiment, implemented the holding touch sensation that transferred to the real hand while the cane held by the hand touched a virtual block of studded pavement [119]. Siu et al. tested out their body-worn cane prototype [108] with participants navigating in an indoor virtual space where they were tasked with detecting and avoiding virtual obstacles. These obstacles were portrayed through large-scale haptic forces through the 3-axis brake mechanism in their prototype that helped in the shape discrimination.

Textural Rendering

Rendering virtual surfaces is an umbrella term for creating haptic-based representations of real-life surfaces through tactile mimicry of physical properties, such as compliance (soft/hard), texture (smooth/rough) dimension, and temperature conveyed through applied contact forces (friction, tension, spring etc) [130]. Rendering realistic textures in VR is crucial for creating an immersive experience, and the current literature, as illustrated in Table 2.5, suggests two types of virtual texture rendering: hand-mediated and foot-mediated.

Hand-mediated virtual textures are conveyed through hand-wearable haptic devices [111] [119] that simulate textures found mostly in indoor spaces, such as denim, plastic, tile, vinyl, paper etc. On the contrary, foot-mediated virtual textures comprise textures from outdoor spaces such as concrete, gravel, marble, bricks, etc., conveyed through prop-based passive haptic mechanisms [121] [48]. Moreover, the rendering of impact vibration and texture frequencies of virtual textures is prominently done through contact acceleration data recorded in the real world of different sample surfaces, which is then used to create models to render realistic textures in VR. This method is evident in the works of Siu et al.[108] that describe impact forces as better predictors of perceptual hardness compared to stiffness. Similarly, Culbertson et al. [34] refined methods for creating realistic haptic virtual textures from tool-mediated contact acceleration data from a previous study of a similar nature by Romano and Kuchenbecker [97] where they developed a recording system that captured motion, force, and acceleration data during interactions with real textured surfaces and then used the data to create real-time frequency-domain texture models.

Challenges to Haptic Feedback in VR

The following challenges are from the works of Wee et al. and Wang et al., who, through their literature review, highlight the gaps in existing research and pinpoint the impediments of haptic devices used in VR.

Adaptability

Wearable haptic systems need size-adaptable wearables because actuators in those systems need close contact with the user's body to convey sensations effectively. This is because the design should also account for skin stretch and deformations that may occur during use [9]. Moreover, the challenge of haptic rendering for wearable haptics is to meet the contradictive requirements of realism and efficiency. For realism, haptic rendering needs to ensure the simulation of the human hand manipulating diverse objects with diverse physical properties while also considering the efficiency prospect of the hardware to handle the requirement of over 20 DoFs for movement that a human hand possesses [129]. On the other hand, handheld haptic interfaces are built to cater to the public without extensive training; however, they must be flexible enough to adapt to various hand sizes while being able to adapt and predict users' unwarranted movements and still perform well without any loss of functionality or damage [130]. Furthermore, a significant research gap is found in double-hand manipulation in hand-held haptic devices from a design perspective.

Compactability

Wee et al. highlight the need to develop multimodal actuators to replace employing different actuators that render distinct sensations [130]. They emphasize the necessity for compact wearable devices to incorporate actuators and sensors that fit into the body's confined spaces. This requirement is particularly critical for gloves, where the limited surface area of each finger poses a challenge. Moreover, for hand-held haptic devices, the compatibility challenge is achieving more abundant haptic feedback patterns within the compact volume of a handheld device, such as localized and diverse spatial-temporal vibrotactile patterns and texture feedback. [129].

Weight Factor

While the need is for the haptic devices to be adaptable and compact, the weight of the devices is an important factor to be considered while developing novel force-feedback mechanisms. Since wearable are worn in the body and hand-holds are held by the hand, the issue to be mitigated is that of user comfort without exposure to fatigue when the device is used for a prolonged period of time. Lack of comfort exacerbated through fatigue in usage is detrimental to a realistic and immersive virtual experience [130]. Moreover, devices that are prop-based or resemble a physical object, such as a standard white cane, need to be of the correct dimension and have proper weight distribution. This is evident in the research [141][108] where haptic devices are built around standard props like the white cane that focus on the center of mass not to be far from the gripping area as it can lead to user fatigue over time [96].

2.6.3 Related Work on Spatial Audio

Spatial Audio

Spatial audio is based on the simulation of sound distribution phenomena that affect the sound emitted from its source. As a result of the interaction of the sound and the environment, the sound received by the human auditory system contains spatial information on the position of the sound source and the surrounding environment [82]. The length and timbre of the reverberation of a room give users information on the approximate size and nature of the environment and also serve as the *memoria loci* – a personalized reminder of the place or event [82]. Spatial audio generated through sound engines can be implemented through ray-tracing or acoustic wave-based simulation. The former is known as 'Ray-based Spatial Audio' and the latter as 'Wave-based Spatial Audio', which is described in detail below.

Ray-based Spatial Audio

Ray-based spatial audio, also known as ray tracing, models the propagation of sound waves similarly to how light rays are traced in computer graphics. This method simulates the paths that sound rays take as they reflect, refract, and diffract around environmental obstacles. Each sound ray is traced from the source to the listener, accounting for interactions with surfaces and objects. This technique is computationally efficient and can handle complex environments with multiple reflective surfaces. Ray-based methods are particularly effective for simulating early reflections, which is crucial for perceiving spatial audio cues as these reflections help users determine the direction and distance of sound sources in a computationally feasible way to achieve real-time performance for dynamic scenes, enhancing the realism of the VR experience [91].

Wave-based Spatial Audio

Wave-based spatial audio, in contrast, models the propagation of sound waves as they interact with the environment based on the physical principles of wave acoustics. This method involves solving the wave equation to simulate how sound waves propagate, diffract, and interfere with each other in a given space. Wave-based methods can capture complex acoustic phenomena such as diffraction around obstacles and detailed sound scattering. Although wave-based methods provide a more detailed and physically accurate acoustic simulation, they are also computationally intensive, and implementation techniques often require significant processing power compared to ray-based techniques. Thus, they are used in scenarios where high fidelity of sound reproduction is essential [99]. Moreover, Siu et al., in their research [108], compare the two types of spatialized audio implementation techniques through sound engines and justify using wave-based spatial audio in the VE due to its computational capability for physics effects such as occlusion, reverberation, and portaling - important for spatial navigation in non-visual conditions.

Acoustic Virtual Environments

‘Audio-games’ or ‘Audio-Only games’ is a concept in game development driven by “complete auditory interfaces” such that it can be played without any graphics [44]. Friberg and Gärdenfors propose an approach to auditory interface design that incorporates three listening modes: casual listening, semantic listening, and reduced listening. This framework helps include meta-level information to achieve a high level of complexity and provide elements of open-endedness to sound objects [44].

‘Soundscape VR,’ an acoustic VE implementation by Fialho et al. [42], talks about spatial navigation and wayfinding. Their experiment demonstrated that participants could effectively form and use spatial representations of the environment, improving their navigation performance over repeated trials. The results indicate the ability of spatial audio to create immersive and ecologically valid VR environments that mimic real-life situations, aiding in the development of cognitive maps and enhancing spatial learning.

Echolocation has recently been proposed as a navigational tool in acoustic VEs [15]. Over the past seventy years, research has shown that echolocation –using self-emitted noises such as mouth clicks and the ambient sounds generated by a person’s cane or shoes– is an effective tool for human navigation [100]. The process of echolocation is aided by sound engines being able to simulate physics effects such as occlusion, diffraction, reverberation and portaling [108]. Moreover, the design implication pertains to directional listening of sound waves' reflections off of virtual surfaces, objects, etc. This skill appears to be a highly relevant mechanism used by blind individuals to understand the configuration of an environment or parts of it [85].

Challenges in Approaches to Spatial Audio

One of the challenges to spatial audio is its inability to independently convey the shape of the virtual geometry or the material properties; more importantly, users lack feedback from their proprioceptive senses for navigation [108]. On the contrary, providing too much audio information can overwhelm players, complicating gameplay. Thus, it is crucial to understand the balance between providing enough sound to support spatial cognition of the surroundings, especially for users with VI, and overcomplicating the process required to understand the game state through excessive audible clutter [65].

In the case of audio games, using arbitrary and inconsistent artificial sounds - 'earcons' [131] (e.g. beeps, musical tones, or repetitive sound patterns) exacerbates confusion and complexity in VE without visuals [16]. Gaver et al. [46] describe that audio mappings with greater articulatory directness [55] may be easier to learn than those which are purely symbolic. In other words, the more a sound is like the sounds experienced in a comparable situation in the real world, the easier it is to learn and remember [16]. Moreover, audio games have been criticized for failing to offer the full range of gameplay experiences; the navigation is often constrained, where the player's character is limited to moving on a coordinate-based grid-like system [47].

Chapter 3

Methodology

3.1 Phase 1: Small Scale Virtual Environment

The plan for the first phase of the implementation was to primarily explore the gaps in the current literature and use the gained perspective to give direction to the research - it was about ideating potential application domains and observing what sticks. Moreover, it helped create a low-fidelity prototype as a ‘proof of concept’ to help learn and create the basics of different facets such as interaction, navigation, audio and haptics that would later be scaled or improved.

3.1.1 Brainstorming

At the very start of my degree, it was clear that I wanted my thesis project to be in the domain of VR and focused on the visually impaired demographics. However, at that time, what aspect or need of the demographics could be fulfilled by using VR as a technology was unknown. The brainstorming process helped in ideating the specific application domain of the project after perusing different contemporary VR applications for the target population. The applied domain included VR, which is used in the fields of gaming and entertainment, assistive technology, skills and training, health care, sports and rehabilitation, architectural design, etc. Among these, due to my previous involvement and interest in assistive technology, it was decided that the project would loosely fall under this domain, focusing on making the experience interactive and game-like using VR.

The literature review helped in understanding the current work in the field and exposed potential gaps that could route the research. Specifically, the works by Zhao et al. [141] on the ‘Canetroller’ were a very big inspiration to my project, along with the research by Kreimeier et al. [64] that used an omnidirectional slide mill for indoor navigation in VR. These, along with other publications, helped identify issues faced by visually impaired people while navigating unknown indoor and outdoor environments. During

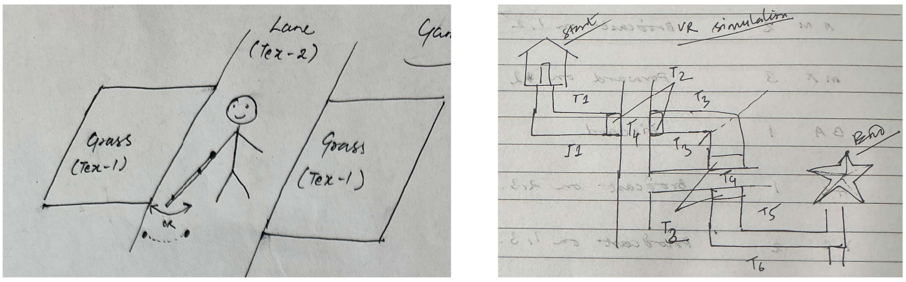


Figure 3.1: Brainstroming sketches during the first phase

the brainstorming process, a list of possible pathways emerged in developing the project that focused on either training visually impaired users to identify low-hanging obstacles or objects that couldn't be easily detected using the cane (slippery surface) in unfamiliar spaces [58] or simulating textures found in outdoor environments with a focus on a large scale explorative techniques using VR. The latter was found to be more appealing and thus was taken forward.

The sketches during this process, as shown in Figure 3.1, contributed to creating the basics of what the gameplay in the simulation would look like. Through the sketches, it was realized that the project as a whole would revolve around the central theme of textural differentiation in a simulation using a virtual cane Figure 3.1 (left) and would support some form of navigation to explore the virtual environment to reach from point 'A' to point 'B' Figure 3.1 (right).

3.1.2 Unity driven City Block

The long-term goal was to create a large-scale city simulation in VR using the Unity GE; however, the plan was first to create a minuscule version of the city that would resemble a city block but have most of the necessary elements. This would provide a platform to test the preliminary implementation of the rendering, mapping, and virtual cane to interact with the different elements in the simulation. Moreover, it assisted in creating a boilerplate framework for different aspects of the game design, such as the traffic system, spawning and travel of vehicles, sound design and navigation that would eventually be expanded or updated to a more scalable version going forth.

As shown in Figure 3.2, the city block made using Unity is made up of two parks surrounded by a strip of pavement/sidewalk on either side of a two-lane road. There is a road crossing between the parks, and in the scene, a few vehicles are spawned to get

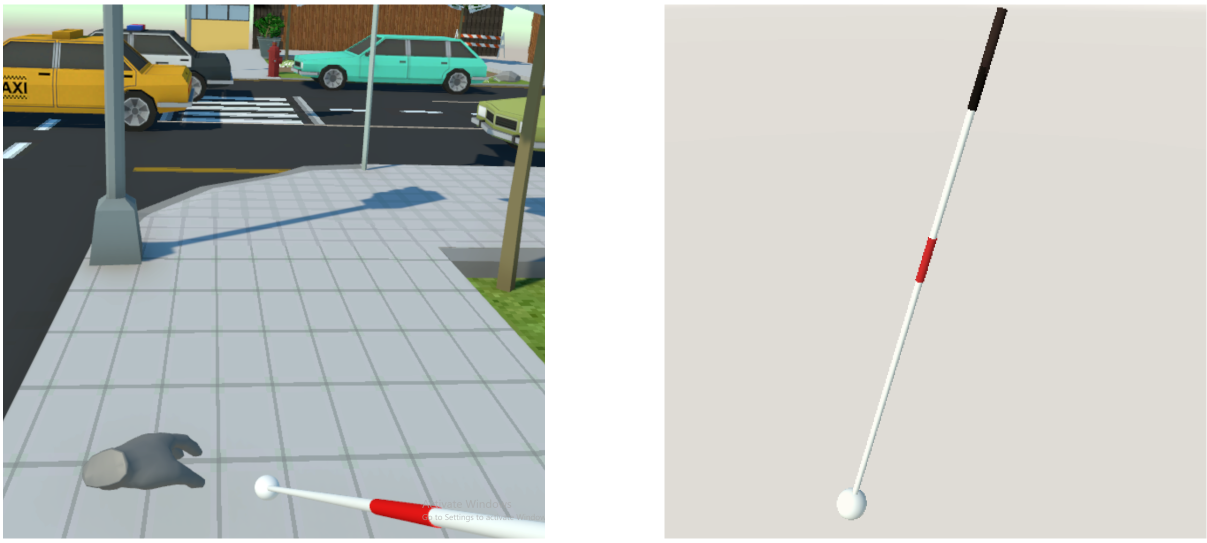
Figure 3.2: Small scale VE model created in Unity for the first phase of implementation

across the road. However, the vehicles are controlled by a simple traffic system, which also signals when it's time to cross the road. The virtual surfaces involve all the primary textures, namely sidewalk, grass, gravel and road, that would, later on, be used in the larger cityscape with the exception of metallic studded pavement blocks. The whole map was created using the Polygon City Pack asset [117] that provided 3D elements such as the floors, city props, buildings, natural elements like trees, etc, along with vehicles such as cars, buses and taxis.

3.1.3 Controller enabled Virtual Cane and Navigation

The virtual cane needed to be mapped to something in the real world, and for exploration in the VE, the user needed to be able to navigate. The implementation had to be quick to implement and easy to test. Thus, the readily available VR controller was used to map the virtual cane and enable navigation.

As outlined in Section 2.4.1, a small-scale VE constitutes exploration interfaces (interaction and locomotion) that are generally controller-based. Hence, as shown in Figure 3.3b, a 3D version of a standard white cane was fabricated in Unity, which was used as a game object (an essential element used to represent any object in a game scene) and mapped to the right-hand controller. The navigation was a mix of toggle-based and gaze-based, where the toggle controlled the movement (left/right, forward/backward) using the joystick button on the left-hand controller, while the gaze (direction of the user's



(a) (b)

Figure 3.3: (a) First-person view in the Small-scale VE while holding the virtual cane. (b) 3D model of a standard white cane used in Unity

head) specified the direction of the movement. The coupling of the cane and the navigation through the controllers were enabled through the built-in scripts with the 'XR Interaction Toolkit' package from Unity [126].

3.1.4 Audio and Haptic Synthesis

During this phase, there was no complex set of auditory feedback; the audio feedback was limited to the interaction sound of the cane with the surfaces and the sound from the traffic light (wait/walk pedestrian crossing sound). All the interaction sounds rendered during the virtual cane interaction were downloaded from the internet, free to use without any credits, and were original sounds recorded by swiping a real cane with different physical surfaces. The same was the case with the pedestrian crossing sound. Each virtual surface had a collider attached to it, and each collider had Unity's built-in audio playback component called 'Audio Source' that played back surface-specific audio clips. Moreover, the pitch, volume, and the start and end of the audio playback from this component were controlled through an external script attached to the virtual cane Game Object (GO). Whenever the tip of the virtual cane collided with any surface collider, for the duration of the collision, the audio clip specific to that collider was played with its volume and intensity adjusted according to the speed of the movement of the cane.

As the VR controller was mapped to the virtual cane, the haptic feedback was also provided through it. The 'XR Interactor' script that comes with the XR Interaction Toolkit was used for this. This script has a built-in function called haptic events that triggers vibrations in the VR controller; however, an external script attached to the virtual cane managed the duration of these haptic events, utilizing the same collision logic described for auditory feedback.

3.2 Phase 2: Large Scale Virtual Environment

Figure 3.4: High-level system architecture: A broad overview showing key components and their interactions

In the second phase of the implementation, the small-scale VE was used as a basis to expand to a large-scale model of the planned city simulation. As before, Unity was used as the GE to create the large-scale cityscape and acted as the central platform for connecting and communicating with various other systems. Moreover, as illustrated in Figure 3.4, significant changes were made regarding the navigation modality and interaction interfaces, along with using a dedicated sound engine and haptic engine for audio and haptic synthesis, respectively. However, while implementing individual systems for tracking, navigation, haptics and audio, the approach underwent an iterative approach built on inclusive design approaches and the accessibility needs of the target demographics.

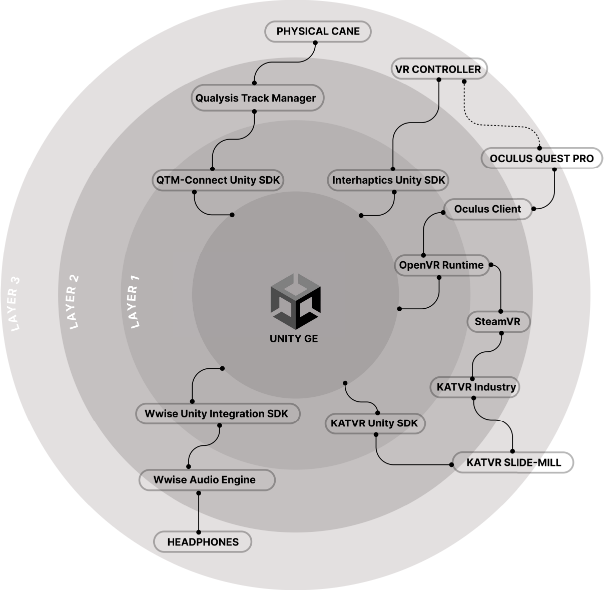


Figure 3.5: Layered architecture of the system showcasing different components integration with the Unity GE

As illustrated in Figure 3.5, the system architecture follows a modular integration framework, with Unity as the central facilitator. Unity acts as a foundational environment where all the development takes place and serves as the central point of coordination for various SDKs and external components, manages performance among these components and handles the real-time rendering of the VR environment.

The architecture can be divided into three layers, and each layer resembles a different level of abstraction and functional dependency with Unity.

Layer 1: Core Integration Layer

This layer represents the foundational components that directly interact with the Unity GE. It includes essential SDKs and run-time frameworks integral for connecting to middleware software and provides implementation utilities and APIs for developmental purposes.

Layer 2: Middleware Layer

This layer is an intermediary that acts as a communication channel between the core integration components, various external applications, and user-facing hardware. This layer mainly comprises applications that share the same system and provide the necessary compatibility and performance management.

Layer 3: External Hardware Layer

This outermost layer includes all external hardware interfaces that users interact with and encompasses the physical devices that act as a modality for conveying VR experiences.

3.2.1 Brainstorming and Story-Boarding

Since a basic framework for the implementation was already in place, the second iteration of brainstorming mostly revolved around engendering ideas about what components from the first phase needed to be scaled, updated or replaced. The main components during the brainstorming process were how to expand the city map - to what size and incorporate what elements, how to map the virtual cane to a physical cane, what types of audio to include and how to make it spatial, how to fit the slide mill in the picture for navigation, through what implementation to render haptics on the physical cane and what tasks could be created to test out the system. However, the process only yielded an outline of all these components to a broad extent as some of the approaches, later on, were completely different from those considered during the brainstorming due to various time-related and technical constraints.

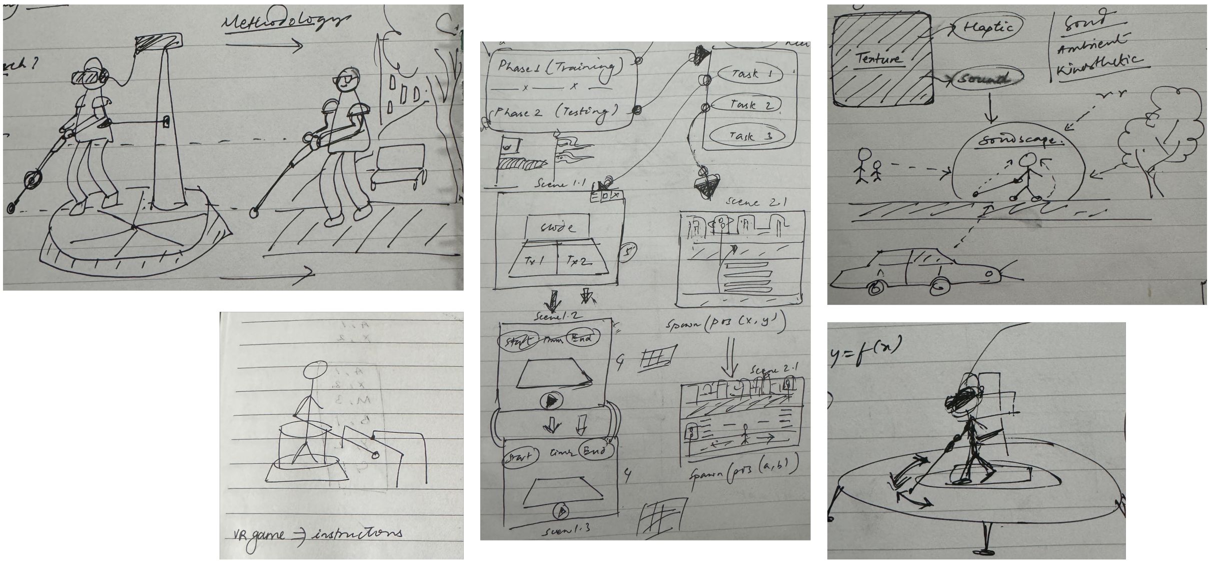


Figure 3.6: Brainstorming and story-boarding sketches during the second phase

This was followed up by creating storyboards for the different tasks that were brainstormed. Specifically, the storyboarding involved creating narrative-based scenes focusing on visualizing the interface and placement of the GOs that would be created in Unity. This approach not only helped determine the flow of the scenes but also acted as a modular representation of the entire project, which ultimately helped in estimating the timeline for the individual components.

3.2.2 Unity driven VR City Simulation

This section talks about the different elements that make up the entire city simulation, which is not just limited to the placement of 3D cityscape renderings but rather a mix of various spatial elements that work in tandem to bring liveliness and add a layer of complexity and realism to the VE. Although the following components will not be visible to a user with VI in the literal sense, they are important aspects during the game design process and act as reference points and containers for auditory and haptic cues that the users will actually be using and interacting with.

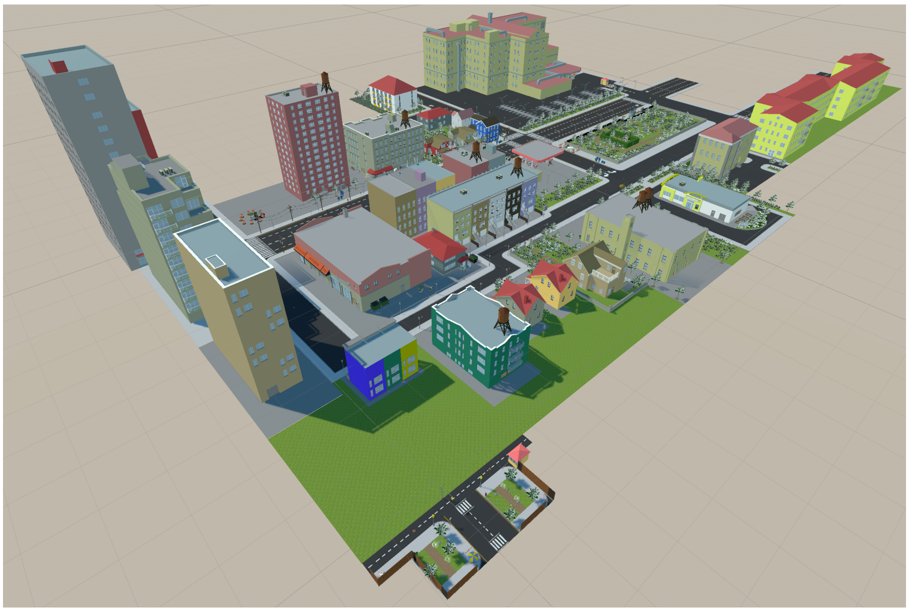


Figure 3.7: Large Scale VE implemented as a cityscape

Rationale of the City Simulation

Implementation of Large Scale VE

Building on the discussion about inclusive design covered in Section 2.5.1, current implementations of 'cane-like' interaction interfaces [108][141] are limited to VEs regulated by the confines of the tracked or available physical space. Moreover, in the research that does employ large-scale VEs [94] [17] to simulate urban settings, the navigation along with interaction is mostly restricted to the use of hand-held controllers. Hence, in the former, although there is physical walking for navigation, there is limited content (Section 2.5.2) as only a portion of a bigger environment is available at any given instance, while in the latter, the prospect of a complex VE is attenuated by the absence of the modality that facilitates real-world-like explorations (Section 2.5.2).

A trade-off exists between the naturalness of the VE and experimental control in such experiments [37], which is something that needs to be balanced for and addressed by systems that focus on rehabilitative assistance and skill training, such as O&M. Thus necessitating the need for VEs that simulate an outdoor environment with complex urban settings to encompass larger exploration areas that can be explored through mediums that support natural navigation techniques, and the components of which can be interacted with 'cane-like' interaction interfaces.

Supporting Spatial Navigation

Spatial navigation refers to the ability to move through and orient oneself from point A to point B in an environment that involves forming, storing, and using the cognitive representation of the space [14] and spatial relationships between objects, locations, and oneself within the environment. Franz and Mallot [43] explain that spatial navigation can be of two types: 'local navigation' and 'wayfinding,' which depends on the location of the navigational goal in or beyond the perceived environment. Moreover, both types are explained through allocentric reference frames (representation of objects relative to the environment) and egocentric reference frames (representation of objects relative to the individual). Therefore, local navigation drives from immediate responses to environmental contingencies, relying mostly on egocentric reference frames as acquired spatial information is related to the observer's body [42]. On the other hand, wayfinding involves both allocentric and egocentric reference frames involving more demanding cognitive operations, such as decision-making about local or remote environments, being also supported in spatial memory representations [101].

The studies conducted by Fialho et al. [42] and Ricci et al. [94] employ spatial navigation for visually impaired users through a mix of egocentric and allocentric methodologies. However, in either case, the navigation in the environment is supported only through specific audio beacons that dictate the travel of users. As discussed in Section 2.6.1, any VR simulation constitutes a 'sound vocabulary' with various source-specific ambient and interaction sounds. Additionally, when users with VI are involved, spatial navigation needs to be driven not only through spatial audio but also through some form of tactile feedback. However, in these studies involving urban simulation, the 'goal to reach' encapsulates an auditory cue that needs to be followed, but other forms of environment sounds and haptic modalities that help form the hapto-acoustic complexity of the VE are missing.

Components

The city simulation comprises a suite of different 3D renderings of real-life elements that make up a cityscape governed by scripts that control their appearance, movements and modality of conveying information. These settings feature architectural elements like apartment buildings, suburban homes, and parks, complemented by detailed streets, roads, and pavements populated by vehicles and pedestrians that add to the dynamicism of the environment. Additionally, the infrastructure includes functional traffic lights and road signs and is enhanced by natural elements such as trees, bushes, and flowers.

Urban Cityscape Elements

'Cityscape' refers to the visual appearance of a city or an urban area, essentially an urban equivalent of 'landscape.' In order to simulate a decent-sized city, most of the readily available elements in a real-world city were taken inspiration from, including but not limited to apartment buildings, sub-urban houses, parks, gas stations, hospitals, malls, etc. Figure 3.8 illustrates some of these architectural examples in the form of 3D renderings.

These renderings are part of the same city pack asset, which was strategically placed

Hospital Building

Children's park

Builidngs and sub-urban houses

Gas Station

Figure 3.8: Different types of cityscape elements in the VE

and combined with other essential urban infrastructure and natural elements. The design incorporated pavements, roads, and various green spaces such as lawns and grass banks to create a realistic outdoor environment. This method enhances the visual appeal and realism of the virtual city and creates a platform to simulate real-world urban dynamics and interactions.

Vehicles: Spawn and Travel

The 3D models of the vehicles came with the city pack asset from the Unity Asset Store, and in total, four types of 3D vehicle models were used, namely police cars, taxis, personal sedans (three colours) and buses (two colours) as shown in Figure 3.9 (right).

A vehicle controller system was created in Unity, which comprised spawner GOs, a set of waypoint GOs and 3D vehicle models (Figure 3.9 (right)). In total, six spawner objects were placed on various points on the city map, and each spawner object had a specific set of waypoints and a few random vehicle model options. The script attached to the spawner object controlled the random instantiation of a vehicle model, its direction, and a randomized interval between each instantiation. On average, each spawner had

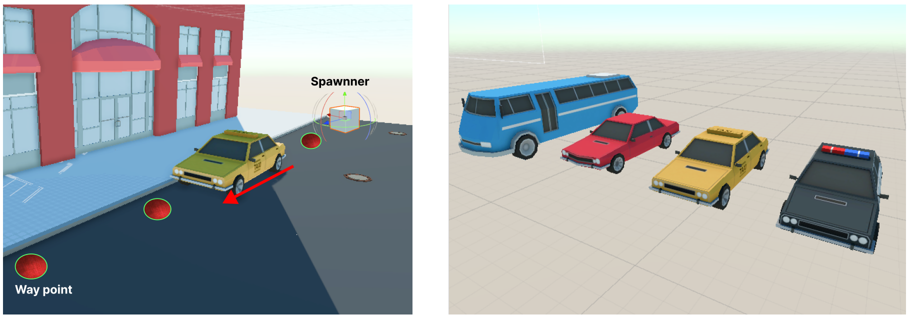


Figure 3.9: Vehicle Control System and the different 3D models of the vehicles used

at least five waypoints, and its main objective was to create pathways by acting as markers for a spawned vehicle to follow. Once spawned, each vehicle had a script that controlled its speed and directed it to follow the direction of the nearest waypoint specific to its parent spawner until the model was eventually destroyed when it reached the last waypoint. However, along the path, when going through traffic controller systems, the vehicles stopped for a fixed period of time, the process of which is detailed in Section 3.2.2. Another script attached to the vehicle model controlled the type of sound to be played according to the model of the vehicle, which had two variations depending on whether it was in motion or stationary.

Traffic Light System

Each crossing on the city map had a dedicated traffic light system to control the traffic flow, comprising a pair of pedestrian crossing lights, a vehicle traffic light, and a set of register/de-register blocks (Figure 3.10). All three components were part of a parent GO that controlled how each component communicated with each other to provide specific cues for pedestrians to cross the road and the approaching vehicles to halt or move.

The parent script controlled the wait/walk sound emitted through the pedestrian crossing traffic lights on either side of the road and the lights on the main vehicle traffic lights. Every ten seconds, the sequence alternated: the pedestrian traffic lights signalled that it was safe to cross, and approaching traffic was stopped, followed by a period where vehicles were allowed to pass, and pedestrians were signalled not to cross. Initially, the issue with timing these events was that the approaching vehicles sometimes stopped at the crossing, making it look unnatural. To resolve this, a set of registering and de-registring blocks were employed that were placed a few meters away from the crossing

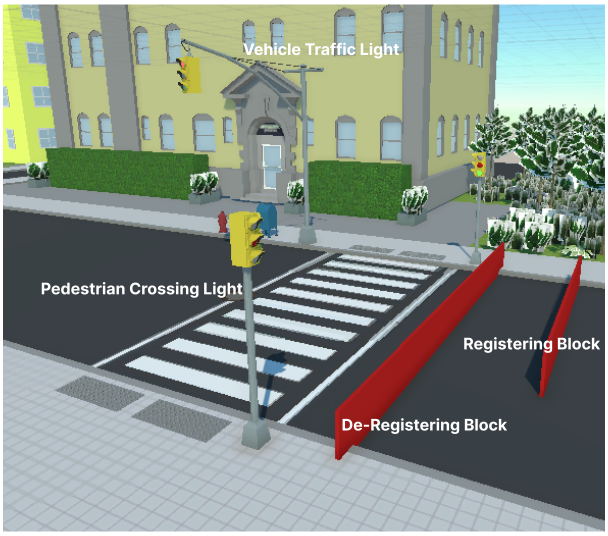


Figure 3.10: Components of the Traffic Controller System deployed at each crossing in the map

towards the direction of the incoming traffic. The idea was to use the 'registering block' to register the vehicle that collided with it by setting a boolean flag to true and use the 'de-registering block' to de-register the vehicle that collided with it by setting the boolean flag to false. Thus, at any given time, if the traffic control system signalled the vehicles to stop, those which were de-registered would be considered to pass through to prevent them from stopping near the crossing area, and any registered vehicle and the vehicles that followed would halt automatically.

Non-Playable Characters

To add more depth to the simulation and make it more similar to a real simulation of a city, Non-Playable Character (NPC)s, also referred to as pedestrians, were spawned at various random locations on the map. All the NPC characters and the animation of the human-like movements were generated using Adobe's online platform - Mixamo [6]. The NPCs were either spawned individually or in clusters and engaged in some form of activity that primarily served the purpose of auditory markers or points of reference for task-specific events.

From the assortment of various characters available on the Mixamo platform, a set of NPC models that resembled people who would be commonly encountered while navigating through a sub-urban cityscape was selected. These included characters such as

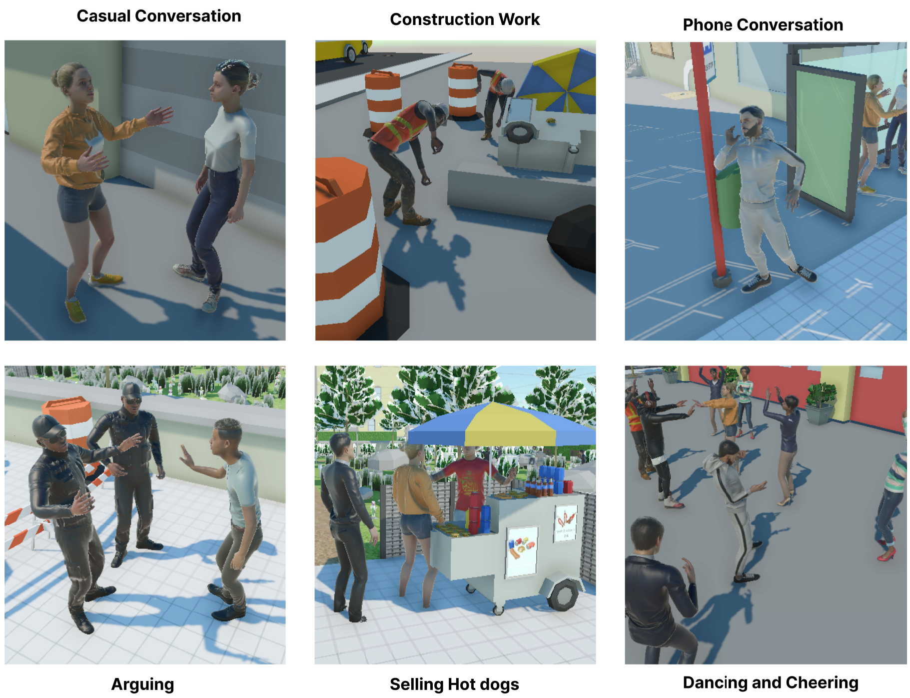


Figure 3.11: Types of NPCs throughout the city with different animated movements

police officers, construction workers, and people dressed casually, in formals or active wear. Moreover, each NPC treated as a GO was attached to an ‘Animator’ component containing a controller that dictated how their body would be animated. This animation controlled their body movements, posture, and hand gestures that created an act of human-like engagement. The NPCs were thus placed in groups or as individuals at various spots of the city, which created a ‘narrative’ around them (Figure 3.11), such as people conversing, arguing, working, talking over the phone, cheering to a group of people dancing, etc. Based on this semblance of narratives, each cluster (not each NPC) had specific sounds that not only added to the richness but also helped covey the visual narrative through audio.

Colliders

Colliders in Unity are used with the physics engine and attached to GOs that define the physical area that the object occupies, detect when other objects enter, stay, or exit the area, and trigger events when objects interact. Among the various types of colliders in Unity, ‘Box Colliders,’ ‘Sphere Colliders,’ and ‘Capsule Colliders’ were predominantly used with different GOs to primarily serve the purpose of triggering an event, initiating a process or logging essential game data. All of the colliders were customized to detect collision between only a specific set of objects through the use of a script, which made the interactions computationally efficient. However, the renderings of all the colliders were disabled as for the scope of the project, their role was purely functional rather than aesthetic. Based on the types of GOs the colliders were attached to, the types of colliders used in the simulation can be classified as follows.

- *User and Cane Colliders*

  User Collider is a sphere collider that encompasses the boundary of the user's avatar, whereas the Cane Collider specifies the contact area of the tip of the virtual cane. Both colliders interact with the Surface Colliders to trigger footstep sound and haptic and audio feedback of the cane, respectively. Moreover, the User Collider also interacts with the Event Colliders to elicit certain events or to log certain game statistics.

- *Surface Colliders*

  The box colliders attached to cover the area of a surface (road, gravel, grass, etc) in the map are referred to as the 'Surface Colliders.' These are the most widely used and one of the most important sets of colliders in the city that interact with the body of the user and the tip of the virtual cane to control surface-specific haptic and auditory feedback. The sounds it emits upon interactions are footsteps and cane audio.

- *Traffic System Colliders*

  As discussed in Section 3.2.2, these are standalone box colliders - not attached to any GO and used to specify an area or periphery. The main purpose of these colliders is to interact with the vehicle GOs and switch between a boolean flag that dictates their movements.

- *Event Colliders*

  These are box colliders typically used for triggering or logging an event when interacted with and, similar to the traffic colliders, are standalone in usage. These are specifically used to detect collisions with the User Collider to trigger events such as notifying the completion of a level and logging the time of specific interactions, such as the time taken to complete a certain task or a set of tasks.

Scenes and User Interface

‘Scenes’ in Unity act as a container that holds a portion of the game-play, often referred to as ’levels,’ and help manage the overall flow of the game and provide structure. A Scene is made of multiple GOs, assets, cameras, and scripts, and multiple scenes are navigated programmatically (automatically) or using menu layouts (manually). For this project, the scenes were controlled with the help of a menu layout that acted as the User Interface (UI). However, when we say ‘UI,’ it is not something directly interactable by the user using the system but by an operator that remotely interacts with it to provide a sense of pseudo-control to the actual user.

As illustrated in Figure 3.12, in the Unity project, there are altogether eight scenes that follow a hierarchical structure with multiple parents and sub-scenes that can be replayed back and forth through the provided interface. The initial scene, or the parent scene, is the ‘Study Start Scene’ that bifurcates to the training and testing scene, each with a set of sub-scenes that occur one after the other.

In order to render the UI elements, the ‘Canvas’ component was used as the foundational layer that acts as a vessel and determines how and when they are rendered. Moreover, the UI is trivial and very generic in terms of usage as it mainly comprises buttons (clickable and toggle) to navigate between scenes, text fields to enter user details, and placeholder texts to convey details about the scene and user information.

3.2.3 Audio Synthesis

Wwise Sound Engine

Contrary to the first phase of the implementation, which used Unity’s built-in ‘Audio Source’ component, for the second phase, all the game audio was rendered using a third-party audio engine from AudioKinetic called ‘Wwise’ [25]. The main reason for this switch was the need of a dedicated sound engine that could support complex audio synthesis and 3D spatial audio capabilities. The Wwise audio engine provides a real-time audio channel that connects to the Unity GE through the Wwise Unity Integration SDK [28] made up of plug-in libraries and API bindings in C#. Specifically, for rendering the cane audio and spatial audio described below, two plugins were used, respectively, the granular synthesis plugin called ‘Sound Seed Grain’ [23], and the spatial audio enabler

Figure 3.12: Different Scenes and UI elements

plugin called ‘AudioKinetic Reflect’ [27].

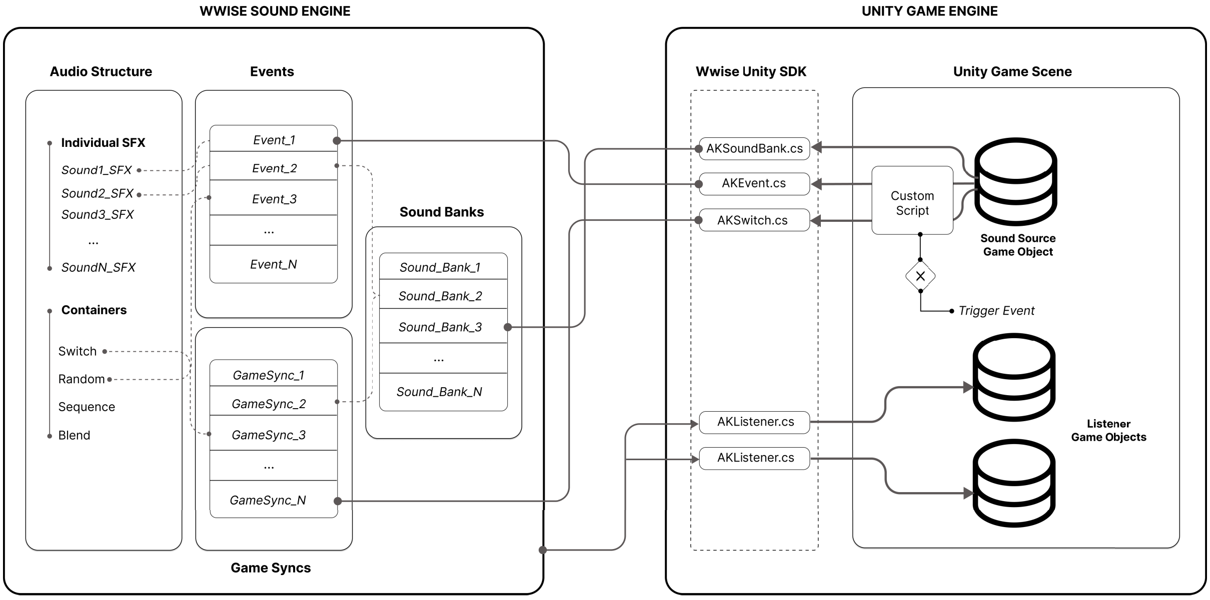


Figure 3.13: Association of different Wwise components with Unity

Primarily, Wwise builds and integrates audio into the Unity GE through the use of the following components:

- Audio Structure
- Sound Banks
- Events
- Game Syncs
- Emitter Objects
- Listeners

Among these components, the Audio Structure and Sound Banks are on the Wwise side, whereas the Emitter Objects and Listeners are on the Unity Side. Events and Game Syncs are configured on the Wwise side but are triggered by emitter objects through the Wwise SDK. Figure 3.13 illustrates the association of these various components. The following descriptions of the various Wwise elements are adapted from the available Wwise documentation on ‘wwise 2012.1: Fundamentals’ [26].

Audio Structure

The audio structure in Wwise comprises ‘Audio Objects’ and a hierarchal collection of these objects called 'Containers.' The various voice and Sound Effects (SFX) in a game are represented in Wwise by special objects called audio/sound objects. An audio object contains an encapsulation layer called an 'audio source,' which is used to apply conversion settings and multi-language development to the original audio file it is linked to.

When audio objects are grouped together to create a hierarchical project structure, it creates containers mainly used to play a group of objects according to a certain behaviour, such as random, sequence, switch, etc. This enables audio properties and behaviours to be applied at different levels in the hierarchy, which gives control and flexibility to build realistic and immersive game experiences.

Sound Banks

A Sound Bank is essentially a file that encapsulates Event and Game Sync data that are dynamically loaded and unloaded at different points in the gameplay. These are created in the Wwise Engine and integrated into Unity using the SDK helper script called 'AKBank.' Basically, these banks map particular events or game syncs to an audio object in the audio structure, which helps utilize platform memory and enhance performance.

Events

Events in Wwise drive audio in the game that apply actions to the different audio objects or containers. The specified actions determine if the objects play, pause or stop an audio. To accommodate as many situations as possible, there are two different types of Events: *Action Events* These events use one or more actions, such as play, stop, pause and so on, to drive the sound, music, and motion in the game.

Dialgoue Events These Events use a decision tree with arguments, basically a set of rules or conditions that dynamically determines what object is played.

After Events are created in Wwise, they can be packaged into SoundBanks, which are then loaded into the Unity GE so that they are "triggered" at the appropriate time in the game. On the Unity side, the Wwise Unity SDK has a helper script named 'AkEvent' that can be attached as a component to a GO to specify what event to call in Wwise. An alternative approach is using a custom script that specifies the events that can be triggered through the game logic.

Game Syncs

Game Syncs help define and control the dynamic aspects of a game's audio experience and can streamline the game development process, balancing quality, memory usage, and time constraints. They enable a more organized approach to managing audio elements, economize memory usage, and significantly enhance the immersive experience for players. There are five primary types of Game Syncs:

- **States** - Changes that affect the properties of existing sounds, music, or motion on a global scale.
- **Switches** - Represent alternatives for specific game elements that may require new sounds, music, or motion.
- **Real-Time Parameter Controls (RTPCs))** - Properties mapped to variable game parameter values, altering the properties themselves as the game parameters change.
- **Triggers** - Respond to spontaneous occurrences in the game, launching a stinger, a brief musical phrase that overlays the current music.
- **Arguments** - Collections of similar argument values grouped to form a category or outcome within the game, dynamically determining the dialogue based on present values and their order.

Emitter Objects

Emitter Objects, also called sound or audio objects, are any GOs or elements in Unity that can emit a sound, including characters, props, animate and inanimate ambient objects, etc. Source objects are essential aspects in Wwise architecture as a variety of information corresponding to the emitter object is used to determine how each sound will be played back in the game. In order to store relevant information, however, the emitter objects need to be 'registered' and, when no longer needed, need to be ''de-registered'. These processes can be explicitly called through a script attached to the object or through the SDK's helper script named 'AkGameObject.' The following types of information may be associated with the GO:

- Property offset values associated with the GO, including volume and pitch.
- 3D position and orientation.
- Game syncs information, including states, switches, and RTPCs.

- Environmental effects.
- Obstruction and Occlusion.

Listeners

A listener represents the output channel constituting a position and orientation in Unity's 3D space. A GO (the user avatar or the main character) is explicitly made a listener by attaching the 'AkListner' component with the Wwise SDK. In a single-player game, there needs to be at least one listener, while in a multiplayer setting, there can be multiple listeners. In any scenario, the listener GO and the source GO work in tandem as during gameplay, the coordinates of the listener are compared with the audio object's position so that 3D sounds associated with GOs can be assigned to the appropriate speakers to mimic a real 3D environment.

Sound Collection and Processing

The different types of audio used in the VE are explained in Section 3.2.3; however, this section explains how the different sounds that make up each audio type were procured and processed for use with the Wwise Sound Engine.

- *Localized Speech*

  Each NPC group animated to converse with one another had a context to talk about that narrated a scene to work as an audio cue in the VE - some talked about construction, some about bus stops and parks, and so on. These conversations formed the basis of the localized speech sound in the VE. ChatGPT 3.5 [84] was used for each scenario to generate context-specific scripts comprising dialogues between two people. These scripts in .txt format were uploaded to an online text-to-speech converter platform called 'Speechify' [112] that generated AI voices for different characters, which could be customized for different genders and emotions. Each conversion yielded a Waveform Audio File (WAV) file for a specific context, which was then loaded into Wwise for further use.

- *Inanimate Object Sounds*

  Sounds for different inanimate objects, such as vehicles and traffic lights, were downloaded from a royalty-free SFX platform called 'Artlist' [19]. The sounds of the vehicles varied depending on the type (police car, taxi, buses, etc.) and the state of the vehicles (mobile or stationary). The sound of the traffic lights refers to the pedestrian crossing wait and walk sounds. All the SFX files downloaded from Artlist were in the WAV format, which didn't require any processing post-download, and were directly used with Wwise.

- *Ambient and System Sounds*
  The ambient sounds comprised various environmental sounds, such as public noise, background traffic noise (honks and brakes), children playing, street vendors, etc., whereas the system sounds only consisted of a level completion prompt. Similar to the object sounds, the audio files were downloaded from an online SFX platform called 'Motion Array' [18]. All the sounds used were from the free-to-use section of the platform, which doesn't require any paid royalty. Moreover, all the downloaded files were in the wAV format and ready to use with Wwise without any post-processing.

- *Interaction Sounds - Cane sounds and Footsteps*
  The interaction sounds are made up of the sound emitted from the cane interaction and the footsteps on different surfaces. The footsteps SFX were used from the 'Footsteps - Essentials' asset from Nox_Sound [83] available on the Unity Asset Store, whereas the cane sounds were digitally recorded using a contact microphone.

  In order to collect cane interaction audio recordings, the data collection cane, described in Section 3.21, was used to swipe different surfaces such as gravel, asphalt, grass, pavements, and metallic studs in a double-swiping motion over the span of ˜2 seconds. A digital microphone (Model: Zoom H5 Digital Multitrack Portable Audio Recorder) shown in Figure 3.22b was mounted on a tripod and placed 0.5 meters from the contact point of the cane and the surface to record the sound. The microphone was configured to store audio in 24-bit/96kHz WAV Format. The experimenter manually swiped each surface six times (Figure 3.22a) while maintaining a constant swipe speed and force, resulting in 12 swipes worth of audio data set. The double-swiping motion was employed as it is a standard technique commonly used by visually impaired people during O&M training [127].

  Once the interaction audio of the white cane with different surfaces was captured, each surface had a set of four files, each with a couple of bi-directional sounds. To make sense of the recorded audio, Audacity, an audio editing tool [21], was used to divide the larger audio files into smaller, more manageable segments. Each segment focused on a single cane swipe, preserving the unique auditory characteristics of each interaction between the cane and a surface. The goal of segmenting the audio in such detail was to set the stage for the next phase - audio synthesis. By having these well-defined audio snippets, an infinite number of interaction sounds can be produced for a single surface leveraged through granular synthesis, discussed in detail in Section 3.2.3.

Spatial Audio Characteristics

Apart from the cane audio, a number of other sounds, as described in Section 2.6.1, were used as part of the acoustic vocabulary in the VE. These sounds were assigned to either a GO or an asset in Unity that worked as the 'emitter objects' (described in Section 3.2.3) to emit audio to the listener. Moreover, each audio was defined with spatial properties using the sound engine, which offered variability in terms of the sound's direction, volume, pitch, etc., relative to the listener's position and orientation in the space. Table 3.1 summarizes the different audio types, associated sounds, emitter objects and defined spatial characteristics for each sound.

Distance Attenuaton

Distance attenuation in Wwise simulates natural sound roll-off as an audio source moves away from the listener or becomes obscured by obstacles. This process is controlled through a 'Driver-Property' pair that uses attenuation curves and an emitter cone to adjust the intensity of audio signals.

The 'Diver-Property' pair adjusts the intensity of audio signals based on 'Driver' parameters like distance, obstruction, occlusion, diffraction, and transmission and maps these to audio properties such as volume, pitch, spread, low-pass, highpass filter etc. Thus, the attenuation curves for each of the pair can be modified to manipulate how sound properties change with a change in the distance between the emitter and listener. For example, as shown in Figure 3.14, attenuation curves help in programming how quickly sound volume increases with a decrease in distance between the car and the listener from

Table 3.1: Summary of Audio Types, their Sounds, associated Emitter Objects and defined Spatial Characteristics

Audio Type	**Sounds**	**Emitter Object**	**Spatial Characteristics**
Localized Speech	Individual/Group conversation, Street Vendors, Cheering	NPCs	Distance Attenuation, 3D Spatialization, Diffraction and Occlusion
Inanimate Object	Vehicle and Traffic Light Sounds	Vehicle models, Pedestrian Traffic Light GO	Distance Attenuation, 3D Spatialization, Diffraction and Occlusion, Dynamic Reflections
Interaction	Cane Sweep and Footsteps	Surface Colliders	Distance Attenuation, 3D Spatialization, RTPCs (Refer Section 3.2.3)
Ambient	Distant Traffic Sound, Public Noise, Children Playing, Dance groups	Localized GO in different parts of the city map	Distance Attenuation, 3D Spatialization, Diffraction and Occlusion
System	Level Completion Sound	Task Colliders	NA

-20 dB to 20dB when it is in front and again drops as the car starts becoming distant.

Moreover, while the attenuation curve focuses on the positional aspect, ‘Cone attenuation’ adjusts audio signal intensity based on the emitter’s orientation relative to the listener. This enables the spread of the sound vertically as per height and horizontally through customizable inner and outer cone angles as depicted in Figure 3.15. Here, the listener experiences the highest intensity of sound when in the inner cone, decreased intensity when in the outer cone angle and minimal when out of the outer cone region. This feature helps simulate the directness of sound sources more realistically, inherently enhancing the spatial audio experience and making it more natural.

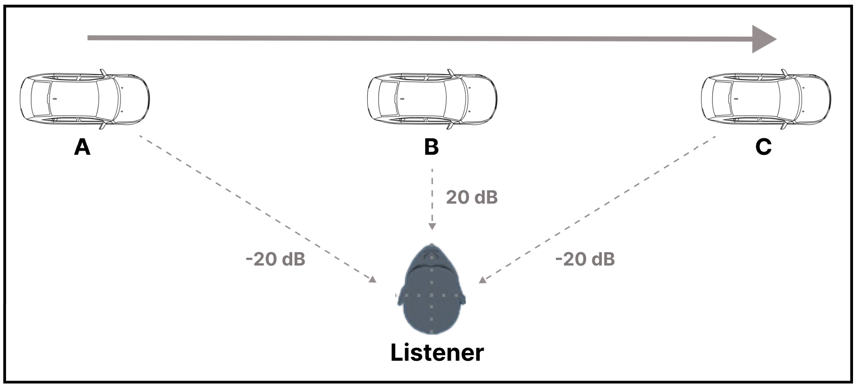


Figure 3.14: Attenuation of sound based on the distance between the emitter and the listener

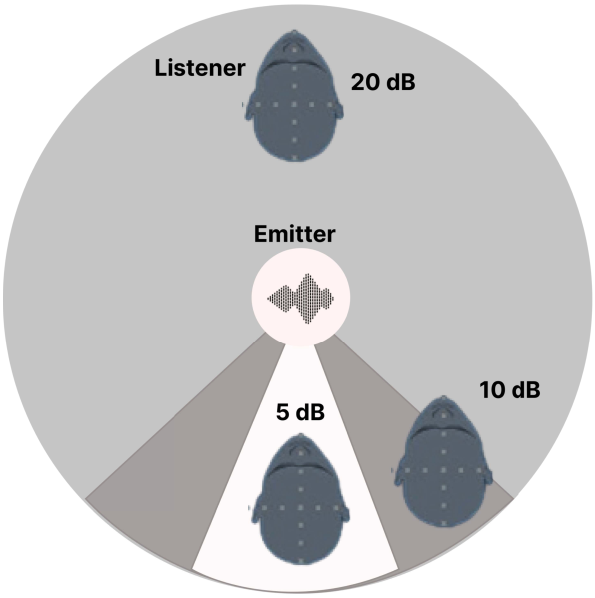


Figure 3.15: Cone Attenuation Feature in Wwise to model directness of sound based on Listener's orientation

3D Spatialization

3D spatialization in Wwise is a listener-relative audio routing technique used for positioning audio objects within a VE to simulate the physical behaviour of sound in a 3D space by defining the spatial attributes of sound emitters and listeners. Key to this system is the ability to utilize either position alone or both position and orientation data to define how audio is projected and heard within the environment. This means that as the positions of emitters (like a character or a moving vehicle) and listeners (typically the player) change, the audio properties, such as volume and pan, are dynamically adjusted to reflect these spatial changes, enhancing the immersion and realism of the game's auditory experience. Figure 3.16 illustrates how the humming sound of a butterfly (emitter) is perceived by the listener relative to its position and how the position and orientation of both the emitter-listener pair create an effect of the sound being perceived binaurally from left to right.

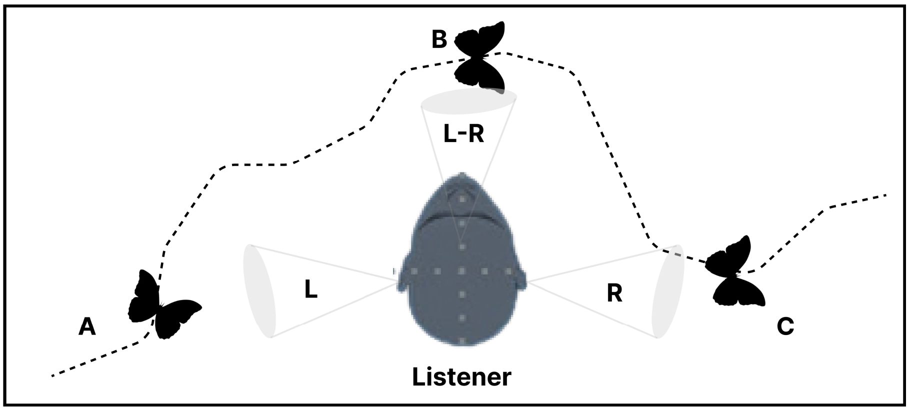


Figure 3.16: 3D spatialization of a butterfly humming sound relative to the listener

Additionally, Wwise allows for intricate control over how sounds are perceived in relation to the listener's position through automation. For instance, sounds that aren't directly associated with an on-screen object can be spatially automated to behave as if they occupy a specific point in space relative to the listener. This is particularly useful for creating ambient effects or simulating sounds that emanate from dynamic but invisible sources.

Diffraction and Occlusion

Diffraction in Wwise is modelled when the direct line of sight between an emitter and a listener is blocked by an object. This creates a path around the obstacle that allows the sound to "bend" around corners or through openings, and depending on the angle of the path around an edge, the sound is attenuated to simulate obstruction while maintaining a level of audio presence even when the direct path is obstructed. The sound's attenuation along these paths is determined by diffraction curves that help simulate the natural behaviour of sound as it encounters physical barriers in the environment.

Occlusion, on the other hand, deals with the blocking of sound by an object that results in a reduction of sound transmission to the listener. In Wwise, occlusion is modelled to

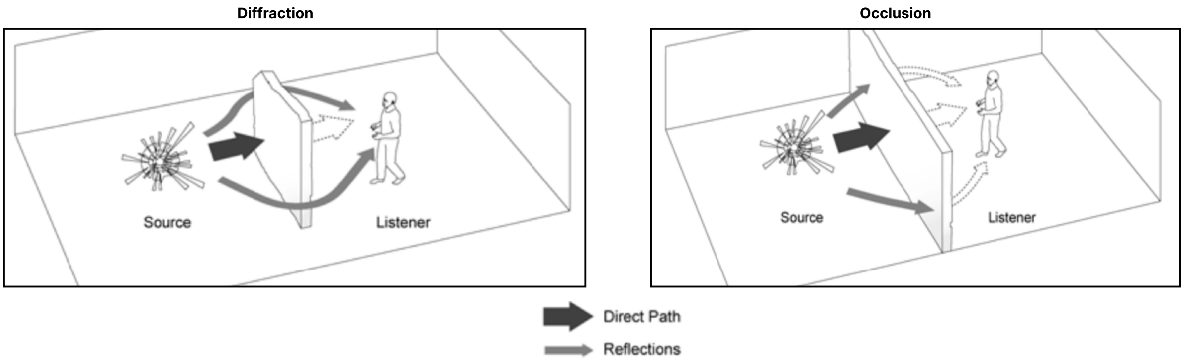


Figure 3.17: Diffraction and Occlusion of sounds in a 3D space
Images reused from [22]

affect both the direct sound path and the reverberated sound through auxiliary sends, depending on the object's properties and the environmental setup. This is achieved by setting occlusion levels that modify the volume, low-pass, and high-pass filters applied to the sound as it passes through or is blocked by materials in the game world. Occlusion can dramatically change the perceived sound by simulating how thick or dense materials affect the sound passing through them.

Dynamic Reflections

Dynamic reflections in Wwise are implemented through the AudioKinetic Reflect plug-in [27], which is used to simulate the way sound interacts with environments in a realistic and dynamic manner. The Reflect plug-in utilizes the geometry of the 3D environment, assessing the proximity of the listener and emitter to various surfaces like walls, ceilings, and floors to generate early reflections. The technology behind this involves the Wwise Spatial Audio Geometry API [24], which calculates image sources representing sound reflections off surfaces. Figure 3.18 how sound travels to the listener from the source when surrounded by walls. Here, the geometry of the walls, along with their surface material, determine in what ways sound would bounce off to reach the listener to simulate how sound would naturally bounce in a real-world setting, providing a richer and more authentic auditory experience.

Granular Synthesis

Conventional methods of producing sound in VR games, often involving the playback of pre-recorded samples, face limitations when it comes to accommodating the various sound

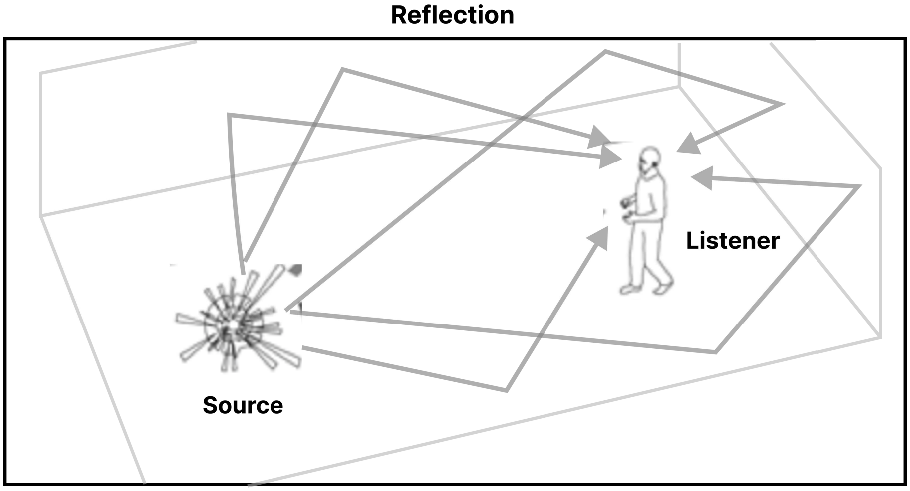


Figure 3.18: Sound reflecting off the walls of a room to reach the listener
Images reused from [22]

variations required for depicting diverse interactions between GOs. This challenge necessitates the use of model-based sound synthesis techniques capable of generating numerous unique sound instances without relying heavily on additional sound samples [72]. For the purpose of the research, the sounds emerging from interactions are typically non-musical, often referred to as 'environmental' sounds encompassing actions like scraping, rolling, and bouncing, among others. Hence, preserving the quality of these sounds is crucial for creating a sense of realism and immersion within the gaming experience. Among various available sound synthesis techniques, granular synthesis was chosen for our cause due to its infinite process of creating ample variations of a given sound file while preserving complex sound textures and characteristics [72].

As illustrated in Figure 3.19, granular synthesis operates by replaying tiny segments of an audio file, known as "grains," while applying an amplitude envelope through a process called "windowing." These grains are then combined in the final output. The intervals between each grain's playback, the length of time for which a grain is audible, the rate at which the original file is read, and the specific position within the source file are all managed separately. As a result, even from a solitary source file, an extensive range of diverse sounds can be produced. For the process of granular synthesis, Wwisewas was used along with the Soundseed Grain plug-in [23]. This granular synthesis tool has a dedicated filter per grain and 3D spatialization capabilities. All of these parameters can be modulated or randomized, using Real Time Controlled Parameters (RTPCs) and/or the embedded

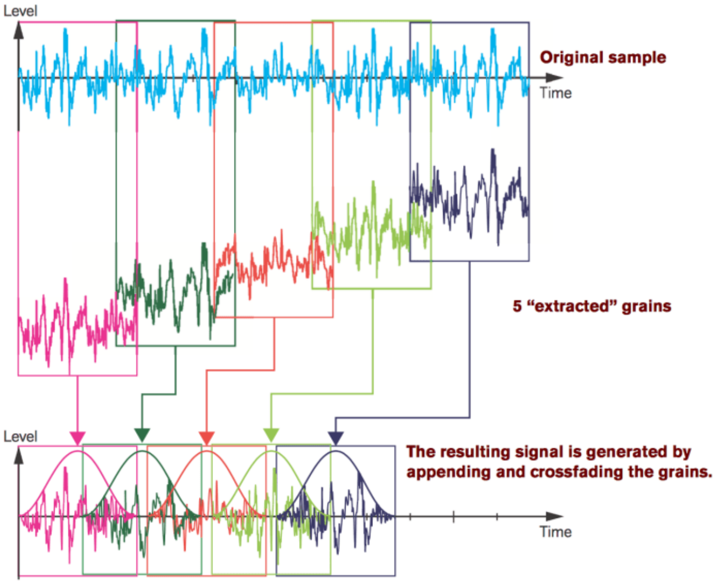


Figure 3.19: General concept of granular synthesis
Image reused from [29]

grain modulators, ultimately enabling the generation of an endless parameterized array of novel sounds.

Wwise enabled Granular Synthesis and Integration with Unity

Granular Synthesis is supported by the Wwise sound engine using the granular synthesis plugin- ‘Soundseed Grain’ [23]. The audio files collected using the data collection prototype and segmented during post-processing, as explained in Section 3.2.3, were used with the plugin to render cane audio by introducing variability and randomness in the sound. The segmented audio files for the respective surfaces were loaded into separate random containers (audio containers that randomly select and play an audio file during run-time, refer to section 3.2.3) in the sound engine. These random containers are controlled by a container switch that governs the initiation and switching between the containers when triggered by a particular event. Any audio file randomly selected by the random container is then fed into the granular synthesizer plugin. The pitch and volume are the two attributes of the granular synthesizer plugin that are exposed as RTPCs whose values are provided by Unity. The plugin then outputs the granularly synthesized sound using the connected audio output system. The integration points for the synthesizer and Unity thus become the plugin’s RTPCs, each of the random containers, and the container switch.

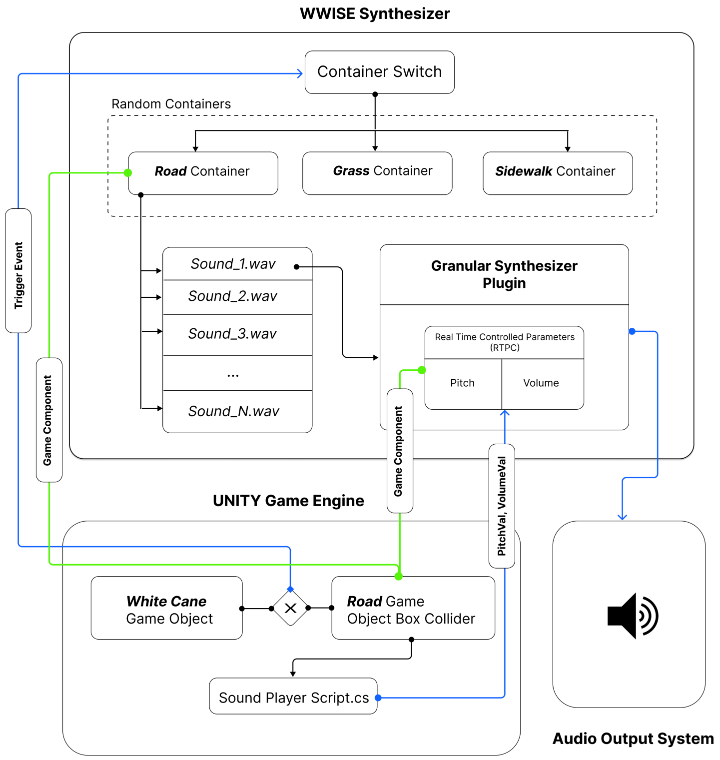


Figure 3.20: Wwise synthesizer and Unity integration block diagram

In the Unity GE, the two primary constituents that are crucial points of contact for its integration with Wwise are the surface box colliders and the sound player script (custom C# sciprt to drive audio event). The box colliders, as explained in Section 3.2.2, are game components that represent the invisible spatial boundaries of a particular GO; in this case, they are the virtual surfaces (roads, sidewalks, etc.). The box colliders are linked with the respective random container and the plugin's RTPCs. Whenever the white cane GO in the game collides with any of these box colliders, a particular event is triggered that activates the container switch in the synthesizer. The switch then decides which container to select depending on the box collider's linkage with the respective random container. Concurrently, the box collider also calls the sound player script that, in return, sends the pitch and volume values depending upon the white cane GO's interaction speed calculated using a lerp function. This whole process is repeated for each frame of the game until the white cane GO is no longer in contact with any of the box colliders. Figure 3.20 illustrates this process as a block diagram, where the green lines represent the linkage between system components, the blue lines represent inter-system calls, whereas the black flow lines represent the intra-system calls.

3.2.4 Haptic Rendering

Contrary to the first phase, where the handheld VR controller served as the medium for conveying haptic feedback, in this phase, the idea was to build a prop-based passive haptic hardware as discussed in Section 2.6.2. The objective was to create a medium to convey high-fidelity vibrotactile feedback of cane interactions with various surfaces in the VE through a handled prop resembling a white cane. An iterative approach was followed where two methods of haptic generation were built and tested; the first was through a voice-coil actuator assembly, and the second was using an audio-to-haptic encoder engine.

Method 1: Custom Actuator Based Hardware

This method collected three-axis contact acceleration data of cane interactions with various physical surfaces, processed it, and mapped it to corresponding virtual surfaces. When the virtual cane interacted with any virtual surface, the idea was to use the mapped acceleration data of the virtual surface to drive voice-coil actuators attached to a physical cane that would render a similar tactile sensation when interacting with a physical surface.

Acceleration Data Collection and Processing

Data Collection

For the collection of acceleration data, a custom data collection system, as shown in Figure 3.21, was created. The hardware assembly consisted of a pair of digital three-axis accelerometers controlled by a microcontroller, which was attached to a standard white cane where data was captured on a PC through a serial channel. The accelerometer sensors captured high-frequency accelerations when the white cane interacted with various surface textures in a double-swiping motion over the span of ˜2 sec. Each surface was manually scanned six times by the experimenter (Figure 3.22a) while maintaining a constant swipe speed and force, resulting in a total of 12 swipes worth of acceleration data set. The double-swiping motion was a standard technique as it is commonly used by visually impaired people during their O&M training [127].

High-frequency vibrations were captured using two accelerometer sensors: an ADXL345

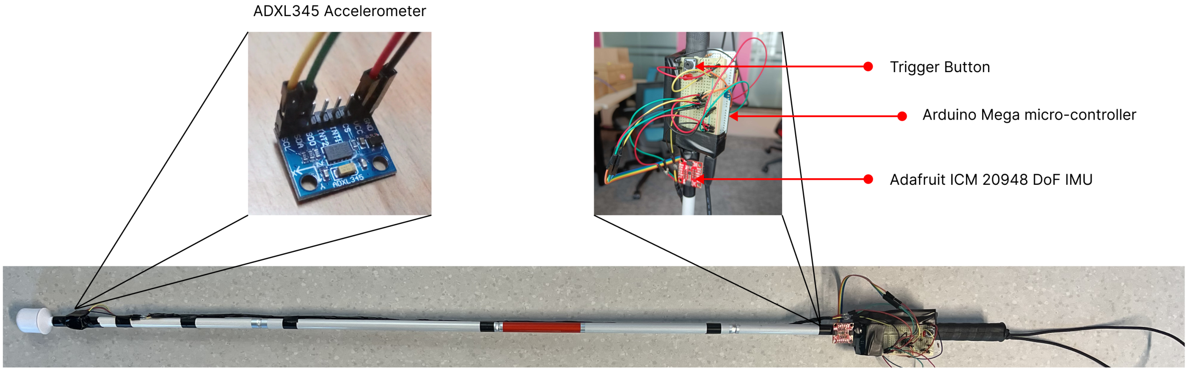


Figure 3.21: Acceleration data collection device using a standard white cane

(a) (b)

Figure 3.22: (a) Acceleration data collection of a white cane's interaction with the Sidewalk's texture (b) Zoom H5 Audio recorder used to collect interaction sound

three-axis digital MEMS-based accelerometer, and a SparkFun ICM-20948 9DoF Inertial Measuring Unit (IMU) affixed to the white cane near the tip and the handle, respectively. These chips were selected for their ability to read accelerations along three Cartesian axes, their high bandwidth, and their configurable range of up to $\pm 157, \mathrm{m/s^2}$ $(\pm 16, g)$. An Arduino Mega microcontroller was used to establish a serial interface, connecting both accelerometer sensors to the PC and transferring data. To reduce redundant data collection when the white cane was not in contact with any surface, a trigger button was configured. When pressed, the trigger button initiated the data collection sequence, and it terminated when the button was released.

A custom data collection program gathered acceleration data from the ADXL and

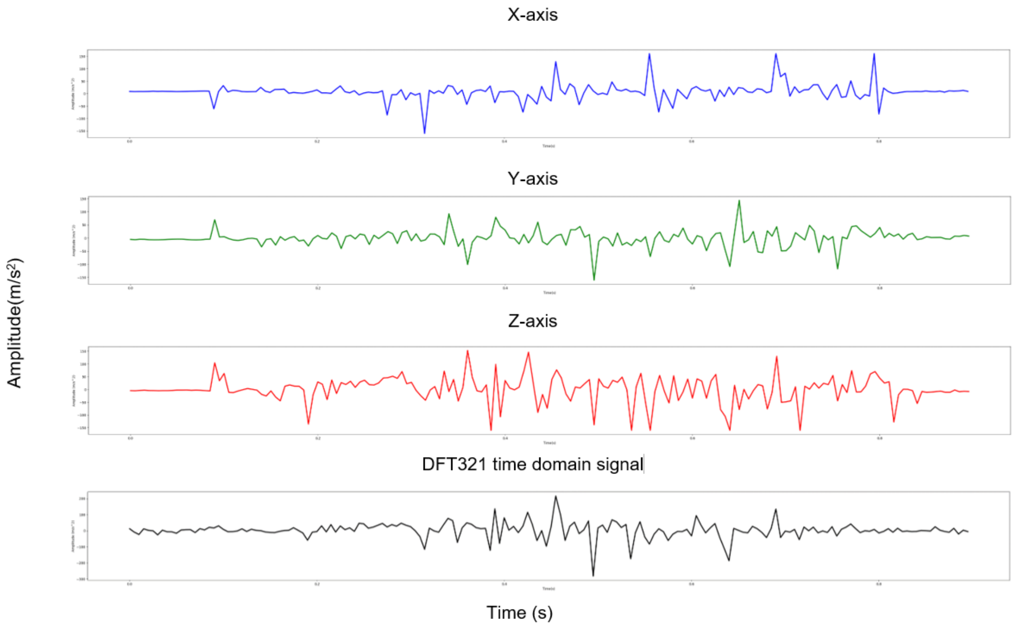


Figure 3.23: Three-axis time-domain signals are reduced to one-axis using DFT321. The depicted signals were recorded during an interaction of the white cane with a sidewalk

IMU sensors using an Arduino C Library whereas a PC-based application named CoolTerm was used to interface the serial connection that also stored the set of recordings into a txt file after each interaction. Upon startup and until the duration of the trigger button being pressed, the program configured both the accelerometer sensors into $\pm 157\,\mathrm{m/s^2}$ ($\pm 16\,g$) mode with a sampling rate of 200Hz.

Data Processing

Numerous researchers have highlighted that human sensitivity to high-frequency vibrations remains largely unaltered by their direction [97]. Furthermore, devising a high-frequency vibration haptic device with just a single output axis is significantly simpler compared to the complexities of crafting a device with three output axes. These factors prove crucial for condensing recorded vibration data, originally captured across three-dimensional Cartesian directions, into a singular dimension using the DFT321 algorithm [97]. The primary objective of this compression is to generate a vibration signal that is nearly indistinguishable from the original three-dimensional signal and effective in capturing both the spectral energy and temporal information from all three axes, as demonstrated in a sample recording showcased in Figure 3.23.

The DFT321 technique employs frequency-domain methodologies to unify the initial three signals into a cohesive signal while upholding the overall spectral power [97]. Initially, the squared sum and square root of the magnitudes of the smoothed Discrete Fourier Transforms (DFTs) are calculated for each of the original three signals:

$$|\tilde{A}_s(f)| = \sqrt{|\tilde{A}_x(f)|^2 + |\tilde{A}_y(f)|^2 + |\tilde{A}_z(f)|^2} \tag{3.1}$$

Here $|\tilde{A}_s(f)|$ is the frequency-domain magnitude of the new DFT321 signal and $\tilde{A}_x(f)$,$\tilde{A}_y(f)$, and $\tilde{A}_z(f)$ are the DFTs of each original Cartesian acceleration vectors. The resultant DFT321 signal is established by computing the inverse tangent of the sum of imaginary parts divided by the sum of real parts of the DFTs of the original signals, resulting in an average phase $\phi(f)$:

$$\phi(f) = \tan^{-1}\left(\frac{\mathrm{Im}(\tilde{A}_x(f) + \tilde{A}_y(f) + \tilde{A}_z(f))}{\mathrm{Re}(\tilde{A}_x(f) + \tilde{A}_y(f) + \tilde{A}_z(f))}\right) \tag{3.2}$$

Subsequently, an inverse DFT operation is carried out on the calculated magnitude and phase to create the new time-domain signal. This algorithm is implemented in Python using the standard Numerical Python (NumPy) and Scientific Python (SciPy) libraries.

Rendering Hardware

For rendering the haptics, a custom-built hardware assembly was created using vibrotactile transducers known as ‘haptuators’ (TactileLabs, model no. TL002-14-R) attached to a white cane. This hardware assembly was controlled by an Arduino Nano microcontroller, which was serially interfaced to the PC for data transfer. The input to the haptuators was provided through a linear audio amplifier module (DKARDU, model no. LM386). The haptic output was driven by the haptuators, which were firmly attached to the white cane near the tip and the handle using plastic brackets, as shown in Figure 3.24. The primary reason for using a haptuator was its low static friction and commercial availability. Additionally, the specific positioning of the haptuators was intended to accurately mimic the acceleration captured through accelerometer sensors positioned in the same locations during the data collection phase, as shown in Figure 3.22a.

A custom Arduino-based C program was used to connect to the Unity GE through the serial connection specified at a particular port and a set baud rate of 96000. As

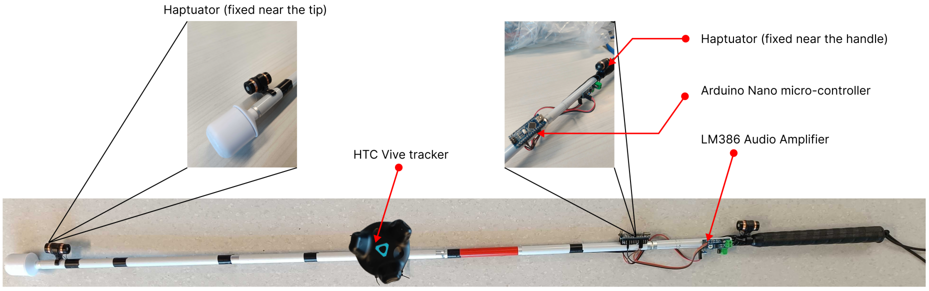


Figure 3.24: Haptics rendering device using haptuators vibrotactile transducers

explained in Section 3.2.3, similar to how Unity triggers surface-specific audio to be synthesized on collision of the cane GO with a surface box collider, the same event sends a start haptics flag appended with the surface tag through the serial connection to initiate the haptics rendering. For example, if the collided box collider was of type - 'road,' the flag - 'startHaptics_Road' was sent over the serial channel to play the road surface-specific acceleration magnitude signal, which was generated and stored during the data processing phase (Section 3.2.4) of the implementation. Similarly, a stop flag was set to indicate the haptics rendering to be terminated once the cane GO was no longer in contact with any surface box colliders.

Rendering Process

Due to certain technical limitations, for each surface, a single DFT321 time domain signal for only one accelerometer sensor (ADXL345 attached near the tip of the cane) was used to render vibrations following the mechanism discussed above. A DFT321 time domain signal was converted into its spectral representation using Fourier transform where only the values of amplitude before the Nyquist cutoff frequency were considered. Thus, since the original acceleration data were sampled at a frequency of 200 Hz, in the frequency domain, 100 Hz was considered the highest attainable frequency as per the Nyquist-Shannon sampling theorem [103].

After calculating the spectral domain, Pulse Width Modulation (PWM) was used to play the magnitude at each frequency through the haptuator for a specified amount of time (~10 ms). Frequency for a given instant was used to calculate the period of square (pulse width) waveform, and a duty cycle of 50% was used, which resulted in a modulated pulse width with high and low values for an equal amount of time. The Fourier transform was implemented using the standard Python NumPy library, whereas the PWM was programmed using an Arduino-based program using standard C libraries.

Method 2: Third-Party Haptic Engine

The issue with the haptic rendering hardware used in the first version of the cane prototype was the fidelity of the force feedback (vibrations). Although audio amplifiers were used in the hardware, the rendered haptic sensation was not on par with the vibrations generated during real-world interactions with the cane. Moreover, attempts were made to refine the hardware assembly with more powerful amplifiers and dedicated audio synthesizing micro-controllers; however, due to time constraints in the project, creating a custom hardware assembly for rendering haptics was eventually dropped.

From the second version of the cane prototype, the Quest Pro VR controllers equipped with programmable haptic motors were used as the source of haptic feedback. These controllers were attached to the end of the cane stem using a custom 3D printed handle, explained in Section 3.2.5. The reason for going forward with the VR controller was that it was readily available, had in-built haptic motors that were programmable and provided a higher degree of perceivable force feedback.

Interhaptics

The haptic rendering on the controllers was made possible through the haptic engine from ‘Interhaptics’ called the ‘Interhaptics Haptic Engine’ [5] which was integrated to work with the Unity GE using the ‘Interhaptic Unity SDK’ toolkit. The rendering process follows an audio-to-haptic encoding method that translates audio waves into corresponding haptic profiles comprising three properties: transients, amplitudes and frequencies. The primary components of the Interhaptics suit detailed below are summarized from the online documentation by Interhaptics [5].

Haptic Engine

The Interhaptics Engine is a haptic rendering framework designed to enhance the tactile feedback capabilities in various gaming applications by translating haptic effects into precise control signals compatible with various haptic devices. It supports real-time haptic rendering, which supports spatial and event-based interactions, allowing users to experience nuanced tactile sensations such as vibrations, textures, and stiffness. The engine’s capability to convert audio signals into haptic feedback, combined with full-body mapping and spatial haptics, addresses the common issue of under-sampling in haptic feedback through hyper-sampling techniques, thereby maintaining consistent and stable tactile responses even with varying position acquisition rates from tracking systems.

.haps file format

The '.haps' files are written in a JSON-based format that serve as a crucial component in storing and managing haptic effects using the Interhaptics Composer and rendered by the Interhaptics Engine. The .haps files describe the haptic signals triggered by specific events or interactions within the application. Designed to be platform-agnostic, these files ensure that the haptic experiences can be translated accurately to various peripherals through the Interhaptics SDK, maintaining consistent haptic feedback across different devices. The .haps file structure includes several key elements organized in a hierarchal structure that contribute to the functionality of haptic rendering. The file structure contains the following components:

- **Keyframes** provide detailed control over haptic signals by specifying the position and properties, such as amplitude and frequency, to be modulated.
- **Haptic Notes** group these keyframes, facilitating the organization and management of haptic effects on a timeline.
- **Haptic Melodies** comprise lists of non-overlapping haptic effects that, when rendered, can deliver combined outputs if they overlap on different melodies.
- **Perceptions** are the haptic effects that can be created and rendered using the Haptic Composer. The currently supported perceptions are Vibrations, Stiffness, and Texture:
    - **Vibration:** Vibrotactile events occurring over time, represented widely in .haps file and stored using a wideband representation and created using transients, effects, and multiple melodies.
    - **Stiffness:** Force in response to displacement, adaptable to device strength and used for adaptive trigger feedback.
    - **Texture:** Vibrotactile feedback relative to spatial displacement, rendered based on the device's movement over haptic surfaces, with application-dependent rendering rates.

Haptic Composer

The Haptic Composer is a graphical tool for designing haptic experiences stored in .haps file format that includes several key functionalities:

- **Audio to Haptics Encoding**: Allows automatically importing a .wav audio file and extracting haptic properties with selectable properties.
- **Note Editing**: Enables precise editing of haptic notes in the note editor, including keyframes that can be dragged, dropped, and modified.
- **Haptic Presets**: Provides pre-designed notes, or Haptic Presets, that can be assembled and edited to create haptic melodies through drag and drop.
- **Testing**: Facilitates rapid iteration and testing of haptic experiences with built-in testing methods through connected peripherals.

Interhaptics Unity SDK

The Interhaptics SDK for Unity integrates the control of the Interhaptics Engine through the Unity Engine. Moreover, it enables cross-platform support for the engine, such as OpenXR and Oculus headsets and mobile devices. The Out-of-the-Box (OOB) scripts and plugins that come with the SDK support the specified haptic sources (presets or .haps files) and control over amplitude and frequency at run time.

Rendering Process

The haptic rendering process has two steps; in the first step, audio recordings are converted into the .haps file format, and in the second step, a script in Unity uses the .haps file to render the haptic feedback in the controllers using the Interhaptics Engine. However, both steps are asynchronous, with the first step needing to be completed manually and the second triggered through event-based logic.

As illustrated in Figure 3.25, all the audio recordings of the physical cane interaction with real surfaces (explained in Section 3.2.3) stored in the WAV format are loaded into the Haptics Composer. Each audio file generates a corresponding haptic profile of transients, amplitudes and frequencies upon loading. These haptic profiles are saved in the .haps files format in the resource directory of the Unity project.

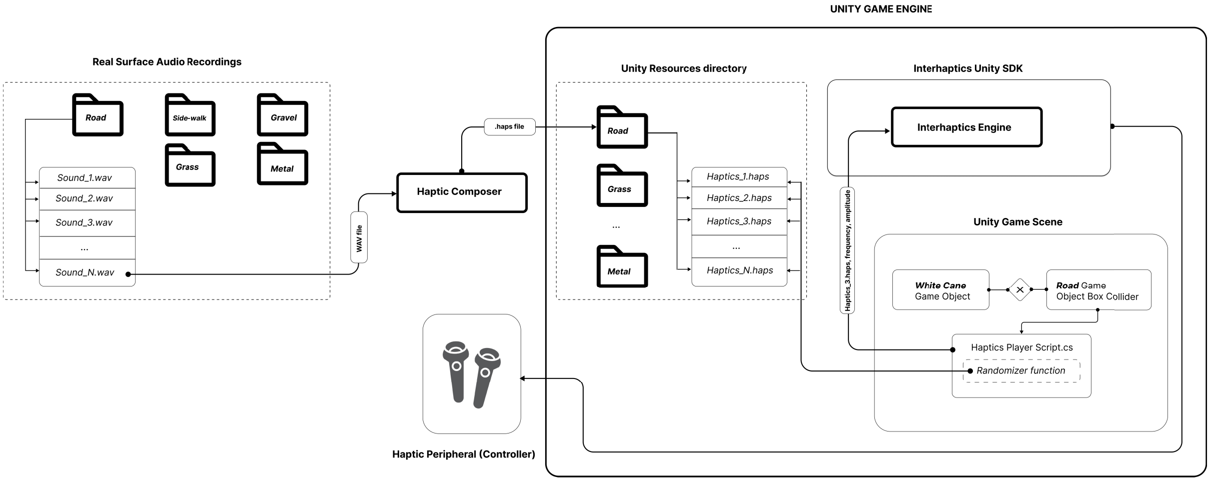


Figure 3.25: Haptics Rendering Process using the Interhaptics Engine

In the Unity project, the Interhaptics Engine is integrated through the SDK toolkit. In the game scene, each virtual surface has an attached collider that specifies what type of surface it is (road, grasses, gravel, etc.). Each surface collider is linked to the respective folder in the resource directory with a corresponding set of .haps files. Whenever the virtual cane in the scene interacts with any virtual surfaces, the attached colliders trigger an event through a custom script that randomly picks a .haps file from the linked haptic profile folder and forwards it to the Interhaptics Engine. Moreover, for the entire period of interaction between the cane and the surface, the script determines the frequency and amplitude of the generated haptics using the velocity of the virtual cane, which is then forwarded to the haptics engine to modulate the properties of the rendered haptics.

3.2.5 Physical Cane Prototypes

As discussed in Section 2.4.2, the 'Object In Hand' interaction interface was used for the large-scale VE by using a physical cane that worked as an augmented prop held by the users in hand. Altogether, three different physical cane prototypes were iteratively designed to map the physical cane to the virtual cane, focusing on making it the primary mediator for haptic feedback to interact with the different virtual surfaces. The prototypes differed in terms of the employed haptic rendering hardware, tracking system, tip of the cane, and the form factor. With every iteration of the cane design, attempts were made to create a prototype that looked and felt more like a standard white cane in terms of weight distribution, length and resemblance. Table 3.2 compares the weight and length of the three prototypes with a standard white cane.

Table 3.2: Comparison of weight and length of cane prototypes with a standard white cane

Cane	**Weight**	**Length (handle to tip)**
Standard	250 gm	123 cm - 125 cm
Prototype 1	424 gm	123 cm
Prototype 2	376 gm	123 cm
Prototype 3	334 gm	126 cm

Prototype 1

In the first version of the prototype, as shown in Figure 3.24, a standard white cane was attached with an HTC Vive tracker in the middle along with two voice-coil actuators on the end of the cane's stem - one near the handle and the other near the cane's tip. The former was responsible for the tracking (explained in Section 3.2.6) and the latter for haptic rendering (explained in Section 3.2.4). The actuators were driven through a hardware assembly of an Arduino Nano mico-controller and a pair of audio amplifiers connected to it. The microcontroller was connected to the Unity GE system through a USB interface that acted as the channel for transferring data.

The placement of the aforementioned components was strategic in terms of equally distributing the weight so that the center of mass of the prototype aligned with that of the actual physical cane without any alterations. However, the issue with this design was the added weight to the cane and the wired connection of the components. The actuators connected to each audio amplifier and the amplifiers themselves connected to the microcontroller resulted in a mesh of jumper wires. Moreover, the controller itself was connected to an external system through a long USB cable, resulting in spatial constraints to the range of the cane usage, along with having to be careful not to disconnect the wirings. The overall weight was a concern, along with the form factor that made the cane look less humane. Moreover, the original tip of the cane was left intact, which made the idea of using a proxy surface (slide mill skirt) unattainable, as whenever the tip was swiped over a surface, the natural interaction masked the rendered haptics and sound.

Prototype 2

Due to the shortcomings of the first prototype, specifically - haptic (explained in Section 3.2.4) and tracking (explained in Section 3.2.6), the major changes to the design were replacing the actuator-based haptic hardware with VR controllers and attaching a set of retro-reflective markers on the stem of the cane in place of the Vive tracker to enable Qualisys based MoCap tracking (explained in Section 3.2.6).

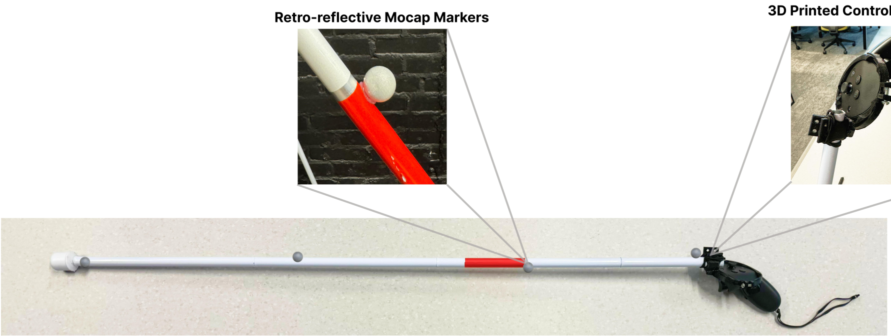


Figure 3.26: Second version of the cane prototype

As shown in Figure 3.26, the cane stem was detached from the handle, and the handle was replaced with a 3D-printed mount that connected it to the Quest Pro VR controller. Moreover, the original tip of the cane was also replaced by a similar-sized styrofoam ball to work as an absorber to inhibit natural vibrations and noise during interactions with physical surfaces. A total of six retro-reflective markers were glued all around the cane stem to ensure that at least three markers would be visible to the MoCap cameras at any point during the tracking.

Prototype 3

As shown in Figure 3.27, the third design of the cane is similar to its previous version, with only a change in the tip of the cane and the replacement of the mount with an actual handle to hold the VR controller. The issue with the 3D printed mount in the previous version was that it made holding the cane less natural as it had to be gripped through the controller as if holding the pistol grip of a gun. This resulted in a structurally weaker design when moved mid-air without support and made the whole interaction with the cane more rigid, as it could only be held in a single position. To address this design flaw, a 3D-printed handle was created to replace the mount with a two-part assembly.

The first part, which attaches to the cane stem, features grooves, while the second part, housing the controller, has a screw design that securely fixes it to the first part.

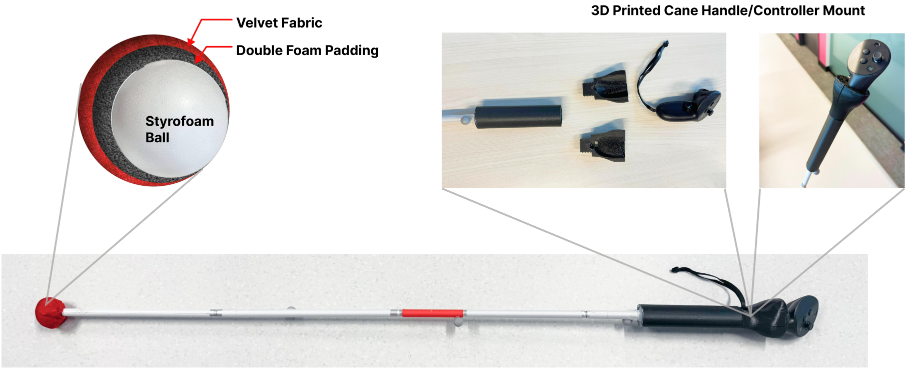


Figure 3.27: Final version of the cane prototype with layers of the tip and parts of the 3D printed handle

The styrofoam tip of the cane from the previous version, when used on the slide mill skirt (detailed in Section 3.2.7), although minimized the external vibration and noise, still produced a perceivable high-pitched sound. To counteract this and muffle the noise completely, the tip was updated with an added layer of foam enveloped in a velvet fabric.

3.2.6 Physical and Virtual Cane Mappings

Contrary to the first phase of the methodology, which utilized VR controller-based virtual cane mapping, the second phase aimed to employ a physical cane as a ‘prop’ mapped to its virtual counterpart that precisely coupled both the canes’ orientation and position. A couple of tracking systems were iteratively tested throughout the design of the cane prototypes. Initially, the most readily available and heavily used tracker-based system in the literature [108][141], which shared similarities with our use case, was used; however, due to a few limitations, a shift was made towards a more efficient motion-capture-based tracking system.

Method 1: SteamVR Tracking System

For the first iteration of mapping the physical cane to the virtual cane, a single HTC SteamVR/Vive Tracker (Model: 99HANL002-00) was attached to the stem of the first cane prototype as shown in Figure 3.24 using a Velcro contraption and was exactly in the middle. Using Velcro enabled the tracker to be easily swapped with another, making charging the tracker much easier. Moreover, since the tracker has weight to itself, its attachment position had to be determined through hit and trial so that the center of mass of the cane would remain undisturbed.

Tracking Process

The HTC Vive Tracking system, also known as SteamVR Tracking [115], operates with three primary components: Base Station, Vive trackers, and Host. For our implementation, we used a single tracker and a couple of base stations facing down and mounted on tripods reaching heights up to 7 meters. The description of the following components is summarized from the online documentation by SteamVR [115].

Base Station

The base stations facilitate the tracking environment as they are outfitted with a 120° multi-axis laser emitter that sweeps the room with multiple sync pulses and laser lines, reaching about 5 meters. The timings between pulses and sweeps enable the base stations to find the location of each tracker attached to an object. However, to permit 360° coverage, two base stations are typically required. Additionally, these are designed to be fully self-contained and don't need cables to connect them to the host or the tracked objects, except their power supply.

Vive Trackers

Any objects that need to be tracked within the tracking system are attached with trackers that ensure accurate and responsive tracking. The number of trackers that need to be used depends on the size or area of the object to be tracked. Moreover, the tracker's position also depends on the nature of the movements that the tracked object is intended to have. These trackers have Application-Specific Integrated Circuit (ASIC) sensors integrated with a 1000Hz IMU (Inertial Measurement Unit), which offers low latency and high-resolution tracking capabilities. Communication between the trackers and the host is facilitated wirelessly via USB dongles connected to the host.

Host

The host component of the tracking system plays a critical role in synthesizing and managing the tracking data. It primarily integrates 3D positional information from all the trackers within the system. Typically, a PC is the host which processes the data received from various trackers linked through USB dongles. Each tracker has its own USB dongle connected to the host. The host uses the SteamVR API, which ensures accurate timing, synchronization, and prediction of the tracked objects.

Issues Faced

Cross-Platform Integration

Although the HTC Vive Tracking system is not made exclusively to work with HTC headsets, there were integration issues when connecting the Vive trackers with any Oculus HWD at run time. The primary issue was that the Oculus HWD and the tracking system worked on completely different coordinate systems, and although the SteamVR application was the intermediary between the two, there was no OOB solution for the issue. However, certain third-party plugins that worked with SteamVR enabled the correct calibration of the Vive trackers with the Oculus system. The downside was that the calibration only existed for a specific session while the SteamVR was active, and thus, re-calibration was required every time SteamVR had to be initialized.

Base Station Synchronization

The base stations also suffered a few issues at times, especially when any external light reflected off a mirrored surface and reached it. This blinding affected the sync pulses and laser lines emitted from the stations, resulting in incorrect tracking of the trackers. Additionally, using only two base stations failed to provide 360° coverage, as there were times when the tracker mounted on the cane was obstructed by the user's body and introduced potential blind spots in the coverage area. These issues required the tracking area to be much more controlled and necessitated additional base stations.

Method 2: Qualisys Motion Capture System

Due to the various limitations of the SteamVR tracking system, an alternative method for tracking the cane using MoCap was used while prototyping the second and third versions of the cane. Specifically, the Qualisys MoCap system [90] was used to track the cane within the same space with six Six Degrees of Freedom (6DoF) using an optical-passive technique through Infra Red (IR) cameras.

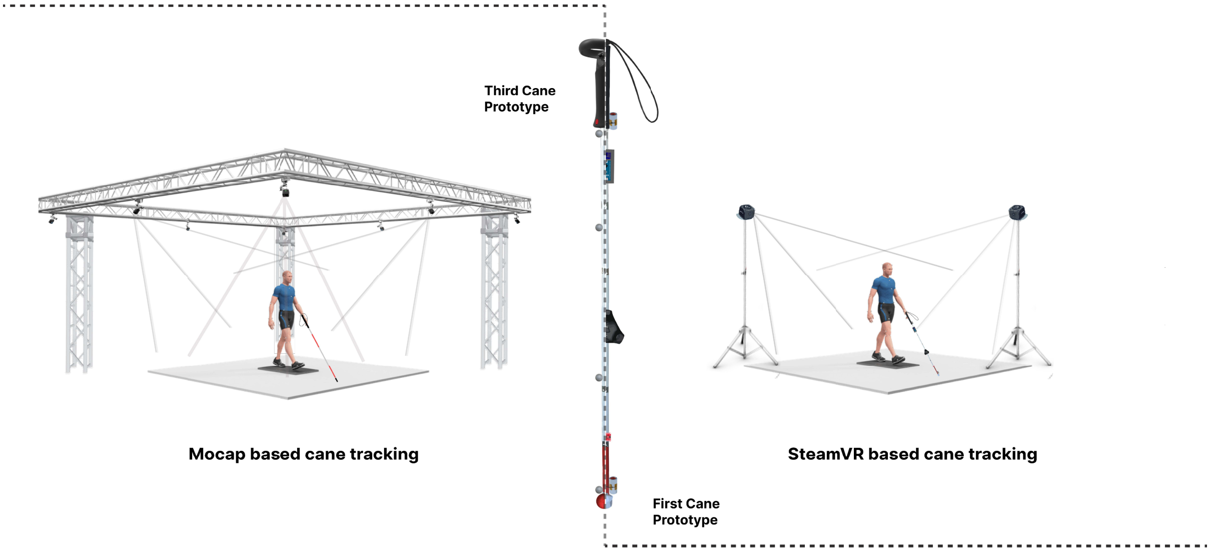


Figure 3.28: Qualysis MoCap based tracking with third cane prototype (left) and SteamVR based tracking with first cane prototype (right)

Tracking Process

For the space in which the user study was conducted, ten IR cameras (Model: Miqus M5) were attached to a metallic truss that tracked the space. The cameras were positioned to prevent blind spots and ensure maximum tracking accuracy. All the cameras were daisy-chained and powered through a camera sync unit that synchronously gathered the tracking data from them and forwarded it to the dedicated PC that hosted the Qualisys utility called the 'QTM Track Manager.'

Before the tracking process, the space had to be calibrated using a set of calibration tools that determined the point of origin in the space along with the x-y-z axis to be mapped in the Qualisys coordinate system. The IR cameras capture the motion of any physical object through a process called 'Passive Optical MoCap' that uses retro-reflective markers that bounce off IR pulses from the cameras to provide the exact position at any given instance. Any physical object requires at least three adequately spaced markers attached to it to be accurately tracked. The requirement of the number of markers is directly proportional to the complexity of the geometry of the tracked object. The cane prototype had six retro-reflective markers glued to its stem between the tip and the handle. The markers were aligned and spaced so that at any given instance, at least three markers were visible to any of the ten IR cameras. Moreover, since the cane constituted a simple linear geometry, it was fairly easy to track.

The ‘QTM Track Manager’ utility is a proprietary application by Qualisys that has OOB functionality to calibrate, define virtual objects and transfer real-time tracking data to other platforms such as Unity GE. Once the markers on the cane were sufficiently visible and tracked in the 3D space of the application, an identifier was assigned that represented the cluster of markers. Later on, the real-time data of the tracked cane was mapped to control the orientation and position of the virtual cane (GO) in Unity using this identifier. However, the real-time data transfer to the Unity GE from the QTM Track Manager required these applications to be either on the same system or two different systems in the same network and on the Unity side required an SDK called ‘QTM Connect for Unity’ from Qualisys.

3.2.7 Omni-directional Slide Mill

For the locomotion of users in the VE, we employed an omnidirectional slide mill from KATVR (Model: KATVR Walk Mini S). The slide mill came with three primary sections: a base plate for foot movement tracking, a rotating back support to track the user’s body’s yaw orientation, and a set of waist and leg straps/harnesses to safely secure the user to the back support. The base plate of the slide mill, as the name suggests, is a convex-shaped metallic plate that, unlike a treadmill, doesn’t move; however, the plate is slippery and needs specialized overshoes to maintain traction and accurately track foot movements. The slide mill was integrated with Unity using KATVR’s native development kit *(KAT Unity Integration SDK)*[128] and needs SteamVR as the runtime environment. Moreover, a specialized PC utility for the slide mill called ‘KAT Industry’ provides a dashboard interface to configure various control parameters and calibration options at runtime. The control parameters mostly comprised controlling the sensitivity and speed of the slide mill’s base plate along with walk-in place and cruise options.

The general procedure for using the slide mill is pretty straightforward. First, the rotating back support is locked in position, and after the user puts on the overshoes, they mount the slide mill. The user is then strapped into the back support using the waist and thigh harness, and once secured, the back support is unlocked. Before starting to walk on the slide mill, calibration is necessary. This is done through the calibration option in the KAT Industry utility, which aligns the direction of virtual movement with the physical direction the user is facing. Once the user starts to take strides on the base plate, at any given instance, only one foot touches the plate, which registers a 'true' boolean value marking that the user is in movement. A ' false ' value is registered whenever both feet are in contact with the plate, marking that the user is stationary. The slide mill thus calculates the speed of the user's movement according to the rate of change of strides detected by the base plate.

On the Unity side, a custom C sharp script that comes with the SDK provides OOB implementation to map the direction and speed of rigid bodies (GOs with a component called 'Rigid body' to enable physics properties) in the Unity GE with the data from the slide mill. However, the OOB script had to be customized to enable calibration of the slide mill tracking direction according to the position of the back support at any given instant to recalculate the orientation offset between the user's HWD and the body's actual yaw orientation, aligning the virtual representation to match the physical direction the user is facing.

Figure 3.29: KATVR Wal Mini S slide mill and closeup of the skirt showing the plastic layer atop the wooden structure

Slide mill Skirt

The slide mill skirt, as shown in Figure 3.29, is a wooden platform made of trapezoid-shaped plywood blocks and covered with a double layer of plastic. It serves two purposes:

acting as a proxy surface to compensate for the elevation of the user due to the height of the slide mill's base plate from the ground and minimizing contact noise, and suppressing vibrations emitted from physical cane interaction on the platform surface. The usage rationale and the design process are explained in detail as follows.

Usage Rationale

Whenever a user got on the slide mill and held the cane in one of their hands, it resulted in an awkward slouch of the user's stance as they tried to rest the cane tip on the ground and move it around. This was more evident and severe for users who had taller stature. Over extended periods, this could significantly strain the user's back and promote an unnatural stance and cane usage, which differs significantly from how the cane would be used in an outdoor environment. In order to mitigate this, the height of the ground had to be alleviated by using a platform that would serve as a proxy for the ground beneath it.

Moreover, whenever the tip of the cane (of the third prototype - Section 3.2.5) interacted with the ground, the tip produced a sweeping sound along with minor vibrations, which were highly noticeable. This indicated that even the padded tip and the velvet covering were insufficient in muffling the external sound and vibration noise. Thus, the slide mill skirt presented as an alternative to introduce an added layer of noise suppressor. This was accomplished by enveloping the entire top surface of the skirt with two layers of thick plastic film, which were stretched and securely stapled in place. The slick texture of the plastic effectively reduced the external noise from cane interactions and created a smooth surface without any natural irregularities.

Design Process

The design process of the slide mill skirt entailed the following steps, which required calculating the optimal distance of the extreme ends of the skirt from the center of the base plate, sketching and iterating different possible shapes of the skirt, finalizing the material to build the skirt with and testing different materials to go on top of the skirt.

Step 1: Dimensions of the skirt

In order to determine the dimensions of the skirt that would surround the slide mill, we performed an experiment with three users of varying heights who would stand on the slide mill and hold and swipe a cane that had its tip replaced with a marker onto a large piece of white chart paper laid out in front of them. The users held the cane in their dominant hand and were asked to swipe it in a left-to-right motion in two different conditions and with a few different 'intents.' The conditions and the intents are as follows.

- Condition 1: The user's body is stationary, and the cane is held with intents: searching nearby surfaces (hands contracted), searching far away surfaces (hands stretched), and searching the surfaces normally (hand rested)
- Condition 2: The user's body is in motion (walking), and the cane is held with intent: searching the surfaces normally (hand rested)

As shown in Figure 3.31, after each user swiped the cane on the chart paper, measurements were recorded with a measuring tape from the center of the base plate (C) to the nearest point of the pattern (L) and the farthest point of the pattern (H). These measurements provide the radius of the inner circular pattern (C-L) and the outer circular pattern (C-H). Among the three users, these radii values were then adjusted using a scale factor (K=5cm), which decreased the minimum inner circle radius (Min C-L) by K units to give (Adjusted C-L) and increased the maximum outer circle radius (Max C-H) by K units (Adjusted C-H). The scale factor (Figure 3.30) was applied to compensate for the height if the paper was higher from the ground (on the same level as the nearest point of the baseplate from the ground), which the skirt would eventually match. Eventually, the difference between these min inner radius and max outer radius gave the required thickness of the skirt (Final L-H).

Table 3.3 shows the height of each user along with the measured circular dimensions and cane patterns, and using the data, the calculation for the exact skirt thickness is as follows.

Min C-L = 77 cm
Max C-H = 151 cm
Base frame height adjustment scale factor (K) = 5 cm
Adjusted C-L = Min C-L - K = 77 cm - 5 cm = 72 cm
Adjusted C-H = Max C-H + K = 151 cm + 5 cm = 156 cm
Final L-H (skirt thickness) = Adjusted C-H - Adjusted C-L = **84 cm**

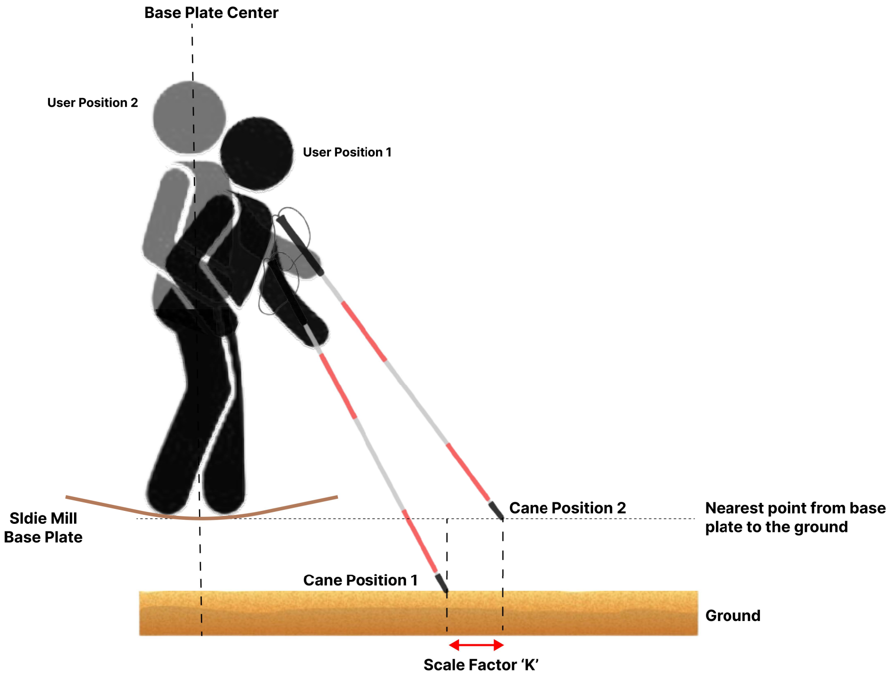


Figure 3.30: Scale Factor Diagram

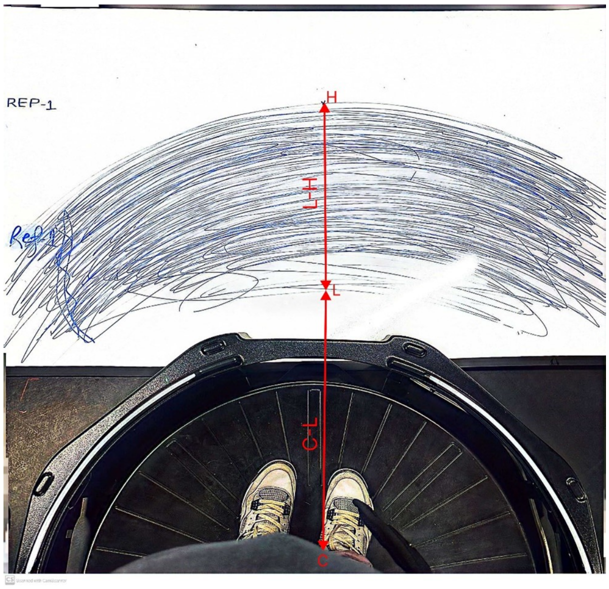
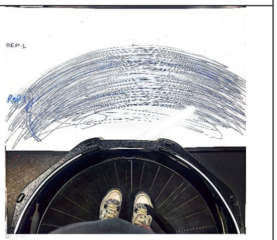
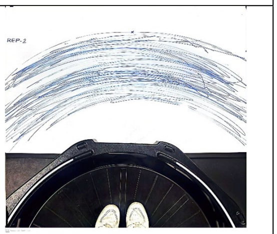
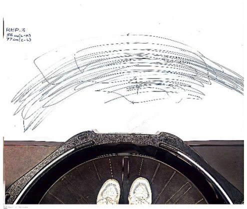


Figure 3.31: Slide mill skirt calculation parameters

Step 2: Material to build the skirt

After determining the shape of the skirt, we found that the feasible options were either Styrofoam or wooden sheets. The former was easy to manage, cut out, and build but failed to provide structural integrity. The latter was comparatively difficult to work with in the absence of specialized tools but was a sturdier option. Since the skirt had a high chance of getting stepped on while users got on or off the slide mill, the material had to be sturdy enough to withstand those events. Thus, wooden sheets were found to be the ideal material for building the skirt.

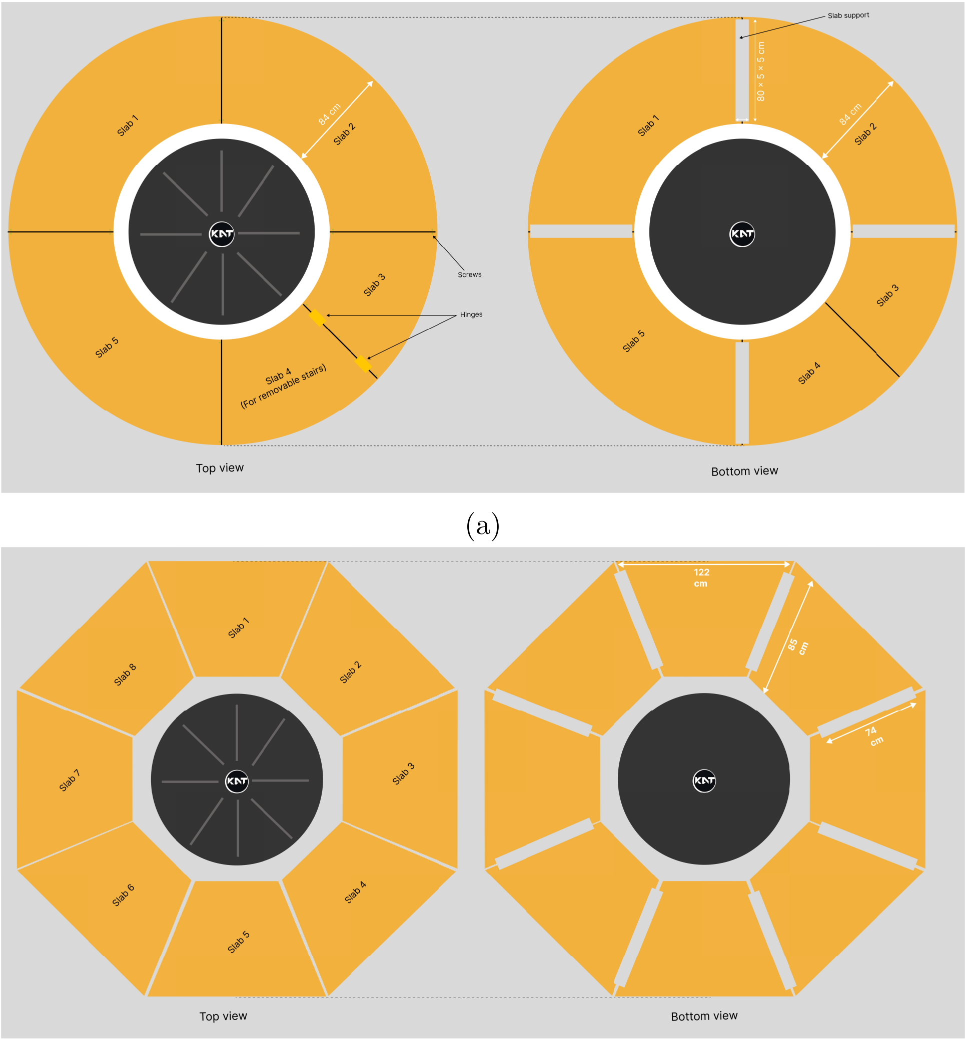
Table 3.3: User heights and corresponding measurements

User height (cm)	C-L (cm)	C-H (cm)	Rep #	Photo
167.64	77	142	1	
172.72	93	151	2	
187.96	77	133	3	

Step 3: Shape of the skirt

As illustrated in Figure 3.32, there were two iterations to determine the shape of the skirt. In the first iteration (Figure 3.32a), the dimensions obtained from the first step were used to sketch a design that loosely resembled the pattern of the cane sweep (Figure 3.31), which was in the form of a circular annulus. Thus, the initial plan was to divide the annulus shape into five circular sectors and cut out each of the five sectors from a sheet of plywood. The annulus was planned to have a width of 84 cm, and each slab was planned to be connected using a two-by-four of 80 cm length (attached breadth-wise).

However, it was later identified that the circular cutout would be difficult to cut out and transport and lead to exorbitant waste of plywood sheets. Thus, in the second iteration (Figure 3.32b), an octagonal annulus was finalized and comprised of eight trapezoids, each with 85 cm legs. Similar to the previous design, each slab was planned to be connected using a two-by-four attached breadth-wise, but with a reduced length of 75 cm. Moreover. in the first design, the plan was to cut one of the slabs in half to create a joint to lift one-half of the slab. This was primarily designed to make room for removable staircases that would be used to help users get on. However, this idea was dropped in the final design when the skirt was deemed sturdy enough for users to use it to get on and off the slide mill.



(a)

(b)

Figure 3.32: Planned Design of the slide mill skirt (a) Design layout from the first iteration, (b) Finalized design layout from the second iteratio

Step 4: Material to go on top of the skirt

As discussed above in Section 3.2.7 (Usage Rationale), the skirt needed to be enveloped in a material that could contain external noise and vibration. Different materials such as carpets, fabrics, and plastic were tested by covering a small space of the skirt with these materials and sweeping the cane over each. It was discovered that carpets worked the best, and the plastic worked equally well, while fabric introduced additional noise and resistance. The issue with the carpet was that not only was it economically a less feasible option than plastic, but it resulted in a lot of waste as big rolls of carpet would have to be cut to fit the skirt's design. Ultimately, vapour barrier plastic was chosen to cover the skirt as it was fairly easy to stretch and affix using a staple gun.

Chapter 4

User Study and Evaluation

4.1 Study Design

The overall aim of the user study was to conduct a controlled lab experiment to evaluate the system's overall feasibility, tolerability, and empathetic efficacy in the navigation of VR spaces without visual cues. Moreover, the study aimed to build on inclusive approaches to accommodate non-sighted users, not only by promoting accessibility as a means to appropriate VR systems but also by providing sighted users with a medium to experience and understand the perspective of their non-sighted counterparts.

4.1.1 Research Questions

This section presents the key inquiries guiding the investigation into the effectiveness of utilizing audio cues and haptic feedback in navigating VEs by users in non-visual conditions. For the research, there were three primary Research Question (RQ):

- **RQ 1: Level of Differentiation Between Virtual Textures**
  To what extent can users differentiate between virtual textures in non-visual conditions using the system?

  The first research question aimed to assess the extent to which sighted individuals could differentiate between virtual textures without relying on visual cues provided by the system. By exploring this question, the study sought to understand the inherent perceptual capabilities of users when utilizing the proposed system.

- **RQ 2: Effectiveness of Audio and Haptic Feedback in VE Navigation**
  How effectively can users navigate an unfamiliar virtual environment using only audio cues and haptic feedback, without any visual information?

  The second research question delved into the effectiveness of audio cues and haptic feedback as navigation aids for sighted individuals within the virtual environment.

By examining users' ability to navigate the VE solely based on these non-visual cues, the research aimed to evaluate the potential of these sensory modalities in enhancing spatial awareness and orientation.

- **RQ 3: Impact of Visual Information on Cognitive Load**
  In the absence of visual information, does navigating the virtual environment reduce cognitive load for individuals accustomed to visual cues, or does their inherent reliance on these cues increase their overall cognitive load?

  The third research question investigated the impact of visual information on the cognitive load experienced by sighted users during VE navigation. By comparing cognitive load between conditions with and without visual cues, the study sought to discern whether the absence of visual information alleviated or exacerbated cognitive demands, shedding light on the interplay between sensory modalities and cognitive processing in immersive environments.

4.1.2 Study Design Overview

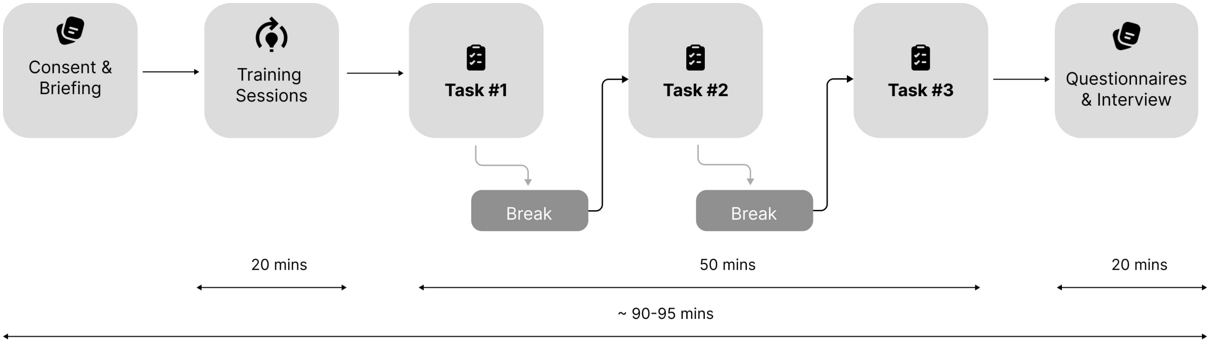


Figure 4.1: Flow diagram showing different steps of the user-study

For the study, I recruited 20 participants from the Dalhousie University community through an online mailing list. The participants underwent a training session for the first part of the study. This session focused on acquainting the participants on how to walk on the slide mill, hold the cane prototype, and effectively use it to understand how the virtual textures feel and sound. Throughout the training, they could see the VR space and no data was logged/recorded. Following this, they were asked to perform three tasks of varying complexities where they could not access any visual information and had to rely solely on their sense of hearing and touch. While the participants performed the tasks, I video-recorded their interactions, screen-recorded their in-game VE exploration performance, and scripts in the Unity GE logged essential game statistics. Once the tasks were completed, the participants completed a post-study survey comprising questionnaires to understand their objective feedback on the system's overall workload demand and usability. Following this was a post-experiment interview to derive feedback on particular aspects of the system the participant may have liked, disliked or suggested improvements on.

4.1.3 Data Collection Instruments

Unity Game Engine

I used C# scripts in the Unity GE to record the participants' game statistics throughout the game. These scripts work in tandem to capture real-time logs from the debug console using tags to categorize the type of statistic being logged. Categorically, two kinds of game statistics were logged: Game Logs and Position Logs using GAMELOG and POSLOG tags, respectively.

At the start of every iteration of the study's evaluation phase, when the participant ID was manually recorded using the game interface, the system generated and appended two files corresponding to these categories of logs. These files, denoted as `'GameLog_ParticipantID.txt'` and `'PositionData_ParticipantID.txt'`, were systematically stored within the persistent data path of the project application.

Upon initiating the logging procedure, each log entry for every task and sub-task undertaken incorporated essential metadata, including the participant ID, task number, and modality condition. This structured approach facilitated subsequent batch data analysis by enabling filtering based on these predetermined parameters. Each log entry adhered to a standardized format, presenting information in either the form of *[Tag]:[Event].info* or *[Tag]:info*, contingent upon the specific requirements of the logged event.

1. **Game logs**

   The logs captured as game logs are either manually triggered or automatic event-based. Manually triggered log entries comprise the tasks completed by participants in the first task of the evaluation phase, where each log entry correspondingly captures the time taken by the participants to identify a given texture. Automatic event-based logs are of two categories: time to complete a given part of the task or the whole task and vehicular collision with the participant's in-game avatar.

2. **Position logs** The logs captured as game logs are continuous and stream-based, where the logging is done between the start and end of each of the navigational tasks (second and third tasks) during the evaluation phase. Each log entry corresponds to a single frame of the game, which captures the position of the participant's in-game avatar 3D position for that frame. The positional values are Vector3 values in the XYZ plane of the 3D space in the game.

Observation Sheet

I used the paper-form observation sheet to record in-situ participant observations while they performed the tasks during the evaluation phase of the study. This observation sheet, detailed in Appendix E, was organized into tasks and sub-tasks to document participant behaviours and interactions within the virtual environment systematically. Specifically, the observation sheet facilitated the documentation of various aspects of participant performance, including their ability to discern textures through verbal confirmation during the first task. Furthermore, it enabled recording nuanced behavioural patterns, errors in directional decisions, navigational choices, and the identification or confirmation of checkpoints during navigational tasks devoid of visual cues.

Post-Study Questionnaires

After the end of the evaluation phase, the participants were asked to fill in a set of post-study questionnaires detailed in Appendix F. I created an online form to collect feedback for these questionnaires, which consisted of three sections, namely: demographic questionnaire, workload assessment questionnaire (NASA Task Load IndeX (NASA-TLX)) [50] and System Usability Scale (SUS) [98].

The NASA-TLX questionnaire had four versions where the questions were identical, but the context was task-based (second and third task), which was further categorized into whether each of the tasks was performed with or without any visual information. All of the versions of the NASA-TLX fall under the "workload assessment questionnaire" section and were filled out by each participant. Similarly, the SUS questionnaire had two versions where the questions were identical, but the context was based on the holistic experience of using the overall system in the presence or absence of visual information.

Interview Questions and Audio recordings

I used a paper-based interview questionnaire to guide an open-ended semi-structured interview with the participants. The primary focus was to elicit feedback on specific components of the system, their comparative and overall experience, and suggestions for improvement. The questionnaire also consisted of certain probing questions, which were used on a conditional basis per the participant's response to the main questions, detailed in Appendix G. The interview sessions were audio recorded using a desk mic (Zoom H5 Digital Multitrack recorder [143]) set to 16-bit 44.1KHz audio quality and stored in the standard WAV format.

Video and Screen Recordings

A Sony 4K camcorder mounted on a fixed tripod was used to collect the video of the participants' performance throughout the evaluation phase. The camera captured video at a resolution of 3840 x 2160 pixels and a frame rate of 30 frames per second. The tripod was set approximately 10 feet from the slide mill, and the camera was used on the widest field of view setting to accommodate the wooden platform surrounding the slide mill.

To record the in-game performance of the participants, a dedicated monitor screen was partitioned into two sub-screens where the first screen displayed the user's point of view (Unity GE camera view), and the second screen tracked the Unity GE camera in the game from up top. The aim was to merge the recordings from these camera angles with the video recording to perform ex-situ observation during the analysis phase. The entire screen was recorded using OBS studio [7], which recorded at a resolution of 1920x1080 pixels, an aspect ratio of 16:9 and a frame rate of 30 frames per second.

4.1.4 Recruitment

Participant Pool

The study population comprised people interested in exploring VR in a non-conventional way without any visual information. Previous studies in this field have recruited participants who were either visually impaired or sighted. The studies with non-sighted participants[62][64][108] had a sample size of 7, whereas the ones with sighted participants [67][94] had a sample size of 39. Since the primary objective of the study was to test the system's performance from a sighted user's perspective, the aim was to collect as much quantitative data as possible. Thus, the estimated sample size of this study was decided to be 20 sighted participants.

Inclusion/Exclusion Criteria

For the required 20 participants, the baseline inclusion criteria were as follows:

1. *Had prior experience or knowledge of VR systems*

   Participants with prior experience or knowledge of VR systems were included to ensure a foundational understanding of VR environments. This criterion was essential to facilitate efficient engagement with the study's VR components, thereby minimizing the need for extensive introductory training sessions. By including individuals familiar with VR, the study focused on exploring nuanced aspects of user interaction and perception within the virtual environment, thus enhancing the depth and validity of the findings.

2. *Had no history of VR-induced motion sickness*

   Participants without a history of VR-induced motion sickness were included to uphold the integrity of the study's experimental conditions. Motion sickness could significantly impact users' comfort and performance within VR environments, potentially confounding the interpretation of study outcomes. By excluding individuals prone to motion sickness, the study aimed to ensure that participants could engage with the VR system without experiencing discomfort or adverse physiological reactions, thereby facilitating more reliable data collection and analysis.

3. *Had no visual or hearing impairments*

   For the system in which participants navigated in the VE without using any visuals, auditory cues played the most crucial role, and auditory impairment hindered participants' ability to fully engage with the VR system. Moreover, visually impaired users were excluded to ensure the integrity of the collected data and prevent any possible form of bias. Nevertheless, individuals with general visual acuity issues such as myopia and hypermetropia could participate.

Additionally, the following set of exclusion criteria was defined to overcome practical issues related to the opportunity to participate in the study itself and refine the participant population further:

1. *Individuals with accessibility needs*

   Excluding participants with accessibility needs, such as motor and cognitive disabilities, that could not be accommodated through the user interface and head-mounted displays (HWDs) used in the study ensured that the study's technology and procedures were suitable for all participants. By excluding individuals with accessibility needs that could not be adequately addressed by the study's equipment, the aim was to maintain the integrity of the study's data collection process and minimize potential barriers to participation.

2. *Individuals who could not attend the study settings in person*

   Since the whole experimental setup was in a designated research facility that required the physical presence of the participants, all the interested individuals who couldn't make it to the study venue were excluded from the pool.

3. *Individuals with any neurological or physiological condition*

   Excluding individuals with neurological or physiological conditions, such as previous neurological illness, heart conditions, or medical conditions significantly restricting full-body mobility, ensured participant safety and data reliability. By excluding such individuals, the study mitigated potential risks associated with participation. It provided an accurate and consistent representation of the target population's experiences without undue influence from medical conditions impacting their engagement with study tasks.

Additionally, all the lab members who were directly or indirectly part of this research were excluded from the participant pool to avoid any possible conflict of interest.

Recruitment Protocol

I implemented a convenience sampling methodology to facilitate participant recruitment. Initially, a recruitment notice was sent via the Dalhousie University Computer Science undergraduate and graduate mailing list. This notice comprehensively outlined the study, encompassing its procedural aspects, eligibility criteria, data collection protocols, and compensation arrangements, while also providing instructions for contacting the principal investigator for further details. Subsequently, interested individuals were furnished with a confidential doodle poll to ascertain their availability for participation. Session invitations were dispatched on a first-come, first-served basis, contingent upon the availability provided. Furthermore, an auxiliary wait-list doodle poll was distributed to accommodate prospective participants beyond the initial capacity. This supplementary poll encompassed all ongoing sessions and offered slots outside the initially designated timeframe. When participants either rescinded their participation or failed to attend scheduled sessions, individuals from the wait list were invited to participate in sequential order of submission. These replacements were executed either in the vacant slot or in slots after the original scheduled times, as per availability.

4.1.5 Study Preparation

The sole paper-based documents required for the study comprised consent forms, observation sheets, the user study script, the interview questionnaire, and the participation compensation receipt. These documents were printed as needed, corresponding to the number of participant slots booked for each day. In total, 20 copies of the consent forms, observation sheets, and compensation receipts were printed (one for each participant), along with two sets of user study scripts and interview questionnaires.

The room for the study matched the layout as depicted in Figure 4.2. The room was divided into four primary sections: workstations, participant section, MoCap tracking section, and performance recording section. The workstations comprised a post-study station, the researcher control hub, and the Qualysis MoCap control station.

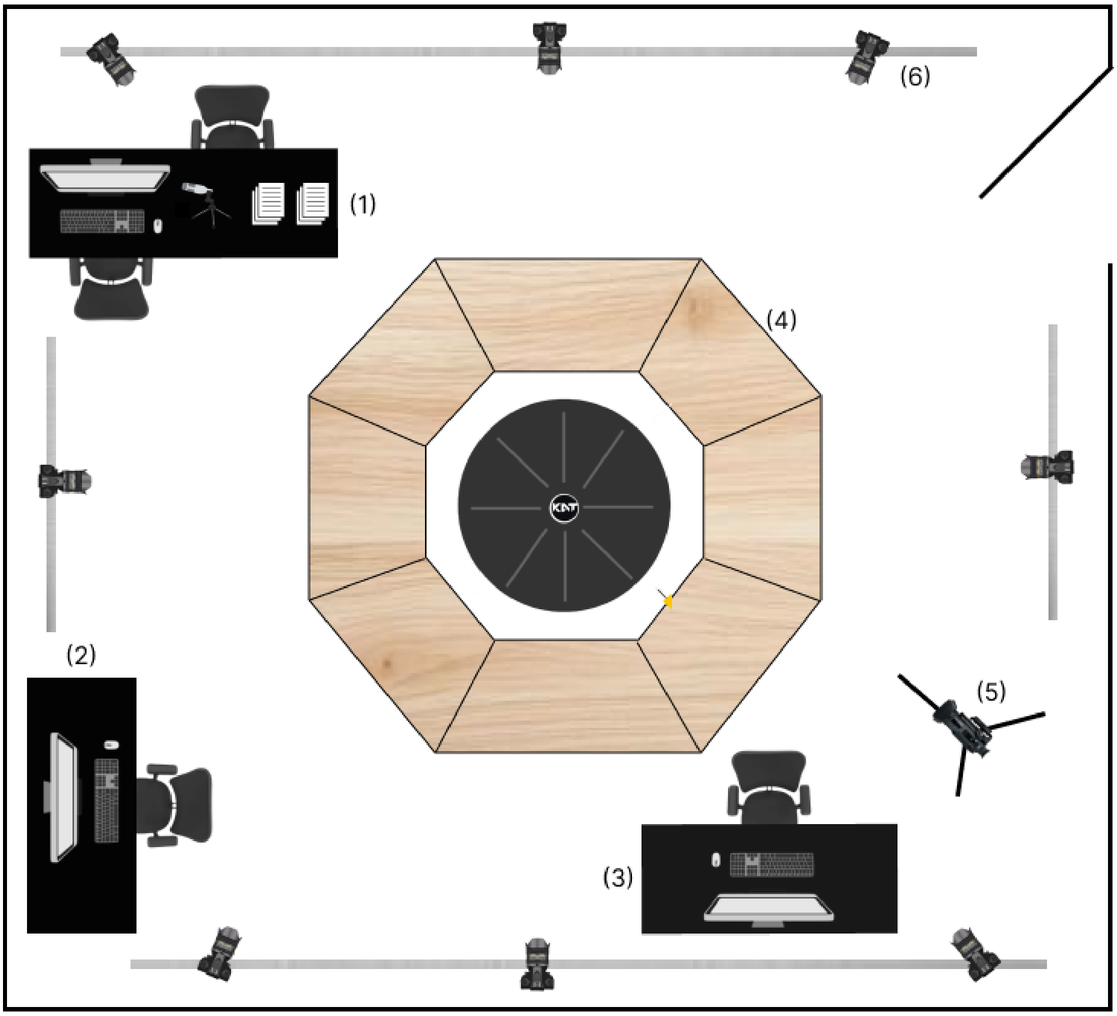


Figure 4.2: Room layout for the study. (1)Post-Study Station, (2)Researcher control hub, (3)Qualysyis Mocap Station, (4)Participant section, (5)Performance recording section, (6)Mocap tracking section

The post-study station was where the participants were briefed and introduced to the study and, later on, where they filled out the questionnaires. The Qualysis MoCap control station and the researcher control hub were solely designated for use by the researchers to control the overall study tasks and calibrate the MoCap bindings of the system, respectively. The entirety of the training and evaluation was performed in the participant section, which consisted of the slide mill and the skirt surrounding it. The tracking section consisted of 8 MoCap IR cameras mounted on metallic turf beams across the room's periphery. The performance recording section of the room was a dedicated fixed place for the tripod and the camera to record the participants' performance.

4.1.6 Study Procedure

Participant Arrival, Briefing and Consent

Participants were instructed to arrive outside the lab building five minutes before the scheduled start time as the main entrance was locked and access restricted. Once the participants arrived, they were greeted and directed into a chair inside the lab room. I then welcomed and briefed the participants on the study's rationale, what they would be asked to do, the consent process, the data collection process, and the pro-rated compensation. Lastly, participants were asked if they had any questions. Once satisfied, the participants were directed to read and sign the consent form.

Slide mill setup

Once done with the briefing and informed consent, the participants were directed to safely get on the slide mill by carefully stepping on the slide mill skirt and simultaneously holding the slide mill's backrest. They were then firmly strapped and buckled into it using the waist and thigh straps that came with the slide mill setup. I was responsible for the whole procedure and ensured it was followed stringently and per the safety protocol outlined by the slide mill manufacturing company to ensure maximum safety and evade any possible injury.

Slide mill and Cane Calibration

Once the participants were strapped into the slide mill contraption, the rotation of the backrest was briefly locked for the calibration of the setup. They were given the Oculus HWD to be worn and the cane prototype to be held in their dominant hand. They were then asked to stand upright and look straight for a few seconds to calibrate the slide mill according to the position of the HWD and the orientation of the backrest. Once completed, the participants were instructed to firmly hold the cane in front of them such that the tip of the cane rested on the skirt. At this point, the MoCap tracker bindings were verified so that all the markers on the cane were visible to the IR cameras. Moreover,

Figure 4.3: Slide mill setup with the wooden skirt

the virtual cane mapped to the cane prototype was manually calibrated in the Unity GE to match the tip of the virtual cane touching the surface plane in the game to the tip of the real cane touching the skirt surface. This was necessary due to the participants' variable heights, arm lengths, and their normal stance of holding the cane at rest.

Training Session

After the calibration, the first session of the study was the training session, where the main objective was to introduce them to using the aspects of the system they were unfamiliar with, namely, the slide mill and the cane prototype. During this session, they could visualize everything in VR and no data was recorded, as it was purely meant for them to acclimate to the system's settings. The entire training session lasted an average of 15-20 minutes and was divided into the following phases.

1. **Phase 1: Slide Mill usage**

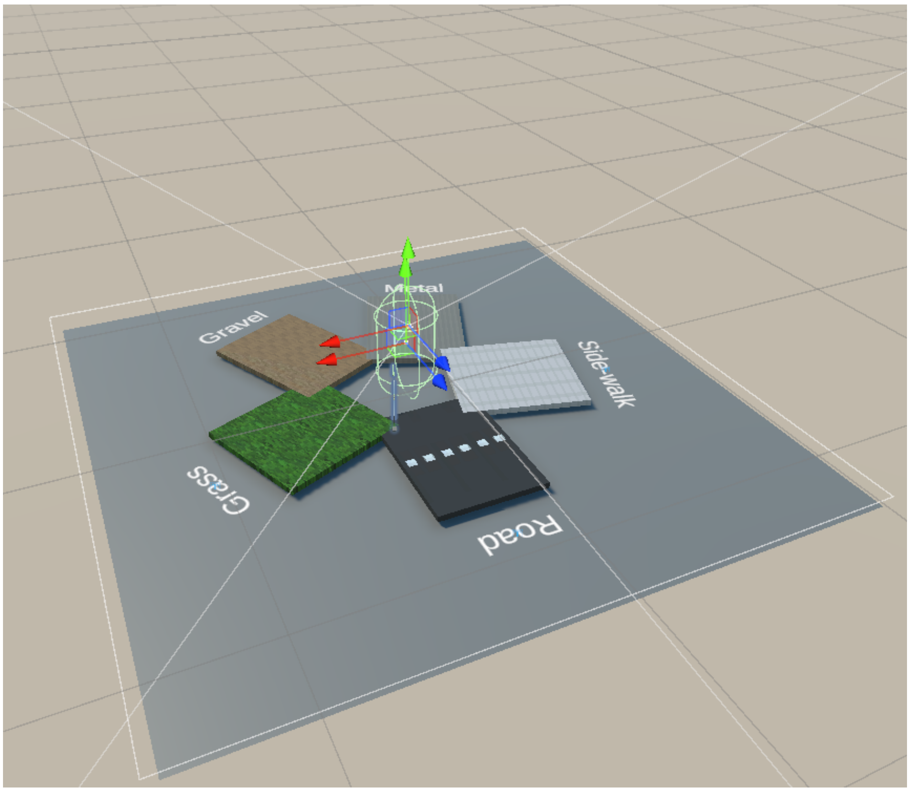


Figure 4.4: Training Session Scene in Unity

Due to almost all participants' lack of prior experience in using the slide mill, it was essential for them to get acquainted with using it in the game. This session lasted roughly ten minutes, or whenever they felt ready, whatever came first.

The participants were instructed to try the whole thing in a way that made them feel comfortable. Afterwards, they were given instructions on the most effective way to navigate in the VE regarding the stance, the stride height and length, and the optimal speed to replicate natural walking speed without having to try very hard or get fatigued easily. Most importantly, I instructed them to lean a bit forward so that the center of gravity of their body is not directly on top of their feet and not to take high strides while maintaining a tip-toe movement to avoid exerting exorbitant effort.

In the game, they spawned in a scene that placed them in the middle of the VR city simulation, the same as the one used for the training session, but without any ambient environmental sound.

2. **Phase 2: Cane usage**

In this training phase, the participants held the cane in their dominant hand with the exact grip as when holding an actual white cane. They were instructed to hold the cane before them, with the tip pointing downwards and touching the skirt.

They were trained on using the cane so that the haptic feedback emitted was maximum perceivable. Primarily, I instructed them not to hold the cane too tight but firm enough so they wouldn't lose hold of it. Moreover, it was about the movement of the cane and the speed of swiping the cane. I provided an analogy of using the cane like a paintbrush where they were supposed to use the cane more like an extension of their arm and try to make sense of the vibrations while maintaining a constant normal sweeping motion and continuous firm contact with the surface of the skirt.

In the game, they spawned in a scene surrounded by five different virtual blocks of surfaces, each reachable using the virtual cane mapped to the cane prototype (Figure 4.4). The slide mill navigation was disabled for the scene as it was supposed to be a stationary task, but they could rotate freely. They were asked to feel the vibrations, hear the emitted sound one surface at a time, and familiarize themselves with each virtual texture's haptic and auditory profiles. This session lasted roughly ten minutes, or whenever they felt ready, whatever came first.

Evaluation Session

After the training session, the participants underwent the evaluation in three parts, each corresponding to a task. Before each task, the participants were briefed about what they had to accomplish to complete them and, in between tasks, they were provided with mandatory and voluntary breaks. The evaluation began with helping the participants put on the blindfold, the HWD on top, and the headphones, followed by handing over the cane. In the task description below, the term 'without visuals' corresponds to tasks done with the blindfold on, while 'with visuals' corresponds to tasks done with the blindfold off.

The overview of the tasks, along with details of the sensory involvement and complexity level, is outlined in Table 4.1 along with the index of the acronyms used in the table.

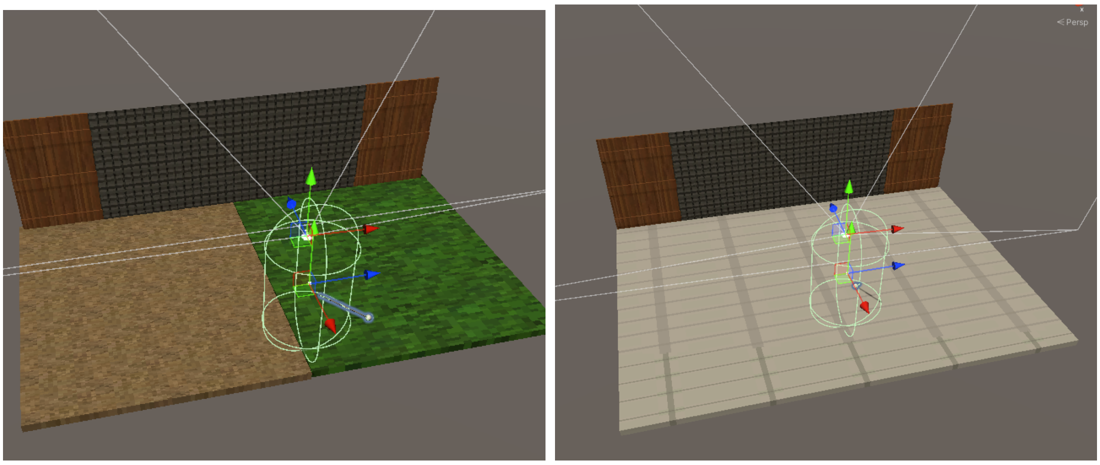
1. **Task 1: Surface Identification**

Table 4.1: Task categories in the study

Task	Category	Level	Description	No. of Reps	Sensory involvement	Complexity
Task 1 Surface Recognition	NV	1	T1.1	5	• Hearing: **Y (IS)/N** • Haptic: **Y (1-2 surfaces at a time)/N** • Visual: **N** • WBM: **N (Only one arm)** • Time to Reach (TTR) count: **NA**	Low
			T1.2			
			T2.3			
Task 2 Road Crossing	NV	2	T2.1	1	• Visual: **N/Y** • Hearing: **Y (IS+OS+SS)** • Haptic: **Y (Multiple Surfaces)** • WBM: **Y (Single DOM, 1-2 turns)** • TTR count: **1-2**	Medium
	NV		T2.2			
	WV		T2.2			
Task 3 Scavenger hunt	NV	3	T3.1	1	• Visual: **N/Y** • Hearing: **Y (LS+NLS+OS+IS+AS+SS)** • Haptic: **Y (Multiple Surfaces)** • WBM: **Y (Multiple DOMs, 2-5 turns)** • TTR count: **2-5**	High
	NV		T3.2			
	WV		T3.2			

NV: No Visuals, WV: With Visuals, WBM: Whole Body Movement,
LS: Localized Speech, NLS: Non-Localized Speech, OS: Object Sounds,
IS: Interaction Sound, AS: Ambient Sound, SS: System Sound,
TTR: Target To Reach, DOM: Direction of Movement

Participants were instructed to perform a series of tasks involving texture perception using the virtual cane interface and verbally confirm the name of the perceived texture. Each task comprised specific sub-tasks to evaluate the participants' ability to perceive textures through a permuted combination of auditory and tactile cues. The possible textures included grass, sidewalk, asphalt (road), gravel, and metal (tactile paving). Since this task was to be performed while stationary, the rotation and navigation function of the slide mill was disabled.



(a) (b)

Figure 4.5: Snapshots of Task 1 Scene in Unity (a) Task 1.1 (NV) (b) Task 1.2

Task 1.1

In this sub-task, participants could hear and touch the texture as their cane moved across the virtual surface. Two surfaces were on either side, and they were instructed to feel their right and left sides using the cane to identify the perceived texture. This sub-task was repeated five times, each with a pair of different textures from the list of possible textures.

Task 1.2

In this sub-task, participants were instructed to identify textures solely based on auditory feedback and without haptics while the virtual cane made contact and moved across the virtual surfaces. This sub-task was conducted five times, with participants encountering a single texture from the aforementioned list during each iteration.

Task 1.3

In this sub-task, participants were instructed to identify textures solely based on haptic feedback and without audio while the virtual cane made contact and moved across the virtual surfaces. This sub-task was conducted five times, with participants encountering a single texture from the aforementioned list during each iteration.

2. **Task 2: Road Crossing**

   For the second task, participants were placed in a city simulation with the primary goal of crossing the road. They were required to utilize immersive spatial audio of the VE and textural perception conveyed through their cane. This task included two sub-tasks, with the second sub-task having two variants: one With-Visual (WV) aids and one with Non-Visual (NV).

(a) (b)

Figure 4.6: Snapshots of Task 2 Scene in Unity (a) User's Point of View (b) World view of user navigation

Task 2.1

In this sub-task, participants were spawned near a road with a pedestrian crossing (traffic light) a few feet before them. Their task was to reach the pedestrian crossing and cross the road upon determining it was appropriate to cross it. The cue to cross the road simulated real-life scenarios, including changes in traffic lights and vehicle sounds, indicating it was safe to cross. Upon successfully reaching the other side, participants received a task completion notification signalling the successful completion. When it was observed that the participants lost track and reached a state where they couldn't recover to finish the task, I manually stopped the task and notified them about it.

Task 2.2

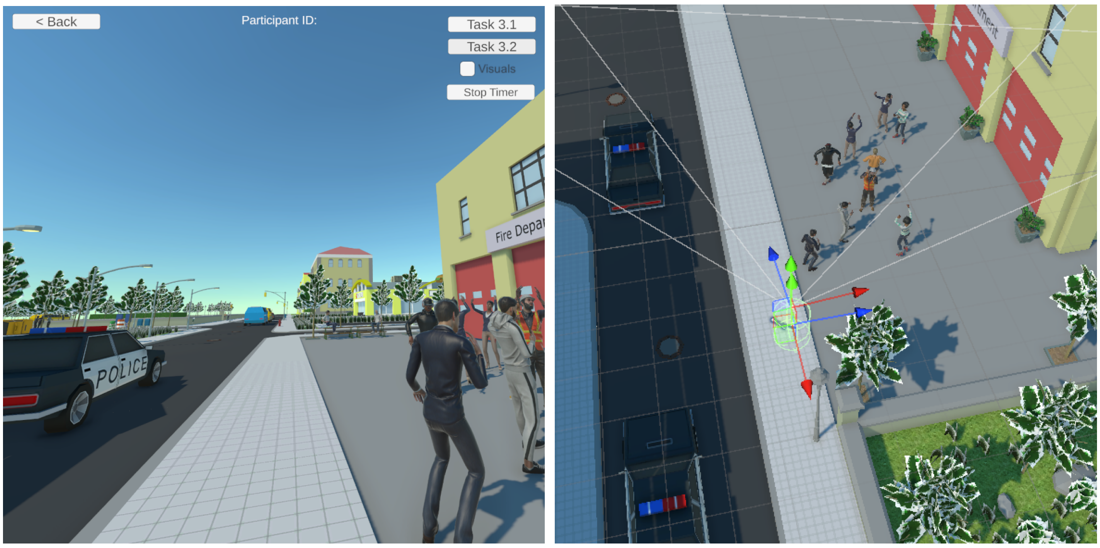
In this sub-task, participants were spawned between two pedestrian crossings; one was to their left and the other to their right. The difference, however, was in the distance from their spawn point, meaning one was nearby and the other farther away. The difference in distance between the crossings and their spawn point was marginal and more discernable through the difference in the sound intensity emitted by the crossings rather than their location being perceived visually.

Thus, based on the sounds emitted from both crossings, participants were required to judge the direction of the farthest crossing and navigate towards it to cross the road safely. The objective and completion criteria were similar to Task 2.1; however, the first variant of this task was performed without any visuals, while the second variant followed with the help of visuals.

3. **Task 3: Scavenger Hunt**

For the third task, participants remained in the city simulation, and their primary goal was to complete a scavenger hunt. The scavenger hunt was for specific auditory cues called targets. These targets were classified as intermediate and final/destination targets, where the intermediate targets paved the way for the final target to be located. Participants had to use the immersive spatial audio of the VE by listening for auditory cues, such as pedestrian conversations or ambient noises, and textural perception conveyed through their cane. Upon reaching the final target, they would hear a completion sound, and then they had to search for a specific texture near it. Road crossings were also involved, requiring participants to navigate through them safely. This task included two sub-tasks, each with two variants: one with visual aids WV and one without NV.

The requirement of 'locating' a target comprised the participants verbally confirming that they were nearby by explaining the direction of the target and what the target sounded like. Once they could locate a specific target, I provided them with directional hints as to where they could expect the next target to be from their current location. At times, when the participants lost track of their sense of direction or location, I assisted them with clues to where they could navigate to get back on track.



(a) (b)

Figure 4.7: Snapshots of Task 3 Scene in Unity (a) User's Point of View (b) World view of user navigation

Task 3.1

In this sub-task, participants were spawned near a pedestrian crossing. The first intermediate target was a construction site located across the road. The second intermediate target was a bus stand to the right of the construction site, which they had to locate using conversational cues. Participants had to listen to conversations from the bus stand to find the final target, a hotdog vendor, and identify the texture of grass near it.

Task 3.2

In this sub-task, participants started in front of a lawn, where the first intermediate target was a local group of people spectating a rap and hip-hop dance battle. The second intermediate target was a nearby traffic light crossing, which participants needed to cross safely. The third intermediate target was to find a group of children playing in a park, the entrance to which they had to find by listening to conversations. The fourth intermediate target was a hotdog stand; next to it was the final

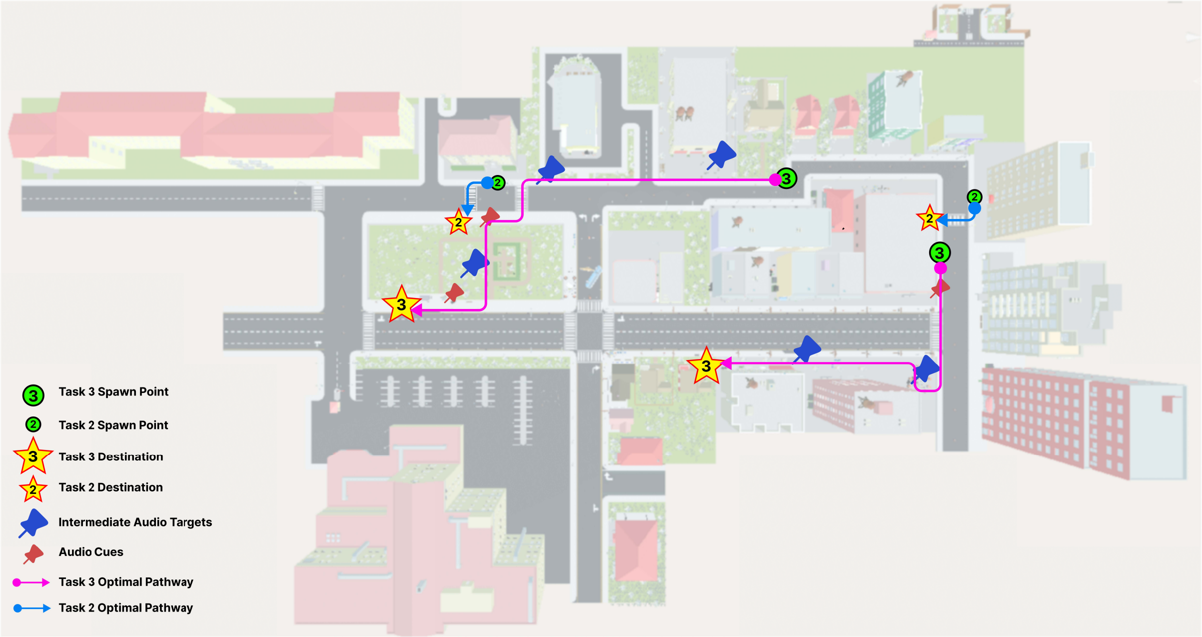


Figure 4.8: Top View of the city map with task routes, start and end points, and targets

target, a bus stop, and they had to identify the gravel texture near it. The objective and completion criteria were similar to Task 3.1; however, the first variant of this task was performed without any visuals, while the second variant followed with the help of visuals.

Interview

Once seated, the participants were briefed on the interview process, and I indicated that I would be using the microphone to audio record the interview. I asked the participants if they were ready, and when confirmed, they started the interview recording.

I started with the first question and asked each in order unless the participant strongly tended toward a different question during their explanation. Priority was keeping the conversation fluid while following the order of the questions was second. If interesting points occurred, I prompted for an additional explanation; I asked the listed questions primarily.

Once the interview was nearing completion, I stopped the audio recording if the participant had no other feedback to add. Once stopped, I provided the participants with the compensation and the receipt to be signed.

Closing

Once the participants completed the post-session questionnaire, they were compensated and thanked for participating in the study. I then led the participant outside of the lab.

Preparation for Next Participant

After the session, I transferred the video recording from the camera and the interview recording from the microphone via a direct USB cable connection to the main PC. These recordings were saved in a participant-specific folder for easy access and retrieval. Concurrently, I transferred all the log files and screen recordings of the participants' in-game performance to this folder. While handling the files, I scanned the observation sheets using my phone and transferred them to the same folder. Once all the files were transferred, I verified each to ensure they weren't corrupt. Once verified, the original data from the data collection instruments were deleted. The whole folder was also copied to another external SSD for a second backup.

After handling the participant-specific data, I inspected the base plate of the slide mill for any accumulated dirt and cleaned it. This step was necessary as a dirty base plate of the slide mill can trick the sensor into recording a false reading. Moreover, I collected the consent form and compensation receipt filled by the previous participant, securely placed it into my locker, and set up the post-study station with the consent form, observation sheet and compensation receipt for the next participant.

4.2 Data Analysis Structure

4.2.1 Post-Study Questionnaires

The participant data from the post-experiment questionnaire were first summarized to provide a clear overview of the study's sample population. This initial step helped to outline the demographic and experiential makeup of the participants. The next step involved applying the Shapiro-Wilk normality [104] test to the responses from the System Usability Scale (SUS) [98] and NASA-TLX [50] included in the post-trial questionnaires. This test was chosen to determine if the data distributions adhered to a normal distribution, a key factor in deciding whether to use parametric evaluation methods.

For data sets that showed normal distribution, I used an Analysis of Variance (ANOVA) to identify any statistically significant differences in SUS scores under different conditions. This analysis was critical for assessing the effect of visual aids on the usability of the system by comparing responses from various visual settings. In instances where the data did not meet the normality assumption, I employed the Kruskal-Wallis Test [66]. This non-parametric test was essential for examining the NASA-TLX scores, which span several parameters such as Mental, Physical, Temporal, Performance, Effort, and Frustration. It allowed for the identification of significant differences between the conditions, providing insights into how visual aids influence various aspects of task load.

4.2.2 Logged Game and Positional Data

The data logged for each participant through the Unity Game Engine comprised of in-game statistics and 3D positional data as explained in Section 4.1.3 (*Unity Game Engine*). The in-game statistics consisted of task-specific logs focusing on performance metrics, whereas positional data focused on the participant's in-game coordinates during navigation tasks.

Logged Game Data

The data or the first task was analyzed through the visualization of the participant's ability to identify a given virtual surface using a confusion matrix. This was repeated for all three variations of this task. This gave an idea of the overall success rate of textural identification for each of the variations of the task. Furthermore, I performed a factorial analysis using a 2-way ANOVA where the presence and/or absence of audio and haptics (sensory modality) along with the Texture of the virtual surface were treated as the independent variables, whereas the success rate of identification was treated as the dependent variable. This statistical analysis helped us understand the significance of each of these variables in the overall success rate and how the interactions between them dictate the successful identification of a given virtual surface.

For the second and third tasks, the data was analyzed in two ways: statistical tests and interpretation through descriptive statistics. For each task, firstly, the normality of the data, which was the task completion time for Task 2 and the Time to Reach (TTR) intermediate targets and destination for Task 3, was assessed through the Shapiro-Wilk

Normality test. Secondly, depending on whether the data was normally distributed or not, parametric (one-way ANOVA) and non-parametric tests (Related-Samples Wilcoxon Signed Rank [132]) were used to assess the statistical difference under the two conditions of visual availability. Moreover, the in-game statistics, such as vehicular collision tr, target location error rate, etc, were analyzed by contrasting and comparing descriptive statistics such as mean and standard deviation.

Positional Data

The positional data for the third task was analyzed through a custom C# script that provided three computational metrics: the closest average distance of participants from an intermediate target, the average time spent by participants within the audible radius of the target and the average Unity Frames per participant within the audible radius. Moreover, it also rendered the navigational paths of each participant from the spawn point to the destination. The data from the metrics helped to analyze the difference in the value of these metrics when the same task was performed under varied visual availability conditions. Moreover, the path rendering helped analyze the navigation patterns and visualize the difference in the path taken by participants with respect to interaction with audio targets.

4.2.3 In-situ and Ex-situ Observations

In the study, observational data was collected and analyzed through both in-situ and ex-situ observations, complemented by interview data to create a comprehensive understanding of participant behaviour in the VE. In-situ observations were conducted live, using observation sheets to record direct responses as participants navigated through tasks. Ex-situ observations were conducted by going through video recordings as a post hoc measure to reaffirm consistencies and identify any possible biases or discrepancies during the on-site observation. All in all, the mix of both of these observations primarily assisted in identifying the following:

- **Navigational Errors**: Recording any navigation errors encountered by participants.
- **Behavioral Observations**: Monitoring how participants used their cane, standing and walking postures, and overall movement patterns.

- **Sensory and Interaction Details**: Assessing participants' abilities to differentiate between various virtual surfaces and noting any recurring navigational challenges.

- **Emotional and Cognitive Responses**: Observing expressions of distress or frustration and documenting participants' thought processes and decision-making strategies.

After completing these tasks, ex-situ observations were conducted through video and screen recordings of participants' performances, analyzing their actions beyond the immediacy of task execution. These observations focused on:

- **Analysis of Cane Movements**: Identifying patterns and themes in how participants managed their canes, providing insights into their adaptation of techniques and approaches.

- **Directional Decisions and Audio Localization**: Examining how participants used auditory cues to make navigational decisions and their ability to pinpoint sound sources within the VE.

Comparative and thematic analyses were conducted that provided insights into the adaptive behaviours and strategies developed by participants in response to challenges in the VE. Interview data focusing on auditory cues and decision-making processes were integrated with observational findings, enriching the contextual understanding and ensuring robust analysis.

Audio sources recalled by participants during navigational tasks were labelled as tags—such as road sounds, traffic sounds, people talking, etc—and classified under distinct audio types in the VE. This process of tagging and categorization allowed for a structured analysis of how different audio cues influenced navigational decisions. A quantitative analysis followed, documenting the frequency of each audio tag alongside the participants who reported them. This data provided insights into which audio cues were most commonly recalled and potentially the most influential in participants' navigation. Furthermore, these auditory tags were integrated with tactile cues and behavioural observations recorded during the tasks, creating a comprehensive dataset. This integration was pivotal in uncovering key strategies participants employed to navigate, particularly in scenarios without visuals.

Integrating qualitative and quantitative data provided nuanced insights into behavioural patterns and strategic developments across different sensory and interaction contexts. It highlighted the interplay of sound and touch, demonstrating the essential navigational strategies developed by participants in response to the multi-sensory inputs of the VE.

4.2.4 Interviews

To analyze the interviews, the audio files first had to be converted into text-based transcripts. These transcripts were automatically generated using an online tool - Revolvdiv [92]. This tool automatically labelled the parts of the transcripts according to the number of speakers; however, this process had some issues at times, which had to be manually corrected. The labels in the transcripts were further converted into “R: ” for the Researcher or “P: ” for the Participant to maintain clarity, and a few misinterpreted words were first corrected using Microsoft Word [78]. The transcripts were then loaded to Nvivo [89] for coding.

After reviewing the data, I constructed three categories of codes in Nvivo for the first round of organizing the data. The first category comprised codes derived from the interview questions; the second included codes that facilitated the comparison of design elements; and the third consisted of emergent codes that did not fit within the previous categories. In all, there were 25 distinct codes. Although these codes were not directly linked to specific research questions, they enabled comparisons across different conditions, helping to identify similarities and differences and ensuring that relevant information within the data was preserved. I then parsed through each of the transcripts and tagged segments of raw text that shared a resemblance to any of the three categories or the set of codes. After completing this process for all the participant interview transcripts, I exported the code book from the Nvivo application and opened it in MS Word. This step was performed as the exported codebook is systematically displayed in a table when exported which is easier to view while analyzing the codes.

In the thematic analysis process, I utilized an online tool called Miro [80] to create an affinity diagram for organizing and refining themes derived from the interview transcripts as shown in Figure 4.9. Starting with the raw texts from the code book, the initial step

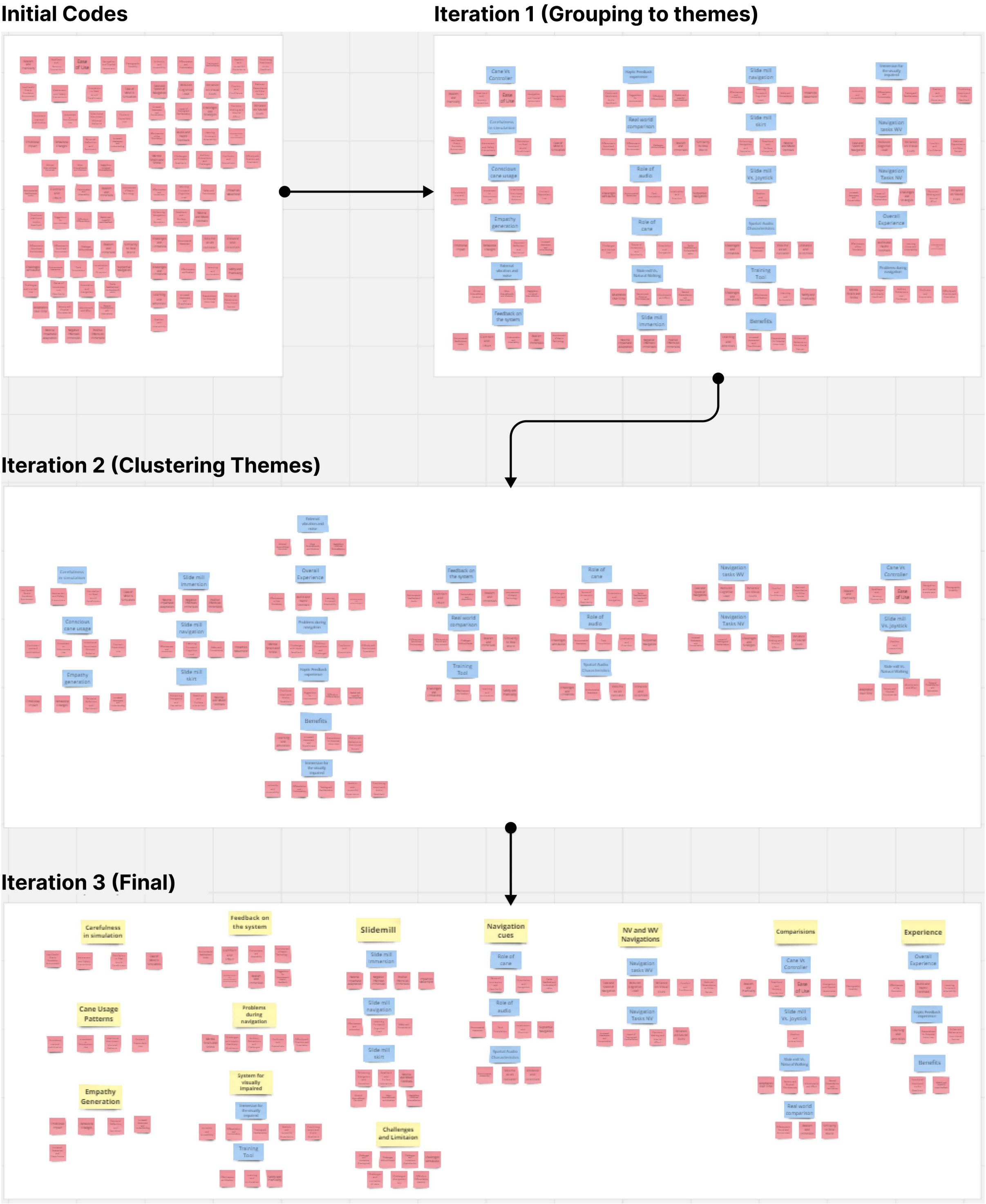


Figure 4.9: Affinity Diagram Iterations

involved transferring the coded data onto individual sticky notes within Miro. Each sticky note contained a specific code representing ideas, phrases, or concepts articulated by participants, labelled to align directly with the raw textual data.

The first iteration of the affinity diagram involved arranging these sticky notes into broader thematic categories. This grouping was driven by the inherent relationships or affinities between codes—where codes that shared similar concepts or topics were clustered together. This phase required iterative adjustments to ensure that each grouping accurately represented a coherent theme capable of encapsulating the essence of the combined codes. The second iteration was more focused on refining these broader categories into more distinct and polished themes. This process involved examination of each category, ensuring that there was internal consistency and that each was clearly distinguishable from others. During this stage, some categories underwent merging, splitting, or further refinement based on deeper insights into the data. Finally, the third iteration aimed at finalizing the thematic clusters and clearly defining them. Each cluster represented a distinct theme that emerged distinctly from the analysis, and these themes were further elaborated with sub-themes that provided detailed insights into specific aspects of the participant responses. The thematic analysis ultimately resulted in the interpretation of interview data into key themes and sub-themes, ensuring coherence and guiding the reader through a logical progression of the findings reported by the users of the system.

4.3 Summary

In summary, this section provides an overview of how the user study was conducted to evaluate the VR system. It begins with explaining how participants were recruited, ensuring a representative sample while avoiding conflicts of interest. Then, it describes the preparation of the study environment and the procedures followed during the study sessions, including participant briefing, system calibration, training, evaluation, and interviews. Finally, it outlines the structure of data analysis, which includes questionnaires, logged data, observations, and interview transcripts, all aimed at understanding participants' navigation skills and their experiences with the VR system.

Chapter 5

Results

5.1 Quantitative Questionnaire Results

5.1.1 Sample Population Demographics

Gender and Age

The sample population of 20 participants had a distribution of 45% female (9/20) and 55% male (11/20). Given that I was not controlling the gender within the recruitment strategy, I achieved a relatively balanced gender distribution, and this near-equal gender distribution provided a balanced perspective on the characteristics being studied.

Moreover, the study included participants aged 18 to 44 years old. The majority fell within the 25-34 age bracket, comprising 65% of the sample, followed by the 18-24 age group at 30% and a smaller representation from the 35-44 age group at 5%.

Visual Acuity

A significant portion of participants reported having no visual acuity (12 out of 20, 60%). Among those who did report some form of visual acuity, myopia was the most common condition (5 participants, 25%), followed by hypermetropia (1 participant, 5%). Two participants preferred not to disclose their vision status. This distribution suggests that while most of the sample does not experience vision issues, a significant minority does; however, for participants with vision issues, navigating the VE for tasks requiring visual interaction was not significantly challenging when using the HWD, as they were able to wear their prescribed glasses throughout the activities.
A significant portion of participants reported having no visual acuity (12 out of 20, 60%). Among those who did report some form of visual acuity, myopia was the most common condition (5 participants, 25%), followed by hypermetropia (1 participant, 5%). Two participants preferred not to disclose their vision status. This distribution suggests that while most of the sample does not experience vision issues, a significant minority does; however, for

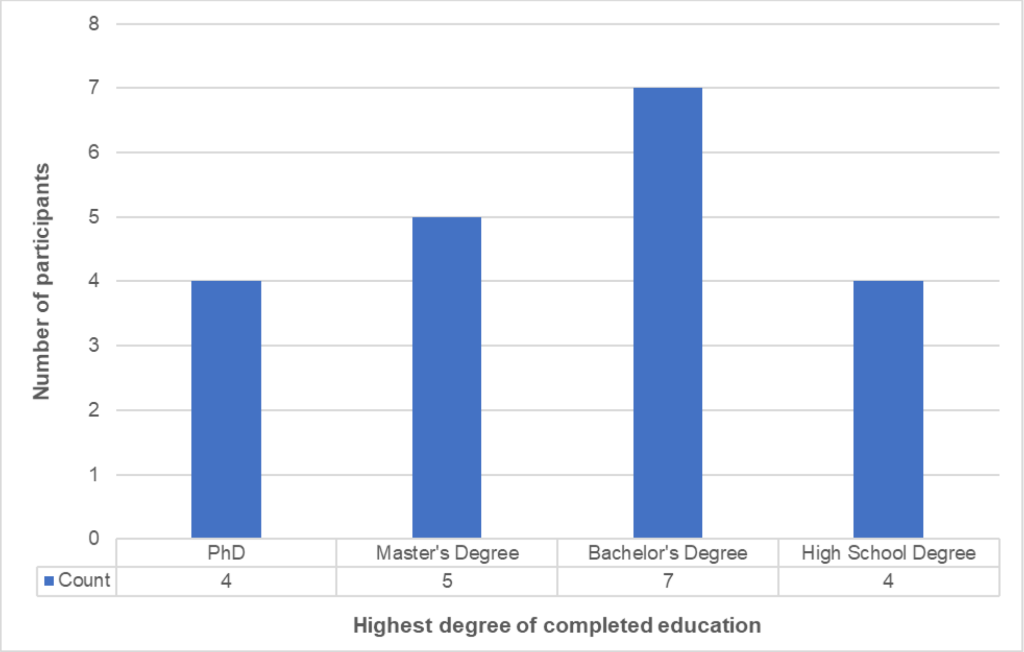


Figure 5.1: Education level of the sample population

participants with vision issues, navigating the VE for tasks requiring visual interaction was not significantly challenging when using the HWD, as they were able to wear their prescribed glasses throughout the activities.

Orientation and Mobility Training

As discussed in Chapter 2, Section 2.2.4, O&M training is an essential and standard learning practice for the visually impaired and their instructors to possess. One of the demographic questionnaires asked the participants if they had any level of exposure to this training; however, all the participants reported that they didn't have any prior knowledge of O&M training.

Education

As the recruitment strategy targeted university students, there was a mixture of undergraduate and graduate-level university students. Since the question targeted the highest degree completed by the participants, it can be derived that 45% (4/20) are graduate students with 4 Ph.D. and 5 Master's degrees, 35% (7/20) are undergraduate students, and 20% (4/20) are high-school students. This statistic is illustrated in Figure 5.1.

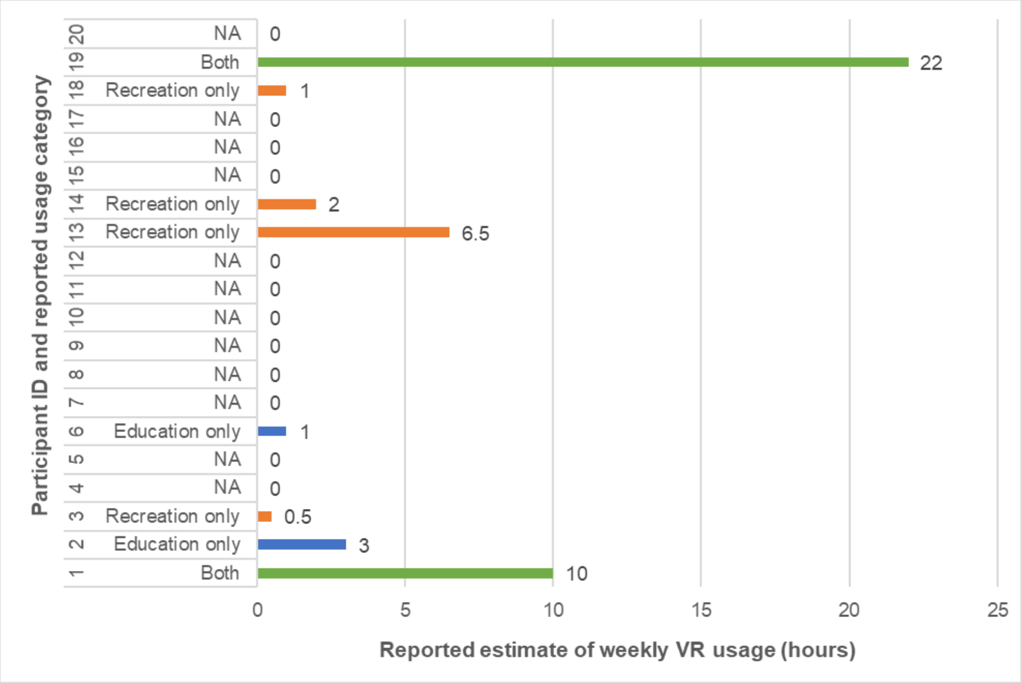


Figure 5.2: Self-reported VR usage and estimates by participants

Prior Experience with VR

Given the question if the participants had any prior experience with any form of VR technology and, if yes, what they use it for and an approximate estimate of time they often spend using the technology, the distribution of the results is illustrated in Figure 5.2.

Based on their responses, the prominent usage category can be categorized into Recreation, Education, Both and Not Applicable (NA). The 'Recreation' category constitutes answers such as *fun, video games, gaming, leisure and entertainment* that revolve around the usage of VR for mostly recreational purposes. Similarly, the 'Education' category constitutes answers such as *research, university, courses* while the 'Both' category constitutes answers that stress the VR usage for both recreation and educational purposes. Moreover, the NA is the usage category for participants who reported having no VR technology experience, and their corresponding weekly usage was automatically equated to zero. The weekly hourly usage of the participants in Figure 5.2 is an average of their weekly usage estimate range.

5.1.2 SUS Results

Descriptive statistics were initially provided, summarizing the mean SUS scores and standard deviations obtained from the study. Specifically, it was found that the mean SUS score for the system without visuals (SUS NV) was 62.63, with a standard deviation (SD) of 16.11. Conversely, the mean SUS score for the system with visuals (SUS WV) was 75.63, with a slightly higher SD of 17.432. These values place the system's usability with the visuals in the 'acceptable category' with an adjective rating of 'good', whereas the system with no visuals in the 'marginal category' with an adjective rating of 'okay'. These ratings [31] offer insights into the perceived usability of the system under these two conditions and conclude that the overall usability reported was higher when the system was used with the availability of visuals depicted in Figure 5.3.

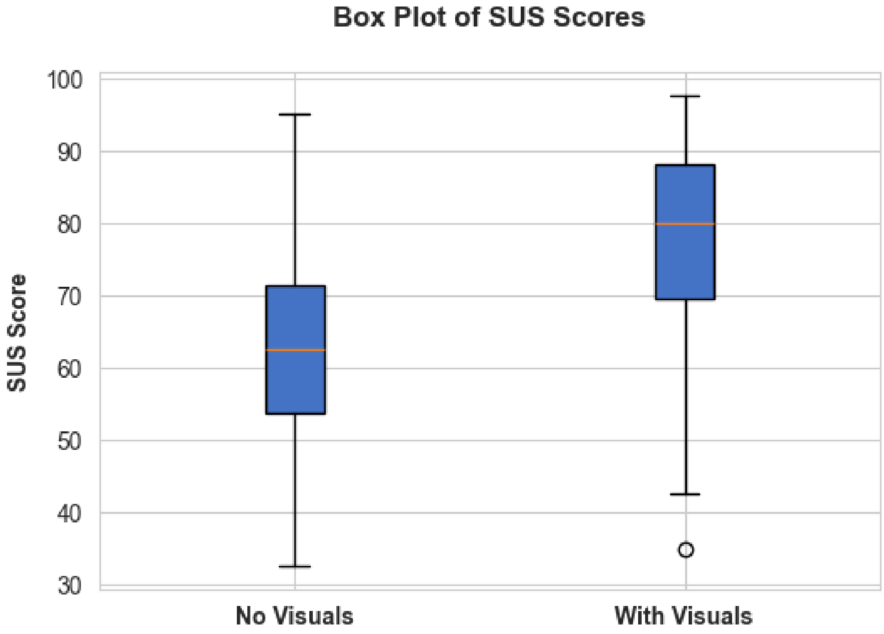


Figure 5.3: SUS score box plots of the system when used with and without visuals

Following the summary of descriptive statistics, the normality of the SUS scores was assessed using the Shapiro-Wilk test. The test yielded the result with the test statistic and corresponding p=value as W(40) = 0.963, p = 0.215. This outcome suggests that the SUS scores may be reasonably assumed to approximate a normal distribution, as the p-value exceeds the conventional significance threshold of 0.05.

Subsequently, a one-way ANOVA (α=0.05) test was performed to examine potential differences in SUS scores between the two conditions (SUS NV and SUS WV). An F-statistic of 5.999 was obtained, with associated degrees of freedom (1, 38) and a p-value of 0.019, i.e. F(1,38)=5.999, p=0.019. This result indicates a statistically significant difference in SUS scores between the two conditions at the chosen significance level. In other words, the presence of visuals appears to have a discernible impact on the perceived usability of the system.

5.1.3 NASA-TLX Results

In analyzing the NASA-TLX ratings for both the navigational tasks completed under two conditions, namely with visuals (WV) and no visuals (NV), the first step involved assessing the normality of the data. This was accomplished by applying the Shapiro-Wilk test to each of the NASA-TLX parameter ratings across both conditions (NV and WV). The normality test is encapsulated in Table 5.1. The test revealed that across all parameters, namely, *Mental, Physical, Temporal, Performance, Effort, and Frustration*, the p-values were below the conventional significance threshold of α=0.05. Consequently, the data was deemed non-parametric, indicating a departure from normal distribution.

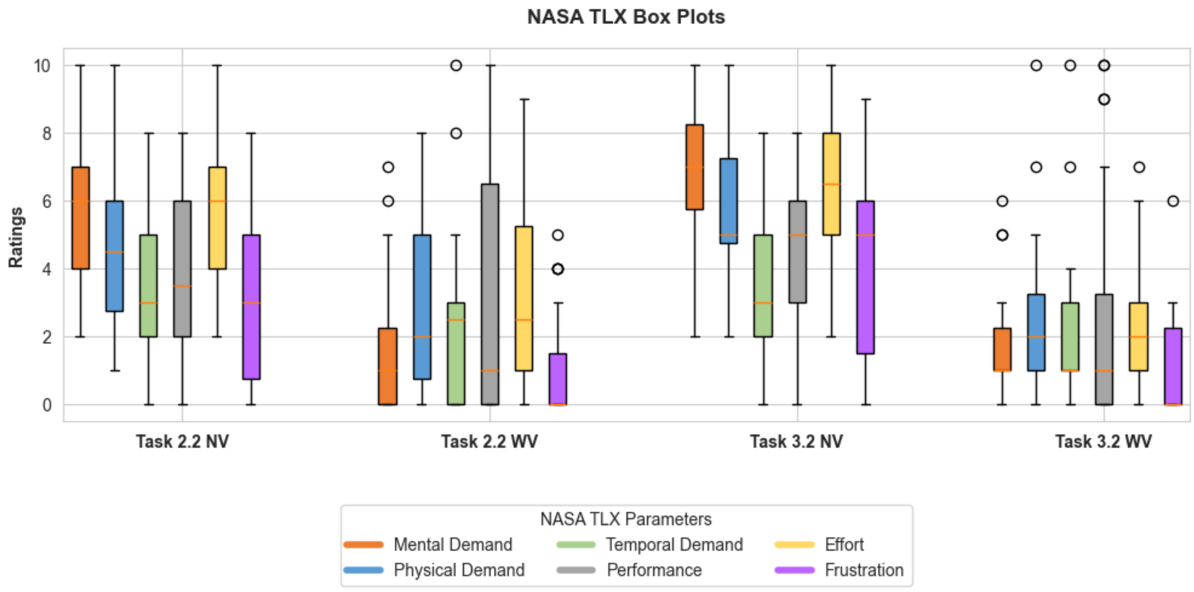


Figure 5.4: NASA-TLX Box Plots for tasks done with and without visuals

Table 5.1: NASA-TLX Normality Test Summary using Shapiro-Wilk Test

Parameter	**Test Statistic (W)**	**p-value**	**Conclusion**
Mental	0.166	$< .001$	Non-Parametric
Physical	0.172	$< .001$	Non-Parametric
Temporal	0.167	$< .001$	Non-Parametric
Performance	0.186	$< .001$	Non-Parametric
Effort	0.136	0.003	Non-Parametric
Frustration	0.225	$< .001$	Non-Parametric

Subsequently, the Kruskal-Wallis test was employed to discern if a statistically significant disparity existed between the ratings for the two conditions. The results, as presented in Table 5.2, demonstrated that for each NASA-TLX parameter, the p-values fell below the predetermined significance level (α=0.05). Consequently, the null hypothesis was rejected, indicating a statistically significant difference in ratings of each parameter between the two conditions.

Table 5.2: NASA-TLX Non-Parametric Test Summary using Kruskal-Wallis Test

Parameter	χ^2	**df**	**p**	**Decision**
Mental	42.51	1	$< .001$	Reject the null hypothesis
Physical	18.025	1	$< .001$	Reject the null hypothesis
Temporal	6.679	1	0.01	Reject the null hypothesis
Performance	6.519	1	0.011	Reject the null hypothesis
Effort	27.407	1	$< .001$	Reject the null hypothesis
Frustration	15.692	1	$< .001$	Reject the null hypothesis

5.2 Task Based Analysis

5.2.1 Task 1: Surface Recognition

Confusion Matrix of Identified Textures

The confusion matrices below demonstrate participants' performance in identifying various virtual surfaces (gravel, grass, road, metal, and sidewalk) across three sub-tasks of Task 1: Surface Identification. These sub-tasks evaluated the participants' ability to perceive textures through a virtual cane interface under different sensory feedback conditions: combined auditory and haptic feedback (Task 1.1), auditory feedback alone (Task 1.2), and haptic feedback alone (Task 1.3). In the matrices, the predicted labels refer to the names of virtual surfaces as verbally anticipated by the participants, while the actual labels represent the names provided to them. In Task 1.1, participants were provided with two instances of each of the five surfaces, totalling 100 samples (calculated as 20 participants * 2 instances * 5 surfaces). Conversely, in Tasks 1.2 and 1.3, participants received one instance of each of the five surfaces, resulting in 50 samples per task (20 participants * 1 instance * 5 surfaces).

In Task 1.1, participants had access to both auditory and haptic feedback. The results, as depicted in Figure 5.5 indicate:

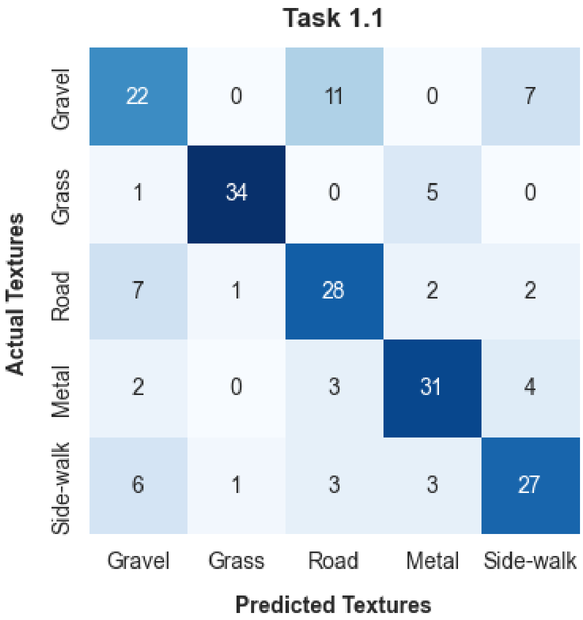


Figure 5.5: Confusion matrix for Task 1.1

- **Gravel**: Identified correctly 22 times, with notable confusion with sidewalk (7 instances) and road (11 instances).
- **Grass**: Correctly identified 34 times, showing high accuracy and minimal confusion.
- **Road**: Correctly identified 28 times, with some misidentifications as gravel (7 instances), sidewalk and metal (2 instances each).
- **Metal**: Identified correctly 31 times, but occasionally confused with road (3 instances) and sidewalk (4 instances).
- **Sidewalk**: Correctly identified 27 times, with confusion mainly with gravel (6 instances), metal and road (3 instances each).

In Task 1.2, participants relied solely on auditory feedback. The performance decreased compared to Task 1.1, as the results depicted in Figure 5.6 indicate:

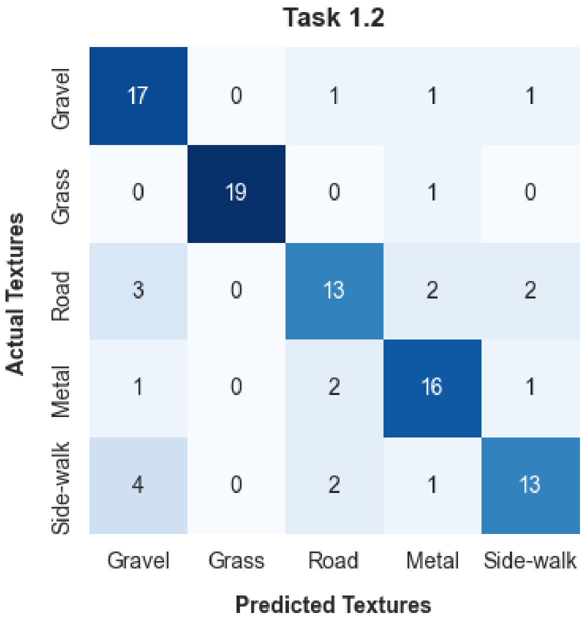


Figure 5.6: Confusion matrix for Task 1.2

- **Gravel**: Correctly identified 17 times, with confusion with road, metal and sidewalk (1 instance each).
- **Grass**: Maintained relatively high accuracy with 19 correct identifications.

- **Road**: Correctly identified 13 times, with confusion mainly with gravel (3 instances), metal and sidewalk (2 instances each).
- **Metal**: Identified correctly 16 times, with occasional confusion with road (2 instances), gravel and sidewalk (1 instance each).
- **Sidewalk**: Correctly identified 13 times, but often mistaken for gravel (4 instances) and road (2 instances).

In Task 1.3, participants used only haptic feedback. This task saw the lowest overall accuracy, as depicted in Figure 5.7:

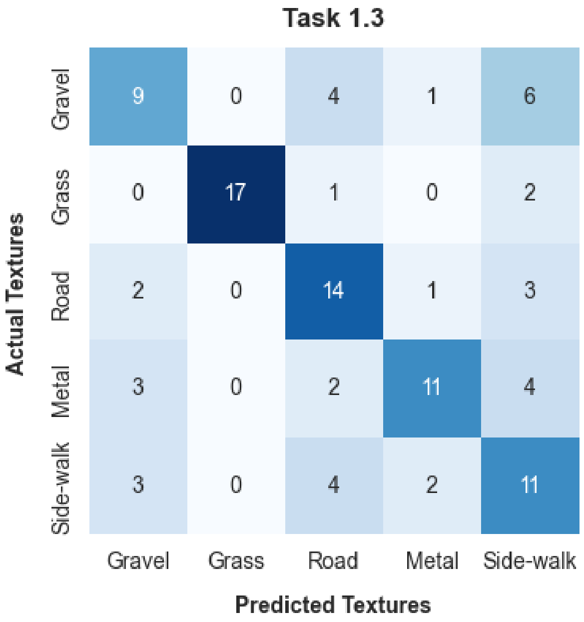


Figure 5.7: Confusion matrix for Task 1.3

- **Gravel**: Correctly identified 9 times, with substantial confusion with sidewalk (6 instances) and road (4 instances).
- **Grass**: Correctly identified 17 times, showing reasonable accuracy despite the absence of auditory cues.
- **Road**: Identified correctly 14 times, with confusion primarily with gravel (2 instances) and sidewalk (3 instances).

- **Metal**: Identified correctly 11 times but confused with sidewalk (4 instances) and gravel (3 instances).
- **Sidewalk**: Correctly identified 11 times, with frequent confusion with gravel (3 instances) and road (4 instances).

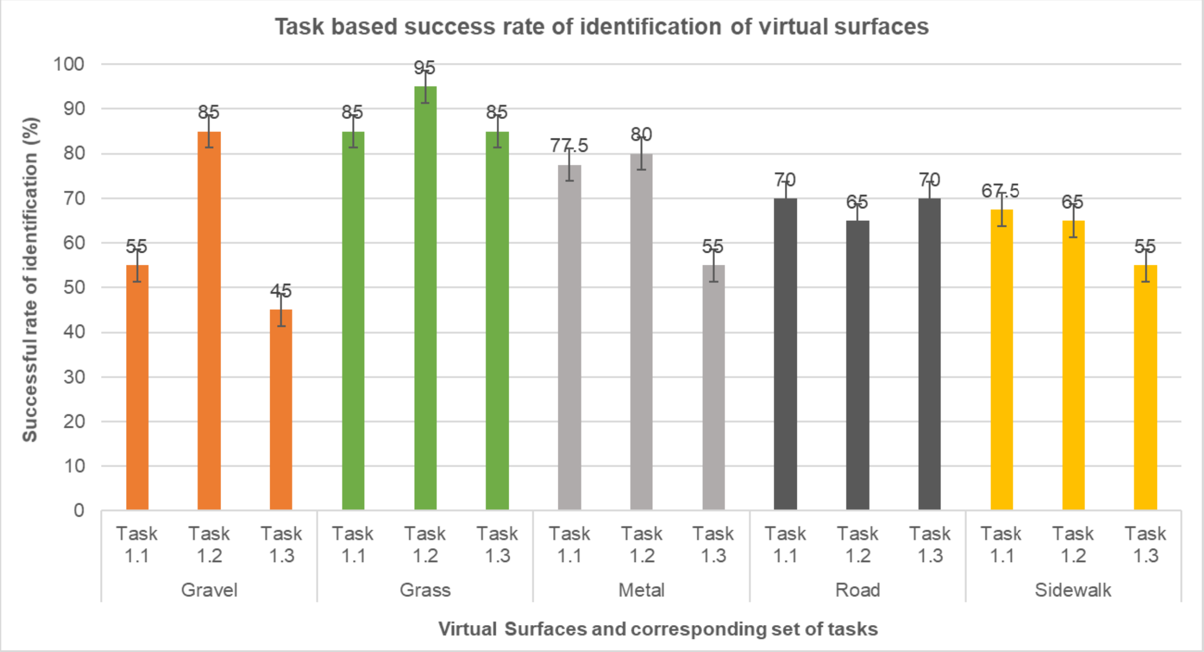


Figure 5.8: Task based success rate of identifying virtual surfaces

As illustrated in Figure 5.8, across all of the tasks, grass persistently showed high identification accuracy and the least variability, while sidewalk was least accurately identified and gravel had the highest variability. When the textures are compared across tasks, the confusion matrices also indicate a pattern where the textures with similar physical textural attributes possess the highest misidentification rate. This is evident in multiple instances where the textures, namely gravel, road and sidewalk, were confused across all the tasks with one another. Similarly, the performance of metal also varied significantly across tasks where it was mostly confused with the sidewalk. This confusion can further be understood by the consistent performance of grass across the tasks, as its physical textural attributes are distinct from the other textures.

The combined results from the three sub-tasks analyzed using the confusion matrices and performance chart demonstrate that participants performed better and achieved marginally higher accuracy when only auditory feedback was available during Task 1.2 ($\bar{x}$

= 78%) as compared to when both the auditory and haptic feedback as available during Task 1.1 ($\bar{x}$ = 71%). Additionally, the decline in performance and reduced accuracy observed in Task 1.3 ($\bar{x}$ = 62%) underscores that relying on a single sensory modality through only haptics significantly hampers the ability to perceive textures accurately.

Two-way Analysis of Variance of Textural Identification Success Rate

I performed a two-way ANOVA to investigate the effects of two independent variables, Sensory Modality' and Texture,' on the dependent variable, the 'Success Rate of Identification.' The sensory modality variable had three categories/levels: *audio only, haptics only, and audio and haptics.* As the name suggests, these categories referred to the first task in three variations with mixed audio and haptic feedback. The texture variable referred to the textures of the virtual surfaces, each with unique auditory and haptic properties, and had five categories/levels: *grass, sidewalk, road, metal, and gravel.* The dependent variable referred to the average of all participants' ability to successfully identify the different virtual textures under a permuted combination of sensory modalities provided.

The statistical analysis revealed a significant impact of the choice of sensory modality on the success rate: F(2,5) = 7.167, p = 0.034. This indicates that different sensory modalities lead to significantly different success rates, suggesting how information is presented—whether through sound, touch, or both- substantially influences participants' ability to correctly identify a given virtual surface. Similarly, the texture of the virtual surface was found to significantly affect the success rate, F(4,5) = 9.294, p = 0.016. This result suggests that different textures also lead to significantly different success rates, highlighting the importance of physical characteristics such as roughness, smoothness, hardness, or softness conveyed through audio, haptics or both in successful identification. However, when examining the interaction between sensory modality and texture, it was found that this interaction was not statistically significant, F(8,5) = 2.410, p = 0.174. This suggests that while both sensory modality and texture individually affect the success rate of identification, their combination does not create a unique impact beyond what each factor contributes on its own. In other words, knowing the sensory modality and the texture independently helps predict the success rate, but their combination does not provide additional predictive power in this context.

In conclusion, sensory modality and texture significantly influence the identification success rate, underscoring their individual importance in the process. However, their interaction does not significantly affect the outcomes, indicating that the individual effects of sensory modality and texture are sufficient to understand their impact on identification success rates.

Time-based task performance

This section presents the analysis of the participant's performance in Task 1, using a box plot to visualize the time taken to successfully identify different textures of the virtual surfaces as illustrated in Figure 5.9. Moreover, this data excludes the time taken by participants to identify the textures wrongfully.

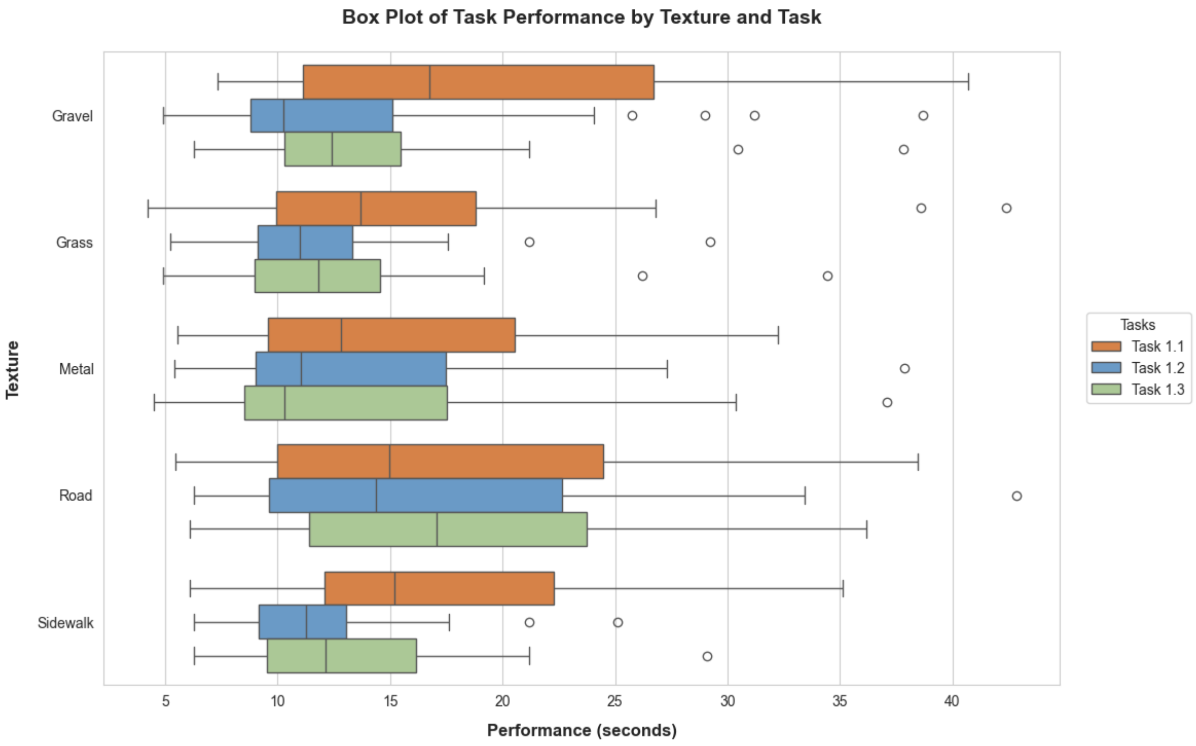


Figure 5.9: Time taken to identify each texture successfully in Task 1

The analysis shows that performance times and variability differ across textures for texture-specific performance. For instance, grass took the least time across each task to be identified with the most consistent variability, outperforming all the other textures, while road took the most time to be identified across each task, and gravel exhibited the highest variability across tasks.

Among the tasks performed for each texture, it is observed that Task 1.1 exhibits the most significant variability among participants, as evidenced by its broadest interquartile range (IQR) and highest median identification time. Conversely, Task 1.2 demonstrates the least variability, with the narrowest IQR and lowest median identification time, signifying a highly consistent performance. The performance metrics for Task 1.3 lie between those of Tasks 1.1 and 1.2, indicating moderate performance variability. This result can also be attributed to the fact that during Task 1.1, the participants had to identify textures pair-wise, resulting in prolonged time for decision-making; as for the rest of the tasks, the identification was for a single texture.

5.2.2 Task 2 and Task 3

Statistical Tests

Task 2: Road Crossing With and Without Visuals

The Shapiro-Wilk test was conducted to assess the normality of the TTR data for Task 2 under the conditions of performing the task with visuals and without visuals. The results indicate that the data significantly deviates from a normal distribution. The Shapiro-Wilk test ($W(40)=0.78$, $p< .001$) indicates a test statistic of 0.78, which is considerably lower than 1, suggesting a poor fit to the normal distribution, whereas the p-value $< .001$ leads to the rejection of the null hypothesis of normality.

Following the identification of non-normal data distribution from the Shapiro-Wilk test, the Wilcoxon Signed Rank Test was employed to compare the TTR under the two conditions ($Z = -3.472$, $p < .001$, $N = 20$). The test yielded a highly significant p-value, indicating a statistically significant difference in the median TTR times between the two conditions. The Z-score's negative value suggests that the NV condition's median value is significantly higher than that of the WV condition. This implies that participants took longer on average to complete the task when not supported by visuals compared to when visuals were provided.

Task 3: Scavenger Hunt With and Without Visuals

As explained in Section 3, Task 3 was about finding intermediate targets to reach a final destination to complete the scavenger hunt, where the targets were auditory sources that needed to be followed. Moreover, Task 3 had two sub-tasks, with the second sub-task having two variations (with and without visuals). Whenever a participant reached nearby a particular target, the system automatically logged the TTR of that target; however, reaching the target always didn't qualify as them being able to locate the target. Therefore, there are two metrics where TTR is a numerical metric that signifies the time to reach the target (in seconds), whereas a binary metric - 'successfully located' indicates if the participants could confirm their presence near a target verbally.

The following statistical test is performed for the numerical metric of TTR, specifically for the two variations of the second sub-task of Task 3, whereas the 'In-game statistics' discusses the binary metric of successful target location.

Table 5.3: Shapiro-Wilk Normality Test Summary for TTR* Task 3 targets

TTR Target	**Test Statistic (W)**	**p-value**	**Conclusion**
Target1	0.938	0.029	Non-Parametric
Target2	0.968	0.302	Parametric
Target3	0.969	0.332	Parametric
Destination	0.953	0.098	Parametric

**Time To Reach*

Non-Parametric Test

The analysis of the TTR for Target 1 under different conditions, WV and NV, was conducted using the Related-Samples Wilcoxon Signed Rank Test. This non-parametric test was chosen to determine if there are statistically significant differences between the median TTR values when participants completed the task under these two conditions. The test results were highly significant, with a Z-score of -3.622 and a p-value of less than .001 ($Z = -3.622$, $p < .001$, $N = 20$).

The negative Z-score indicates that the median TTR for the condition NV is greater than the median TTR for the condition WV. This suggests that participants generally took longer to reach Target 1 when they performed the task without the aid of visuals compared to when visuals were provided. The statistical analysis clearly supports the conclusion that visuals played a significant role in enhancing performance, effectively reducing the time required to reach the target.

Parametric Test

The analysis of the TTR for Targets 2 and 3 and the destination under the two different conditions (WV and NV) was conducted using one-way ANOVA to reveal varying levels of impact by these conditions. For Target 2, the one-way ANOVA results yielded [F(1, 38) = 0.001, p = 0.980], whereas for the destination target, the ANOVA results were [F(1, 38) = 0.383, p = 0.540]. This suggests no statistically significant difference in the time between the two conditions to reach the second and the destination target. The high p-value indicates that the observed differences in TTR between conditions are likely due to random chance rather than a systematic effect of using visuals. The results for Target3, however, showed a significant difference, with an ANOVA yielding [F(1, 38) = 43.592, $p < 0.001$]. This significant finding indicates a substantial effect of the condition on the time required to reach this target. The very low p-value suggests a strong influence of visuals on performance, where its presence or absence markedly affects the speed with which the target is reached.

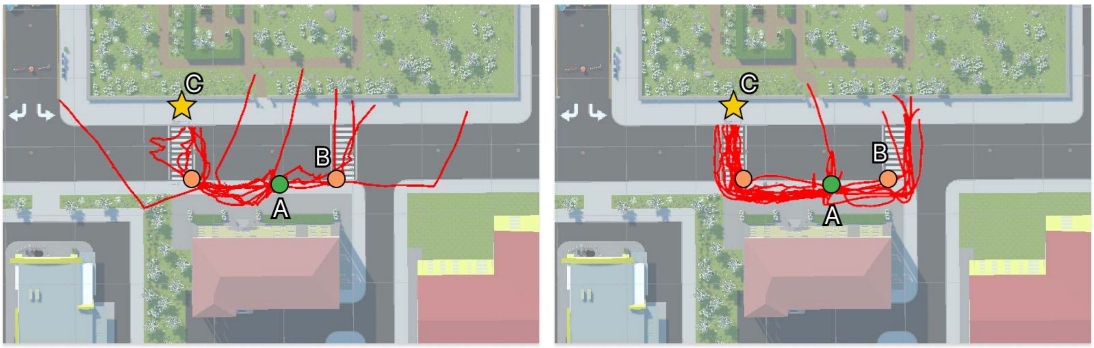
In-game statistics

Task 2: Road Crossing and Time To Complete

For the second sub-task, participants had to choose between two traffic lights on either side based on their distance and cross the road from the crossing they thought was farther away from them. This task was performed with and without visuals. In either case, the traffic light on the left was the farthest away, as shown in Figure 5.10.

In the variation of this sub-task without visuals, 12 out of 20 participants successfully chose the correct traffic light. Despite this, everyone managed to reach the other end of the road. However, only 11 out of 20 participants crossed the road from the designated area of the crossing. It took them around 121 seconds on average to complete the task successfully. Additionally, only 50% of the participants could avoid incoming traffic without any vehicular collision.

In contrast, for the variation of this sub-task with visuals, 10 out of 20 participants successfully chose the correct traffic light, and 19 out of 20 reached the other end of the road. A notable 18 out of 20 participants crossed the road from the designated area of the crossing, and it took them around 61 seconds on average to complete the task successfully.



(a) (b)

A: spawn point, B: Traffic Light/Crossing, C: Destination

Figure 5.10: Navigation paths of participants for Task 2.2 under different conditions, (a) Without visuals, (b) With visuals

Moreover, only 20% of the participants had any account of a vehicular collision.

The results show that the road crossing tasks performed in the absence and presence of visuals significantly impact the success rate and total time taken on the task. Without visuals, participants took longer (121 seconds on average) and had more collisions (50%) compared to when visuals were present, where they took half of the time on average and had 30% less collisions.

Task 3: Target Location and Time To Complete

For the first sub-task, the data reveals that all participants successfully located the first intermediate target, with a couple of participants not fully confident about it. Only 16/20 participants successfully identified the second intermediate target, whereas only 18/20 participants successfully located the final destination target. The average total time taken by participants who successfully located all the targets to complete this sub-task was 314 seconds (SD = 138.7).

For the second sub-task, where there was an increased number of targets, the variation in which there were no visuals indicated that all the participants were able to locate the first three intermediate targets successfully, but only 18/20 participants were able to locate the destination target. On the contrary, for the variation of the sub-task with visuals, all participants could locate all of the targets. The average total time taken by participants who successfully located all the targets to complete the second sub-task with and without visuals was 226.4 seconds (SD = 29) and 165 seconds (SD = 34.8), respectively.

Vehicular Collision across tasks

Since each sub-task in Task 2 primarily focused on assessing participants' performance on safely crossing the road from a designated area, the highest chance of vehicular collision was when they attempted to cross the road. Similarly, each sub-task in the third task had at least one scenario where they had to cross the road while searching for targets. This made pedestrian crossings the hotspot of vehicular collision, and the chance of a collision particularly depended upon the participants' ability to time their movement to reach the other end when the traffic stopped and the traffic signals signalled a safe passage.

As shown in Table 5.4, it's evident that sub-tasks in the third task without visuals show the highest average collision rates at 2.5. Additionally, the standard deviations for these tasks are the highest among the data at approximately 2.54, indicating a wide range of participant outcomes. This variability suggests differing levels of adaptability in safely crossing the road without visual cues, with some participants struggling more than others.

Table 5.4: Average Collision with Vehicles and Standard Deviations Across Tasks

Tasks	*3.1*	*3.2 NV*	*3.2 WV*	*2.1*	*2.2 NV*	*2.2 WV*
Average vehicular collision	2.5	2.5	0.9	1.7	1.4	0.3
SD	2.5	2.5	1.3	2.1	1.6	0.9

Conversely, when looking at tasks conducted under visual conditions for both the second and third tasks, there is a noticeable decrease in both the average collisions and the variability among participants. For instance, Task 3.2 WV shows a collision average of 0.9 with a lower standard deviation of 1.34, suggesting that participants found this task easier to manage. Similarly, task 2.2 WV records the lowest collision rate at 0.3, with a minimal standard deviation of 0.9, highlighting it as the task with the best overall performance and least participant deviation.

The data for sub-tasks in the second task conducted without visuals shows intermediate collision averages of 1.7 and 1.4, respectively. These figures and their moderate standard deviations suggest a better performance than the sub-tasks in Task 3 without visuals but still point to a certain level of challenge and inconsistency among participants.

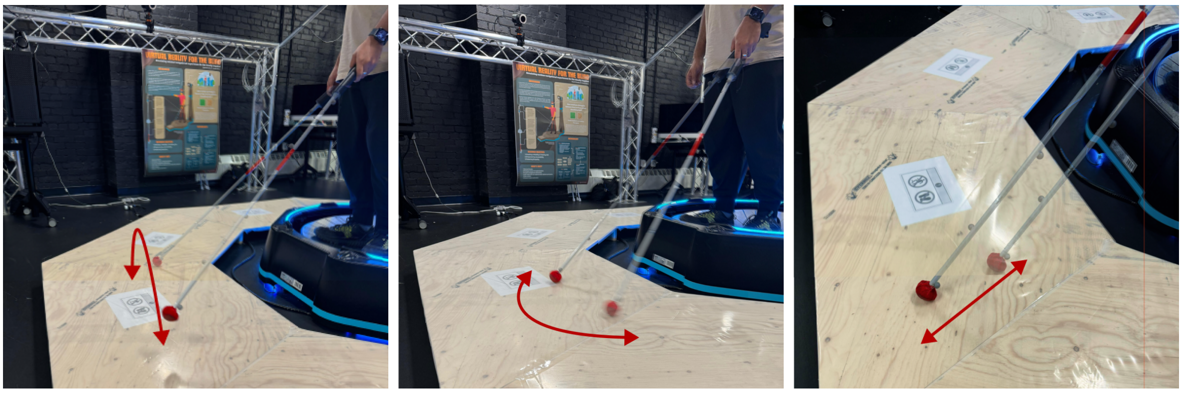
5.3 Behavioural Analysis

This section is based on the analysis performed using the logged positional data of participants along with the in-situ and ex-situ observations made using the observation sheets and video recordings.

5.3.1 Cane Statistics

Cane Movement and Handling

Although the participants were trained only on the "side-to-side sweep" method of using the cane, where they swept the cane from one side to the other in a pendulum motion, it was observed that they instinctively used other forms of cane movement. Figure 5.11 illustrates that participants employed a mix of cane movements along with the sweep method, where one movement was a sweep towards and away from them *(up-and-down sweep)* and one movement was a conventional *point touch* where they hit the cane on the floor move it in the shape of a parabolic arc to touch the floor on the other end.



(a) (b) (c)

Figure 5.11: Different cane movement (a) Point touch, (b) Side-to-side sweep, (c)Up-and-down sweep

As illustrated in Figure 5.13, the observed data shows that all participants employed the side-to-side cane sweep method. While most of them stuck to the training and solely used the method, a handful of participants (20%) also used the up-and-down sweep at times. However, only one participant (P7) used the side-to-side sweep and point touch method.

The analysis of cane handling methods among participants revealed that they demonstrated three distinct methods: using a mix of dominant and non-dominant hands and both hands simultaneously. The categorization of cane usage was divided into three primary methods:

- **Dominant Only:** Participants who exclusively used their dominant hand to maneuver the cane. This method was the most straightforward and emphasized ease and habitual use.
- **Dominant/Non-Dominant:** Participants alternating between their dominant and non-dominant hands. This covers participants either switching hands during the activity or using the non-dominant hand to assist periodically. This method suggests balancing physical exertion or enhancing precision in certain situations.
- **Both Hands:** A minority of participants used both hands simultaneously to grip the cane. This method suggests greater stability and control, particularly in situations requiring greater detail from feedback.

(a) (b)

Figure 5.12: Different cane handling style of participants. (a)Dominant/Non-dominant hand grip, (b)Both hand grip

The data showed that 60% of participants preferred using only their dominant hand, while 40% alternated between their dominant and non-dominant hands. Interestingly, a smaller group, representing 10% of the participants, used both hands simultaneously for additional stability and control. Specifically, one participant combined their dominant hand with both hands occasionally, and another alternated between dominant/non-dominant hands while also employing both hands. It is important to note that the percentages exceed 100% due to the overlap in handling methods, as some participants employed multiple techniques concurrently, thus contributing to more than one category.

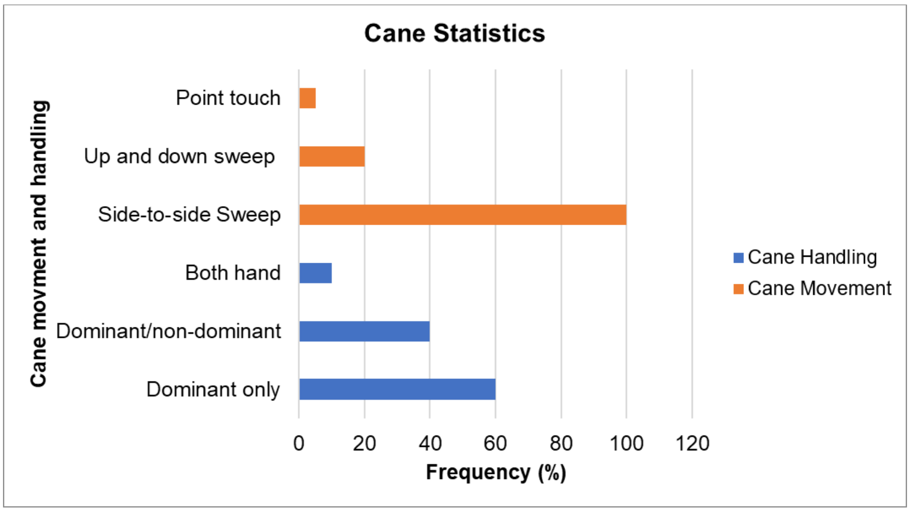
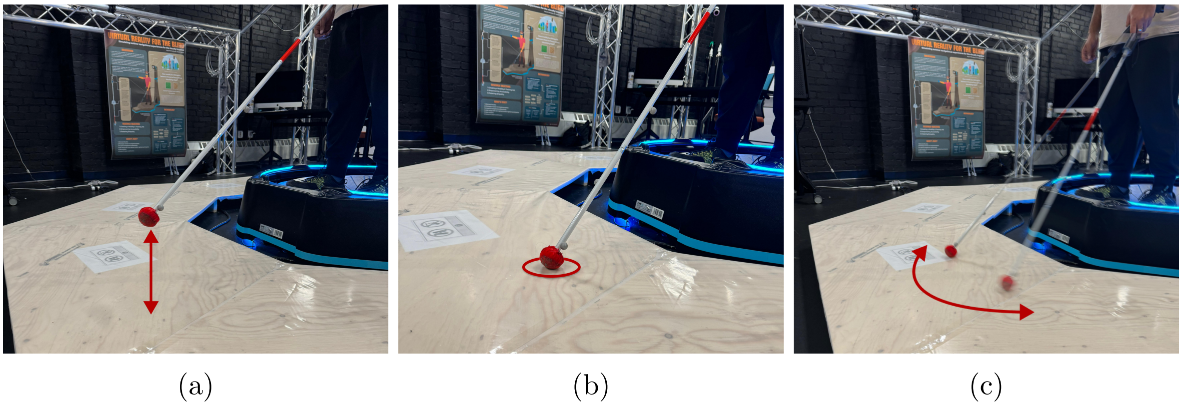


Figure 5.13: Different cane movement and handling frequencies of participants

The results reveal participants' adaptive and instinctive behaviours when using the cane despite being trained solely on a single method. The observation that participants employed a mix of cane movements highlights the versatility and personalized adaptation of cane techniques in real-world scenarios. This finding underscores the importance of considering multiple cane techniques in training programs, as it demonstrates that users naturally gravitate towards methods that best suit their navigation needs and preferences.

Cane Usage

The participants used the cane in several different ways throughout the tasks. Upon analyzing the video recordings of their task performance, the cane usage can be broadly divided into three primary forms/states: *Stationary, In-use and Not-in-use* as shown in Figure 5.14.



(a) (b) (c)

Figure 5.14: Different cane states (a) Not In Use, (b) Stationary, (c) In Use

- **In-Use State:** The cane is actively used and moved across the slide mill skirt to elicit feedback from the virtual surface, indicating active engagement with the task.

- **Stationary State:** The cane rests on the slide mill skirt without any explicit movement, reflecting a conscious decision to pause movement, possibly to process information or make decisions.

- **Not-In-Use State:** The cane is not in contact with the slide mill skirt and hovers above it, typically reflecting an unconscious decision where the participant may be focusing more on auditory or other sensory inputs.

Data analysis reveals distinct patterns in cane usage among participants across visual conditions. When deprived of visuals (Tasks 2 and 3), 85% of participants utilized the cane in each of the Stationary and In-Use states, compared to only 40% in the Not-In-Use state. In contrast, during tasks with visual assistance, cane usage remained prevalent without explicit instructions: 70% of participants used the cane in the Stationary state, 75% in the In-Use state, and 55% in the Not-In-Use state. As mentioned earlier, the percentages exceed 100% due to the overlapping nature of participants' cane usage, as individuals often transitioned between different states during the tasks.

Figure 5.15 illustrates the duration of cane usage across these three states during navigational tasks, focusing specifically on periods when participants were in motion. A significant disparity is observed in the In-Use state, where the duration of cane usage without visuals averaged 154.3 seconds—approximately five times longer than with visuals, where the average was 34.1 seconds. Similarly, in the Stationary state, participants

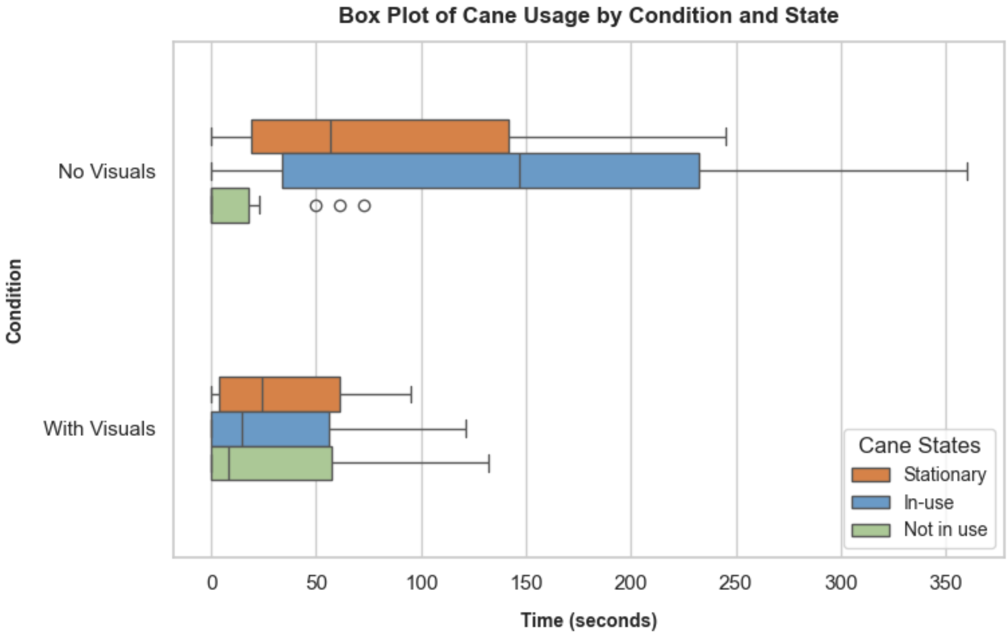


Figure 5.15: Box Plot of Cane Usage by Condition and State

spent an average of 74.7 seconds using the cane without visuals, which is 2.5 times longer than with visuals ($\bar{x}$ = 31.15 seconds). Conversely, in the Not-In-Use state, the duration was greater during tasks with visuals ($\bar{x}$ = 32.3 seconds) compared to those without ($\bar{x}$ = 13.05 seconds).

The data stresses the reliance on the cane in tasks devoid of visuals and highlights the importance of its presence as a medium for interaction in such situations. The importance of the cane persists even with visual assistance, reflecting its integral role in navigation where the cane sub-consciously became habitual and oftentimes acted as an extension of their arm. Further insights into participants' perceptions and behavioural justifications are discussed in Section 5.4.8 (*Role of Cane*) and 5.4.9 (*Cane vs Controller/Joystick*).

Participant Experiences with Cane Usage

Participants' interactions with the cane varied, reflecting different strategies and levels of reliance on the cane for navigation.

Consistent Use and Habituation Several participants reported using the cane consistently and even becoming habituated to its use, indicating that it felt natural to incorporate it into their navigation routines. P11 described the cane's use as second nature:

"But also just because I had grown so used to using it, it just felt like second nature to start you know brushing it around." P2 acknowledged the constant helpfulness of the cane: *"No I was aware that the cane is there and it is anytime helpful."* P18 emphasized frequent use: *"I think I used it for the most part."*

Inconsistent or Subconscious Use Some participants used the cane inconsistently or subconsciously. They did not always remember to use it but would occasionally realize its utility during navigation. P10 noted a decrease in usage over time: *"I think I used it like a little, but not as much as like I probably did at first."* P13 and P19 described moments of forgetting the cane, only to remember its importance later. P13: *"I think, yeah, I think sometimes I just forgot about it and then when I was like walking forward I'd realize, oh yeah, I could probably have the cane down and it might be helpful."* P19: *"I was like occupied with the spatial surrounding (audio) so I was like in between at some point I forgot about the cane but instantly when I was going closer to the target then I was thinking okay now I think it's the time then it clicks in my head okay it's the time to find surface am I on."*

Intentional Non-Use or Minimal Reliance Other participants intentionally did not use the cane or relied on it minimally, preferring to depend on other senses such as audio for navigation. P16 discussed minimal reliance on the cane: *"I did use it, but it was, I will say, maybe 10% only. I didn't rely on it."* P9 expressed a preference for relying solely on audio cues: *"I completely lifted off the cane off the surface because I didn't want to hear it, it was more sort of a distraction because I only relied upon the audio, spatial audio."*

Context-Dependent Use Participants also highlighted that their use of the cane varied depending on the context or specific task they were performing, with some forgetting about it in certain situations. P15 and P3 discussed situations where they relied more on audio. P15: *"Maybe (forgot about the cane) at times I was only relying on the audio of the environment, especially for the last one when I have to find the gate of the park or the construction spot."* P3: *"I would say I tried to use it because that was important because like even if I compare it with real life scenario like even I am working on a side walk and I just went to the road I would not be able to know."* P6 noted frequent use in specific contexts: *"I realize I use the cane a lot, like when I'm trying to find my left and right."*

5.3.2 In-game Navigation Patterns

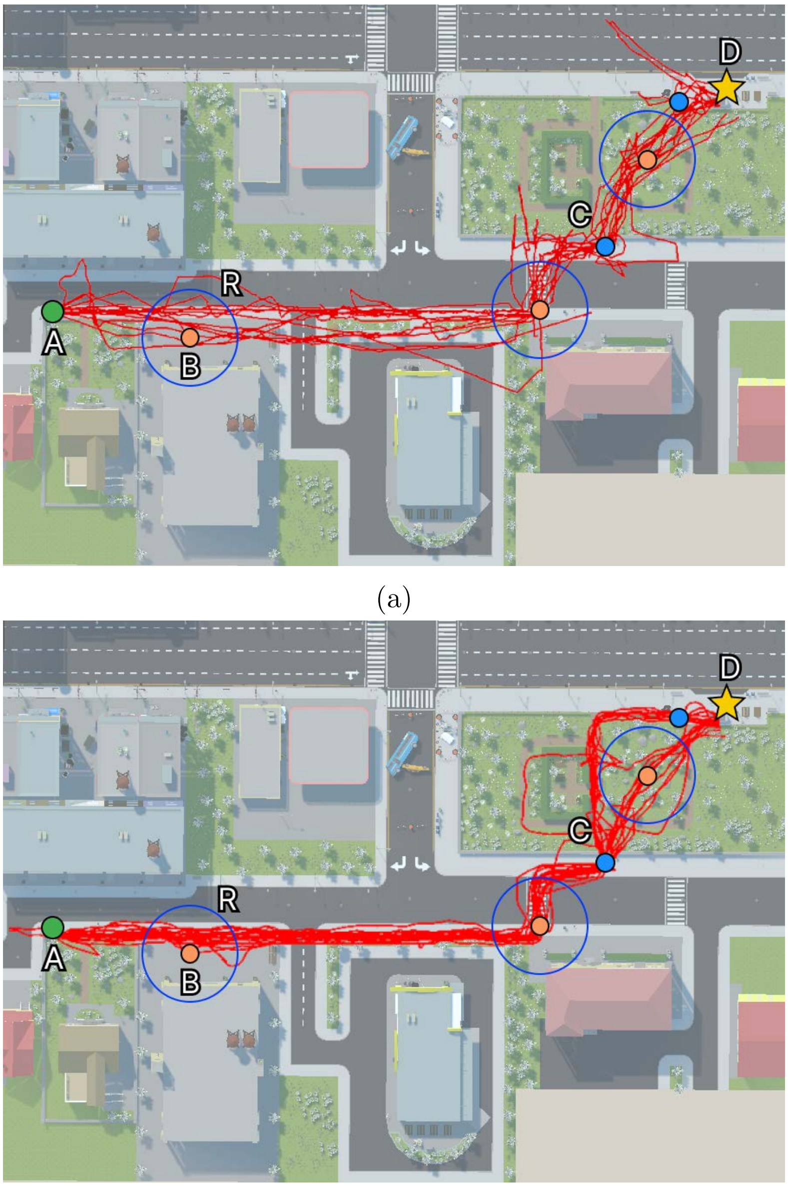
This section analyzes the navigation patterns of users performing Task 3 (Scavenger Hunt) under two conditions: NV and WV. As depicted in Figure 5.16a and 5.16b, the orange dots represent the center of the 'target,' an intermediate audio source and the blue circle represents the area up to which the target is audible - 'target radius.' Thus, the metrics, such as the closest distance (CD) to the target and time spent (TS), elucidate the navigation patterns of participants in terms of efficiency, quickness and accuracy, respectively.

Average of **Targets**	**CD_NV (Unity units)**	**CD_WV (Unity units)**	**TS_NV (sec)**	**TS_WV (sec)**
Task 3.2				
Intermediate Target 1	3.3	4.3	16.9	9.1
Intermediate Target 2	1.7	2.2	22.1	8.8
Intermediate Target 3	1.9	4.8	24.0	11.2
Task 3.1				
Intermediate Target 1	2.3	NA	35.8	NA
Intermediate Target 2	1.8	NA	32.9	NA
Mean($\bar{x}$)	**2.2**	**3.8**	**26.3**	**9.7**

CD: Closest Distance (to the target), TS: Time Spent (within Target radius)

Table 5.5: Game statistics of user interaction with targets during Task 3

Aggregating the data for Task 3 in Table 5.5 reveals apparent differences in navigation performance between the two conditions. However, it is evident that the paths taken by participants in situations devoid of visuals closely resemble those taken when the visual cues were present. In the NV condition, users appeared to adopt a more cautious approach, resulting in prolonged interaction times and increased precision in reaching targets. However, this cautiousness translated into participants exhibiting less direct



(a)

(b)

A: spawn point, B: Intermediate Audio Target, R: Audible radius of Intermediate Audio Target, C: Audio Queue, D: Destination Audio Target

Figure 5.16: Navigation paths of participants for Task 3.2 under different conditions. (a) Without visuals. (b) With visuals.

navigation toward targets, as indicated by higher average time spent ($\bar{x} = 26.3$ seconds) within the target radius. In contrast, visuals facilitated more efficient and precise navigation, with users spending significantly less time ($\bar{x} = 9.7$ seconds) within target radii and navigating more directly to the targets. Moreover, the users also exhibited confidence in reaching the targets with visuals as they didn't show tendencies to reach the exact source of the sound and rather confirmed it from a distance. This is evident from the fact that the closest distance from the center of the targets on average was 3.8 Unity units, double the distance when the same targets were reached without visuals (2.2 Unity units). This highlights the importance of the audio source being precisely discerned by the users in the absence of any visuals and their tendency to reach the center of the audio source.

5.3.3 Directional Decisions

As part of the post-study interview, when asked about the different audio participants recall hearing during the navigational tasks, they reported a mix of audio sources in the VE. These reported audio sources were treated as tags and then classified under respective types of audio integrated into the VE, namely, Ambient, Object, Speech, Interaction and System. The number of times the participants reported different tags for a particular audio type and participant number(s) is illustrated in Table 5.6. Similarly, participants referred to using various audio and tactile queues and strategies when asked about the directional and navigation decisions made during navigational tasks without visuals. Both of these aspects (audio tags and navigation queues), combined with the observations made while the participants performed the tasks, revealed a sophisticated use of sound and touch to help identify key strategies that facilitated participants to make critical directional and navigational choices, as summarized below.

Ambient Feedback

- **Directional Guidance:** Participants reported having harnessed animate and inanimate object sounds (reported in 15 instances), such as the noise of cars and traffic, to locate and navigate towards or away from specific areas. For example, P1 used the direction of car sounds to locate the road, while P10 and P11 determined the exact positions of road crossings through the sounds of vehicles stopping. As noted by P14, the sound of traffic lights was crucial for determining safe crossing times, reflecting the instrumental role of specific object-related sounds in navigation. P14 explained, *"If I need to cross a road, I would make sure to listen to the flow of traffic and signal lights when they turn green to make my decision to cross it."*.

Table 5.6: Summary of Categories and Instances in Virtual Reality Simulation

Category	*Ambient*	*Object*	*Speech*	*Interaction*	*System*
Tags	Road sounds, traffic sounds, construction site, churning machine sound, sirens or honks	Bus stand, traffic light, hot dog stand, cars, buses, vehicles crossing the road, emergency vehicle sirens	People talking, children playing, hot dog guy, hip hop people, bus stop people talking around, voices of the people, people dancing, people speaking	Sound of walking, sound of different texture on the road while you move your cane around, my walking steps, different surfaces (footsteps), my own footsteps, goals at the end	
Instances	11	15	21	9	1
Participants	P5, P6, P8, P9, P10, P12, P13, P14, P16, P18, P20	P1, P3, P5, P6, P9, P10, P12, P13, P14, P15, P16, P17, P18, P19, P20	P3, P5, P6, P9, P10, P13, P14, P15, P16, P17, P18, P19, P20	P1, P5, P6, P9, P10, P13, P14, P15, P18, P19, P20	P10

- **Safety Cues from Traffic Infrastructure:** Traffic-related auditory cues, particularly from traffic lights, were highlighted as pivotal for assessing the safety of crossing roads. P14 utilized these cues to make informed decisions about when it was safe to cross, indicating a strategic use of sound to mitigate environmental risks.

- **Broadened Spatial Awareness:** Ambient sounds (reported in 11 instances), encompassing a range of background noises from traffic ambiance to construction sites, were noted to have provided a more extensive spatial awareness. Participants reported that these sounds assisted them in forming mental maps of their surroundings, which were essential for understanding their movement and direction within the simulation. P8 highlighted the significance of ambient sounds in enhancing navigation by compensating for visual limitations: *"the intensity of the sound was heightened because the sense of hearing was heightened due to not being able to see."*

Conversational Feedback

The analysis underscores the critical role of speech sounds (reported in 21 instances) in VR navigation, particularly through their use in conversational feedback, which enables participants to locate targets, confirm directions, and interact more effectively with the virtual environment.

- **Target Location Identification:** Participants such as P3, P6, and P15 highlighted how they utilized conversational sounds to pinpoint targets within the VR environment. The use of human voices and dialogue not only helped in identifying the direction but also in locating specific environmental targets. This was crucial for navigation as it provided a reliable auditory cue in scenarios where visual information might be limited.

- **Directional Confirmation:** P13 exemplified the use of conversational feedback for confirming navigational directions. Participants' reliance on the interactive sounds of conversation highlights how speech can serve as an effective tool for orienting oneself within a VR setting. The feedback from conversations helped P13 to confirm and adjust their navigational decisions based on the sounds of human interactions.

- **Enhanced Interaction for Decision-Making:** Speech sounds were reported to have been integral to navigation through their role in conversational feedback. For example, P14 noted the strategic use of speech to navigate, stating, *"if it's near people talking, I would try to listen to that,"* which illustrates how participants used conversational sounds to assess their proximity to targets and make informed decisions about their movements.

Interaction Feedback

- **Guidance for Orientation:** P18 and P19 utilized the sidewalk as a primary navigational guide, relying on its consistent presence to maintain correct orientation within the simulation. This behaviour underlines a strategic use of environmental features for navigation, where the physical boundary of the sidewalk provides a tactile cue to stay aligned with safe walking paths. P17 exemplified the use of tactile feedback for error correction in navigation. After sensing different road surfaces underfoot, P17 adjusted their walking direction, showcasing a reactive strategy to tactile inputs.

- **Confirming auditory feedback:** Participants were observed to rely on tactile feedback from the ground textures to confirm auditory cues from interaction sounds (reported in 9 instances) to navigate effectively within the VR environment. P3's experience highlights this dual reliance, as they used both the sensation of textures underfoot and the sounds generated from interactions to orient themselves when visual cues were lacking. Similarly, P17 highlighted the dual reliance on sensory inputs by stating, *"the first things that came to my mind was what are you hearing? So the sound was the first thing I paid attention to and then what are you feeling?"* This comment underscores the intertwined use of auditory and tactile feedback to make real-time adjustments, enhancing the efficacy of navigational strategies within the simulation. Moreover, P14 noted, *"Understand where I am - what surface I am standing on - listen to sounds to understand my position - make a directional decision based on ambient sound."* This statement illustrates how tactile feedback, supplemented by auditory information, forms a comprehensive sensory input system that enhances spatial awareness and navigational accuracy in the VE.

5.4 Interview Analysis Results

5.4.1 Preference of sensory input in the absence of visual cues

When participants were asked about their preference for sensory modality for performing the tasks, the data reflected a diverse range of choices, some of which were situational. The preferences split across audio only, haptics only, and a combination of both audio and haptics.

9/20 participants (45%) expressed a strong preference for audio feedback only. 7/20 participants (35%) valued the combination of both audio and haptic feedback, stating that it maximized their ability to interact with and understand the virtual environment. 4/20 (20%) participants gave mixed responses, stating the situation would influence their choice. Among them, three participants (P2, P7, and P13) stated that their preference was based on the situation or the nature of the task at hand; specifically, if it were just surface identification (Task 1), they would choose audio-only, but if it were navigation in the city simulation (Task 2/3), they would prefer to have the maximum scope consisting of both haptics and audio feedback. This reason is largely attributed to the presence of varied audio engagements within the cityscape that, at times, muffled the actual interaction sound between the surface and the cane. This case is explained by P13, *“if it was like in a city setting where I had to detect what the surface was like at the end of task three, each of the sub-tasks for task three, then I’d say both were important.”*. Moreover, P6 mentioned that upon being given a choice, they would choose both sensory inputs for maximum effectiveness but, if they had to let go of anyone, would stick with audio feedback only as the bare minimum.

5.4.2 Overall empathetic impact after the experiment

Many participants reported an increase in empathy towards visually impaired individuals after experiencing the challenges of navigation without sight.

Increased Awareness and Understanding

Participants noted a significant increase in their understanding of the challenges faced by visually impaired individuals. This was evident as P1 described, *“I think whatever they do they do a great job. Yeah, it’s very difficult task to navigate without the vision.*

I would say I feel more awareness." Similarly, P10 added insight into their respect for visually impaired people's navigational skills, stating, *"I'm always really impressed by the way that people can navigate around the world with less sight."* P11 underscored the educational impact of the simulation, mentioning it was an *"exercise in empathy,"* which highlighted the importance of understanding the navigational challenges in busy urban areas.

Personal Reflection and Sentiment

Reflection on personal sentiments was common among participants after the simulation. For instance, P12 shared, *"Yeah I did like especially when you are blindfolded and when you walk you can realize how much you rely on visual cues on your daily life."* This comment points to a sudden awareness of how dependent we are on sight. P15 connected this realization to broader issues, suggesting, *"I guess it's also really helpful to people experience how people with VI can experience the world surrounding them."* Moreover, P16 discussed the depth of their empathy, stating, *"(I have) empathy now with blind people. I mean, of course, I had it before, but it's not like, but now I can tell how, like, how mind-demanding this task is for people who can't see."*

Behavioral Changes and Emotional Impact

The potential for behavioral changes post-experiment was evident in participants' intentions to act differently. P13 expressed a newfound commitment, saying, *"In the future now I'd definitely be much more, I'd try and be much more helpful."* P5 also indicated a willingness to assist, noting, *"I think this gives me like an opportunity that if I see some visually impaired person I would jump in to help them."*

The emotional impact of the experiment was reported to be profound. P18 reflected on the difficulty of basic tasks without sight, stating, *"And being in that situation definitely gives you a lot more clarity. Just how difficult it is to just rely on your auditory senses."* Additionally, P9 described their emotional response, saying, *"I kind of felt helpless once I was blindfolded. I had to only rely upon my audio and the cane and nothing else."* These statements and emotional accounts emphasize how the experiment could deeply influence participants' feelings and empathy towards the challenges faced by visually impaired individuals and could foster a more considerate and active approach to assisting them in everyday situations.

5.4.3 Carefulness in the simulation

Mixed Carefulness and Safety Precautions

Many participants indicated a noticeable decrease in caution within the simulated environment, attributing this to their awareness that the situation was not real. This perception of safety led them to behave with less caution than they would in real-world scenarios. P1 explicitly stated this difference: *"Less careful than the real world scenario because you already know that you are in a simulated world."* Similarly, P13 admitted a reduction in carefulness due to the virtual nature of the setting: *"I think I probably was (less careful), yeah, just because I knew it was virtual reality."* P3 also highlighted the absence of real-world dangers as a reason for reduced caution: *"Obviously like I would be less careful because there is no fear of like car hitting me or anything."*

On the contrary, some participants remained mindful of the safety precautions. They responded actively to auditory cues that were designed to simulate real-world dangers, such as the sound of approaching vehicles. P10's reaction to these cues showed a high level of engagement with the simulated safety mechanisms: *"I was very aware of like the big sound of like the trucks or like feeling like I'm getting squished by one and I would stop. In an effort to exude safety precautions."* P15's cautious behavior further stresses this point: *"I guess I was careful. I was trying to pay attention to the sounds of cars especially because I didn't want to get hit by a car even if it's in the simulation."* P6 also noted a sense of security provided by the controlled environment, yet still maintained a careful approach: *"I was careful, but I know that like, okay, I'm in a closed space. I'm strapped in. I'm good. Rather than, you know, out there and about."*

Simulation vs. Real World Carefulness

Participants often drew comparisons between their behaviour in the simulation and how they would act in real life, noting significant differences in their levels of caution. P16 remarked on the distinction between simulated and real-life responses: *"Not exactly. No, I was really trying to (be careful), but I understand this is not real, so, of course, this is not like a life-saving situation, so you will not give this attention as you were in the* *real world."* P14 observed that the lack of physical obstacles in the simulation affected their level of caution: *"I was less like I knew the surface was like similar there were no slopes or inclinations or any other things like step down step ups but yeah it was like straight and level so I wasn't that much cautious but like in real life I would be way more cautious."* P11 compared the experience to playing a video game, which influenced their approach to navigation: *"I was kind of just treating it like a video game, like you know, point A to point B without, I mean I did heed your warnings on, you know, wait for the crossing to signal."*

Ease of Mind in Simulation

The simulated environment also fostered a sense of security that allowed participants to feel more relaxed and less concerned about potential dangers. P2 expressed a certain level of carefulness but acknowledged the lack of physical or mental effects in virtual reality: *"I was similarly careful but I know that in the VR world nothing is affecting me physically or mentally."* P6 mentioned feeling comfortable in the simulated space: *"Like with the simulation, I feel more like in a comfortable space. So I feel like I won't actually get hurt or injured."* P18 summed up the general sentiment by noting a greater ease of mind in the simulation compared to real-life situations: *"I was not less careful, but you can say I had more ease of mind. Whereas in real life I might be too scared to even just take a little step."*

5.4.4 The good, bad and ugly of slide mill based navigational technique

Impact on Immersion by the slide mill

Participants provided mixed reviews on how the slide mill affected their sense of immersion during the VR experience. Opinions varied from enhancing realism to negatively impacting their perception of movement.

Positive Effects on Immersion The slide mill was seen by some participants as enhancing their connection to the virtual environment. P14 mentioned its positive impact on immersion, stating, *"It positively affects it (immersion)."* Similarly, P2 felt that it *"makes you more immersive,"* while P9 described the convenience it offered: *"I think it has definitely had a positive effect because I mean if I had to use the system without the* *slide mill I would have to actually walk. But here I was stuck on standing. I mean I was at a constant space and I was just moving my footsteps."* P8 and P6 also noted improvements in their spatial awareness and forgetfulness of the device's presence, enhancing their immersive experience.

Negative Effects on Immersion Conversely, other participants found the slide mill detracting from their experience due to its unnatural feel and the effort required to simulate walking. P10 questioned the realism of the experience: *"It probably didn't enhance it just cause I was like, am I walking? Am I in a garbage can? Why is it moving like this?"* P11 and P17 mentioned the limitations in movement and the fatigue it caused, respectively, with P17 explaining, *"I would say negatively, unfortunately, because if you have to put in so much effort, I think you get tired quickly."* P19 reflected on the lack of freedom compared to real walking: *"You are immersive because you have a headset right but in real world when you're walking then you are not actually attached to any of the thing because you are freely walking right so it's negatively affects immersion."*

Neutral Impact and Adaptation Some participants experienced initial disorientation with the slide mill, which diminished as they adapted to the technology. P3 found that after initial adjustment, it felt *"exactly the same (like walking),"* while P5 and P20 indicated that they became accustomed to the system, with P20 noting no significant effect on their immersion.

Slide mill Based Navigation

The use of the slide mill as a navigational tool within the system was met with varied insights focusing on its impact on natural movement and overall experience.

Impact on Movement The slide mill was noted by several participants to restrict natural walking movement, leading to a less realistic experience. P1 commented on the unnatural feel of walking with the slide mill: *"Slide mill is good but it doesn't give you the more natural way of walking as if in real there are a bit constant on the walking."* P11 and P20 discussed its restrictive nature and the physical discomfort it caused during extended use.

Safety and Convenience Despite these restrictions, the slide mill was appreciated for its safety and the novel experience it provided. P3 highlighted its safety benefits over traditional VR setups: *"The slide mill is less risky than like using a room because like in a room there is a chance of like falling and just getting hit by a wall."* P15 reaffirmed this sentiment, emphasizing the safety for visually impaired tasks: *"So you feel you're tight, you're tight in one space, you can't move. But for sure it's safer, especially for the task that you cannot see."*

Learning Curve and Cognitive Load Participants also discussed the cognitive load and learning curve associated with the slide mill. P10 mentioned the cognitive challenge: *"It was definitely one of the most, the highest cognitive load I think I had was,"* while P2 highlighted the adaptation process: *"Once you get used to it it's very convenient. The only thing is like it takes time to get used to it."* The initial challenge of using the system was noted, but participants also acknowledged becoming more comfortable and proficient with practice. P10 mentioned needing more practice: *"The use of the machine for someone like me could have used like either a little more practice."* P15 spoke about acclimating to the system: *"At first it was a bit scary to cannot see anything and you have to accomplish a task. But as time goes on, I get used to it and that was easier."* P13 experienced improvement over time: *"I could feel that I was getting better as we got to task three and like later on."*

Effectiveness of the Slidemill Skirt in Simulation

Participants discussed the effectiveness of the slidemill skirt in enhancing their experience during the simulation. The feedback varied, focusing on how it facilitated their interaction with the virtual environment, especially with the use of a cane.

Enhancing Navigation and Sensation The slide mill skirt was noted to help participants maintain balance and provide a realistic sense of surface and space, thus enhancing the overall immersive experience. P10 felt that it aided their forward movement during the simulation: *"I guess it was like, helped me kind of remember that I had to keep going forward, especially with the cane and the walking. It was like, you need to feel both kind of in order to keep going forward in this simulation thing."* P11 appreciated its help in maintaining the cane at a steady level, while P13 found it useful as a physical boundary: *"It was helpful to have like a physical kind of boundary for when my cane would be on the ground."* P14 and P18 commented on its smooth integration with the cane, enhancing their tactile feedback. P14: *"It was very perfect with the cane if I wanted to get a sense of the surface so yeah it was really smooth too."* P18: *"Definitely gave me a lot of surface area to explore. So that was definitely very helpful and very close to the realistic simulation when you're walking on a road."*

Feedback and Surface Interaction The feedback provided by the skirt was reported as a crucial factor in shaping participants' perceptions of the virtual environment and enhancing their navigational abilities. P16 expressed how it improved their sensory experience: *"It helped. I mean, to have the stick and feeling the floor."* P19, P2, and P3 discussed how it modified their interactions with the environment, with P2 noting: *"When you're touching the cane to that surface, it gives you a different experience altogether."* P3 mentioned its impact on vibration feedback: *"It helps to make the vibration a bit lower because if I would use on like without the smooth surface or in real so then it would like to vibrate a lot more."*

Neutral and Mixed Feedback While many found the slidemill skirt beneficial, others gave neutral or mixed feedback, suggesting variations in perception among different users. P6 highlighted its utility in understanding the environment and enhancing safety: *"In terms of knowing the level of the ground and feeling around, like where things are, that was very helpful. I couldn't imagine not having it."* P9 appreciated the ease it provided during use: *"It definitely helped because whenever I used the cane I was easily able to reach it or slide without having to put a lot of physical exertion. I mean I didn't have to bend a lot or move my body a lot. I was just able to reach the surface pretty easy with the cane."*

Minimal or Negligible External Factors Perceived through the Slidemill Skirt Many participants reported minimal to no perception of external noises or vibrations, indicating a generally smooth and controlled simulation environment through the slide mill skirt. P10, P12, and P14 remarked on the lack of external disturbances, with P10 noting: *"Didn't feel any external sound other than the rolling of the cane at times."* P19 experienced minor external sounds, describing them as *"I felt some sort of sound like a clothes sound or something like vibration. Very little vibration, not actually vibration but some very little frictionless sound."*

5.4.5 Potential of the system to be used with visually impaired demographics

Immersion for Visually Impaired

Participants discussed the level of immersion the system provides for visually impaired individuals, emphasizing the importance of combining haptic and audio feedback to enhance immersion, alongside the need for training and the benefits of realistic sensory experiences.

Realistic and Impactful Experience The system was found to provide a realistic and impactful experience, closely simulating real-life scenarios, which helped users understand their environment better. P3 described the realism of the system: *"It is actually kind of like real life scenario so if someone is using it like they are getting everything like that is in real life so it is completely immersive."* P6 reflected on the immersive quality: *"Like in terms of like how I feel it is kind of similar to the actual outside world. I feel like it definitely be immersive. Like hearing the different distances and sound and then how trying to see like if I'm right there in front of the crosswalk or if I'm right there in front of the park entrance and stuff like that."*

Integrating both haptic and audio feedback was highlighted as crucial for creating an immersive experience. Participants felt that relying on both senses together made the simulation more effective and realistic. P1 discussed this integration, saying, *"Combining both haptic and sound would be more immersive than just having one feedback either haptic or sound."* Similarly, P14 mentioned, *"I think it would be really immersive on both audio and the vibration like the haptic part."*

However, the necessity for training and familiarization was noted, with P11 stating, *"I feel like it would require a lot of training, but after a while it would be impactful. And you know as I said before I would definitely look into improving the audio fidelity if possible."* P20 emphasized the importance of familiarity with the system: *"Like once they get to know about it, how it operates then I think that will be comfortable for them."*

Need of better Differentiation and Customization Suggestions were made to improve the differentiation of haptic feedback based on different surfaces, which would enhance the immersive experience by making it easier to distinguish between various textures. P14 suggested, *"If we could differentiate (haptics) a bit more based on the surface like if it could be like if there would be a more differentiation between the audio and haptics of let's say grass and gravel that would like in the beginning be really helpful for them."* P19 expressed concerns about the immersion being affected by the treadmill: *"If I leave (out) the slide mill so I think it's really really immersive right but if we attach the treadmill I don't feel it's that immersive so that thing degrades the quality and the overall setup of the environment."*

System as a Training Tool

The system's potential as a training tool for visually impaired individuals was discussed, highlighting its benefits for learning navigation, enhancing confidence, and providing a safe environment for practice.

Promotes Safety and Practicality The system was noted to offer a safe alternative to real-world training, which can be risky for visually impaired individuals. P1 discussed the practicality of VR-based training for novice cane users: *"Makes more sense giving the training in the VR rather than giving the training in the real world scenario."* P16 elaborated on the safety benefits: *"As I mentioned, it (real world) can be dangerous, it (the training on the system) can be life-saving, but if you just keep trying and train yourself before that, I believe that will be a good option."*

Mediator for Learning and Acclimation Participants reported the system's effectiveness in helping users acclimate to their new circumstances, learn to rely on other senses, and gain confidence in navigation. P11 noted the potential for adaptation: *"If you just improve the fidelity of the audio a little bit I feel like it could be a very useful adapting tool for those who had just recently become visually impaired."* Similarly, P18 praised the system's ability to help users become more acquainted with their senses: *"It's a really good one because it would certainly allow them to be more acquainted with their other senses apart from the visual one. And it would definitely help them gain confidence* *that they can indeed navigate without sight as well."*

Effectiveness and Realism The realism and effectiveness of the system were highlighted, with participants noting its similarity to real-world scenarios and its utility in improving judgment and navigation skills. P14 explained the utility of the system for new users: *"I think it would be really useful to get a sense because like people who have just lost their sight know like how it feels how different surfaces feel and it this system really gives you a sense of how it should feel without the visuals."* P9 and P5 also commented on the training potential and the realism of the system, with P9 stating, *"I think the system would be very effective mainly because it would help them in training their audio with exactly how the real world is and it would also help them with a cane like to hear and to feel the haptics and audio from the cane and improve the judgment on how to walk or which direction to follow."*

Inclusive and Accessible Approach The system was seen as inclusive and accessible, providing essential tools for visually impaired individuals to navigate and understand their environment effectively. P17 noted the system's inclusiveness: *"They're not just walking. It's very inclusive and accessible for them because they have the board which in the virtual reality is actually like the textures they feel from the stick. So I think it's immersive for them."* P9 also commented on the system's comprehensiveness: *"It would be pretty immersive because the system has everything what a visually impaired person would need. Mainly the audio which was very very effective for me and also the cane which is actually essentially a real cane and the surface built around it."*

5.4.6 Problems faced during navigational tasks

Participants discussed the challenges and problems they encountered while performing navigational tasks without visuals. The feedback emphasized the difficulties with direction, confusion, auditory distractions, and the overall complexity of navigation.

Difficulty with Direction and Orientation

Many participants reported difficulties in maintaining a sense of direction and orientation without visual cues, often leading to confusion and uncertainty about their path. P10 expressed concerns about navigating correctly: *"The trouble always is like, am I going in the right direction? Am I walking into something? Am I going the right way?"* Similarly, P11 mentioned getting lost frequently: *"Without visuals, I found myself getting lost a lot. So that was a little bit puzzling and confusing."* P14 also found it challenging to get a sense of direction: *"Without the visuals, it was really hard to get a sense of direction."*

Confusion and Overwhelm

Participants reported frequently experiencing confusion and feeling overwhelmed by the lack of visual information, making it challenging to make decisions and navigate effectively. P18 described the initial challenge: *"In the beginning it was a bit challenging because I was very overwhelmed. I didn't know where to focus, what to focus on."* P2 discussed the difficulty in self-location: *"You are not able to locate yourself are you going in the right path or not and locating yourself in that particular situation is quite difficult."* P15 felt lost within the environment: *"So at some point, I felt I was lost in the environment, which I don't know which direction should I go."*

Auditory Distractions and Challenges

Navigating without visuals heavily relied on auditory cues, but participants often found it challenging to distinguish relevant sounds due to background noise and other auditory overlaps. P13 noted the overpowering noise of cars: *"The cars were really loud, so it kind of made it harder for me to hear other things that could be more helpful."* P6 and P19 mentioned the difficulty with auditory congestion: *"... So it was hard to hear the crosswalk or even like anything else around me or the sidewalk or the gravel." "That was a bit of an annoying. There was a lot of congestion in the sound."*

Challenges with Haptic Feedback

Some participants noted difficulties with haptic feedback as well, finding it hard to differentiate between surfaces and relying more on auditory cues for navigation. P12 expressed a challenge in estimating proximity using haptic feedback alone: *"For example, one time I was near the hot dog stand it was difficult for me to identify how far I was standing from the hot dog stand but that was not the case when I was doing visual."* P8 and P3 discussed the indistinguishable nature of different surfaces: *"I found like the sidewalk and* *road to be very similar and then the gravel and the metal to be quite similar to each other and then grass for some reason I could not feel it at all." "One problem is the surface, because that is creating a vibration. Like this surface, that overlaps with the haptic."*

Mental Strain and Stress

The lack of visual information was reported to have increased mental strain and stress, as participants had to rely solely on other senses to navigate. P18 discussed the mental confusion caused by relying on auditory cues: *"At one point I was just very confused about the sense of direction I was facing because, yeah, I was trying to align myself based on the sound of the traffic, but it didn't really help because there was a lot of noise in the background."* P19 and P17 also highlighted the anxiety and the focus required to navigate by sound alone: *"I was a bit scared not finding the right position or not finding the right destination." "Sometimes I'll be like, you know, I'll be straining my ears for specific sounds or cues."*

5.4.7 Difference in Navigation Experience Based on Visual Cues

Participants shared insights into their navigation experiences, delineating the stark contrasts between navigating with and without visual aids. These discussions highlighted the facilitative role of visual cues in easing navigation and reducing cognitive load, as well as the challenges and increased mental efforts required when relying solely on non-visual senses.

Ease and Speed of Navigation With Visuals

Visual cues significantly enhanced the ease and speed of navigation. Participants reported that visuals allowed for quick assessment of surroundings and more intuitive decision-making. P10 noted the stark reduction in cognitive load when visuals were available: *"The cognitive load was like way different because when I could see, I feel like I didn't have to think as much and it was more intuitive just because that's what I'm used to."* Similarly, P11 observed a faster completion of tasks: *"I noticed that I got everything done a lot faster when I had the visuals on."* P2 also highlighted the ease brought by visual aids: *"With visuals obviously it was like quick easier for me to do it."*

Reduced Cognitive Load Through Visual Information

The presence of visual information drastically reduced the cognitive load on participants, making navigation feel almost automatic and significantly easier. P10 emphasized the ease: *"Like very different. The ones where I could see were like immensely easier."* P18 reflected on the natural ease and familiarity of using visuals: *"The difficulty level certainly decreased by quite a lot. And yeah, it was much more natural because I'm so used to relying on the visual feedback."* This sentiment was reaffirmed by P6, who felt that navigating with visuals was nearly effortless: *"With visuals, it feels like almost automatic."*

Challenges of Navigation Without Visual Cues

In stark contrast, navigating without visuals was noted to have introduced significant challenges. Participants reported that they needed to rely heavily on auditory cues and other sensory information, which required greater mental effort and careful decision-making. P10 detailed the increased cognitive demands: *"I guess more decisions to be made and it was I guess more mental power to be like where are you? Are you going in the right direction? Are you getting hit by a car? Are you in a trash can?"* P2 and P9 spoke about the difficulties in orientation and the heightened mental load: *"You are not able to locate yourself are you going in the right path or not and locating yourself in that particular situation is quite difficult."* *"Without visual, it was a little hard because I had to rely, I had to, the mental demand was high."*

Difference in Reliance on Sensory Cues

Participants' experiences highlight that the reliance on sensory cues varied significantly between navigating with and without visuals. When visuals were present, participants largely ignored other sensory inputs, relying almost exclusively on what they could see. P12 highlighted this reliance: *"It's visible then I 100% rely on the visual context rather than on audio or the feedback on the cane."* In contrast, without visuals, participants mentioned to have been primarily dependent on auditory cues. P10 described their strategy: *"It was mostly just going where the sound were or weren't."* Both P11 and P2 emphasized the use of sound to guide their navigation: *"I would mostly use the distance of the audio to navigate from a traffic light or whatever."* *"I was completely depending* *on the sound."*

Increased Awareness and Concentration Without Visuals

While navigating in non-visual conditions, participants reported heightened awareness and increased concentration, making them more attuned to their surroundings. P6, P7, and P8 noted the heightened sensory awareness necessary when visual information was absent. P6: *"Usually outside of the simulation, it's like you know without even thinking about it, it's time to cross at a certain sound. But now you're actually like really paying attention. Like, okay, is this one crossing or is this one waiting?"* P7: *"The distance between you and the source of sound helped you to make decisions."* P8: *"The intensity of the sound was heightened because the sense of hearing was heightened due to not being able to see."*

5.4.8 The Role of Cane and Spatial Audio for Navigation

Role of Cane

Participants discussed the role of the cane in navigating their environment, emphasizing its importance in providing tactile feedback, aiding in orientation, and enhancing the sense of immersion.

Tactile Feedback and Surface Identification The cane was highlighted as essential for identifying different surfaces and textures, providing participants with additional information about their surroundings. P10 noted a significant change in tactile sensation between surfaces while crossing the road: *"I could notice that there was like a big change in like the feel. So I think that maybe I was aware if I was in traffic or on gravel."* P2 found the cane helpful when audio cues were unclear: *"When I was not able to figure out the audio so the cane was helpful in identifying which material I am on."* Similarly, P15 relied on the cane for specific navigational tasks: *"I was able to find the crosswalk based on the haptic mainly and also where you asked me to find the grass or I can't remember the other... Gravel."*

Orientation and Navigation Participants reported using the cane to help maintain their orientation and navigate through different environments, especially when visual cues were absent. P13 used the cane to distinguish different walking surfaces: *"Use the cane to like figure out if I was on the road versus on like the sidewalk."* P20 described how the cane informed them of their position relative to the road: *"So when I used to move the cane I got to know if I am on the off side of the road or I am walking in the pavement area."* P5 noted the cane's role in correcting their path: *"The cane also helped because I think that I was meant to walk on the sidewalk but then sometimes I felt that I am going off track with the cane with from this like the feedback from the cane and that I guess helped me to get on track."*

Sense of Immersion and Familiarity Participants highlighted the use of the cane enhanced their sense of immersion, making the experience more realistic and familiar. P11 emphasized the immersive experience provided by the cane: *"Definitely provide the sense of immersion in the sense that you know most people don't experience VI on a day-to-day basis."* P18 spoke about the role of familiarity in navigation: *"When you're trying to navigate through a new system. That kind of familiarity would play an important role."* P15 felt fully engaged in the experience, even when visuals were available: *"I guess cane was useful. Even when I could see, I'm not sure if you noticed, but for the task that I could see I was also using the cane the same way. Because I feel like I was really into the experience."*

Role of Spatial Audio

Participants highlighted the significant role that spatial audio played in navigating and understanding their environment. The feedback emphasized how spatial audio cues were essential for direction, location awareness, and task completion.

Supportive Navigation Participants reported the importance of audio cues in helping them navigate their surroundings, providing necessary information about direction and proximity to various elements in the environment. P1 mentioned the supportive nature of navigation: *"More of supportive navigation."* Similarly, P14 and P16 found audio cues vital for providing directions and understanding their proximity to streets: *"They were very helpful to provide me directions and the distance from where I was. So it was the main thing which guided me towards the target." "The cars voices is helping me to* *understand if it's like a street or something or I'm close to the street."*

Localization and Direction Participants noted that audio played a vital role in helping them localize their position and determine the direction they needed to go. P13 used audio to know his position: *"It was how I like localized where I was or how I knew where I was at like a certain, like, goal, certain area."* P20 and P18 relied on audio for finding directions and making navigational adjustments: *"Like to know the direction, like where I am going, where I am heading to. They helped me to find that. Find my goal or destination." "It was a cue to sort of readjust myself based on the traffic lights, the traffic sound, the people speaking. So definitely a huge role."*

Task Completion Audio cues were also noted to have played a crucial role in indicating task completion and helping participants stay on track. P19 appreciated the confirmation provided by completion sounds: *"I think the sound, the completion sound really helped me to acknowledge myself. I was like, okay, I actually found the place."* P20 and P5 discussed how sounds marked task completion and guided their navigation: *"But when I hear the completion sound I like, yeah, then I came to know the task is done." "The sound played a really important role while navigating through the tasks because there were audio cues that had to be kept in mind to achieve the checkpoints."*

Environmental Awareness The ambient sounds and specific audio cues were reported to have helped participants understand their surroundings and make informed decisions about their movements. P11, P16, and P6 highlighted the role of ambient sounds in navigation. P11: *"Less so the texture sounds and more of the sounds, you know, the ambience in the world was helping me get around."* P16: *"The surrounding environment audios, it was very clear to help me navigate."* P6: *"Trying to figure out where I'm going in terms of whether the crosswalk, like listening to the beeping. And then also hearing the difference in the ground."*

5.4.9 Comparisons

Cane vs Controller/Joystick

Participants discussed their experiences and preferences for using a cane versus a controller or joystick in the simulation. The feedback generally highlighted the realism, sensory feedback, and ease of use associated with using a cane.

Realism and Practicality The cane was reported as more realistic and practical for the target audience, providing an authentic interaction with the environment. P10 emphasized the realism suitable for the intended users: *"I think it's more realistic, especially for like the audience that you're looking to kind of work with."* P18 highlighted the practicality for those with sensory limitations: *"It's much more closer to reality. That's what people use. And for somebody who does not have a lot of their senses in place, I think they would prefer to rely on something more traditional rather than being acquainted to something entirely new."*

Feedback and Sensory Connection Participants highlighted the importance of tactile feedback provided by a cane, which enhanced the ability to sense and navigate surroundings effectively. P14 discussed the connection with the environment offered by the cane: *"The cane is like a better thing because it gives you a sense of connection with your surface it gives you how it feels in your hand you don't have to like look for it in a VR controller which wouldn't be like practical."* P3 emphasized the importance of tactile feedback: *"Definitely have an effect on the like haptics because I will not like now I am actually filling a floor that is like I am navigating in so without that like I can't fill the floor so that would completely miss the realism."*

Ease of Use Many participants reported the cane was easier and more comfortable to use, particularly for simple tasks within the simulation. P11 preferred the simplicity of the cane: *"For the sake of this exercise, I think I would have just preferred the cane honestly because of you know How simple the game was."* P20 discussed the comfort of using a cane: *"It's like it was quite comfortable rather than using that VR thing. Okay. Like it was a kind of personal thing for me so I can use it wherever I want."*

Slide Mill Comparison with Natural Walking

Participants expressed general dissatisfaction with the naturalness of walking on a slide mill compared to real-world walking, often noting differences in comfort and walking mechanics.

General Dissatisfaction with Naturalness Many participants pointed out the unnatural aspects of walking on a slide mill. P1 noted the lack of natural feel: *"Comparing with the real world scenario it's not that much natural."* P10 discussed the unnatural effort required to walk: *"The way that you take steps was different. I don't really normally have to think about walking a lot, but I had to really be like, do you hear your steps? Are you moving?"*

Adjustments and Effort Participants highlighted the need to make various adjustments to their walking style and exert more effort when using the slide mill. P3 compared the effort to real-life walking: *"It takes a bit more like hard work to work on a slide mill than real life but otherwise it is almost same."* P20 and P15 described the physical adjustments needed: *"Like I was not able to lift my leg a bit faster, like I want to move faster but I was not able to." "In normal walking, I guess we move more muscles. We have, yeah, it felt differently. But at first I need to lean a bit and try to walk, which is not how I normally walk. So that part was a bit different."*

Sensory and Physical Discrepancies Participants highlighted discrepancies in physical strain and sensory feedback, detracting from the realism of the slide mill. P13 mentioned discomfort in walking: *"I normally take like larger strides and my feet certainly hurt more than like normal walking around."* P5 and P2 discussed the strain and differences in surface feedback: *"It was more strain-full because of the action that needed to be performed." "When you're walking on a road you feel the surface with your feet. That is the only difference I think."*

Adaptation Over Time Despite initial challenges, some participants noted that they gradually got acquainted with the slide mill's mechanics, finding it more comfortable and similar to natural walking over time. P6 described the adaptation process: *"It's different from natural walking. But after doing it, it kind of like just became automatic."* P9 and P14 discussed getting accustomed to the slide mill: *"The walking stance is a little different in slide mail wherein I will have to like lean forward a little and then walk. But as I got used to it it was pretty comfortable." "It was really accurate and useful while turning it was just like it felt like I was walking in like real life."*

5.4.10 Feedback on the system

Improvement of Haptic Technology

Participants suggested enhancements to the haptic technology used in the system, advocating for more robust and distinct haptic feedback that closely mirrors real-life sensations. P13 highlighted the need for specificity and robustness in haptic feedback to align more closely with actual experiences: *"I think if like the haptic technology would probably have to be more robust and like more specific and so it could be like more aligned with what the real haptic like experience would be."* P2 and P3 expressed difficulty in distinguishing haptic signals, with P2 stating, *"Only thing sometimes I've felt was you know like the haptics well I'm not able to figure out the haptics,"* and P3 adding, *"If the differences of the haptics were a bit more, then it would be better."* Similarly, P5 and P19 noted challenges in differentiating between surfaces based on haptic feedback alone, with P19 commenting on the subtlety of the differences: *"In haptic, it was very hard to find out the difference between the surfaces. It was really minor difference. Sometimes I could not find the difference but the major part was the sound."*

Customization and Adaptability

The feedback highlighted a demand for greater customization and adaptability of the system to accommodate the specific needs of different users, particularly those who are visually impaired. P16 stressed the importance of customization: *"I think it can be helpful, but it's also depending on, like, if it can be, like, really customized through what they need, the blind people."* P9 pointed out the potential for more varied tasks: *"I mean if the task wasn't like walk from point A to point B which is straight if it was something like feel something and go to a certain spot then I think I would have used the cane more."*

Comfort and Effort

Participants also commented on the physical demands of using the system, suggesting the need to reduce effort and enhance comfort for a more natural experience. P17 addressed the effort involved in walking: *"The only thing I could say is maybe if it could require less effort in walking, making it more natural."* P20 highlighted the importance of comfortable standing postures: *"The one thing like the standing posture, like if you can make it more comfortable then that will be good."*

Environmental Feedback and Audio

Clear and distinct audio feedback was noted as crucial for better navigation and understanding of the environment. P3 suggested improvements to audio distinctions: *"If there was a headphone for the texture and there would be a speaker for the surroundings, then it would be better to distinguish the things better."* P8 and P6 discussed the effectiveness of spatial audio, with P6 commenting on toning down some of the sounds: *"Like the loudness of the cars. I think that might, like personally, like how loud they were. And just, I couldn't really, like, it kind of mixed in everything else."*

Impressions and Recommendations

Overall, participants noted a positive impression of the system, with recommendations aimed at enhancing its utility and effectiveness. P2 and P12 discussed the initial learning curve and the effectiveness of vibrations, with P2 saying, *"Probably after multiple chances you are able, it is a good thing to do it but for the first time is I would not go for it,"* and P12 noting, *"The level of vibration it gives on different surface I was not able to feel it."* P5 praised the system's ability to visualize the world for those with impairments: *"A good visualization of the world that people with impairments cannot see."* Suggestions included better surface differentiation and incorporating omnidirectional movement and environmental elements like water and rain. P1 suggested improvements for immersion: *"The slide mill, if you had more omnidirectional side mill, that would be more good in terms of immersion."* P14 and P19 both noted the importance of realistic textures and environmental factors. P14 observed, *"In the beginning as well it was difficult for me to like differentiate with different surfaces if it could be like level and surfaces could be made* *more like subjective and different from each other,"* while P19 mentioned, *"In terms of improvement I think if you have if you have other textures like water you know rain spray surroundings then I think it will increase the immersivity."*

5.4.11 Challenges and Limitations of different facets

System as a training tool

While the system was reported beneficial, some participants mentioned challenges and limitations, suggesting that improvements are needed for it to be fully effective as a training tool. P19 highlighted the initial suitability but questioned the system's long-term training efficacy: *"I think the system is good to start with but definitely not good to train people you know for the person who actually lost their eyesight."* P1 and P20 discussed the real-world applicability and adaptation process, with P1 stating, *"Real world scenario which might have a bit danger to navigate around the real scenario,"* and P20 highlighting *"Initially they will find a bit challenging but when they get acquainted with the system, I think it will be easy for them."*

Shortcomings of Audio

Some participants faced challenges with spatial audio, noting that certain sounds were too loud or drowned out other important cues. P6 and P13 expressed issues with the overpowering noise of vehicles: *"The vehicle sounds kind of like drowned over everything else." "The cars were really loud."* Further issues with audio cues were highlighted by P3, P7, and P9, who discussed the ineffectiveness and inaudibility of certain cues. P3 noted, *"Yes, it played a role, but that's not that significant, "* while P7 reaffirmed, *"I cannot hear anything from the headphone, I cannot find the way,"* and P9 stated, *"Used it to navigate towards it or away from it depending on the tasks."*

Pain points of Haptic Feedback

Many participants reported difficulty in differentiating the various haptic feedback profiles, finding the differences too subtle. P11 expressed frustration with the lack of distinctiveness: *"Could be a little bit more distinguishing factors from the actual textures because even in the tutorial I struggled to differentiate them."* P2 and P18 confirmed this sentiment, with P2 stating *"The haptic feedback was I was not able to figure out between the gravel and the road,"* and P18 noting, *"For me, I think I couldn't really differentiate between the different haptics"*. Moreover, several participants found the haptic feedback realistic and helpful only when combined with audio cues. P1 described the synergy between haptic and audio feedback to be effective: *"So when it combines with the sound it feels like more natural. But if you separate from the sound it would be very difficult to know whether it's more natural or not."* Similarly, P10 and P3 noted specific examples of realistic feedback when combined with audio: *"But I think the gravel one was pretty realistic, I would say, especially with like the noise at the beginning. And maybe the sidewalk one too."*

Limited Use of Cane

While the cane is intended to be a primary navigational aid, some participants faced challenges using it effectively, often opting to rely more on audio cues. Participants used sound-based strategies to navigate effectively in the absence of visuals, demonstrating adaptability but also indicating a need for improved tactile feedback through the cane to promote its overall usage. P16, P19, and P9 discussed their limited use of the cane, focusing instead on environmental sounds. P16 stated their reliance on spatial audio more: *"I tried, but it was not that of an important role, honestly, because the sounding voices were so loud, so I was relying on the voices more than the texture."* P19 reported not to have used the cane for the majority of the time: *"Cane I did not use at all, mostly I would have used 10% of the time of all the tasks combined. But other than that, I only relied on the environmental audio rather than the cane audio."* P9, on the other hand, reported using the cane for specific purposes: *"I was sliding my cane to find out if I was walking on the road or the sidewalk."*

5.5 Summary

In assessing the effectiveness of the system through the System Usability Scale (SUS) and NASA Task Load Index (NASA-TLX), significant differences emerged based on whether visual aids were available. The SUS results indicated a statistically significant higher usability score when visuals were integrated, suggesting that visuals substantially enhance the ease of use and overall system effectiveness. This preference was consistent across the demographics. Conversely, the NASA-TLX results highlighted increased cognitive and physical demands in the absence of visual cues, as participants found navigating and completing tasks more challenging without visual assistance. This indicates that visual aids not only reduce the effort required to interact with the system but also significantly improve user performance and satisfaction. Furthermore, qualitative feedback from participants reinforced the quantitative findings, with many expressing a preference for the visual condition, which they found less taxing and more intuitive. Participants noted that the absence of visuals not only made the tasks more challenging but also heightened their awareness of other sensory inputs, which although educational, led to a more strenuous experience.

The task-based analysis revealed that the type of sensory feedback provided crucially impacts task performance, particularly in texture recognition and navigation efficiency. Tasks performed with visual aids showed markedly higher accuracy and reduced completion times, underscoring the essential role of visual information. In contrast, reliance on auditory and haptic feedback without visual support did not fully compensate for the absence of visuals, as evidenced by lower accuracy and increased task times. This suggests a need for improved sensory feedback designs to better support users in environments where visual cues are restricted or absent.

Behavioral analysis further deepened the understanding of user interaction under varied sensory conditions. Participants adapted their strategies by relying more on auditory cues and moving more deliberately when deprived of visual information, which not only heightened their spatial awareness but also increased their empathy for visually impaired individuals. These behavioral shifts are crucial for navigating complex environments and underscore the importance of designing sensory feedback that can adapt to the needs of users with sensory limitations. This analysis highlights the significant impact of sensory experience on user engagement and the potential of virtual environments as empathetic training tools.

Chapter 6

Discussion

The overall research aimed to create and test a system enabling large-scale VE exploration, especially textural identification and a mix of egocentric and allocentric navigation through sensory modalities other than visuals, namely, hearing and touch. The research constituted an evaluation of the overall technical efficacy and usability of the system through a comparative user study involving the performance of sighted participants in visual and non-visual conditions. Moreover, the study acted as a pilot to gain insights into empathetic literacy engendered among participants and identify key implementational shortcomings of various components to make it more robust for future study.

Here, in the thesis discussion chapter, I summarize the findings by discussing the implications of the results obtained through the study to answer the three primary research questions. I detail the scope of the current work and talk about the degree of reproducibility for anyone interested in replicating the study or taking it forward. I present the notable strengths of the study in terms of how it contributed to the domain of VR research that focuses on the foundational ramifications of visual independent usage by identifying gaps in the current literature and trying to fill them using accessible and inclusive practices. Moreover, I point out certain shortcomings observed throughout the research regarding the study design and various aspects during the prototype development. Lastly, I propose directions for future work, including a structured study that involves domain experts from the start and involves evaluation with actual non-sighted participants.

6.1 Implication of Results

This section narrows down the findings to answer each research question that formed the basis of the user study design as explained in Section 4.1. Specifically, it summarizes the qualitative and quantitative data collected for each of the research questions, mixed methods for analysis and the implications drawn from it that reveal usage patterns and experiential findings.

RQ 1: Level of Differentiation Between Virtual Textures

To what extent can users differentiate between virtual textures in non-visual conditions using the system?

The study collected quantitative data through the surface identification task (Task 1), which involved participants using the cane to perceive and identify virtual textures based on three conditions: audio and haptics, only audio and only haptics. The analysis utilized confusion matrices and box plots to visualize participants' ability to identify different textures and the time taken to correctly identify them, respectively. The matrices showed the overall success rate of texture identification and the specific textures that were more easily or less easily identified. Moreover, the statistical analysis for the study data employed a 2-way ANOVA to examine the impact of sensory modalities (audio and haptics) and texture types of the virtual surfaces on identification success rates.

To answer the research question, even when devoid of visuals, the participants successfully identified virtual textures in all variations of the tasks, using only audio, haptics or both. However, the identification success rate varied across the task and among the virtual textures within a variation. Participants performed marginally better when they had only audio to work with than when both the auditory and haptic feedback were available and suffered a decline in performance when haptics was the only available modality. Statistical analysis revealed that the individual effects of virtual surface textures and the presence or absence of audio and haptic feedback significantly influenced the success rates of texture identification, more so than the interaction between these factors. Additionally, the success of textural identification followed a pattern where texture that resembled soft surfaces (grass) was easier and most accurate to identify, while texture that shared similar physical properties, such as gravel, road and sidewalks, were often mixed and confused.

The performance (success rate and time taken) and inferential (observations and self-reported) results for identifying textures indicate the reliance on and preference for auditory feedback is more apparent and natural than haptic feedback for sighted participants in non-visual conditions. The availability of both modalities, however, accounts for an overall balanced outcome, with one modality compensating for the other, especially in conditions where auditory cues overlap and become difficult to discern, such as during navigatory tasks (Task 2 and 3).

RQ 2: Effectiveness of Audio and Haptic Feedback in VE Navigation

How effectively can users navigate an unfamiliar virtual environment using only audio cues and haptic feedback without any visual information?

The road crossing (Task 2) and scavenger hunt (Task 3) tasks were specifically designed to evaluate and compare participants' ability to navigate the VE under two conditions: one with visuals, audio, and haptic feedback and the other with only audio and haptic feedback. Data on task completion time, target location errors, and vehicular collision rates were collected and analyzed. The statistical test involved testing the normality of the task completion time, followed by parametric and non-parametric tests to assess their statistical significance. Moreover, the descriptive statistics provided insights into participants' navigation performance. Positional data from Task 3 allowed for a detailed analysis of navigation routes opted by participants and helped to understand how participants utilized available modalities to make navigational decisions.

The statistical test and descriptive statistics indicate significant differences in performance metrics between conditions with and without visual information. Without visual information, the participants significantly took more time to complete the tasks and experienced a higher rate of vehicular collisions. Path analysis revealed during the third task that participants without visuals exhibited a cautious approach and spent more time interacting with targets along the way, translating to prolonged task completion time. On the contrary, with visuals, they appeared more confident, precise and direct, contributing to an overall decreased time to complete the task. Despite this, the majority of the participants succeeded in completing the navigation tasks, with only a few outliers as an exception. Moreover, Spatial audio aided in both egocentric and allocentric wayfinding toward targets, while haptic feedback about textures guided orientation.

In summary, while sighted individuals can effectively navigate the VE using only audio cues and haptic feedback, their performance is slower and more prone to errors compared to navigation with visual information.

RQ 3: Impact of Visual Information on Cognitive Load

In the absence of visual information, does navigating the virtual environment reduce cognitive load for individuals accustomed to visual cues, or does their inherent reliance on these cues increase their overall cognitive load?

The study used the NASA-TLX questionnaire to assess cognitive load and the SUS questionnaire to evaluate the usability of the system under different conditions (with and without visual information) for Tasks 2 and 3. The normality of NASA-TLX scores was assessed using the Shapiro-Wilk test, followed by appropriate statistical tests to identify significant differences in cognitive load between conditions. The SUS score was obtained through standard procedure, and box plots were used to visualize and compare scores across conditions.

The analysis revealed that the absence of visual information did impact cognitive load, with participants reporting higher cognitive and physical demands when relying solely on audio and haptic feedback. The results suggested that while the system was usable, the increased cognitive load in non-visual conditions highlighted participants' reliance on visual cues for navigation. Statistical analysis confirmed non-parametric data distributions, further supported by higher median task completion times and more frequent vehicular collisions. Experiential data from interviews reinforced these findings, with many describing heightened mental strain and difficulty in maintaining orientation and direction using only auditory and haptic feedback. They noted that the lack of visual cues required them to rely heavily on other senses, which increased cognitive load and made navigation more demanding.

In conclusion, the absence of visual information in VE navigation increases the overall cognitive load for sighted users, as their innate reliance on visual cues necessitates greater mental effort to compensate with auditory and haptic feedback.

6.2 Notable Strengths of the Research

Large Scale VE with Standardized Interaction Interface

The literature review identified significant gaps in the current approaches to VE exploration as a means of accessible technology for visually impaired users. These gaps primarily stem from the VE scale being limited by the available physical area or the use of interaction interfaces like VR controllers, which are quite different from the standard white cane used in real life. Consequently, the disparity was that the research either utilized a limited physical area, driving exploration in a small-scale VE, or, if a large-scale VE was employed, it lacked 'cane-like interfaces.' The current work successfully addressed both of these gaps. Not only was the VE large-scale, featuring a considerable-sized cityscape, but the mapping of the physical cane to a virtual cane in the VE also facilitated exploration using a standard and familiar tool from real life.

Accurate Depiction of the Real World

As evident from the participant interviews, the simulation of the cityscape shared notable similarities with the real world. Since the 3D model of the city was based on a modern urban environment, architectural features along with various infrastructural and natural elements were present, adding to the dynamicism of the city. The presence of a variety of city-specific audio, spatialized with characteristics that simulated physics in terms of binaural sound, occlusion and attenuation, contributed to the depth and complexity of the experience. In addition to inanimate objects, the presence of avatars engaged in various activities and the interactiveness of various components made possible through the use of the cane and navigation that closely resembled natural walking all enhanced the realism and dynamic quality of the simulated environment.

Exploration through the Slidemill

Navigation in VR involving modalities such as physical walking or redirection increases the realism of the experiences but comes with certain caveats. Firstly, there is a spatial limit to the area that can be covered, and secondly, there is an omnipresent risk of tripping and hitting objects or walls. When highly immersive systems for visually impaired individuals are to be tested, the risk is even more significant. The use of the omnidirectional slide mill in the current work was a novel approach for many of the participants, among whom few reported fatigue and discomfort over prolonged usage; however, the trade-off in the positive sense was the safety factor. Moreover, using the slide mill enabled the exploration of a VE that could not have been mapped within the confines of an indoor space. The immersiveness of any VE is highly susceptible to comparisons made between the real world and the simulation. In this regard, the in-place walking methodology employed by the slide mill aided in increasing the immersive quotient of the overall experience.

Single-trial Success of the study

Since there was a time constraint to the duration of the study, primarily as a precaution to avoid overburdening participants with physical activity that might negatively bias their experience due to fatigue, the tasks designed for the study comprised only a single trial of each task. For the repeated tasks, each trial was considered single as it differed in conditions, making them more like variations.

One of the strengths of the current research is the high completion rate and accuracy of participants in all tasks under non-visual conditions, with results comparable to the same tasks performed with visuals. Even with limited training and the system's novelty for most participants, the results favor the single-trial success of the study. This is a promising indicator for future research, suggesting that task accuracy may improve if the study design incorporates multiple trials.

Practical Relevance

The empathetic awareness reported among participants demonstrates the significant potential of the current system to raise consciousness about the challenges faced by visually impaired individuals. The increase in the overall empathy is a testament to the system's capabilities as a pedagogical tool. Moreover, the current system's prospects for further improvement highlight its practical validity as an assistive technology. Enhancements in haptic and auditory feedback, inclusive practices through user-centred design, and the inclusion of diverse user groups during the ideation and iterative development phases can transform the system into a more robust and effective tool for O&M training. This potential for empathy generation and practical training aligns the system with real-world needs, making it a valuable contribution to both assistive technology and empathetic education.

6.3 Limitations to the Current Approach

This section addresses the shortcomings of the research in terms of the overall study design process and prototype development. It highlights key areas of concern and explains how remediating these issues could have enhanced the overall quality of the research.

Lack of Domain Experts and People with Visual Impairments

Section 2.5.2 explains the need for accessibility considerations to be part of an application development process from the start; however, it is also important that the target population for whom accessibility is considered part of this process from the start. The UN Chronicle [125] explains this principle of participation through the motto “Nothing About Us Without Us,” which emphasizes the importance of including individuals with disabilities in decision-making processes that affect their lives. One of the limitations of this research is that although it focused on accessibility from the beginning, the needs and requirements of the target user base were inferred from contemporary literature that focused on congruent research fields. The depth and complexity of usage scenarios and nuanced experiences from actual visually impaired users could have provided a better understanding of certain aspects, such as variability in cane usage, handling techniques, various styles of cane tips, and specialized navigational and tactile perception techniques, that the literature may fall short of fully capturing.

Hersh and Johnson [52] emphasize the role of domain experts in assistive technologies to ensure they meet the actual needs and preferences of visually impaired users with ecological validity. These domain experts, mainly O&M specialists, are vital in providing indispensable insights into designing and implementing effective training programs for visually impaired individuals that help bridge the gap between theoretical design and practical application. Expert interviews could have been one way of consulting domain specialists in the research, which would have proved helpful in better structuring the training phase of the study and also during the audio design of the VE. As mentioned in the research by Siu et al. [108], where they were able to accommodate various gripping styles and usage techniques of their cane prototype, in this research, experts and individuals from the population would have been able to provide better insights into introducing variability of different types of cane tips and cane usage techniques.

Simulation of the city with infrastructure supporting differently abled

One of the participants pointed out that the cityscape VE accurately depicted an accessible infrastructure readily available in modern cities. However, they also highlighted the need for the VE to be customizable as 'standards' for an accessible infrastructure might differ drastically from one region to the other. This applies to O&M training as well, where navigational techniques and cane usage vary significantly based on infrastructural availability or inadequacies. Our research involved a city simulation with infrastructural elements such as studded pavement blocks, audio-enabled pedestrian crossings and disabled-friendly bus stops, which are loosely based on accessible features of a modern city. This approach can be seen as very rigid as it, in any way, doesn't support accessibility in a universal sense.

Potential Observer Bias

The study employed noting down observations during and after the study through video-recorded participant interaction. This approach ensured the integrity of the observations so that key events wouldn't be missed, important nuances could be reaffirmed, and any particular observational biases could be mitigated. However, the issue with this approach was that for both the iterations of the in-situ and ex-situ observations, only a single observer was responsible for data collection and interpretation. This exposes the research to the risk of potential observer bias.

In the study, as suggested by Levit et al. [74], several mitigation strategies could have been employed, such as the 'blind review method' and 'multiple observer method' to circumvent observational biases. Using the blind review method, an observer is unaware of the specific conditions or hypotheses being tested, whereas in methods using multiple observers, observations are noted, and recordings are reviewed by independent analysts. The same could have been implemented in the research where an independent observer noted down key nuances of participant interactions and multiple observers performed video coding to point out ex-situ observations.

Limited Aspects covered during Training

In the study design, the training constituted around twenty minutes of the study time, whereas the testing made up more than fifty minutes. Moreover, during the training phase, participants were trained under ‘visual conditions’ on how to use the system in terms of holding, handling and sweeping the cane, along with getting acquainted with how to use the slide mill for navigation. There were no complex sets of spatial environmental audio involved during this session, mostly to prevent any possible learning effect that could have altered the data in the testing phase. However, the testing phase involved a mix of visual and non-visual conditions in settings that were unchartered to the participants and involved a drastic spatial and auditory shift comprising complex navigatory tasks.

As noted by Lathan et al. [71], comprehensive training is crucial in VR applications to reduce user frustration and enhance the overall effectiveness of the system. Thus, speculation exists that the disparity between these training and testing scenarios could have resulted in an increased reported cognitive load among participants, as their performance could have suffered due to a lack of comprehensive preparation. Thus, the training could have been better modelled to align it more closely with the test environment so that it would encompass the full range of tasks and the scope of requirements for better participant preparedness.

Limited Environmental Variables

The VE visually comprised a range of obstacles commonly found in a real cityscape; however, for the participants during non-visual conditions, the interaction using the cane was limited to just identifying virtual surface textures. One way to better exploit cane usage would have been to introduce some force-feedback methodology to define the tactile periphery of these virtual obstacles. Moreover, the assumption that participants would mostly stick to only pavements and roads led to collisions being recorded only for vehicles and not with any other obstacles such as buildings, walls or props, which constricted the scope of data that could have been collected.

The lack of auditory complexity is an issue highlighted by a few participants where the audio quality was either not complex enough to accommodate nuances of a real city or was too loud, leading to sound overlaps. Specifically, the richness of the audio lacked depth in terms of the audio through the localized speech of human interactions, and the pitch of passing by vehicles at times enveloped other sound sources. Additionally, the sound design of audio targets to guide allocentric navigation should have been modelled to blend with the environment but rather appeared to stick out and sound a bit too direct.

Limitation due to hardware constraints

Commercially available hardware that was used in the system posed technical constraints in terms of fidelity and the overall experience. The VR controller motor, responsible for rendering haptic feedback, struggled to deliver differentiated tactile sensations, making it challenging for participants to discern between various virtual textures. The controller, however, is partly to blame as the haptic rendering engine had limited implementations in modulating haptic sensations.

Furthermore, the omnidirectional slide mill introduced its own set of challenges. Participants reported difficulties with the slide mill's functionality, particularly while maintaining a natural walking stance. The safety of the slide mill came at a cost as the safety harnesses constrained movement, often resulting in awkward and less realistic walking patterns, fatigue after prolonged usage, and potentially affecting the contextual authenticity of the VE navigation tasks.

Additionally, the design of the slide mill skirt also resulted in issues such as the cane occasionally falling off the edge or getting stuck in the slide mill skirt. This was mainly due to the gap intentionally left behind between the slide mill base and the skirt to provide a safe space for facilitator movement during the study protocol. Although the frequency of these mishaps was minimal, they disrupted the simulation's continuity and potentially affected participants' performance and focus.

6.4 Future Direction

In this section, I discuss the directions of future work in two categories: enhancements and prospective research paths. Enhancements talk about the overall improvements that can be applied to address the current limitations of the research, along with a focus on including experts and a diverse user base from the start. Prospective paths of the research talk about the ways in which the applicability of the research can be extended, specifically in terms of it being used in the realm of skills transfer, assistance mediation, and pedagogical potential.

6.4.1 Overall Enhancements

User-Centered Design

As an improvement to the limitations discussed in Section 6.3, for future research, including domain experts and visually impaired users from the preliminary development phase would be a vital aspect of user-centric accessible design. Looking at the current implementation, the direct implication of the lack of stakeholder input is evident in terms of limited cane usage techniques, defining tactile fidelity of haptic feedback, and provisioning specific environmental cues for critical navigation decisions.

Expert interviews with O&M specialists and trainers would help introduce a range of different cane-holding and handling techniques, which are currently limited to only the cane sweep method. Additionally, their expertise could help integrate essential aspects of the O&M training to standardize the training in the VE. Moreover, individuals from the visually impaired community can be an integral part of the ideation phase and the feedback loop as testers. Their insights into navigation, the personalized techniques they use to detect virtual surfaces and obstacles, their adaptation of unknown spaces, and their use of audio cues to guide themselves can prove to be important assets for making informed design decisions. Creating a feedback loop where user input directly influences ongoing development, their perception of things around them and how they determine specific spatial characteristics to their advantage and a means to make everyday decisions can help model the system can evolve to meet the needs of its users better.

Inclusion of Diverse User Groups

The simulation of visual impairments to sighted participants through blindfolds was a calculated decision that was made at the very start of the user study design. As explained before, the idea was to use it as a basis for evaluating the current system by treating it as a vanilla product so that future design decisions can be inferred from it to guide a study with actual participants with VI. It was patent that in the conducted study, although valuable, the VE experience of sighted users lacked practical accuracy compared to their visually impaired counterparts if they were to perform the same set of tasks and share their experience. It is thus necessary that, in the future, the research strives to include participants with various levels of VI. Not only does this help translate the system's overall effectiveness in terms of the target population, but it also helps gain and compare accurate and authentic insights regarding the system's perspective from one participant with a certain level of VI to the other.

Enhanced Training Protocols and Addressing Limitations

Building on the discussion covered in Section 6.3, the inconsistency between aspects covered in the training and the actual testing impacted participant preparedness and ultimately hindered their performance. To mitigate this, future research should employ balanced training methodologies that realistically reflect the optimal amount of testing scenarios without the cost of any learning effect. More specifically, for instance, if participants are expected to perform in a scenario with highly complex audio feedback, then the training should present some semblance in terms of introducing a version of it in a simplified way. This is a crucial step as it provides contingency in handling any novelty introduced by the system while participants transition from training to testing.

Addressing the limitations introduced by environmental variables and hardware inadequacies is a crucial area for future research enhancement. The introduction of multi-layered auditory feedback is proposed to make the audio in the VE more realistic, which involves adding depth to the sounds of pedestrian interactions, adjusting the volume of various elements, and adding more depth to the spatial characteristics of the sound design. This would create a richer, more immersive auditory soundscape, allowing for a more nuanced perception of the VE. Replacing the haptic feedback mechanisms of the VR controller with custom voice-coil actuators using a dedicated audio synthesizing micro-controller assembly that was ditched mid-way in the current work (Refer to Section 3.2.4) could offer greater control over rendered haptic sensations. Further enhancing the interactivity of the VE, the addition of force feedback through the cane would provide a tangible sense of urban obstacles and props. An upgrade involving detachable filler layers is suggested to address issues with the slide mill skirt that would prevent the cane from slipping off the edge and allow facilitator movement in between sessions.

6.4.2 Prospective Research Paths

Longitudinal Studies on Skill Transfer

The current work focuses on training and testing a user in the VE all within a defined length of time; however, the model of the study can be extended to training the users in the VE and assessing their engendered skills in real-life scenarios. This type of skills transfer evaluation can be based on rendering particular areas of a real city into corresponding 3D models to maintain high fidelity to the actual environment. Through the use of such familiar environments, the immersiveness can be perpetuated through a more robust assessment of how well skills acquired in the VE translate to practical, everyday use.

Furthermore, this approach can be supported through long-term longitudinal studies to assess the long-term effectiveness of VE training on real-world navigation skills. This could involve tracking participants' performance over extended periods and in various real-life contexts to provide valuable data on the durability and impact of the training. Current O&M training framework can benefit from these studies as skills transfer studies would prepare trainees for the varied and unpredictable nature of real-life urban navigation in a safe VE, and longitudinal studies would help understand how well the skills acquired in the VE are retained and applied over a period of time.

Collaborative Research with Specialized Institutions

One of the primary deterrents in accessibility research is recruiting differently-abled individuals to participate in a study. Going forth, for visually impaired individuals to be a part of our work would only be possible through collaboration with rehabilitation centers and institutions specializing in VI. This is where national-level institutions like the Canadian National Institute for the Blind (CNIB Foundation) would become pivotal in connecting the research with potential participants from the community. Moreover, evaluation of future versions of the current work could also be facilitated through partnership with these specialized institutions by integrating VE training into existing rehabilitation programs. CNIB Access Labs [11] is an example of such a platform that supports lived experience accessibility testing through the involvement of testers from the visually impaired community, allowing them to share their personal experiences with accessible products and services. Thus, collaborative efforts could lead to developing tailored training modules that align with current best practices in O&M training.

Pedagogical Evaluation of Emotional Impact

An overall increase in the empathetic aptitude of almost all the participants was observed in the current research. Since these are inherently subjective experiences, it is difficult to quantify these emotional shifts and channel them as a pedagogical measure. Unlike our approach of using blindfolds to simulate absolute darkness and complete blindness, future research can benefit from introducing VI in a phase-wise manner.

As discussed in Section 2.2.2 and Section 2.5.3, the peril of VI simulation is the false projection of trauma, as most simulations fall short of accurately depicting the level of the impairment. This is particularly because VI is not definite but is a spectrum, and even among people with the same type of VI, the impairment is a snowflake condition. It is prudent that future research balances both of these aspects to gradually introduce different levels and types of visual impairments while maintaining clinical accuracy to convey contextual accuracy and inhibit misguided conclusions. The correctness of simulations can be iterated and improved upon through the feedback loop. The ultimate pedagogical benefit is enabling participants to experience a broader range of visual challenges and mitigating the risk of overwhelming participants while providing a more comprehensive understanding of VI experience.

Additionally, contrary to the current research using qualitative feedback to assess empathetic gains and emotional shifts, incorporating mixed methods in future research could further elucidate evaluation. Quantitative measures such as the Interpersonal Reactivity Index (IRI) [35][36] or the Toronto Empathy Questionnaire (TEQ) [113] are standard questionnaires that can be employed to gauge changes in emotional (emotional reaction an emotional response) and cognitive (intellectual apprehension of an emotional state) empathy levels pre- and post-simulation [113].

Chapter 7

Conclusion

This research addresses accessibility issues related to content and interaction in commercial VR applications, highlighting significant gaps in the current literature. It underscores the need for large-scale VEs that enable navigation beyond the confines of tracked physical space and the incorporation of interaction interfaces resembling the standard white cane. A hapto-acoustic VR system was created that employed an urban virtual environment using the Unity Game Engine, a physical cane prototype for interaction and an omnidirectional slide mill for navigation. A user study involving 20 sighted participants evaluated the system's effectiveness in textural identification and navigation tasks of varying complexity in conditions with visual cues and solely through audio and haptic feedback.

The study's hapto-acoustic implementation of the VE demonstrates the system's potential to facilitate quick adaptation of navigation skills among sighted individuals in non-visual conditions, even with minimal training and limited VR exposure. However, their inherent reliance on visual information underscores increased cognitive load and error rates in the absence of visual cues. While open questions persist regarding further refinement of haptic feedback and the potential for dynamically balancing audio levels with haptic intensity based on situational preferences, the study findings point out the critical role of sensory substitution in compensating for the absence of visual cues and future validation through studies involving visually impaired users.

Conclusively, this research contributes to the development of accessible VR systems through a novel white cane prototype, large-scale complex cityscape simulation, realistic spatial audio effects and a comprehensive evaluation demonstrating the system's potential to aid non-visual navigation in VR and engendering empathetic literacy among sighted users.

Bibliography

[1] ACM Digital Library. `https://dl.acm.org`. [Accessed: 23-06-2024].

[2] Google. `https://google.ca`. [Accessed: 23-06-2024].

[3] Google Scholar. `https://scholar.google.ca`. [Accessed: 23-06-2024].

[4] IEEE Xplore. `https://ieeexplore.ieee.org`. [Accessed: 23-06-2024].

[5] Interhaptics. `https://www.interhaptics.com/`. [Accessed: 2024-07-05].

[6] Mixamo. `https://www.mixamo.com/`. [Accessed: 2024-07-05].

[7] Open broadcaster software. [Accessed: 2024-05-15].

[8] Springer Link. `https://link.springer.com`. [Accessed: 23-06-2024].

[9] Actuator 2018 : 16th international conference on new actuators : 25-27 june 2018, bremen, germany. Berlin, Germany, 2018. Institute of Electrical and Electronics Engineers, author., VDE Verlag.

[10] Revision of visual impairment definitions in the international statistical classification of diseases. *BMC Medicine*, 2023.

[11] CNIB access labs. `https://cnib-accesslabs.ca/`, 2024. [Accessed: 2024-07-08].

[12] Emily Ackerman. Virtual reality has an accessibility problem. *Scientific American Voices*, September 2021.

[13] Sun Joo (Grace) Ahn, Jeremy N. Bailenson, and Dooyeon Park. Short- and long-term effects of embodied experiences in immersive virtual environments on environmental locus of control and behavior. *Computers in Human Behavior*, 39:235–245, 2014.

[14] Samantha L Allison, Anne M Fagan, John C Morris, and Denise Head. Spatial navigation in preclinical alzheimer's disease. *J Alzheimers Dis*, 52(1):77–90, February 2016.

[15] Ronny Andrade, Steven Baker, Jenny Waycott, and Frank Vetere. Echo-house: exploring a virtual environment by using echolocation. pages 278–289, 12 2018.

[16] Ronny Andrade, Melissa J. Rogerson, Jenny Waycott, Steven Baker, and Frank Vetere. Playing blind: Revealing the world of gamers with visual impairment. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–14, New York, NY, USA, 2019. Association for Computing Machinery.

[17] Allan Armougum, Éric Orriols, Alexandre Gaston-Bellegarde, C. Joie-La Marle, and Pascale Piolino. Virtual reality: A new method to investigate cognitive load during navigation. *Journal of Environmental Psychology*, 65:101338, 2019.

[18] Motion Array. Motion array - unlimited downloads for creatives, 2024. [Accessed: 2024-07-04].

[19] Artlist. Artlist - royalty free music & sfx for video creators, 2024. [Accessed: 2024-07-04].

[20] Md Mahfuz Ashraf, Najmul Hasan, Lundy Lewis, Md Rashadul Hasan, and Pradeep Ray. A systematic literature review of the application of information communication technology for visually impaired people. *International Journal of Disability Management*, 11:e6, 2016.

[21] Audacity Team. Audacity. `https://audacityteam.org/`, 2024. [Accessed: 2024-07-15].

[22] Audiokinetic. Audiokinetic. `https://www.audiokinetic.com/en/`. [Accessed: 2024-07-05].

[23] Audiokinetic. SoundSeed grain. `https://www.audiokinetic.com/en/wwise/plugins/soundseed-grain/`. [Accessed: 2024-07-05].

[24] Audiokinetic. Spatial audio API geometry. `https://www.audiokinetic.com/en/library/edge/?source=SDK&id=spatial_audio_apigeometry.html`. [Accessed: 2024-07-05].

[25] Audiokinetic. Wwise. `https://www.audiokinetic.com/en/wwise/`. [Accessed: 2024-07-05].

[26] Audiokinetic. Wwise: A fundamental approach. `https://www.audiokinetic.com/download/documents/Wwise_Fundamental_Approach_en.pdf`. [Accessed: 2024-07-05].

[27] Audiokinetic. Wwise reflect. `https://www.audiokinetic.com/en/wwise/plugins/reflect/`. [Accessed: 2024-07-05].

[28] Audiokinetic. Wwise unity integration SDK. `https://www.audiokinetic.com/en/library/edge/?source=Unity&id=index.html`. [Accessed: 2024-07-05].

[29] Audiokinetic. Wwise soundseed grain plug-in, 2024. [Accessed: 2024-07-24].

[30] Joyce Ballemans, Gertrudis A. R. Zijlstra, Ger H. M. B. van Rens, Jan S. A. G. Schouten, and Gertrudis I. J. M. Kempen. Usefulness and acceptability of a standardised orientation and mobility training for partially-sighted older adults using an identification cane. *BMC Health Services Research*, 12:141, 2012.

[31] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123, 2009.

[32] Dolly Bhargava and Mike Offord. Vision impairment definitions an identification for professionals. `https://www.trinity.edu/sites/students-vision-hearing-loss/vision-impairment/definitions-identification-professionals`, 2024. [Accessed: 2024-07-24].

[33] Bruce B. Blasch, Richard L. Welsh, and William R. Wiener. *Foundations of Orientation and Mobility*. AFB Press, New York, NY, USA, 2 edition, 1997.

[34] Heather Culbertson, Joseph M. Romano, Pablo Castillo, Max Mintz, and Katherine J. Kuchenbecker. Refined methods for creating realistic haptic virtual textures from tool-mediated contact acceleration data. In *2012 IEEE Haptics Symposium (HAPTICS)*, pages 385–391, 2012.

[35] Mark Davis. A multidimensional approach to individual differences in empathy. *JSAS Catalog Sel. Doc. Psychol.*, 10, 01 1980.

[36] Mark Davis. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personalilty and social psychology*, 44:113–126, 01 1983.

[37] Francine L Dolins, Christopher Klimowicz, John Kelley, and Charles R Menzel. Using virtual reality to investigate comparative spatial cognitive abilities in chimpanzees and humans. *Am J Primatol*, 76(5):496–513, January 2014.

[38] Niklas Elmqvist. Visualization for the blind. *Interactions*, 30(1):52–56, jan 2023.

[39] Chris Esposito. User interface issues for virtual reality systems. In *Conference Companion on Human Factors in Computing Systems*, CHI '96, page 340–341, New York, NY, USA, 1996. Association for Computing Machinery.

[40] Arthur H. Everett and Phillip Ponder. *Orientation and Mobility Techniques: A Guide for the Practitioner*. American Foundation for the Blind, New York, NY, USA, 1976.

[41] Fabio De Felice, Floriana Renna, Giovanni Attolico, and Arcangelo Distante. Hapto-acoustic interaction metaphors in 3d virtual environments for non-visual settings. In Jae-Jin Kim, editor, *Virtual Reality*, chapter 2. IntechOpen, Rijeka, 2011.

[42] L. Fialho, Jorge Oliveira, A. Filipe, and Filipe Luz. Soundspace vr: spatial navigation using sound in virtual reality. *Virtual Reality*, 27, 11 2021.

[43] Matthias O. Franz and Hanspeter A. Mallot. Biomimetic robot navigation. *Robotics and Autonomous Systems*, 30(1):133–153, 2000.

[44] Johnny Friberg and Dan Gärdenfors. Audio games: new perspectives on game audio. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, ACE '04, page 148–154, New York, NY, USA, 2004. Association for Computing Machinery.

[45] Joseph L. Gabbard and Deborah Hix. A taxonomy of usability characteristics in virtual environments. Technical report, Virginia Polytechnic Institute and State University, November 1997.

[46] William W. Gaver. Auditory icons: Using sound in computer interfaces. *Human–Computer Interaction*, 2(2):167–177, 1986.

[47] Neveen I. Ghali, Omar Soluiman, Nashwa El-Bendary, Tamer M. Nassef, Sara A. Ahmed, Yomna M. Elbarawy, and Aboul Ella Hassanien. *Virtual Reality Technology for Blind and Visual Impaired People: Reviews and Recent Advances*, pages 363–385. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[48] Bruno Giordano, Stephen Mcadams, Yon Visell, Jeremy Cooperstock, Hsin-Yun Yao, and Vincent Hayward. Non-visual identification of walking grounds. volume 123, page 3412, 06 2008.

[49] Mira Goldschmidt. *Orientation and Mobility Training to People with Visual Impairments*, pages 237–261. 01 2018.

[50] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, 1988.

[51] Fernanda Herrera, Jeremy Bailenson, Erika Weisz, Elise Ogle, and Jamil Zaki. Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PLoS One*, 13(10):e0204494, October 2018.

[52] Marion A. Hersh and Michael A. Johnson, editors. *Assistive Technology for Visually Impaired and Blind People.* Springer London, London, 2008.

[53] Sangsun Han Hojun Aan and Kibum Kim. A multiplayer vr showdown game for people with visual impairment. *Human–Computer Interaction*, 0(0):1–21, 2024.

[54] HTC. Vive tracker 3. `https://www.vive.com/ca/accessory/tracker3/`, 2024. [Accessed: 23-06-2024].

[55] Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. Direct manipulation interfaces. *Hum.-Comput. Interact.*, 1(4):311–338, dec 1985.

[56] Dhruv Jain, Sasa Junuzovic, Eyal Ofek, Mike Sinclair, John R. Porter, Chris Yoon, Swetha Machanavajhala, and Meredith Ringel Morris. Towards sound accessibility in virtual reality. In *Proceedings of the 2021 International Conference on Multimodal Interaction (ICMI '21)*, page 12 pages, New York, NY, USA, 2021. Association for Computing Machinery.

[57] Waranuch Jeamwatthanachai, Mike Wald, and Gary Wills. Indoor navigation by blind people: Behaviors and challenges in unfamiliar spaces and buildings. *British Journal of Visual Impairment*, 37(2):140–153, 2019.

[58] Watthanasak Jeamwatthanachai, Mike Wald, and Gary Wills. Indoor navigation by blind people: Behaviors and challenges in unfamiliar spaces and buildings. *British Journal of Visual Impairment*, 37(2):140–153, 2019.

[59] Lingling Jin, Ruining Gong, Yaojun Yang, and Xiaochen Zhang. Design of audio-augmented-reality-based o&m orientation training for visually impaired children. *Sensors*, 22(23):9487, 2022.

[60] Alexis Kirke. *When the Soundtrack Is the Game: From Audio-Games to Gaming the Music*, pages 65–83. Springer International Publishing, Cham, 2018.

[61] Evelyne Klinger, Patrice L. Weiss, and Pierre-Alain Joseph. *Virtual reality for learning and rehabilitation*, pages 203–221. Springer Paris, Paris, 2010.

[62] Julian Kreimeier and Timo Götzelmann. First steps towards walk-in-place locomotion and haptic feedback in virtual reality for visually impaired. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*, pages Paper LBW2214, 1–6, New York, NY, USA, 2019. Association for Computing Machinery.

[63] Julian Kreimeier and Timo Götzelmann. Two decades of touchable and walkable virtual reality for blind and visually impaired people: A high-level taxonomy. *Multimodal Technologies and Interaction*, 4(4), 2020.

[64] Julian Kreimeier, Pascal Karg, and Timo Götzelmann. Blindwalkvr: Formative insights into blind and visually impaired people's vr locomotion using commercially available approaches. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '20)*, pages Article 29, 1–8, New York, NY, USA, 2020. Association for Computing Machinery.

[65] Rynhardt Kruger and Lynette van Zijl. Rendering virtual worlds in audio and text. In *Proceedings of International Workshop on Massively Multiuser Virtual Environments*, MMVE '14, page 1–2, New York, NY, USA, 2014. Association for Computing Machinery.

[66] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621, 1952.

[67] Katharina Krösl, Dominik Bauer, Michael Schwärzler, Henry Fuchs, Georg Suter, and Michael Wimmer. A vr-based user study on the effects of vision impairments on recognition distances of escape-route signs in buildings. *The Visual Computer*, 34, 2018.

[68] Katharina Krösl, Carmine Elvezio, Laura R. Luidolt, Matthias Hürbe, Sonja Karst, Steven Feiner, and Michael Wimmer. Cataract: Simulating cataracts in augmented reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 682–693, 2020.

[69] Orly Lahav. Virtual reality systems as an orientation aid for people who are blind to acquire new spatial information. *Sensors*, 22(4), 2022.

[70] Orly Lahav and D. Mioduser. Blind persons' acquisition of spatial cognitive mapping and orientation skills supported by virtual environment. *Virtual Reality & Assoc. Tech*, 4, 01 2004.

[71] Corinna E Lathan, Michael Tracey, Marc M Sebrechts, David M Clawson, and Gregory A Higgins. Using virtual environments as training simulators: Measuring transfer. *International Journal of Human-Computer Interaction*, 14(3-4):375–385, 2002.

[72] Jung-Suk Lee, François Thibault, Philippe Depalle, and Gary P. Scavone. Granular analysis/synthesis for simple and robust transformations of complex sounds. In *49th International Conference: Audio for Games*, London, UK, February 2013. Audio Engineering Society, AES.

[73] So Yeon Lee, Bharat Gurnani, and Fassil B Mesfin. Blindness. In *StatPearls*. StatPearls Publishing, Treasure Island (FL), January 2024.

[74] Heidi M Levitt, Michael Bamberg, John W Creswell, David M Frost, Ruthellen Josselson, and Carola Suárez-Orozco. Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA publications and communications board task force report, 2018.

[75] Anthony J. Maher and Justin A. Haegele. The authenticity of disability simulations through empathetic imaginings: the perspectives of visually impaired people. *Educational Review*, 0(0):1–18, 2022.

[76] Anthony J. Martingano, Faviola Hererra, and Sara Konrath. Virtual reality improves emotional but not cognitive empathy: A meta-analysis. *Technology, Mind, and Behavior*, 2(1), 2021.

[77] Daniela Massiceti, Stephen Lloyd Hicks, and Joram Jacob van Rheede. Stereosonic vision: Exploring visual-to-auditory sensory substitution mappings in an immersive virtual reality navigation paradigm. *PLoS One*, 13(7):e0199389, July 2018.

[78] Microsoft. Microsoft word. `https://www.microsoft.com/en-us/microsoft-365/word`, 2024. [Accessed: 2024-07-15].

[79] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, E77-D(12):1321–1329, 1994.

[80] Miro. Miro: The visual workspace for innovation. `https://miro.com/`, 2024. [Accessed: 2024-07-15].

[81] M. Mott et al. Accessible by design: An opportunity for virtual reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 451–454, 2019.

[82] Vladislav Nemec, Adam J. Sporka, and Pavel Slavik. Haptic and spatial audio based navigation of visually impaired users in virtual environment using low cost devices. In Christian Stary and Constantine Stephanidis, editors, *User-Centered Interaction Paradigms for Universal Access in the Information Society*, pages 452–459, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[83] Nox_Sound. Footsteps essentials. `https://assetstore.unity.com/packages/audio/sound-fx/foley/footsteps-essentials-189879`. [Accessed: 2024-07-05].

[84] OpenAI. ChatGPT 3.5, 2023. [Accessed: 2024-07-04].

[85] Lorenzo Picinali, Amandine Afonso, Michel Denis, and Brian F.G. Katz. Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge. *International Journal of Human-Computer Studies*, 72(4):393–407, 2014.

[86] Domenico Prattichizzo, Miguel Otaduy, Hiroyuki Kajimoto, and Claudio Pacchierotti. Wearable and hand-held haptics. *IEEE Transactions on Haptics*, 12:227–231, 07 2019.

[87] Texas Deafblind Project. Orientation and mobility training: The way to go. 2020.

[88] Benedikt Pätzold, Alexander Rochow, Maximilian Schreiber, Rasmus Memmesheimer, Christopher Lenz, Michael Schwarz, and Sven Behnke. Audio-based roughness sensing and tactile feedback for haptic perception in telepresence. *arXiv preprint arXiv:2303.07186*, 2023.

[89] QSR International. Nvivo. `https://help-nv.qsrinternational.com/20/win/Content/welcome.htm#`, 2024. [Accessed: 2024-07-15].

[90] Qualisys. Miqus motion capture camera. `https://www.qualisys.com/cameras/miqus/`, 2024. [Accessed: 23-06-2024].

[91] Nikunj Raghuvanshi, Rahul Narain, and Ming C Lin. Efficient and accurate sound propagation using adaptive rectangular decomposition. *IEEE Trans Vis Comput Graph*, 15(5):789–801, September 2009.

[92] Revoldiv. Revoldiv: Convert video/audio files to text. `https://revoldiv.com/`, 2024. [Accessed: 2024-07-15].

[93] Fabiana Sofia Ricci, Alain Boldini, Mahya Beheshti, John-Ross Rizzo, and Maurizio Porfiri. A virtual reality platform to simulate orientation and mobility training for the visually impaired. *Virtual Reality*, 27(2):797–814, 2023.

[94] Fabiana Sofia Ricci, Alain Boldini, Xinda Ma, Mahya Beheshti, Duane R. Geruschat, William H. Seiple, John-Ross Rizzo, and Maurizio Porfiri. Virtual reality as a means to explore assistive technologies for the visually impaired. *PLOS Digit Health*, 2(6):e0000275, 2023.

[95] J. C. Roberts, K. M. Franklin, and J. Cullinane. Virtual haptic exploratory visualization of line graphs and charts. `https://tdl.libra.titech.ac.jp/journaldocs/en/recordID/article.bib-01/ZR000000229851`, 2002. [Accessed: date-of-access].

[96] Mark D. Rodgers and Dr. Robert Wall Emerson. Human factor analysis of long cane design: Weight and length. *Journal of Visual Impairment & Blindness*, 99(10):622–632, 2005.

[97] Joseph M. Romano and Katherine J. Kuchenbecker. Creating realistic virtual textures from contact acceleration data. *IEEE Transactions on Haptics*, 5(2):109–119, 2012.

[98] Jeff Sauro and James Lewis. When designing usability questionnaires, does it hurt to be positive? pages 2215–2224, 05 2011.

[99] Lauri Savioja and U. Peter Svensson. Overview of geometrical room acoustic modeling techniques. *The Journal of the Acoustical Society of America*, 138(2):708–730, 08 2015.

[100] Bo N Schenkman and Mats E Nilsson. Human echolocation: Blind and sighted persons' ability to detect sounds recorded in the presence of a reflecting object. *Perception*, 39(4):483–501, 2010.

[101] Victor R Schinazi, Tyler Thrash, and Daniel-Robert Chebat. Spatial navigation by congenitally blind individuals. *Wiley Interdiscip Rev Cogn Sci*, 7(1):37–58, December 2015.

[102] Nicola S. Schutte and Emma J. Stilinović. Facilitating empathy through virtual reality. *Motivation and Emotion*, 41(6):708–712, Dec 2017.

[103] C.E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, jan 1949.

[104] Samuel S Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611, 1965.

[105] William R. Sherman and Allan B. Craig. *Understanding Virtual Reality.* Elsevier, Amsterdam, The Netherlands, 2019.

[106] Alex M. Silverman. The perils of playing blind: Problems with blindness simulation, and a better way to teach about blindness section. *Journal of Blindness Innovation and Research*, 5(2), 2015.

[107] Tania Singer. The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience & Biobehavioral Reviews*, 30(6):855–863, 2006. Methodological and Conceptual Advances in the Study of Brain-Behavior Dynamics: A Multivariate Lifespan Perspective.

[108] Alexa F. Siu, Mike Sinclair, Robert Kovacs, Eyal Ofek, Christian Holz, and Edward Cutrell. Virtual reality without vision: A haptic and auditory white cane to navigate complex virtual worlds. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*, pages 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[109] Brian Smith and Shree Nayar. The rad: Making racing games equivalently accessible to people who are blind. pages 1–12, 04 2018.

[110] Gordon P. Soong, Jan E. Lovie-Kitchin, and Brien Brown. Does mobility performance of visually impaired adults improve immediately after orientation and mobility training? *Optometry and Vision Science*, 78(9):657–666, 2001.

[111] Giovanni Spagnoletti, Leonardo Meli, Tommaso Lisini Baldi, Guido Gioioso, Claudio Pacchierotti, and Domenico Prattichizzo. Rendering of pressure and textures using wearable haptics in immersive vr environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 691–692, 2018.

[112] Speechify. Speechify - the text to speech app, 2024. [Accessed: 2024-07-04].

[113] R Nathan Spreng, Margaret C McKinnon, Raymond A Mar, and Brian Levine. The toronto empathy questionnaire: scale development and initial validation of a factor-analytic solution to multiple empathy measures. *J Pers Assess*, 91(1):62–71, January 2009.

[114] Mandayam A. Srinivasan and Cagatay Basdogan. Haptics in virtual environments: Taxonomy, research status, and challenges. *Computers & Graphics*, 21(4):393–404, 1997.

[115] Steam. Steam VR tracking. `https://partner.steamgames.com/vrlicensing`. [Accessed: 2024-07-05].

[116] Patrick L. Strandholt, Oana A. Dogaru, Niels C. Nilsson, Rolf Nordahl, and Stefania Serafin. Knock on wood: Combining redirected touching and physical props for tool-based interaction in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[117] 255 Pixel Studio. City package. `https://assetstore.unity.com/packages/3d/environments/urban/city-package-107224`. [Accessed: 2024-07-05].

[118] Jaime Sánchez and Mauricio Lumbreras. Virtual environment interaction through 3d audio by blind children. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, 2:101–11, 02 1999.

[119] Hisayuki Tatsumi, Yasuyuki Murai, Iwao Sekita, Shinji Tokumasu, and Masahiro Miyakawa. Cane walk in the virtual reality space using virtual haptic sensing: Toward developing haptic vr technologies for the visually impaired. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 2360–2365, 2015.

[120] Lauren Thevin, Carine Briant, and Anke M. Brock. X-road: Virtual reality glasses for orientation and mobility training of people with visual impairments. *ACM Trans. Access. Comput.*, 13(2), apr 2020.

[121] Luca Turchet, Rolf Nordahl, Stefania Serafin, Amir Berrezag, Smilen Dimitrov, and Vincent Hayward. Audio-haptic physically-based simulation of walking on different grounds. In *2010 IEEE International Workshop on Multimedia Signal Processing*, pages 269–273, 2010.

[122] Dimitrios Tzovaras, Konstantinos Moustakas, Georgios Nikolakis, and Michael Strintzis. Interactive mixed reality white cane simulation for the training of the blind and the visually impaired. *Personal and Ubiquitous Computing*, 13:51–58, 2009.

[123] Dimitrios Tzovaras, Georgios Nikolakis, Georgios Fergadis, and Stratos Malasiotis. Design and implementation of haptic virtual environments for the training of the visually impaired. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 12(2):266–278, 2004.

[124] UltraLeap. Getting started with virtual hands. `https://docs.ultraleap.com/xr-guidelines/Getting%20started/virtual-hands.html`. [Accessed: 2024-06-18].

[125] United Nations. Nothing about us, without us. `https://www.un.org/esa/socdev/enable/iddp2004.htm`, 2004. [Accessed: 2024-07-15].

[126] Unity Technologies. Unity xr interaction toolkit. `https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@2.2/manual/index.html`, 2022. [Accessed: 2024-06-30].

[127] Gianni Virgili and Gary Rubin. Orientation and mobility training for adults with low vision. *The Cochrane Database of Systematic Reviews*, 2010(5):CD003925, 5 2010.

[128] KAT VR. KAT VR SDK. `https://www.kat-vr.com/pages/sdk`. [Accessed: 2024-07-05].

[129] Dangxiao WANG, Yuan GUO, Shiyi LIU, Yuru ZHANG, Weiliang XU, and Jing XIAO. Haptic display for virtual reality: progress and challenges. *Virtual Reality & Intelligent Hardware*, 1(2):136–162, 2019. Haptic Interaction.

[130] Chyanna Wee, Kian Meng Yap, and Woan Ning Lim. Haptic interfaces for virtual reality: Challenges and research directions. *IEEE Access*, 9:112145–112162, 2021.

[131] Gareth R. White, Geraldine Fitzpatrick, and Graham McAllister. Toward accessible 3d virtual environments for the blind and visually impaired. In *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts*, DIMEA '08, page 134–141, New York, NY, USA, 2008. Association for Computing Machinery.

[132] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

[133] Michele A. Williams, Caroline Galbraith, Shaun K. Kane, and Amy Hurst. Just let the cane hit it: How the blind and sighted see navigation differently. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '14)*, pages 217–224, New York, NY, USA, 2014. Association for Computing Machinery.

[134] Claes Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, EASE '14, New York, NY, USA, 2014. Association for Computing Machinery.

[135] World Health Organization. Blindness and vision impairment. 2023.

[136] Jackie (Junrui) Yang, Hiroshi Horii, Alexander Thayer, and Rafael Ballagas. Vr grabbers: Ungrounded haptic retargeting for precision grabbing tools. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, page 889–899, New York, NY, USA, 2018. Association for Computing Machinery.

[137] Tina Yao, Soojeong Yoo, and Callum Parker. Evaluating virtual reality as a tool for empathic modelling of vision impairment: Insights from a simulated public interactive display experience. In *Proceedings of the 33rd Australian Conference on Human-Computer Interaction*, OzCHI '21, page 190–197, New York, NY, USA, 2022. Association for Computing Machinery.

[138] Wai Yu and Stephen Brewster. Multimodal virtual reality versus printed medium in visualization for blind people. In *Proceedings of the Fifth International ACM Conference on Assistive Technologies*, Assets '02, page 57–64, New York, NY, USA, 2002. Association for Computing Machinery.

[139] Wai Yu, Rameshsharma Ramloll, Stephen Brewster, and Beate Ridel. Exploring computer-generated line graphs through virtual touch. volume 1, pages 72 – 75 vol.1, 02 2001.

[140] Yuxiang Zhang and Ricardo Codinhoto. Developing a visually impaired older people virtual reality (vr) simulator to apply vr in the aged living design workflow. In *2020 24th International Conference Information Visualisation (IV)*, pages 226–235, 2020.

[141] Yuhang Zhao, Cynthia L. Bennett, Hrvoje Benko, Edward Cutrell, Christian Holz, Meredith Ringel Morris, and Mike Sinclair. Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, pages Paper 116, 1–14, New York, NY, USA, 2018. Association for Computing Machinery.

[142] Yuqi Zhou and Voicu Popescu. Dynamic redirection for vr haptics with a handheld stick. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2753–2762, 2023.

[143] Zoom Corporation. Zoom H5 Handy Recorder. `https://zoomcorp.com/en/ca/handheld-recorders/handheld-recorders/h5/`, [Accessed: 2024-05-15].

Appendix A

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for master's in computer science degree from the Dalhousie University, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the Dalhousie University in any scholarly use which may be made of any material in my thesis. Requests for permission to copy or to make other use of the material in this thesis in whole or part should be addressed to:

Head of the Faculty of Computer Science
6050 University Ave,
Dalhousie University,
Halifax, Nova Scotia, Canada B3H 1W5

Appendix B

Research Ethics Board Approval

Social Sciences & Humanities Research Ethics Board

Letter of Approval
February 08, 2024
Aayush Shrestha
Computer Science\Computer Science

Dear Aayush,
REB #: 2023-6992
Project Title: Haptics and spatial-audio based navigational approach in VR for the visually impaired
Review Type: Delegated Review

Effective Date: February 08, 2024
Expiry Date: February 08, 2025
The Social Sciences and Humanities Research Ethics Board has reviewed your application for research involving humans and found the proposed research to be ethically acceptable in accordance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans. This approval will be in effect for 12 months as indicated above. This approval is subject to the conditions listed below which constitute your on-going responsibilities with respect to the ethical conduct of this research.

Sincerely,
Dr. John Cameron
Chair, Social Sciences and Humanities Research Ethics Board
Dalhousie University

Appendix C

Consent Form

Project title: Haptics and spatial-audio based navigational approach in VR for the visually impaired
Lead researcher: Aayush Shrestha, Master's student
`ashrestha@dal.ca`
Faculty of Computer Science
Dalhousie University
Phone: (902) 399-4845
Other researchers:
Dr. Joseph Malloch (Supervisor) `jmalloch@dal.ca`
Matthew Peachey `peacheym@dal.ca`
Funding provided by:
NSERC RGPIN-2021-03845

Introduction

We invite you to participate in a research study conducted by Aayush Shrestha, a Master's student at the Faculty of Computer Science, Dalhousie University. Your decision to take part in this research is entirely voluntary. The information provided below outlines the details of the study, including your involvement, potential benefits, risks, inconveniences, and any discomfort you may experience. If you have any inquiries or concerns, please discuss them with Aayush Shrestha. You are encouraged to ask as many questions as you need to make an informed decision. If you have any questions later on, please contact him at `ashrestha@dal.ca`.

Purpose and Outline of the Research Study

The main goal of this study is to explore how good sighted individuals can become in navigating a virtual reality-based environment (VE) using only audio and haptic feedback without relying on any visual information. Additionally, we aim to test whether the absence of visual information increases mental effort or if reliance on visuals increases the mental load during VE navigation. Insights gained from this investigation will help to make changes to our current approach. This will form a basis for the future version of the system to be tested out with actual participants who are non-sighted or visually impaired. The purpose of this research and future researches based on this one is to explore the possibility of developing ways to make virtual reality accessible to people who are not able to use it due to some form of vision-related disability.

The study comprises two parts: training and testing. During the training phase, participants will get to know about how the system functions and what they can expect in the form of feedback. Later on, in the testing phase, participants will engage in three main tasks using the system with increasing levels of mental demand. Each task will have certain repetitions, and alternating tasks will increase in the level of complexity. The study will end with a couple of online surveys to collect data about the experience with the system and an interview session to gather feedback on what can be improved upon.

The study will require a maximum of 90 minutes to complete. Participants will be compensated with CA$20, which includes the time spent on the survey, interview, breaks, reading of the consent form, and completing the study. The payment will be given to the participants before they leave the lab upon completion of the study. In case of a withdrawal from the study in-between sessions, they will receive a pro-rated compensation based on the amount of time or study phases completed (specified under "Compensation/Benefits").

Who Can Take Part in the Research Study?

Prospective participants with familiarity or understanding of Virtual Reality (VR) and no prior experience of VR-induced motion sickness are eligible to participate in this study. However, participants will not be eligible to participate if they have visual or hearing impairments. Nevertheless, individuals with general visual acuity issues such as myopia and hypermetropia are eligible to participate. Moreover, other exclusion criteria in this study are defined to overcome practical issues related to the opportunity to participate in the study itself such as:

- Individuals with accessibility needs such as motor and cognitive disabilities that cannot be met through the user interface and head-mounted displays (HWDs) used to run the study.
- Individuals who cannot attend the study settings in person.
- Individuals who have a previous neurological illness, heart condition, or any medical condition that significantly restricts full-body mobility inhibiting them from standing and walking comfortably.

What You Will Be Asked to Do

If you decide to participate in this research, you must visit the lab on a specific date and time to participate in the user study. Firstly, you will be given a consent form to read and sign, following which the study will start. The study will take a maximum of 90 minutes divided into three parts - 20 minutes of training, 50 minutes of testing, and 20 minutes of post-study questionnaire and interview.

Our VE has two facets not commonly found in commercial VR experiences apart from head-worn displays (HWDs) and VR joysticks. These are a VR slide mill for in-game locomotion and a novel white cane prototype for virtual surface differentiation through haptic feedback. To familiarize you with these components, you will receive training on using them in the training phase.

During the testing phase, you will be required to use both devices to complete three tasks of different natures and varying cognitive loads. However, during this phase, you will not have access to any visual information and will have to rely solely on your sense of hearing and touch. While you perform these tasks, we will video record your interactions, screen record your in-game VE exploration performance, and log essential game statistics. This data will help us analyze our system's usability from a sighted user's perspective and identify any implementation shortcomings that may form the basis for refined future studies involving non-sighted participants.

Once you have completed the experiment, you will be asked to provide feedback about your overall experience through a survey and an interview. Your input and insights will be valuable in helping us understand the effectiveness of the study and identifying any areas that need improvement.

Possible Benefits, Risks, and Discomforts

Most commercially available VR experiences prioritize visual feedback over other senses such as touch and sound, making contemporary VR systems highly visually dominant. This has been found to exclude a diaspora of users who are in some way visually impaired. Your participation in this study helps us assess the usability of our system. Moreover, it helps to better understand the possibility of a VR system that can substitute visual dependency with other sensory modalities to create an equally immersive and inclusive VR experience. On a personal level, your participation may benefit you by helping you attain a level of empathetic literacy toward understanding how people with visual impairments navigate in real-life scenarios.

The risks associated with this study are minimal; there are no known physical or psychological risks for participating in this research. However, participants may experience boredom, exertion, frustration, or fatigue during the study, which is common in these studies requiring full-body interactions. Additionally, participants may feel less confident due to perceived performance. To mitigate these risks, participants are encouraged to take breaks and complete the study at their own pace.

Compensation / Reimbursement

In thanks for participating in our study, you will receive a compensation of CA$20, including the time spent on the survey, interview, breaks, reading the consent form, and completing the study. The payment will be provided to you upon completion of the study and before you leave the lab. In case of a withdrawal from the study in-between sessions, you will receive a pro-rated compensation based on the amount of time or study phases completed specified as follows:

Participation level	**Study hours invested (minutes)**	**Compensation Amount (CA$)**
Baseline participation	0-44	10
Partial participation	45-59	15
Completed user study	60-90	20

How your information will be protected:

During the testing phase, your interactions will be recorded on video, and your in-game performance will be logged as part of the study. However, we are committed to ensuring your anonymity so that all recordings will be de-identified.

We understand the importance of keeping your information confidential and secure. Only the research team will have access to identifying information, and any data that could reveal your identity will be removed before sharing it with anyone outside the team. All team members are required to maintain confidentiality.

All the digital data collected in this user study (logged game data, screen recordings, video/audio recordings, online questionnaires, and demographic surveys) will be initially stored in an access-restricted private folder on Dalhousie University's institutional SharePoint. Any physical data (signed consent forms) will be stored separately from any de-identified data inside a locked file cabinet in the lead researcher's office space, which has restricted access. Later on, to back up the digitally stored user study data, all the SharePoint folder contents will be backed up onto two separate encrypted external hard drives, one kept in a locked filing cabinet within the researcher's office and another stored offsite.

Your identity will remain confidential, and individual results will not be disclosed in any reports or publications. However, de-identified quotes might be used as part of data analysis to support and contextualize quantitative findings. After the study, identifying information will be deleted from everywhere within two years.

If You Decide to Stop Participating

If you decide to withdraw participation during the user study, you can inform the researcher conducting the user study at any point. Any collected data up until that point will be discarded from all the mediums. If you wish to have your data removed after the completion of the user study, you must make a request within two weeks from the study completion date. After this time, the study data will be used for analysis, and we will no longer be able to remove it. Thus, if you submit a request within the two-week time frame, we will delete your data from all stored mediums.

How to Obtain Results

We will share a summary of the study's group results once completed. However, individual results will not be provided. Approximately 12 months from now, you can access these results by visiting the lab's website (`gem.cs.dal.ca`).

Questions

We are here to assist you with any questions or concerns about participating in this research study. Please feel free to contact Aayush Shrestha (`ashrestha@dal.ca`) with any inquiries, comments, or concerns regarding the study. If you have any ethical concerns regarding participating in this research, you can also contact the Research Ethics Office at Dalhousie University. They can be contacted at (902) 494-3423 or `ethics@dal.ca`. Please mention the REB file # 20XX-XXX in your communication.

Consent

By signing below, you acknowledge that you have read and understood the information provided in this consent form and agree to participate voluntarily in the research study described above.

Participant Name: ____
Participant Signature: ____
Date: ____

Appendix D

User Study Script

Welcome to the user study! As you may be aware, this study is about exploring a different and perhaps an alternative approach to navigation in a virtual (VR) environment using sensory modalities other than visuals, namely, your sense of touch and sound.

My name is Aayush, and I am the lead researcher for this user study. I will accompany you throughout the period of your study where you will be asked to:

- Train on the system to get acquainted with it.
- Perform three primary tasks.
- Fill in a few survey questions.
- Participate in a post-study interview.

The compensation for this study is a total of 20 CAD which is pro-rated, the details of which you can find on the consent form. You are allowed to withdraw from the study at any point which is also explained in the consent form.

Additionally, your task performance will both (in-game and physical) be video-recorded, and the interview will be audio recorded. Please thoroughly read the details of the consent form and sign it so that we can proceed with the study.

Provide the participant with the consent form to sign.

Help the user to wear the shoe slide-ons, help them to get on the slidemill, strap them firmly and lock the rotation.

Help them to put on the:

- Blindfold
- Headset
- Headphones

- Hand them the cane.

Firstly, we will start with a training session where you can are expected to get acquainted with the system as a whole, but primarily on how to use the cane and walk on the slidemill. The duration for this is 10 mins or when you feel you are ready. Whatever comes first!

Proceed with the training scene and help with the session.

Now since you are done with the training, let's start with the testing scenario.

Fill in the participant ID and start with the following tasks in order.

Task 1

For the first task, you need to swipe the cane across the virtual surface(s), depending upon what you perceive, tell me the name of the textures. The possible textures are grass, sidewalk, asphalt (road), gravel, and Metal (Tactile Paving).

Task 1.1

For this sub-task, you can hear and touch the texture when your cane is in contact with and moving with the virtual surface. You need to tell me what kind of texture you perceive. This is repeated 5 times, and every time there will be a pair of different textures from the list of possible textures.

Task 1.2

For this sub-task, you can only hear the texture when your cane is in contact and movement with the virtual surface. You need to tell me what kind of texture you perceive. This is repeated 3 times, and every time there will be a single texture in front of you from the list of possible textures.

Task 1.3

For this sub-task, you can only feel the texture (haptic profile) when your cane is in contact and movement with the virtual surface. You need to tell me what kind of texture you perceive. This is repeated 3 times, and every time there will be a single texture in front of you from the list of possible textures.

Ask for a break.

Task 2

For the second task, you are in a city simulation where your primary goal is to cross the road. For this task, you need to take help of the immersive spatial audio of your surrounding as well as the textural perception conveyed using your cane. This task has two subtasks where the second task has two variants - one with visuals and one without any.

Task 2.1

For this sub-task, you are spawned near a road/pedestrian crossing. Your task is to reach the pedestrian crossing (traffic light) and when you think it is appropriate to cross the road, you may do so. The queue to cross the road would be similar to a real-life scenario, i.e., change in traffic light sound and haunting of vehicles to let you go through. Once you successfully reach the other end of the road, you will be greeted with a task completion notification.

Task 2.2 (NV)

For this sub-task, you are spawned between two road/pedestrian crossing. One is near to you and the other is far from you. Judging the distance between you and the nearest road crossing, you need to reach and proceed to cross the road. Once you reach the crossing, the objective and the result are the same as the first sub-task.

Task 2.2 (WV)

For this sub-task, your objective is the same as the previous task. The only difference being that you can visualize the surrounding. You may or may not use the cane for this task.

Ask for a break.

Task 3

For the third task, you are in the same city simulation but now your primary goal is to complete a scavenger hunt. For this task, you need to take help of the immersive spatial audio of your surrounding as well as the textural perception conveyed using your cane. This task has two subtasks where the second task has two variants - one with visuals and one without any.

From the moment you start the tasks, you can be on the lookout (figuratively) for auditory cues, which can be pedestrian conversation or ambient noise that will give you an idea of where you should be heading to. These cues are intermediate targets that will lead you to the final target. Once you reach the final target, a completion sound will be played, and then you need to look for a particular texture near that final target. You will come across road-crossings as well so be sure to safely cross the road!

Task 3.1

For this sub-task, you are spawned near a road/pedestrian crossing. Your final target/destination is a hotdog stand and the texture you need to find near the hot-dog stand is that of grass.

From where you are, the first intermediate target is a construction site the direction to which you need to find using the conversation as a cue.

The second intermediate target is a bus-stand the direction to which is right of the construction site.

While on the way to the bus-stand hear for any pedestrian conversation that may help you locate the final target/destination from the bus-stand.

Task 3.2 (NV)

For this sub-task, you are spawned in front of the lawn of your neighbor. Your final target/destination is the bus stop, and the texture you need to find near the bus stand is that of gravel.

From where you are, the first intermediate target is a local rap and hip-hop dance battle the direction to which you need to find using the depending upon your auditory perception.

The second intermediate target is a traffic light crossing nearest to you. The third intermediate target is to find a group of children playing in the park. You need to find the entry to the park by searching for any auditory cues in the form of conversation.

The fourth intermediate target is a hotdog stand which is just beside the bus stop (the final target/destination).

Task 3.2 (WV)

For this sub-task, your objective is the same as the previous task. The only difference being that you can visualize the surrounding. You may or may not use the cane for this task.

At this point, stop the screen recording, end the gameplay, and stop the video recording.

Un-strap the user and help them get off the slidemill and direct them towards the interview seating area.

Ask them to fill in the demographic survey followed by the SUS and NASA-TLX.

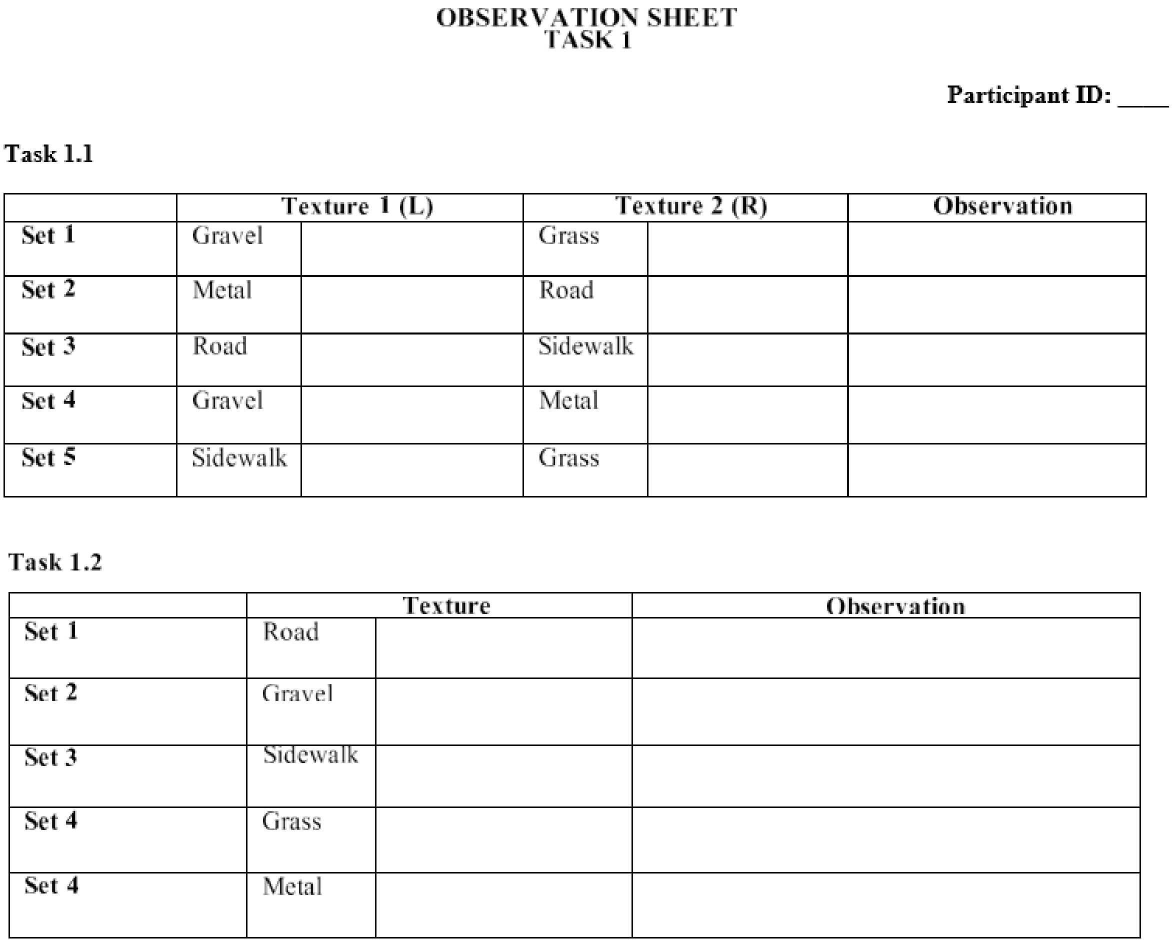
Once the surveys are complete, turn on the microphone.

Proceed to ask the interview questions.

Once the interview is done, hand them the receipt to be signed, cash envelope, and thank them for their participation.

Appendix E

Observation Sheets



OBSERVATION SHEET
TASK 1

Participant ID: ____

Task 1.1

	Texture 1 (L)		Texture 2 (R)		Observation
Set 1	Gravel		Grass		
Set 2	Metal		Road		
Set 3	Road		Sidewalk		
Set 4	Gravel		Metal		
Set 5	Sidewalk		Grass		

Task 1.2

	Texture		Observation
Set 1	Road		
Set 2	Gravel		
Set 3	Sidewalk		
Set 4	Grass		
Set 4	Metal		

Task 1.3

	Texture		Observation
Set 1	Road		
Set 2	Gravel		
Set 3	Sidewalk		
Set 4	Grass		
Set 4	Metal		

TASK 2

Task 2.1

Task 2.2 (NV)

Task 2.2 (WV)

TASK 3

Task 3.1

Task 3.2 (NV)

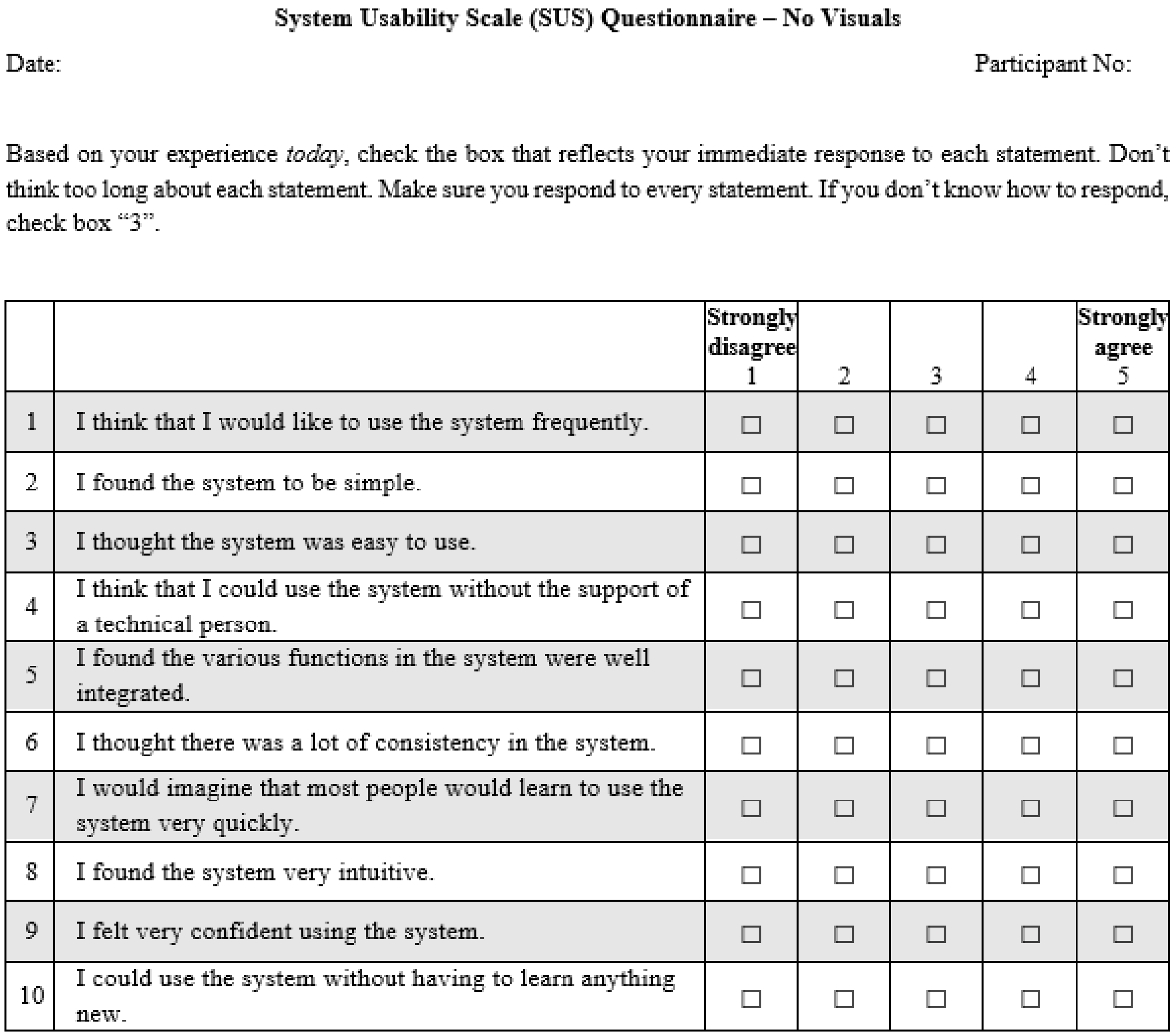
Task 3.2 (WV)

ADDITIONAL OBSERVATION

Appendix F

Post-Study Questionnaires

F.1 System Usability Scale (SUS) Questionnaire

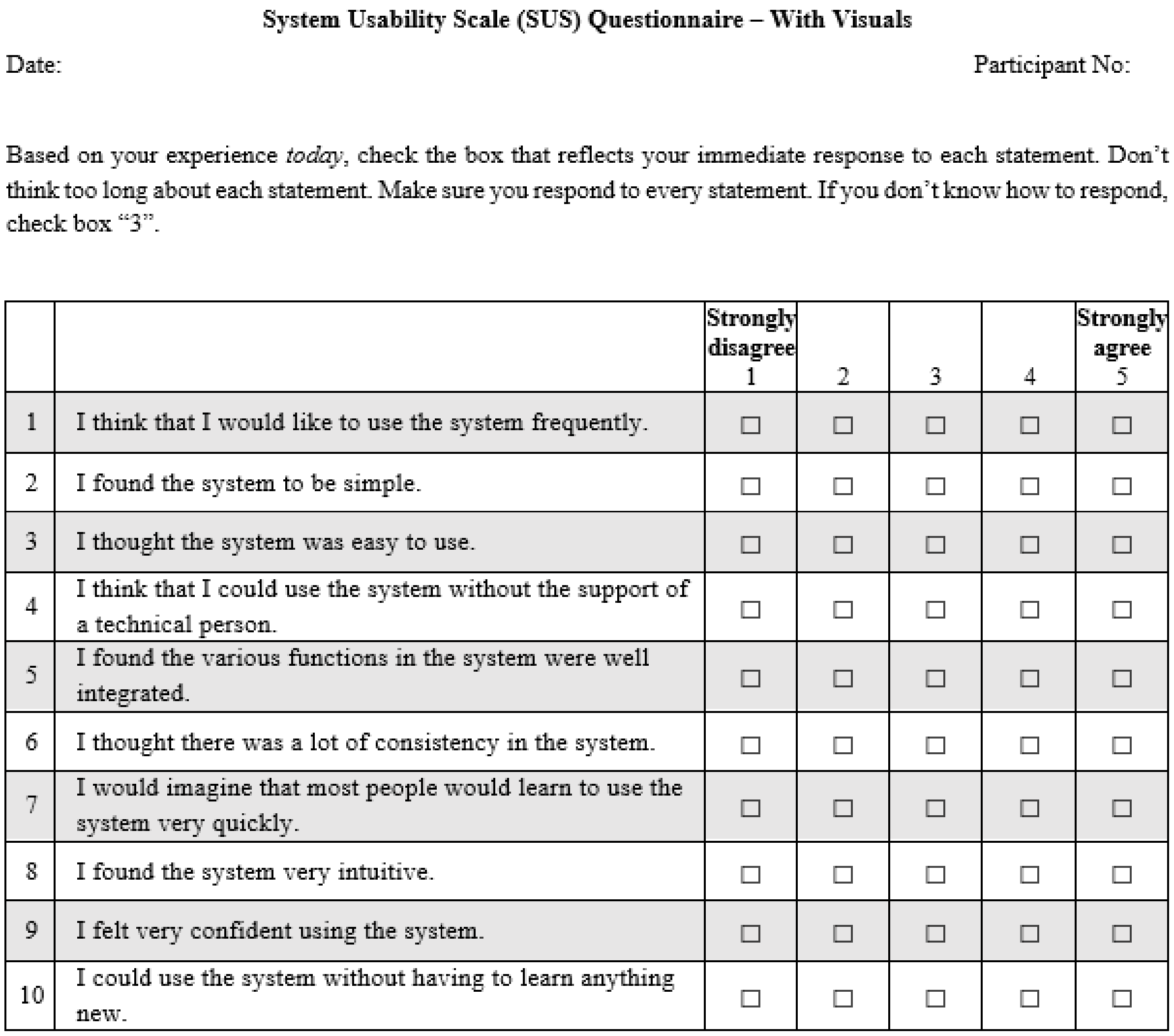


System Usability Scale (SUS) Questionnaire – No Visuals

Date: Participant No:

Based on your experience *today*, check the box that reflects your immediate response to each statement. Don't think too long about each statement. Make sure you respond to every statement. If you don't know how to respond, check box "3".

		Strongly disagree 1	2	3	4	**Strongly agree** 5
1	I think that I would like to use the system frequently.	☐	☐	☐	☐	☐
2	I found the system to be simple.	☐	☐	☐	☐	☐
3	I thought the system was easy to use.	☐	☐	☐	☐	☐
4	I think that I could use the system without the support of a technical person.	☐	☐	☐	☐	☐
5	I found the various functions in the system were well integrated.	☐	☐	☐	☐	☐
6	I thought there was a lot of consistency in the system.	☐	☐	☐	☐	☐
7	I would imagine that most people would learn to use the system very quickly.	☐	☐	☐	☐	☐
8	I found the system very intuitive.	☐	☐	☐	☐	☐
9	I felt very confident using the system.	☐	☐	☐	☐	☐
10	I could use the system without having to learn anything new.	☐	☐	☐	☐	☐



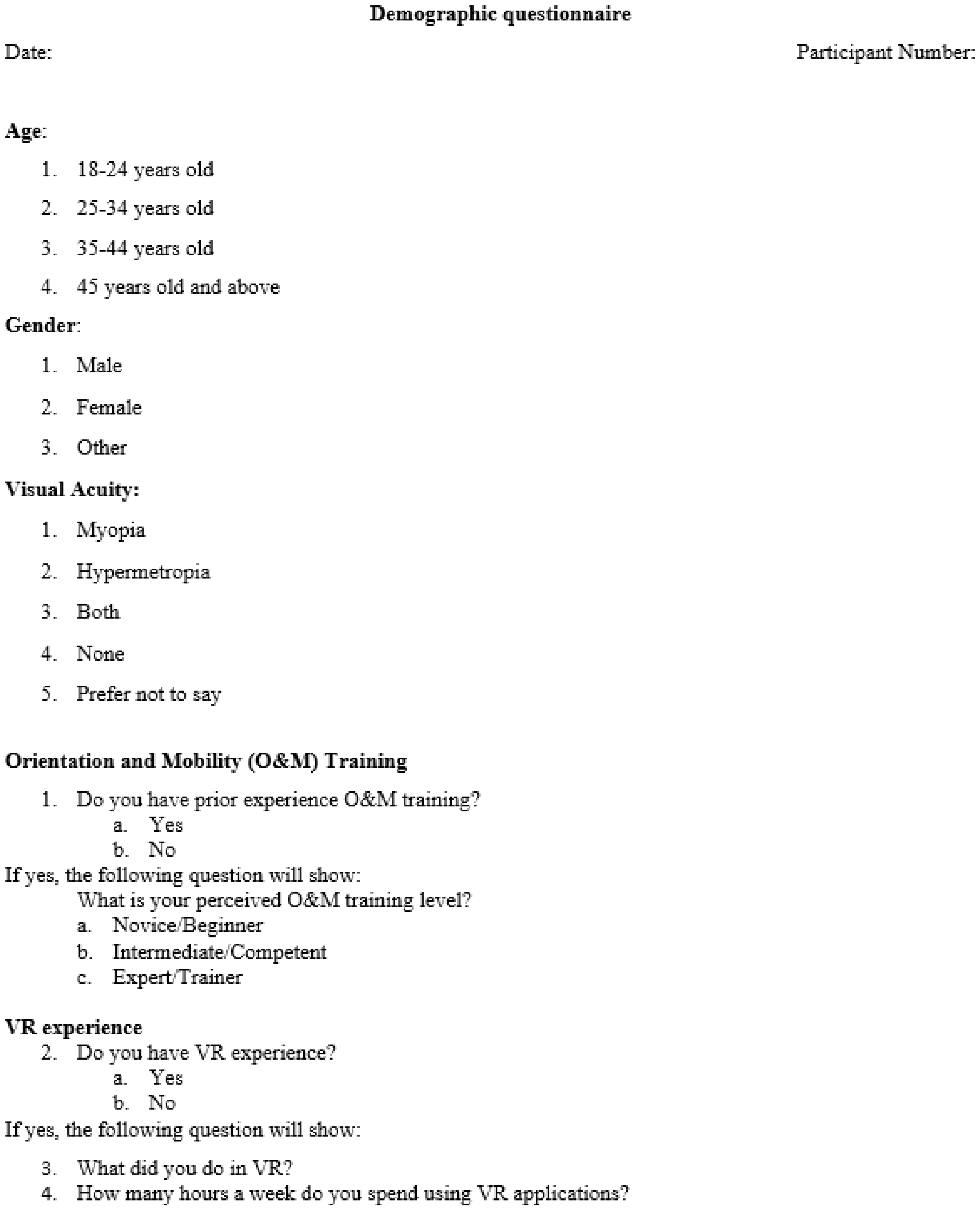
System Usability Scale (SUS) Questionnaire – With Visuals

Date: Participant No:

Based on your experience *today*, check the box that reflects your immediate response to each statement. Don’t think too long about each statement. Make sure you respond to every statement. If you don’t know how to respond, check box “3”.

		Strongly disagree 1	2	3	4	**Strongly agree** 5
1	I think that I would like to use the system frequently.	☐	☐	☐	☐	☐
2	I found the system to be simple.	☐	☐	☐	☐	☐
3	I thought the system was easy to use.	☐	☐	☐	☐	☐
4	I think that I could use the system without the support of a technical person.	☐	☐	☐	☐	☐
5	I found the various functions in the system were well integrated.	☐	☐	☐	☐	☐
6	I thought there was a lot of consistency in the system.	☐	☐	☐	☐	☐
7	I would imagine that most people would learn to use the system very quickly.	☐	☐	☐	☐	☐
8	I found the system very intuitive.	☐	☐	☐	☐	☐
9	I felt very confident using the system.	☐	☐	☐	☐	☐
10	I could use the system without having to learn anything new.	☐	☐	☐	☐	☐

F.2 Demographic Questionnaire



Demographic questionnaire

Date: Participant Number:

Age:

1. 18-24 years old
2. 25-34 years old
3. 35-44 years old
4. 45 years old and above

Gender:

1. Male
2. Female
3. Other

Visual Acuity:

1. Myopia
2. Hypermetropia
3. Both
4. None
5. Prefer not to say

Orientation and Mobility (O&M) Training

1. Do you have prior experience O&M training?
    a. Yes
    b. No

If yes, the following question will show:

What is your perceived O&M training level?
a. Novice/Beginner
b. Intermediate/Competent
c. Expert/Trainer

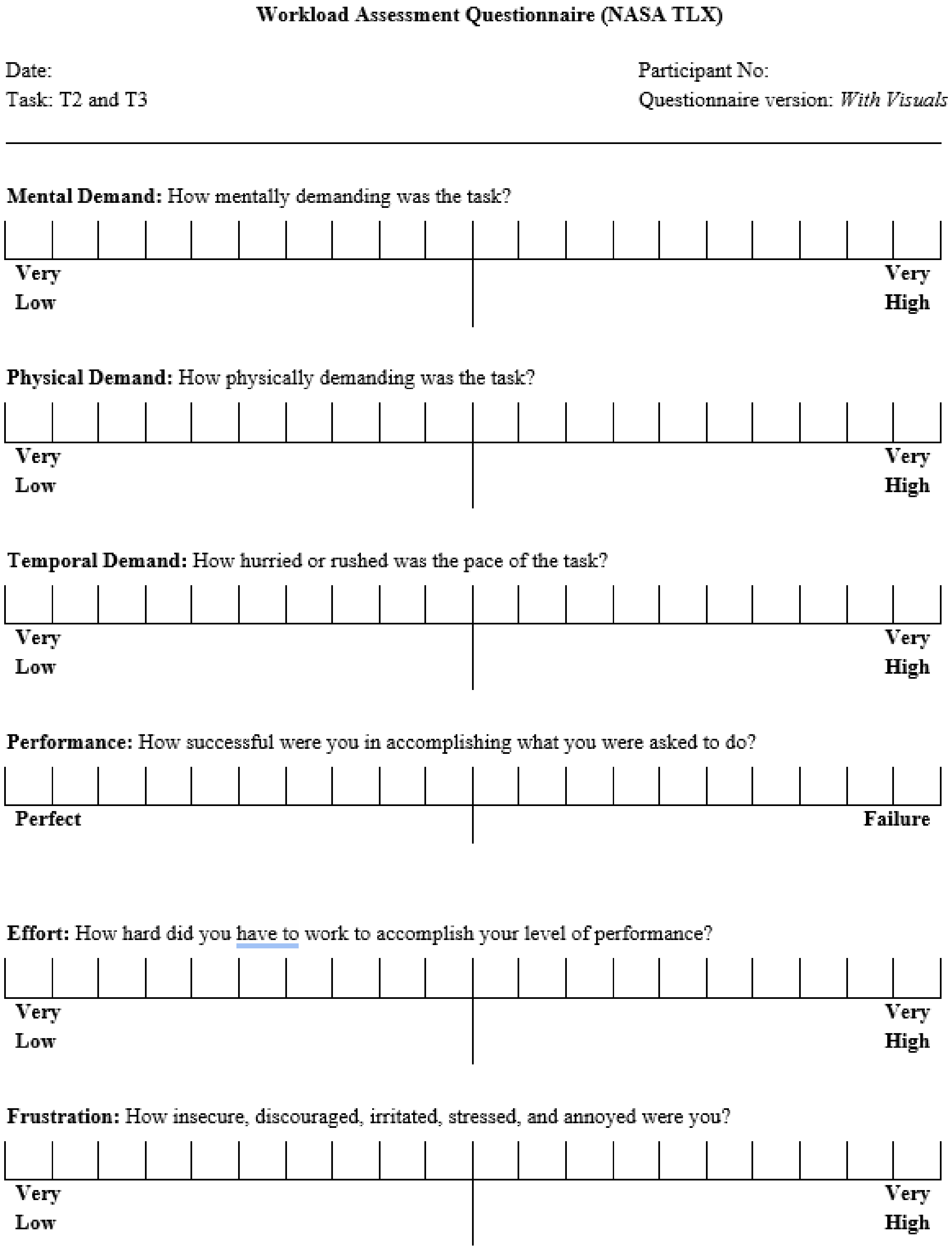
VR experience

2. Do you have VR experience?
    a. Yes
    b. No

If yes, the following question will show:

3. What did you do in VR?
4. How many hours a week do you spend using VR applications?

F.3 Workload Assessment (NASA-TLX) Questionnaire



Workload Assessment Questionnaire (NASA TLX)

Date:
Task: T2 and T3

Participant No:
Questionnaire version: *With Visuals*

Mental Demand: How mentally demanding was the task?

Very Low **Very High**

Physical Demand: How physically demanding was the task?

Very Low **Very High**

Temporal Demand: How hurried or rushed was the pace of the task?

Very Low **Very High**

Performance: How successful were you in accomplishing what you were asked to do?

Perfect **Failure**

Effort: How hard did you have to work to accomplish your level of performance?

Very Low **Very High**

Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low **Very High**

Workload Assessment Questionnaire (NASA TLX)

Date:
Task: T2 and T3

Participant No:
Questionnaire version: *Without Visuals*

Mental Demand: How mentally demanding was the task?

Very Low **Very High**

Physical Demand: How physically demanding was the task?

Very Low **Very High**

Temporal Demand: How hurried or rushed was the pace of the task?

Very Low **Very High**

Performance: How successful were you in accomplishing what you were asked to do?

Perfect **Failure**

Effort: How hard did you have to work to accomplish your level of performance?

Very Low **Very High**

Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low **Very High**

Appendix G

Interview Questions

1. What role did various audio experiences play for you to get around in the virtual environment?
    (a) Could you share the different kinds of audio you encountered in the VR space?
    (b) What characteristics of the spatial audio can you recall from all the tasks performed?
    (c) How would you compare the spatial audio of the VR city simulation with a real-life scenario?
2. Can you walk me through your decision-making process during the navigational tasks (T2 and T3)?
    (a) How differently do you think you were thinking while making critical navigational decisions?
    (b) How differently would you have done the tasks had you been able to see?
3. Particularly during task 1, did you feel and hear any external natural vibrations and/or sound emitted from the real cane interacting with the actual surface?
    (a) How realistic were the haptic profiles of the virtual surfaces?
4. Regarding task 1, do you prefer feedback from any particular sensory modality, such as audio or haptic? Or do you believe both are equally important for maximum effectiveness?
5. Regarding tasks 2 and 3, how would you compare your experience between tasks done with and without visuals?
    (a) What problems (if any) did you encounter doing tasks 2 and 3?
    (b) Do you think there were any benefits to performing these tasks without visuals?

    (c) How would you relate this experience to performing tasks 2 and 3 (without visuals) if you had to do it in the real world?
        i. Would you be more cautious and more aware?
            A. Does that mean you were less careful in the VR world?
    (d) Based on your experience, what are your views about this system being useful as a training aid for novice visually impaired users?

6. For tasks 2 and 3, what was the cane's role for you?
    (a) Did you forget about it due to other auditory engagements, or did you think you used it?
7. What do you think about using the cane in terms of immersion?
    (a) Do you think the use of a cane appropriates the task or not? Or would you have preferred a conventional VR controller?
8. What are your thoughts on using a slide-mill for navigation in VR?
    (a) Do you think it affects the level of immersion?
    (b) How do you find it different from walking?
9. On realism, what do you think about the proxy base/skirt around the VR slide mill?
10. How immersive would the system be for someone with visual impairment?
11. Did this experiment change your empathy or awareness towards people with visual impairments? If yes, could you explain how?
12. What is your overall feedback on the system? What improvements do you think could be made?