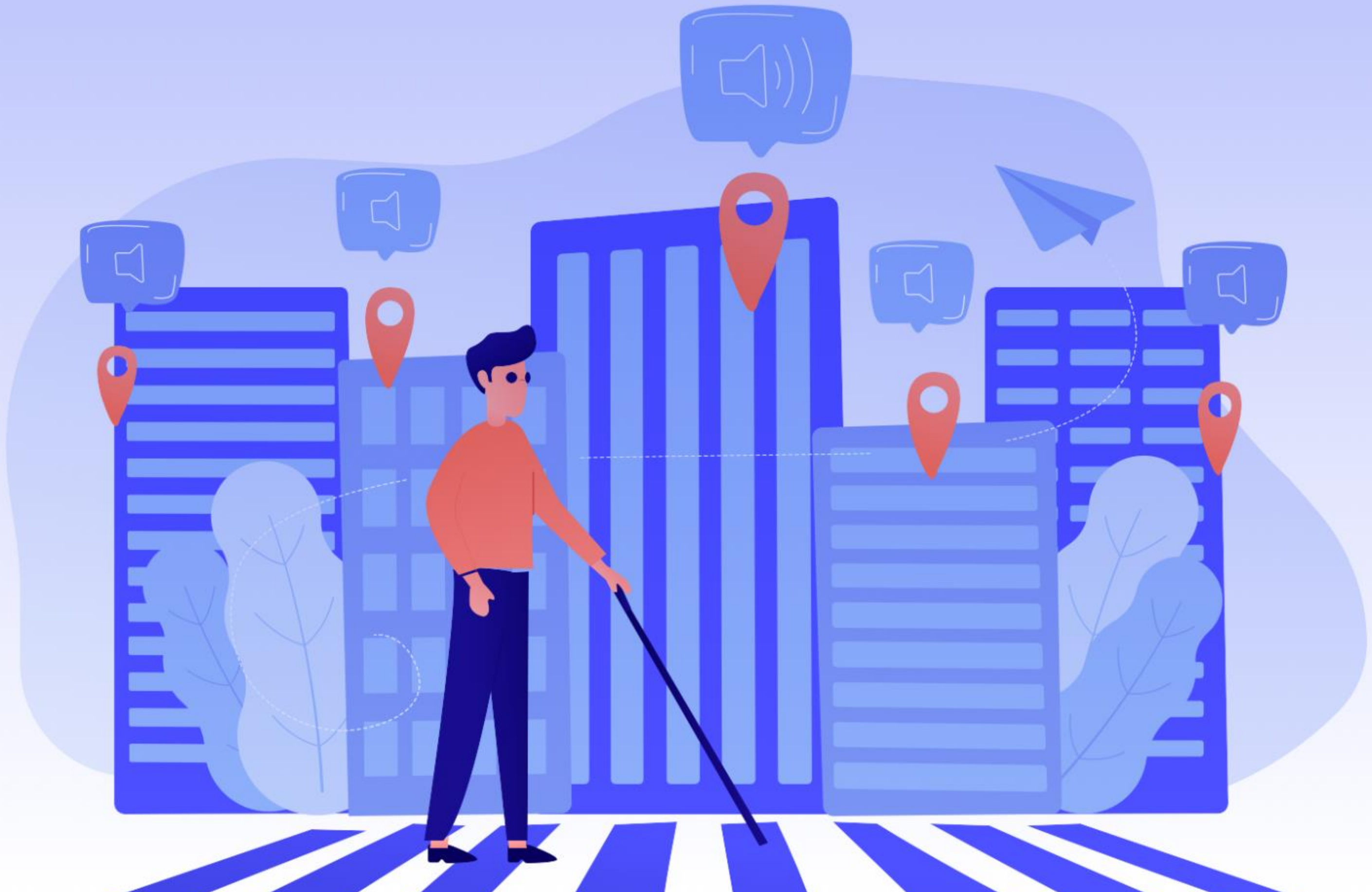
Wonder Vision
Wonder Vision

Wonder Vision



WONDER VISION

A HYBRID WAY-FINDING SYSTEM TO ASSIST PEOPLE WITH VISUAL IMPAIRMENT

By

Jessy Zhang

A thesis exhibition presented to OCAD University in partial fulfillment of the requirements for the degree of

Master of Design

in

Digital Futures

Rooms 104, 205 Richmond Street West

Toronto, Ontario, M5V 1V3 , 12th-14th April, 2022

Toronto, Ontario, Canada, April, 2022

Jessy Zhang 2022

Wonder Vision: A Hybrid Way-finding System to Assist People with Visual Impairment

Jessy Zhang

Master of Design in Digital Futures, 2022

OCAD University

Abstract

We use multi-sensory information to find our ways around environments. Among these, vision plays a crucial part in way-finding tasks, such as perceiving landmarks and layouts. People with impaired vision may find it difficult to move around in unfamiliar environments because they are unable to use their eyesight to capture critical information. Limiting vision can affect how people interact with their environment, especially for navigation. Individuals with varying degrees of vision will require a different level of way-finding aids. Blind people rely heavily on white canes, whereas low-vision patients could choose from magnifiers for amplifying signs, or even GPS mobile applications to acquire knowledge before their arrival. The purpose of this study is to investigate the in-situ challenges of way-finding for persons with visual impairments. With the methodologies of Research through Design (RTD) and User-centered Design (UCD), I conducted online user research and created a series of iterative prototypes towards a final one: Wonder Vision. It is a hybrid way-finding system that combines Augmented Reality (AR) and Voice User Interface (VUI) to assist people with visual impairments. The descriptive evaluation method suggests Wonder Vision as a possible solution for helping people with visual impairments to find their way toward their goals.

Key Words: Augmented Reality, Accessible Mixed Reality, Natural Language Processing,Voice User Interface, Way-finding technologies, User-center Design, Research through Design

Acknowledgements

I would like to express my big gratitude to my supervisors, Dr. Alexis Morris and Dr. Haru Hyunkyung Ji for their continuous support and guidance through my graduate years. Thank you for your patience, motivation, enthusiasm, and broad knowledge.

To Miss Lin, the coordinator of the WeChat online community of people with visual impairment for distributing the questionnaires.

To all online volunteers, thank you for your participation and for providing so much helpful information.

To my family for the unconditional love and support. Special thanks to my husband, I am so grateful for your encouragement and reassurance during my toughest times.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Motivation

People comprehend their surroundings through vision, as it allows us to perceive visual representation, such as color, line, and symbol. It is hard for us to imagine life without our vision even just for a short period of time, beacaue we use it instinctively. Due to High Myopia[1], I often fear misplacing my glasses, because everything gets blurred without clear vision. A few years ago, My mother had eye surgery due to an Idiopathic Macular Hole (IMH)[2]. The treatment could only prevent the vision from worsening rather than a full recovery. That means she has to deal with the visual impairment problems resulting in aging for the rest of her life. From her story, I realized that adapting to live with restricted vision within a short period was very challenging, especially when in unfamiliar places. The limited vision affects her to do routine tasks such as grabbing a glass of water, or finding keys. Besides, there are even some situational disabilities. For example, a person recovering from cataract surgery may require a screen reader for a couple of months, or a person replacing glasses might not read things clearly at that moment. Even though these problems come up with specific context, it impacts painfully on people's lives. These unmet needs motivate me to investigate better tools for visually challenged people.

---

[1] A common vision disorder that can see objects well near but have trouble viewing things that are far. ("Myopia,"2022)

[2] A a kind of common macular diseases, involves tissue defects including the retinal internal limiting membrane (ILM) and even the photoreceptor (PR) layer.(Li et al., 2019)

Maps and indicating signs are provided in complex architectural areas to assist people in arriving at their destination. However, people with limited vision have difficulty seeing thesis signages. In addition, in crowded areas, there are more variables: walking pedestrians and moving traffic, which bringing distraction for getting the most essential signals. Technologies, such as satellite maps, 3D city models, offer people more comprehensive assistance in way-finding. In particular, Mixed Reality (MR) provides a hybrid interface that blends physical and digital worlds. It applies technologies such as computer vision, image processing, rendering technologies, input systems, and cloud computing. The recent studies from the Microsoft research lab (Grayson et al.,2020) have shown that mixed reality can reduce a variety of accessibility barriers, and mobile Augmented Reality is the most mainstream solution for mixing virtual and physical experiences. I was inspired to propose that if there were reliable aids, visually impaired people could feel more confident and independent during traveling.

1.2 Research Goal

Despite a large body of work in accessibility concerning the design of advanced navigation technologies, my intention is to create hybrid interfaces that enable people to have multisensory and inclusive experiences in way-finding activities. This study investigates the impact of varying degrees of vision and the limitations of existing navigation tools to discover potential solutions for way-finding technologies that can support people with varying degrees of vision, both blind people and low-vision people.

1.3 Research Summary

1.3.1 Problem Statement

At least 2.2 billion people worldwide have a near or distance vision impairment, and it affects people of all ages.(World Health Organization, 2021) Aging is one of the reasons that causes these problems. Besides, people have situational vision impairment, such as blocked vision after surgery. Navigation is a multi-faced task but heavily relies on vision.(Taylor & Francis, 2021) For people who live in urban area, it is normal to walk through various walkways.(Figure 1) These walkways are usually designed with steps. Blind people with proper training could use a white cane to track the steps. However, for people who recently become vision challenged, they are less skilled in using a white cane and would easily miss the step and fall. In rush hours, a moving stream of people will pack the space and make it even harder to decide how to move next.

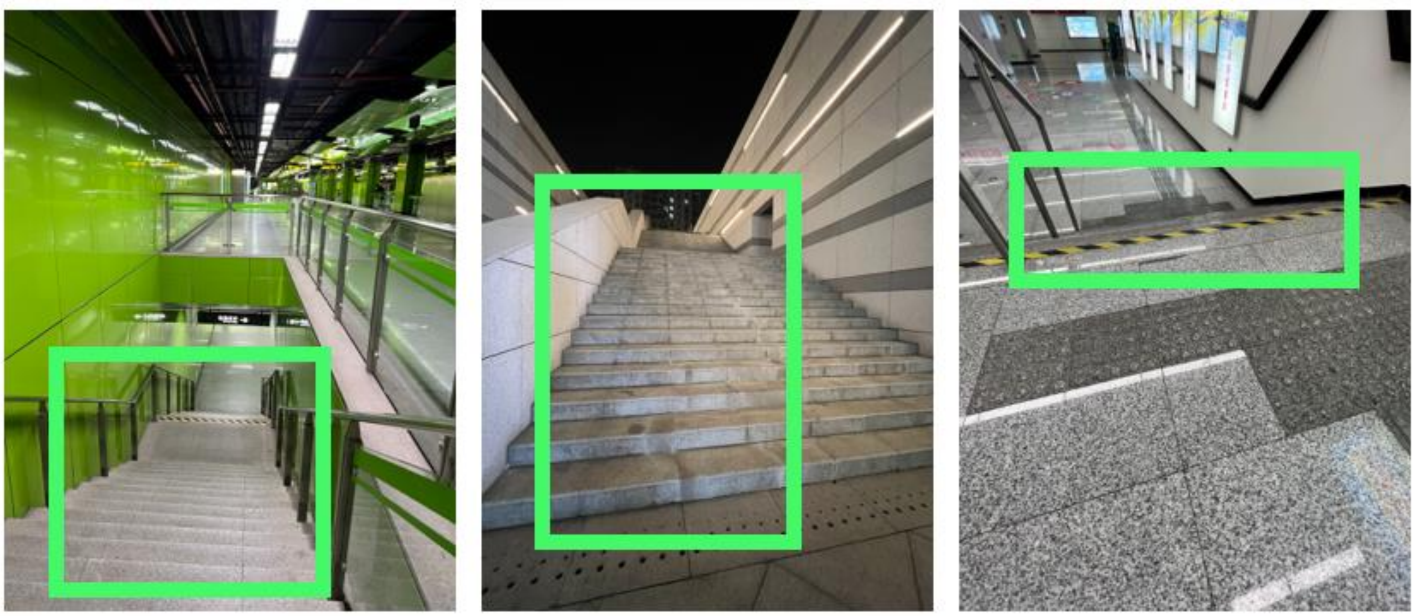


Figure 1. Walking on these staires is dangerous for people with reduced vision, with poor light and a chaotic flow of people. (Images of metro stations in Shanghai, China, taken by the author)

The majority of external navigation aids, such as maps and mobile GPS, are designed for route planning. It does not assist people with limited vision in detecting obstacles or finding useful information. With limited vision, any minor mistake is extremely risky. If a person misses one step, or is tripped by an obstacle, he/she could fall over and get hurt physically. The way-finding process will be risky without effective tools that are tailored to the vision impairment. Therefore, my study focused on the visual searching tasks in way-finding.

1.3.2 Research Questions

The main research question is “how combined AR with VUI can help people with visual impairment in way-finding?” Given this, a series of sub research questions are raised as follows:

- What kinds of assistance do users need in navigation?
- How does limited vision affect users in navigation?
- What are the in-situ difficulties that users encounter in daily travel?
- What kinds of existing tools and technologies do they use?
- What are potential solutions that could improve their navigational experience?

1.3.3 Hypothesis

With the basis of the problem discussed above, my assumption is that, advanced technologies such as artificial intelligence, will offer supportive information for people with limited vision on the move. It scans the environment and captures useful information in the surroundings then this information will be useful clues for users. User interfaces become more hybrid by providing users with a multisensory experience that improves spatial cognition. The system should work as a virtual assistant to allow users of low vision and blindness to interact more freely with it. The latest Augmented Reality(AR) technology featuring hybrid user interfaces and advances in calculating distance between virtual content and the user's location. In addition, advances in machine learning enable the computer to communicate with human in natural languages. My proposed method would combine AR with Voice User Interface(VUI) to help users find their way.

1.3.4 Research Scope and Objects

This study's theoretical development focused primarily on navigational behaviors and the tools that can help people. Taking accessibility as its starting point, the research looked at a group of people who suffer from visual impairment and clarified the differences in vision degree among them. In the meantime, mixed reality and voice interfaces are being studied extensively. It went through their related work in-depth, looking for the adaptations that can be employed as potential solutions. As part of the practical portion of this research, I conducted interviews and online surveys to identify the user’s needs. Specifically, the RTD and UCD employed iterative designs to generate useful solutions. To evaluate my design, I applied descriptive method and the ‘Furtues Cone’ model to critique this design contributions.

1.3.5 Research Contribution

In this research, cognitive science, accessible design, and computer science are connected. The online user research identified the in-situ problems of way-finding; the iterative prototypes demonstrate how they adapt unique features from existing tools. The integration of AR and VUI navigation demonstrates cross-framework compatibility (ARKit and Speech Framework). It also features a hybrid interface that provides audio and speech instructions. Additionally, Wonder Vision emphasizes the importance of innovative technologies in enhancing the power and reach of this platform through speculative and descriptive design.

1.4 Chapter Overview

In the first chapter, I present the background story of this research. The second chapter discusses navigation within the behavioral and psychological framework, followed by a discussion of conversational interfaces and accessible mixed reality as emerging technologies with possible applications. In Chapter three, I introduce the research methods, and in Chapter four, a series of findings from user research is discussed. Chapter five presents a series of iterative prototypes, along with insights and findings from this phase. In chapter six, I uses the perspective design method and 'Futures Cone' model to evaluate its contribution. Chapter seven discusses the thesis' limitations and its future directions.

2. Literature Review

In this chapter, I introduce the concept of navigation and its related tools to emphasize the importance of vision perception. Following that, it discusses potential sulutions in terms of conversational interfaces and accessible mixed reality to support the hypothesis proposed in this thesis.

2.1 Way-finding Concept

2.2.1 Definition

Navigational behavior is ubiquitous for both humans and animals. Animals may rely on magnetite particles as tiny compasses to guide their movements and find food. People will search for recognizable features for landmarks when creating a mind map. Most navigational behavior is a coordinated and goal-directed process. (Montello & Sas, 2017) It is a mechanical movement, walking from one point to another. And in this process, the multiple sensors will help us to receive useful information from the surroundings like a target's location can be determined by its sound, light or smell.

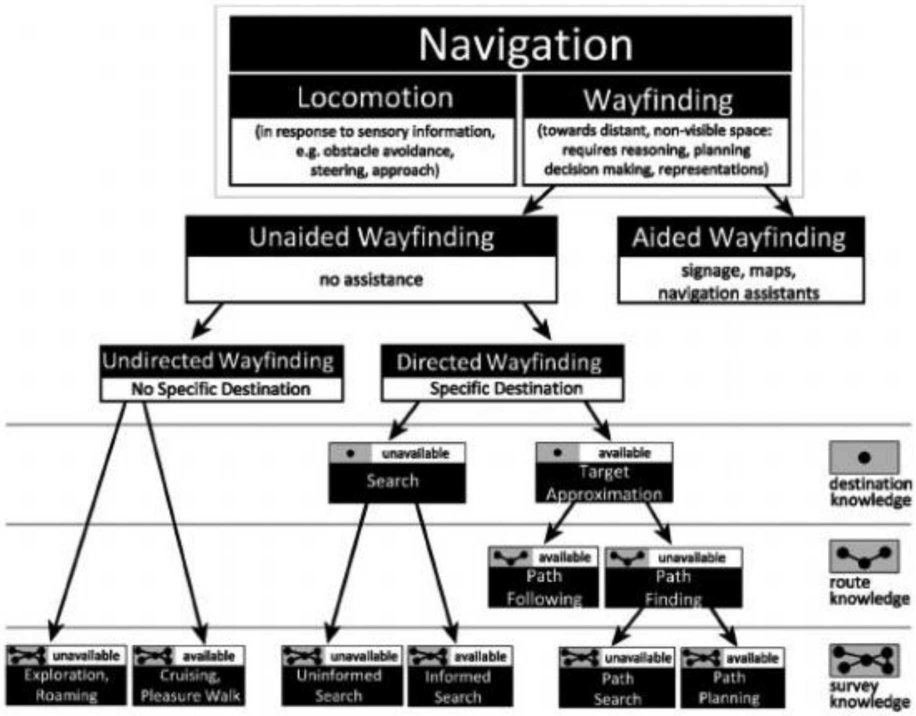


Figure 2. The navigational behaviors have the fundamental layer consisting of locomotion and way-finding.(Wiener, 2009)

Fundamentally, navigation consists of two components: way-finding and locomotion (figure 2). Way-finding is a process of solving problems, such as route planning and making decisions. This term can be extended into two ways depending on whether external assistance is used. Following signs, route instructions are considered aided-finding, while un-aided way-finding requires considerable cognition efforts, such as a mind map or personal experience. The locomotion entails a series of tasks, such as avoiding obstacles, detecting people, and watching for steps. As so the locomotion emphasizes the ability to cope with the surrounding, which is relevant to spatial learning ability.

Learning abilities vary among individuals. Suburbanites may find it hard to concentrate in cities due to traffic jams or get lost in a complex architectural structure. The most common method of finding the way is through the memory of landmarks and dead reckoning. Landmarks are remarkable features of places that help us to have location relatives to the destination. Dead-reckoning updating involves keeping track of components of locomotion, including velocity or acceleration (which, as vector quantities, include heading information) and travel duration (dead reckoning is sometimes called path integration)(Montello & Sas, 2017). Designers also made different kinds of signs to present the architectural features in order to help people better understand the space and layout. Using the patterns and symbols on the map (Figure 3), we can find our location or identify the features of a street and plan our routes. Way-finding and locomotion are closely related to external aids. Some are designed signs and others are features of surroundinngs. We see the curves in a street to make predictions or to avoid falling when walking. We use these our eyes to see these external information.

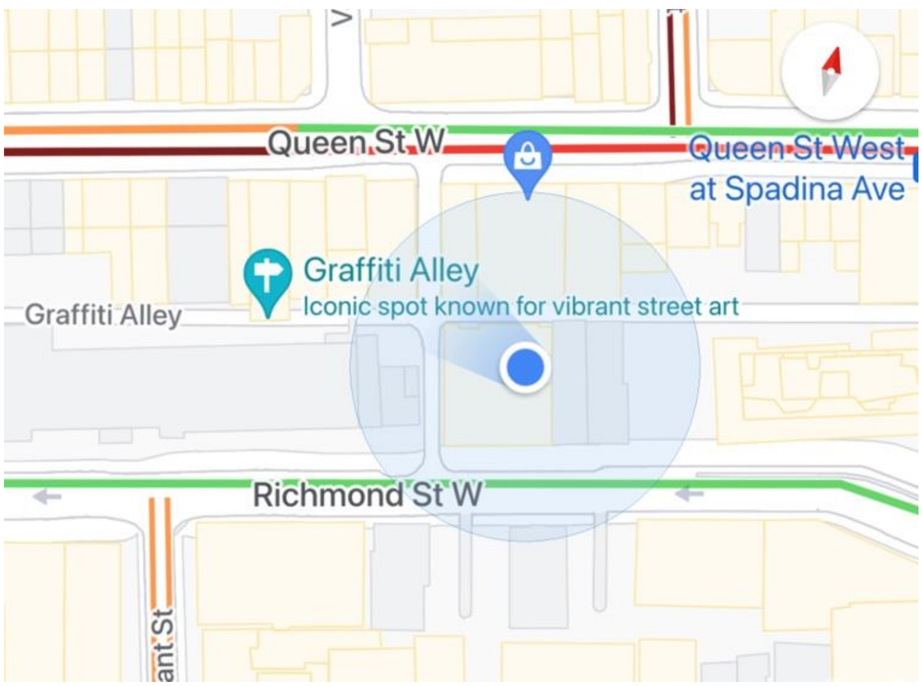


Figure 3. A screenshot from Google Maps.

2.2.2 Tools for Way-finding

The development of digital signage has made way-finding tools extremely digital and interactive. Messages are displayed in loops and ambient images are displayed. The information will be distributed visually, audibly, as well as tactilely. Furthermore, together with Global Positioning System (GPS)[3] and inertial sensors for heading, as well as digital Geographic Information Systems(GIS)[4], digital signages can support in-Vehicle Navigation Systems (IVNS)[5], part of the broader topic of Intelligent Transportation Systems (ITS)[6].(Montello & Sas, 2017). Even though digital technologies are pushing navigation information into a hybrid interface, it is essential to eliminate the barriers between users and their environment. An intelligent informational system is not just an elaborate term for signage, but a system that responds to people's characteristics (age, mobility, gender, etc.) and offers them redundant cues no matter how well-versed they are in spatial knowledge.

2.2 Visual Impairment

2.2.1 Definition

Visual impairment (VI) refers to a wide range of conditions and levels of impairment. It includes low vision and blindness, referring to the degree of impairment to a person's ability to

[3] GPS: a satellite-based radionavigation system owned by the United States government and operated by the United States Space Force.(GPS.gov:GPS Overview, 2022)

[4] GIS: a system that creates, manages, analyzes, and maps all types of data. GIS connects data to a map, integrating location data (where things are) with all types of descriptive information (what things are like there).(Wikipedia, 2022)

[5] IVNS: An automotive navigation system is part of the automobile controls or a third party add-on used to find direction in an automobile.(Wikipedia, 2022)

[6] ITS: is an advanced application which aims to provide innovative services relating to different modes of transport and traffic management.(Wikipedia, 2022)

see. People with uncorrectable, reduced vision are classified as visually impaired. In other words, their uncorrectable vision cannot be improved further by spectacles or contact lenses. Elderly people face this health problem frequently. With advancing age, the normal function of eye tissues decreases, and there is an increased incidence of ocular pathology. (Loh KY;Ogle J, 2020) We may have these problems as aging, such as increased nearsightedness and deteriorating vision. But visual impairment happens occasionally. Jessica Alba [7] had an incident that left her visually impaired for an entire week – all because of an allergic reaction to some shampoo. (How Situational Disabilities Impact Us All, 2020) It reminds us the visual impairment can happen to anyone.

The term low vision ("What Is Low Vision," 2021) is subdivided according to the type of visual field loss, such as central and peripheral vision loss, peripheral vision loss, extreme light sensitivity, and blind spots. The table below (Table 1) lists four types of low vision according to the degree of visual loss. Their eyes can partially perceive shapes and general appearances. Identifying details of objects, like faces, is a common challenge. ("Low Vision and Blindness," 2022) Particularly, varying sight degrees affect life activities differently. People losing their central vision will not be affected when traveling outside, but may not be able to see the details of the environment. People with night blindness may not have these problems if they have sufficient lighting. In a sense, varying levels of vision require different types of tools when traveling.

[7] Jessica Alba: a famous American actress.

Table 1. The table lists the common types of low vision and the impact on life accordingly.

Common types of low vision	**The impact on life**
Loss of central vision The loss of central vision creates a blur or blind spot, but a person's side (peripheral) vision remains. ("Low Vision and Vision Rehabilitation," 2022)	**Difficulties in** ▪ Reading ▪ Recognizing faces and ▪ Distinguishing details in the distance ***Mobility is usually unaffected***
Loss of peripheral (side) vision People who lose their peripheral vision cannot distinguish anything to one side or both sides, or anything directly above and/or below eye level. Central vision remains, however, making it possible to see directly ahead, read and see faces. This is sometimes referred to as "tunnel vision." ("Low Vision and Vision Rehabilitation," 2022)	**Difficulties in** ▪ Mobility ▪ Be able to do slow reading
Blurred vision(Myopic & Hyperopic) With blurred vision, both near and far vision is out of focus, even with the best possible correction with eyeglasses. ("Low Vision and Vision Rehabilitation," 2022)	**Difficulties in** ▪ Reading ▪ Recognizing faces ▪ Mobility ▪ Distinguishing most details in the distance
Reduced contrast sensitivity People with loss of contrast sensitivity have a loss of vision quality. They tend to feel that there is a generalized haze with a sensation of a film or cloudiness. ("Low Vision and Vision Rehabilitation," 2022)	*Depending on the haze area* *Depending on the haze area*
Night blindness People with night blindness cannot see outside at night or in dimly lighted interior areas such as movie theaters or restaurants. ("Low Vision and Vision Rehabilitation," 2022)	**Difficulties** ▪ Mobility in darkness

2.2.2 Case Study among People with Visual Impairment

Maha[8] created series of videos to raise awareness of disability and accessibility issues. One of her videos, 'See-through my eyes' [9] simulates her central lost eyesight. A circle area like TV noise is in the central of her eyesight (Figure 4). It was still pretty noisy and difficult to identify the objects behind even though this area would change according to the time of day. Her attention was also diverted by the central noisy area, as she had difficulty recognizing faces in the crowd. Shopping in the supermarket was also challenging. Using a magnifier, she was able to read labels on products. She had to move her eyes scaling from side to side without aided tools to get a full image of surroundings. It would be worse if there wasn't enough light: the central white circle would blind her eyesight in the evening when there was little light. Occasionally, the reflection on the floor after rain made her confused as to which ones were real.

Figure 4. Screenshots of 'See-through my eyes'9: (a) simulating the scenario of how does Maha see in daytime, (b) simulating the scenario of how does Maha see in night, (c) simulating the scenario of reading

But for those who are legally blind, things will be distinctive. Wisconsin Council of the Blind & Visually Impaired[10], is a private and non-profit agency. It posted a video called 'A day in

---

[8] A visually impaired person living in Egypt, and the creator of this channel of Legally Blind

[9] https://www.youtube.com/watch?v=xG7d-kIlnT8&t=2s

[10] https://wcblind.org/

the life of a person who is blind or visually impaired'[11]presents a typical day of people with visual impairment. Denise[11] was a brilliant speaker and presenter. She was born with an undeveloped optic nerve and consequently legally blind since birth. She has no light perception in the left eye, and the right eyesight is 20 over 400 with limited peripheral vision. She did a lot of computer work with the assistance of a computer program called ZoomText Fusion, which can magnify the texts and read them back.

Cory Bellard[11] is a director of techonolgoy at the Company of Forward Association. He lost his sight due to the detached retinas. The left eye is blind, while the right eye has some light and color perception. He usually takes public transportation or Uber to get to the destination. While walking to the bus stop, Cory usually uses Voice-over [11]to gain access to the GPS[12] to find the bus stop. Crossing the street was the most dangerous and risky thing for them. They use a white cane or trained dog for walking to the bus stops, and they need to judge the situation according to the traffic sounds. In addition, Denise advocated for an accessible pedestrian signal outside the building.

2.2.3 Way-finding Tools for Visual Impairment

One difference between low vision and blind people is the navigational tools they use in daily lives. The majority of blind people use a white cane to detect obstacles in traveling, but the majority of low vision people who have usable vision don't use non-visual aids(Yuhang Zhao, etc.

---

[11] https://www.youtube.com/watch?v=L-AUf0Ky1GM&t=378s

[11] A screen reader that gives audible descriptions of mobile screen

[12] https://en.wikipedia.org/wiki/Global_Positioning_System

2019)To better understand how people with limited vision navigate in practice, the first step is to clarify what navigation technologies are currently used. In the apple store, there are many applications designed to help people who are blind to locate their way (Table 2).

Table 2. Way-finding mobile applications for visually impaired people on Apple Store

Applications found from Apple Store
Be My Eye A free app that connects blind and low-vision people with sighted volunteers and company representatives for visual assistance through a live video call. (“Be My Eyes - See the world together,” 2022).
Soundscape A product from Microsoft Research uses innovative audio-based technology to give people a better sense of their surroundings, thereby enabling them to feel more confident and empowered on the road. Instead of step-by-step navigation apps, Soundscape uses 3D audio cues to enrich ambient awareness and provide a new way to interact with the environment
Seeing AI Microsoft's artificial intelligence app for iOS system. People with visual impairments are able to hear the descriptions of people and objects that the app identifies with the camera of their devices.
RightHear A virtual accessibility assistant that helps users to easily orient themselves in new or casual environments.
Clew Clew is a navigation app designed for blind and visually impaired individuals. The device can save a location, such as a seat or a room, so that you can return to it after exploring your surroundings

GPS systems, camera vision, and artificial intelligence are used with these digital applications to assist users with visual detection. Novel technologies like spatial sound and Bluetooth beacons were also used to indicate users' relative locations of landmarks. A classification of these tools is presented (Figure 5). They can be divided into low-tech and digital devices based on whether they are equipped with digital cues or not. Low-tech tools, such as a white cane, guided dog, or sighted people, are true, but less flexible. White canes can help blind people sense an obstacle without delay, but they cannot support reading signs. It is true that sighted people provide more comprehensive assistance, such as offering conversational instructions, so that users can build an image in their minds. It is difficult to have sighted people around. The term digital tools can be further extended into digital applications and experimental tools. (Figure 5) Some digital devices are mainstream applications, widely used by normal sighted people. For example, Google Maps, which uses a GPS system to provide satellite imagery, aerial photography, street maps, 360-degree interactive panoramic views of streets (Street View), real-time traffic conditions, and route planning for travel by foot, car, bicycle, air (in beta), and by public transportation. Furthermore, other applications like Be My eyes, Soundscape, RightHear, Seeing AI, Lookout, Clew are also designed for the blind. These applications mainly rely on computer vision and artificial intelligence to help users identify objects. Specifically, RightHear, Seeing AI, and Lookout are basically computer vision to convert spatial information, such as objects in front, into audio guidance. They are indeed improving the affordance of way-finding.

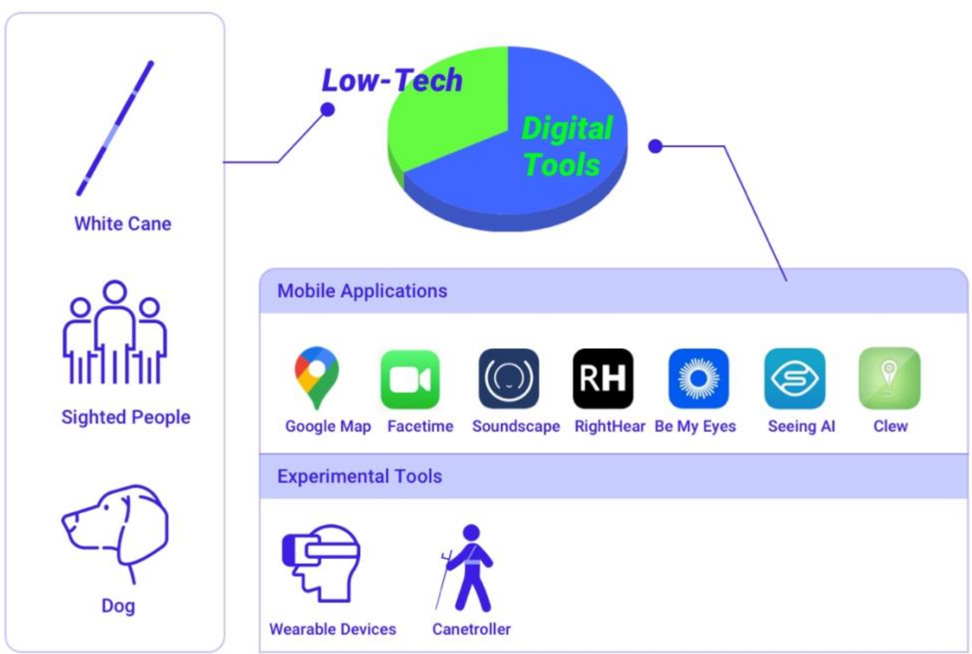


Figure 5. According to the classification of existing way-finding tools for people with visual impairment, it can be seen that digital tools occupy a major proportion.

Furthermore, digital tools are not limited to the mobile application. Researchers also implemented applications on mixed reality platforms, such as HoloLens[13] and Canetroller[14]. Specifically, the Canetroller transforms the white cane into a digital device that facilitates multiple sensors so that users can receive depth data and tactile feedback, compensating for vision cognition. However, most digital tools mentioned above can only support either of indoor or outdoor way-finding. Particularly, indoor navigation systems relying on RFID tags[15], Bluetooth beacons, and other means for localization remain experimental.(“Guided Meditation VR on Oculus Rift Oculus”, 2022)

2.3 Mixed Reality

2.3.1 Definition

The term ‘Mixed Reality’ (MILGRAM & Fumio KISHINO, 1994), was defined as a subset of Virtual Reality and Augmented Reality which comes in hybrid form. It blends the physical and digital worlds, based on advancements in computer vision, graphical processing, display technologies, input systems, and cloud computing. (qianw211, 2022) A Mixed Reality spectrum (figure 6) demonstrates that one end of the spectrum is a physical space while from the other end, we have the corresponding totally digital world, the virtual environment.

---

[13] A smart wearable device created by Microsoft

[14] An interactive cane controller created by Microsoft research team, providing sound and vibration to map out virtual space

[15] Radio-frequency identification uses electromagnetic fields to automatically identify and track tags attached to objects.(Wikipedia, 2022)

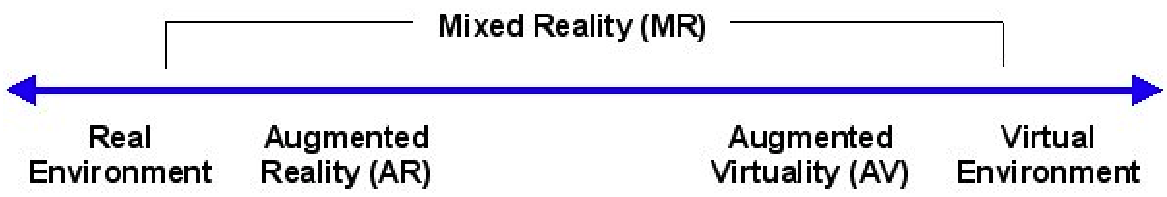


Figure 6. Spectrum of extended reality shows a taxonomy of Mixed Reality Visual Displays. (MILGRAM & Fumio KISHINO, 1994)

In the real world, objects are visible, physical, and touchable, we feel the world with multiple senses. Virtual Reality(VR) is a system that simulates a virtual environment using cutting-edge graphics, artificially rendered. Computer-generated environments allow users to experience an immersive multi-sensory experience that is separated from the real world. People are traveling less because of the epidemic. Many people are eager to escape from their existing living environment and experience the outdoors or exotic scenery. VR Meditation is one of the applications on the Meta Quest platform16, that allows users to wear VR headsets to do mediation in their desired places. There are a variety of options, such as forest areas or waterfalls. Users have a 360-degree view of the scene and can escape the physical world with spatial sound.

In Augmented Reality (AR), information is overlayed on the world in a way which enhances the user's sensory experience, by overlaying visual, auditory or other types of information. Mixed Reality (MR) goes beyond Augmented Reality (AR) by providing users with a hybrid interface where virtual content and real space are seamlessly combined. VR devices block users' view of the real world while holographic MR devices allow users to see through the display with digital content. Other than the visual display, the MR technologies also allow for more human interaction, such as hand-tracking, eye-tracking, and speech input. The experts wear MR devices,

---

[16] A platform of VR game

for example, to deconstruct the virtual models to review the surgical plan, or to overlay the surgical plan on the patient at a glance. Microsoft indicates that MR also incorporates spatial sound, object recognition(AI), and capturing distances between users and virtual objects in real time. These features indicate that MR technology evolves human interaction continuously. Individuals are able to consume content or access digital assets differently through a hybrid interface. When designing mixed reality systems, it is important to take accessibility as a foundational principle. Designing for aceessibility is not hard(Stanley, 2018), User-centered Design gives us the power and responsibility to ensure that everyone has access to what we have created, regardless of ability, context, or situation There are more than a billion people worldwide with some type of disability. Integrating accessibility into a system's iterative design process is important for everyone.

2.3.2 Accessibility Mixed Reality and Related Works

Designing in both AR and VR projects has a similar focus based on the interface. User interfaces are the access points where users interact with systems, specifically in terms of Graphical User interface(GUI)[17], Voice Controller Interfaces(VUI)[18], Gesture-based Interface.[19] GUI design usually concentrates on the 2D surfaces, specifically making the layout hierarchical and readable, keeping the buttons and labels informative and predictable. A well-designed interface is also accessible to users of all abilities, whether they have low vision, blindness, hearing impairment, or motor impairments.

[17] A form of interface allowing users to control interact with electronic devices through graphical icons

[18] A form of interface allowing users to use voice to interact with electronic devices

[19] A system allows people use gestures to control the devices

MR interfaces come in a hybrid form of a digital control panel, sound controller, and interactive space. Users of all sorts of abilities can undoubtedly take full advantage of its strengths. For instance, immersive sound, text, and image magnification functionality can remedy the berries of needs among blind and visually impaired users. While the MR industry is rapidly growing, accessibility is still an afterthought, as early VR experiences rely heavily on visual perception. Much as the two-dimensional digital interface has become an integral part of human life and work, the features of MR may suggest a similar trend. Accessibility is a human right (Aviv Elor, Joel Ward, 2021). The following projects address accessible design perspectives in MR technologies. As we move into a hybrid world, we should ensure the users include people with different abilities.

A Haptic and Auditory White Cane

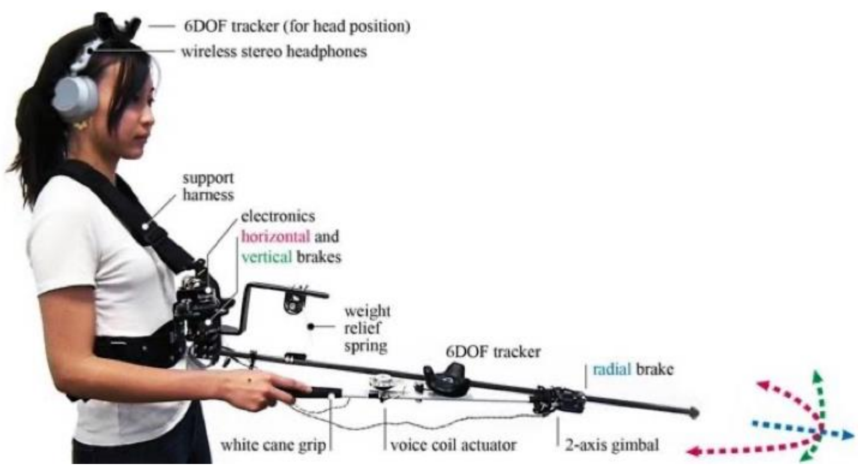


Figure 7. This picture [20]features the components of cane controller and the way of wearing, as well as a hybrid user interface.(Siu et al., 2020）

Microsoft's Accessible Mixed Reality team designed the haptic and auditory white cane (Figure7) in 2020. This is a lightweight cane controller with axis brake mechanism sensors. As the

[20] Permitted to copy this image, copyrights owned by authors.

device is lightweight, users are able to move freely from one point to another. The scale and shape of virtual objects can be felt by people who are blind. Based on contact vibration, a voice coil actuator renders textual information, while spatial audio is determined by the progression of sound through geometry surrounding a user. Moreover, this device has been approved to be a useful tool in helping blind people locate virtual targets without any collisions.

AR Visualizations to Facilitate Stair Navigation

Research team [21] designed an AR visualization stair navigation tool on a projection-based device and smart glass. (Figure 8) In the projection-based AR system, participants held a handheld projector that displayed animated highlights on the stairs, along with auditory feedback. Through the optical see-through smart classes application, the staircase shows a virtual path that is animated with a glow effect. Based on user testing, the glow animation was helpful, while some users appreciated that the path graphic provided a clear indication of the stairs' direction. Though people have preferences for the display of visualization, this experimental project approves the effectiveness of a hybrid interface in navigation.

---

[21] Yuhang Zhao, Elizabeth Kupferstein, Brenda Veronica Castro, Steven Feiner from Jacobs Technion-Cornell Institute, Cornell Tech Cornell University; Shiri AzenKot from Department of Computer Sceince, Columbia University

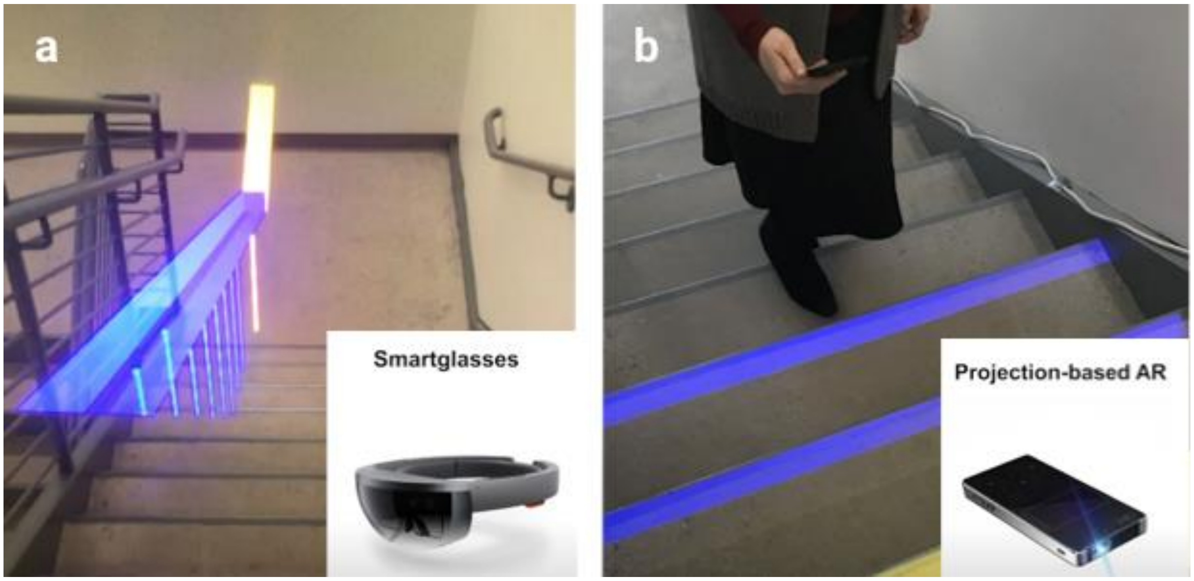


Figure 8. Screenshot shows AR Visualizations: (a)smart glasses,(b) projection-based AR to facilitate stair navigation for low vision people[22]

2.4 Conversational User Interface

2.4.1 Definitions

Language is a means of human communication. Some words are abstract, but people can understand when these words appear in different contexts. Conversation provides a context and allows people to engage naturally and intuitively, but it also uses social relationships, like talking before a meeting. Intentions also accompany conversational behavior, such as calling to book a flight. A natural conversational interaction also conveys the meaning of the words and includes peoples' personalities or emotional states. (McTear et al., 2016) For example, a face-to-face conversation covers the literal content and non-linguistic languages, such as gestures and personal emotion. (Figure 9)

---

[22] https://www.youtube.com/watch?v=TdgYSpXJAbs

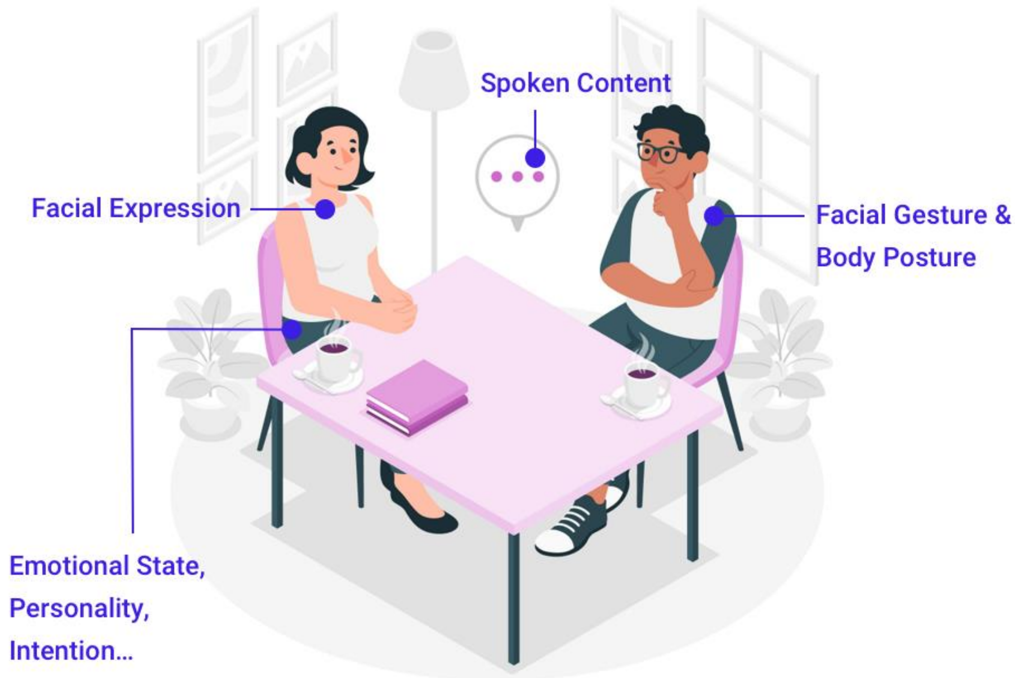


Figure 9. A face-to-face conversation usually comes along with behavioral languages.

Computer systems cannot directly understand human language; the programming language is designed for mechanical devices or computers, which is more systematic. Unlike human langue, which has morphology or specific context, it expresses the particular meaning of things. Though programming languages are artificial creations, humans design their rules and definition. Within strict sets of rules, there is no space for evolving as the human language develops.

Conversational User Interface (CUI) is the access point that enable people to interact with computational devices with conversational language. The speech act theory 23 defines the utterance not only presents information but performs an action as well. Typically, a conversational interface consists of several components (Figure 10) and operates the input dialog in the following

[23] It was introduced by Oxford philosopher J.L. Austin in *How to Do Things With Words* and further developed by American philosopher J.R. Searle.

way: recognizing the words-interpreting the words to commands that the computer can understand-executing the command-generating a response-sending the response to the user. From this formulation, we can see a general model that employs an intention-based conversational interface.

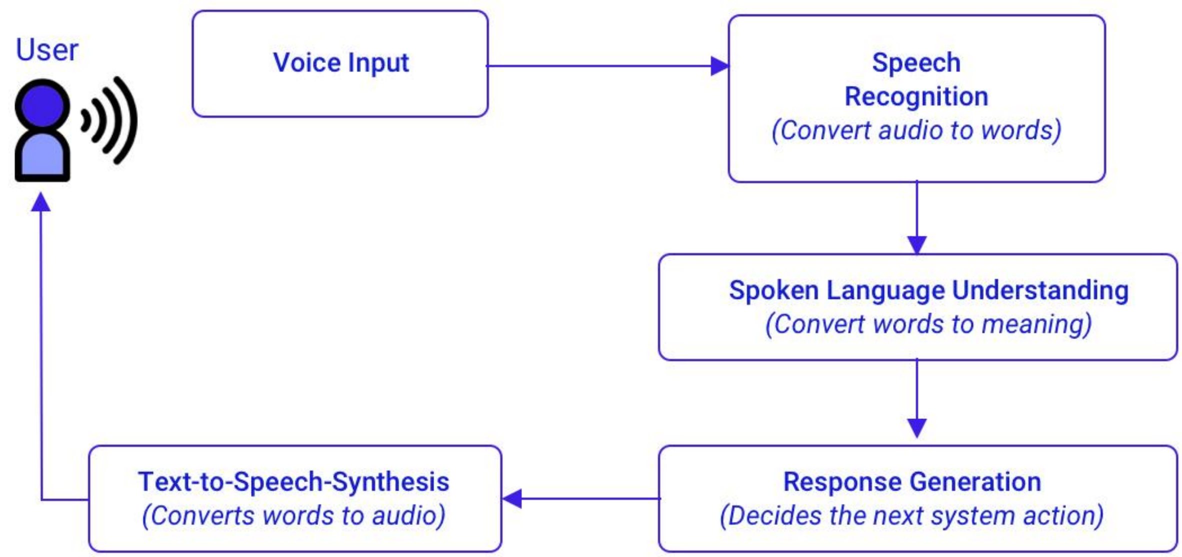


Figure 10. A process of conversational interface and working flow.

2.4.2 Types of Conversational Interfaces

Through the development of Artificial Intelligence (AI), computer systems become more human-like, allowing conversational interfaces to complete more complex tasks, such as setting alarms and finding restaurants. Up to this point, conversational interfaces can be categorized into three types: basic bots, text-based assistants, and voice-based assistants. (Cem Delugegani, 2018)

Robots and chatbots are examples of basic-bots, which are computer programs or machines. People can chat with them in the open domain, and have a spiritual companion by offering emotional comfort. The communicational function of chatbots supports language learning (e.g., conversation practice; Fryer et al., 2017). It offers students a daily learning environment without a campus. It could answer students' questions (e.g., storybook reading; Xu, Wang, Collins, Lee, &

Warschauer, 2021) and thus build an interactive study atmosphere. Among online websites, text-based conversational interfaces are widely used. A typical example is virtual customer service. Usually, we go to customer service with a specific purpose in mind, such as returning or replacing an item. A schemed conversational model allows the advance virtual customer service to understand requests by extracting keywords from input text. Voice user interfaces (VUI) are also widely adapted to the modern world. Users can use voice commands to interact with voice-based assistants. The Google Assistant and Amazon's Echo, for example, have been extensively adapted for online shopping. Customers can use voice commands to order products or add items to wish lists, and these virtual agents will assist them. People with limited vision can now use voice-based assistants, erasing barriers to accessibility. VUIs provide users with a hands-free and eye-free way to interact with a product while focusing their attention elsewhere. It offers intuitive instruction that is more accessible to people with limited vision.

2.5 Summary and Research Gap

Vision is the most intuitive way to distinguish key features about the space and understand the surroundings. For blind and visually challenged people, tools or more advanced assistance are essential for their navigation. Blind people might adapt to white canes. But in urban areas where staircases, ramps, and bends are ubiquitous, using traditional tools only brings more safety risks. Besides, partially sighted people with less or no proper training might not harness white canes at all. The existing tools are designed to assist route-finding in a very basic manner, and it has been falling behind the ranks compared to the development of cities. Consequently, it leaves a big gap in way-finding tools for both low-vision and blind users.

Mixed reality provides a possible solution for filling the gap. It is rooted in computer vision and rendering technologies, and it comes up with a hybrid interface by merging virtual rendering with reality observation. A more enhanced visualization can be seamlessly overlayed over real space, allowing environmental information to be accessed more easily. The latest MR projects indicate that they can effectively support people with vision impairments. Multi-sensory feedback, such as auditory and tactile feedback, can provide more accurate spatial information. The interaction pattern is more natural; with the help of AI, it can understand the users' voice commands or read the user's gestures. So the user could freely speak or use body language to express their thoughts in route planning, just like what they did with a human assistant. The COVID pandemic impacts people's interaction patterns dramatically, and it brings several accessibility issues. Nowadays, it is difficult for people with visual impairments always to have human assistance. But with the help of AI, MR, and VUI, users can interact with computer systems freely, with more safety insurance. This also eliminates accessibility issues for people with varying degrees of limited vision. Therefore, more research into the domains of way-finding tools for vision-impaired people is necessary. And combining MR with cutting-edge technologies like AI and VUI, more accessible tools looks promising.

3. Methodologies

This research employs User-centered Design (UCD) (Lillemaa, 2004) in user research practice. It emphasizes immersing into the user perspective to understand the problems in specific context. Within the UCD framework, I conduct user research among target users to identify problems. In addition, the design work also follows the Research through Design (RTD)(Savic et al., 2014) method by iteratively developing several prototypes.

The UCD process consists of five phases: empathize, define, generate ideas, prototype, and evaluate. Empathizing is fundamental and helps designers have a comprehensive understanding of the problems. Particularly immersing yourself in the community where the design outcomes can make a difference, reaching out to target users, finding in-situ problems, and observing behaviors to learn from them. UCD design emphasizes problem-solving and generates solutions by involving the human perspective in all steps of the problem-solving process. (Berg, 2022) Empathizing is a way to jump out of the self-perspective, especially facing problems that you never had before. My mothers' experience only represent one type of use cases. To make effective solutions for a group of users, it is crucial to learn from individuals experience with different ages, genders, or living areas. From narrative, I could put myself in their giving context and understand the impact of problems. The typical way is to observe their behaviors and have in-person talks. However, due to the ongoing COVID-19 epidemic, observation may bring people being exposed to the virus due to the epidemic. In addition, some areas have travel restriction at this moment. Meeting with them in person was not allowed. So my backup plan was to build connections by the online communities such as WeChat[24] groups in China. Inviting people to participate in online surveys, and having

[24] A social communicate app in China

audio interviews. The findings from the user research are a condensed reality that will be helpful to extract user needs and preferences.

RTD is a model for interaction design research. Following this framework, designers can generate new technological products and develop improved solutions. The study usually emphasizes potential solutions through deep reflection and an iterative understanding of people, problems, and the context around a situation. This process involves developing a prototype iteratively to better understand complex cases. From the prior prototype, the designer can obtain insights. My study includes the design work which is artifice-led and practice-based. Research knowledge is a standard component for ideation. With the overlapped area in MR and VUI, I combine insights from theoretical studies with online user research findings to make prototypes iteratively.

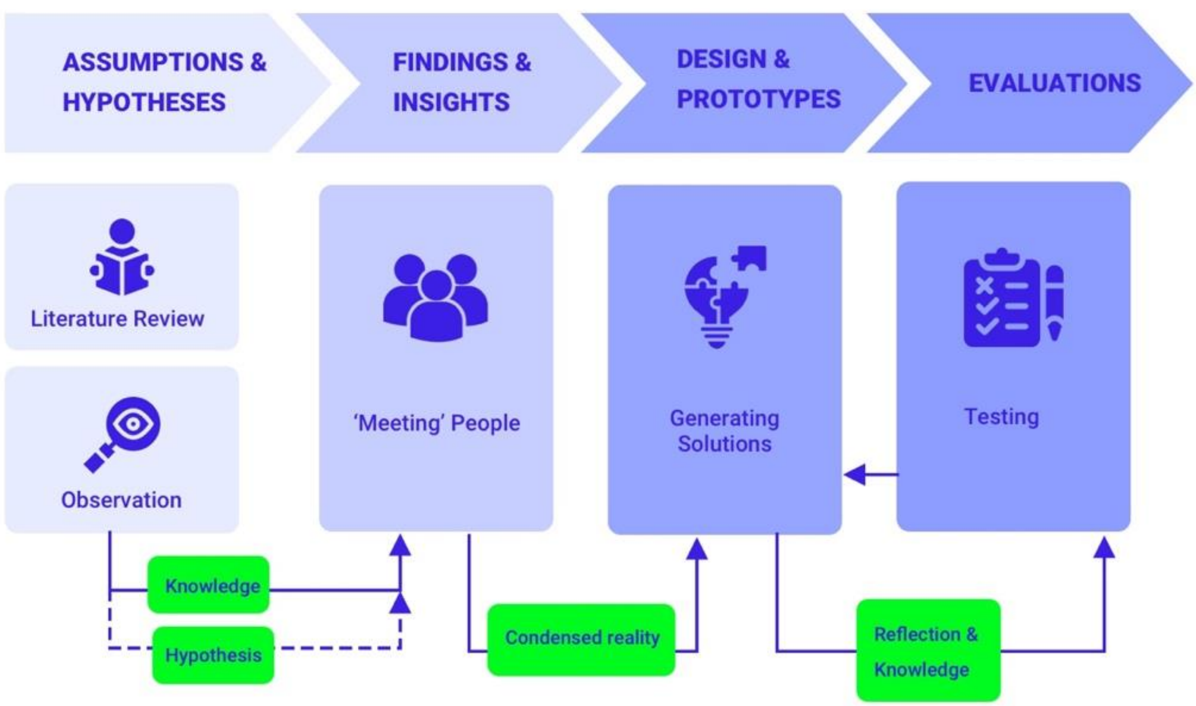


Figure 11. A diagram demonstrates the working process.

The whole process can be concluded into the following steps(Figure 11):

- Doing literature Review, along with the personal experience, to raise the hypothesis
- Needs Finding form online survey and audio interview/problem Identification
- Ideation in design solutions and making iterative prototypes
- Using knowledge-based design methods to evaluate the outcomes

4. User Research Online

In this chapter, we introduce the process of user research. The questionnaire was distributed to a WeChat group that has 100 people. A voice interview invitation was attached to it. All the activities are voluntary. In total, I received 53 questionnaire feedback forms and had a chance to talk with 5 volunteers to learn their navigation experience.

4.1 A Semi-structured online survey

A Semi-instructed questionnaire (Appendix A) is designed to identify general characteristics of target users. The basic questions such as age, occupation, and living area provide an initial background of the user. Way-finding related questions were followed after the basics, they were raised from the following aspects:

- What are the most common ways of finding their way?
- How do they use these tools?
- What kinds of information is information for way-finding?
- How much do they know about new technology tools?
- How do they think about the way-finding task?

A questionnaire was distributed to the group of people with visual impairments in China. It is a 100 sized online group that contains peoples of different ages, genders, and occupations. This ensured the diversity of data samples. To make sure each member can access the questionnaire, I used the program called Wenjuanxing[25], which is a survery tool from WeChat. After uploading questions, it generated a link for participants. (Figure 12) The group coordinator helped me post

---

[25] An online platform allows researchers to upload lists of questions and generate the questionnaire link

the link to the WeChat group, individuals could open it from any device. This link was activated for 2 weeks, and 53 pieces of response were collected in the end.

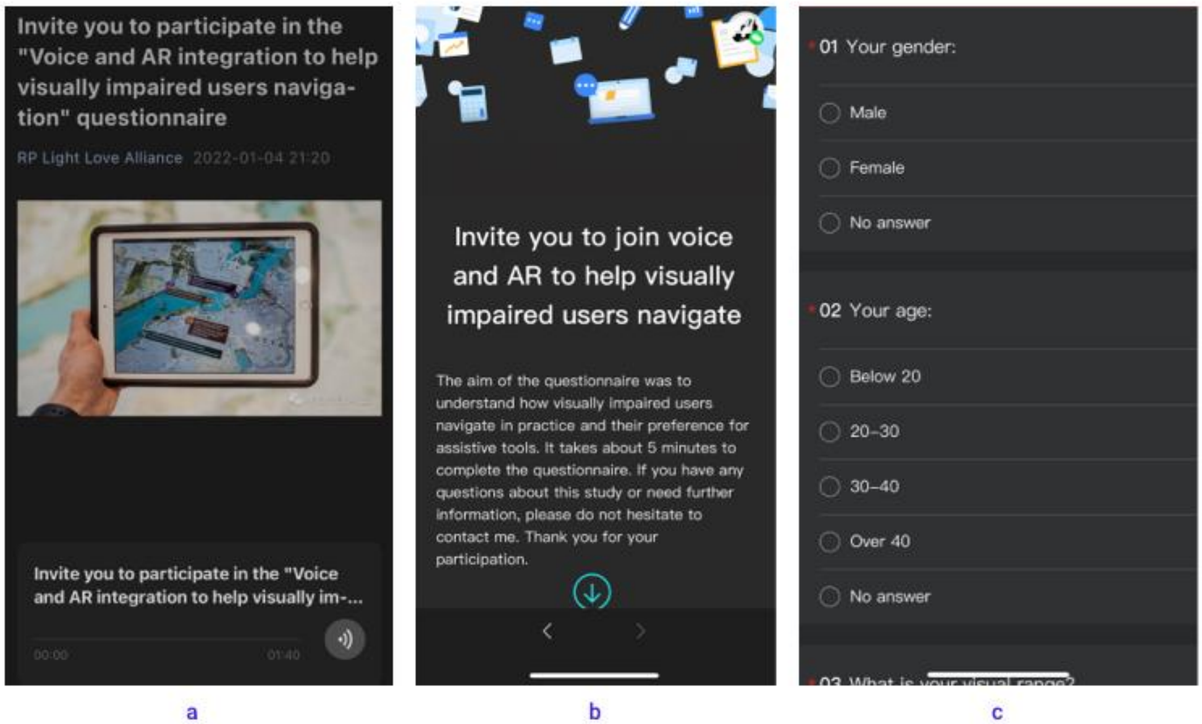


Figure 12(a)(b). Cover pages of questionnaire displayed in mobile device (c)the display of questions on mobile device, users used screen reader to get questions and gave responses

4.2 Audio Interviews

During the opening period of the questionnaire, several participants contacted me through WeChat and expressed their willingness to participate in the voice interview. Participants in the audio interview were categorized into three types: blind, low vision, and normal sight person who experienced situational vision problems. Four interviewees are from the Wechat group who joined the online survey. Another participant was a volunteer that helped blind people go hiking in Hong Kong. This audio interview follows a basic script (Appendix C) to raise the questions in sequence. Data from the prior online survey provides a general knowledge of the tools they used and what kinds of information were significant to them. Based on their daily routines, I went through in-situ challenges they had before. I took notes during each round of interviews and summarized the answers into a form. (Appendix D)

From their narrative description, I found differences in their transportation and living areas. However, they use the same tools, such as Google Maps and white cane. However, they used these tools in a different situations with the individual's intention. All these finding, along with the data from the online survey were concluded in the following sections.

4.3 Need-findings from User Research

4.3.1 User characteristics

The primary user information shows that (Table3) more people tend to have vision problems after 40 years. One of the interviewees also mentioned that he felt his eyesight deteriorate more rapidly after turning 40 years old. Meanwhile, aging will bring other access issues in mobility and cognition. It may take longer for older people to walk through the stairs. Or they can not catch up with the evolving progress in digital tools, having operational barriers to using new tools. The aging of society may result in more visually severe issues. From a foresight perspective, navigation tools should consider user mobility and their knowledge of digital devices. Older people may need more time to learn the instrument. Therefore, the interface design should present the information more directly.

Table 3. Basic information collection from online survey

Your Age:		
Under 20	3	5.70%
20-30	11	20.80%
30-40	13	24.50%
More than 40	26	49.10%
Not to answer	0	0
Which of the following mobile system do you use:		
iOS	26	49.10%
Android	26	49.10%
Other	1	1.90%
Degree of eyesight:		
Normal	1	1.90%
Low vision	17	32.70%
Legally Blind	24	45.20%
Totally Blind	11	26.20%
Which of the following area do you live in?		
Urban Center	29	73.80%
City Suburbs	13	21.40%
Suburbs	9	17.00%
Other	2	3.80%

4.3.2 Sound cue is efficient in spatial cognition

People use sound to analyze the surroundings, such as hearing a car approaching from behind or estimating the direction of a bird in a forest. Even vibrations can be used to locate objects. Auditory cues are crucial to people with visual impairments (Table 4). Additionally, tactile feedback and smells provide spatial cues. One of the interviewees described that the traffic sound is helpful for him to recognize the direction and distance of the vehicle and the pedestrian. He also mentioned that the surface texture was an important cue, especially walking on stairs. He felt the height of the steps with his feet, and by touching the structure and shape of the handrail, he could roughly determine the actions.

Table 4. Demonstrating the sound is one of the important cue for people with visual impairment

Which of the following information is important to you in way-finding? You may select more than one.		
Smell	20	37.70%
Sound	49	92.50%
Tactile information	17	32.10%
Other, please specify	7	13.20%

4.3.3 Avoiding overwhelmed information

Blind users depend primarily on their hearing to navigate, and two of the interviewees stated they use Google Maps only for route planning, traffic updates, or finding nearby stores. Blind people do not use Google Maps for walking. Hearing is a lower bandwidth source of information than vision. One interviewee described that when he was in a noisier environment, such as a crowded street, he could hear many people talking, along with vehicle sounds from a different direction. He became distracted and lost the last cue quickly. Sometimes, he analyzed the surroundings according to the smell because places such as the bakery, the cafe, or the restaurant have a unique scent. Designing the voice interface should avoid too much speech instruction; otherwise, it may overwhelm users.

Table 5. Data statistics reflects people with visual impairment prefer asking help from sighted people nearby.

Which of the following tool do you use for wayfinding? You may select more than 1		
White Cane	20	40%
Guided dog	2	2.40%
Sighted people nearby	34	66.70%
Face-time with sighted people	14	33%
Mobile application	38	69%
Other	2	3.80%

4.3.4 Existing tools cannot support way-finding adequately

Google Maps is a web mapping platform that offers users geographical information through satellite images, aerial photography, street maps, and street views. GPS maps are quite helpful when traveling in unfamiliar areas. The online survey shows that digital tools are used a lot for way-finding. Table 6 reflects that users will use several tools in actual travel. This phenomenon is also consistent with the descriptions from the interview. One of the interviewees described that sometimes he took the destination as the virtual landmark and searched the stores, bus stops, and coffee shops on Google Maps. But when he went out, he preferred to ask for other people's help. People are more trustworthy than digital devices and will give comprehensive instructions.

Table 6. People use complementary aids from different tools

How many tools did you use for wayfinding?		
Only 1	27	51%
2~3	22	41.50%
More than 3	4	7.50%

Table 7. The exiting tools do not meet the most users' needs

On a scale of 1-10, where 1 is "not at all" and 10 is "extremely", how confident do you feel when you go out independently?		
1	4	7.50%
2	1	1.90%
3	2	3.80%
4	4	7.50%
5	8	15.10%
6	7	13.20%
7	11	20%
8	3	5.70%
9	6	11.30%
10	7	13.20%

4.4 Summary

The semi-structured questionnaire was more accessible to statistical characteristics of the user base, and the interview is an effective way of understanding problems through other people's narratives. The common challenge in way-finding is identified as a failure to read visual information through user research. It resulted in lots of inconvenience and safety issues. Existing navigation tools can not support partially sighted people in visual searching. Object detection by AI, AR visualization, and VUI could provide a higher level of assurance. This system should consider more facts, such as the user characters of vision, aging, and living areas. Additionally, the VUI design should also use tactile and sound as additional feedback instead of continuously giving speech instruction to help users discover the surroundings.

5. Design and Prototypes

5.1 Persona

A user persona (Rikke Friis Dam & Teo Yu Siang, 2022) is a way to understand groups of potential users empathically. Users with different backgrounds and preferences have different expectations from the final product. By organizing the different needs and merging them into fictional and desired user roles, the user persona can be used to refine the group's typical demands. As a result, personas can help distill design concepts and shape the decision-making process. According to prior user research, I created three personas.

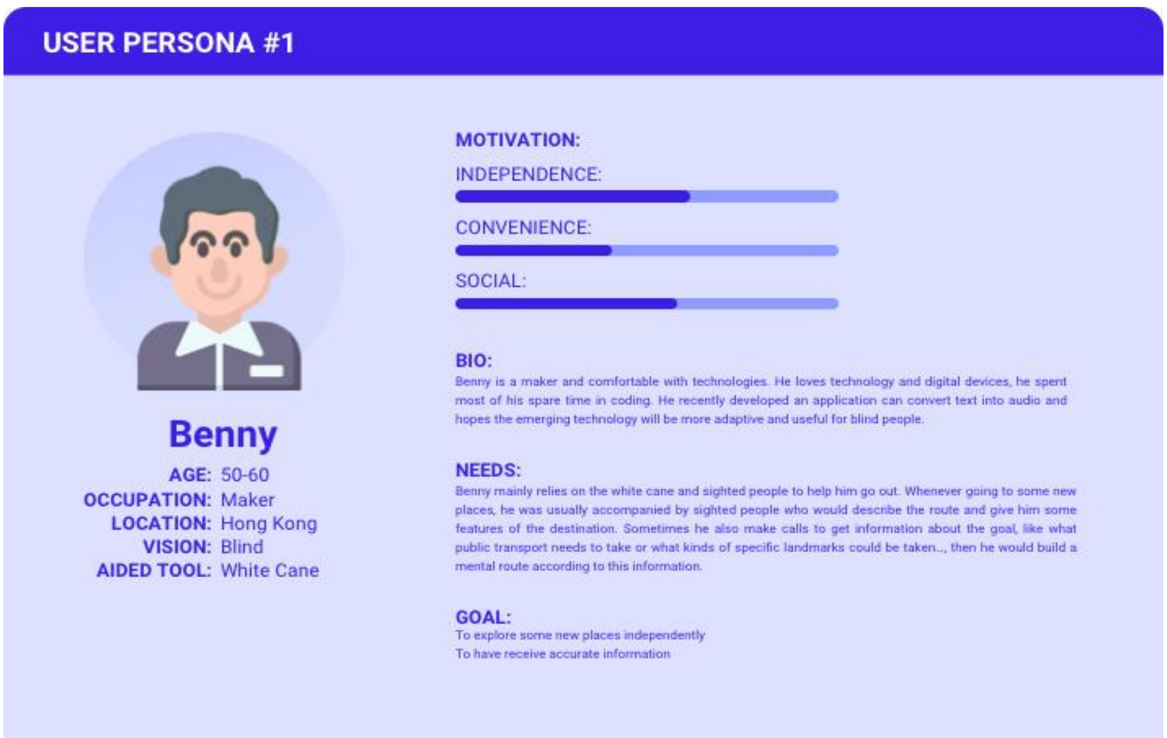


Figure 13. User Persona1 A blind person who use white cane mostly, and interested in digital tools

Figure 14.User Persona 2 A normal sighted user who had high myopia and hope to have a smart voice system to help her navigation.

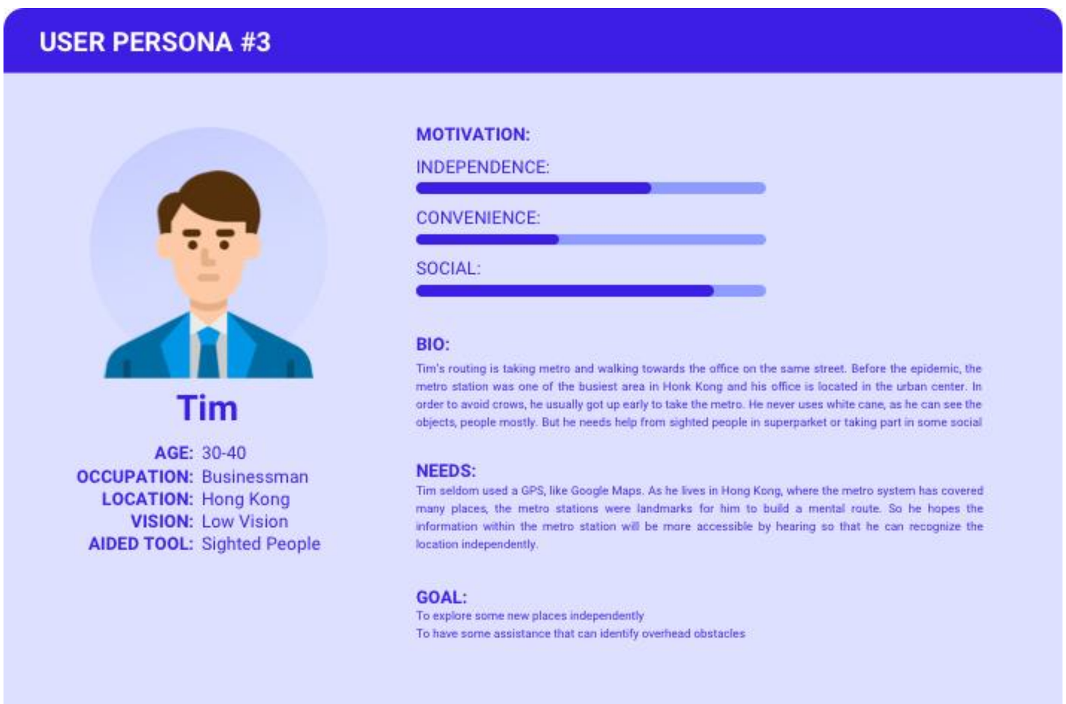


Figure 15. User Persona 3 A low vision user who has difficulty in details reading

5.2 Prototypes

This prototype map illustrates how the first four iterations contribute key elements to the final version: Wonder Vision. The crucial features are object (plane) detection, distance measurement, speech-to-text conversion, spatial audio. Wonder vision is the final version including all these modules.

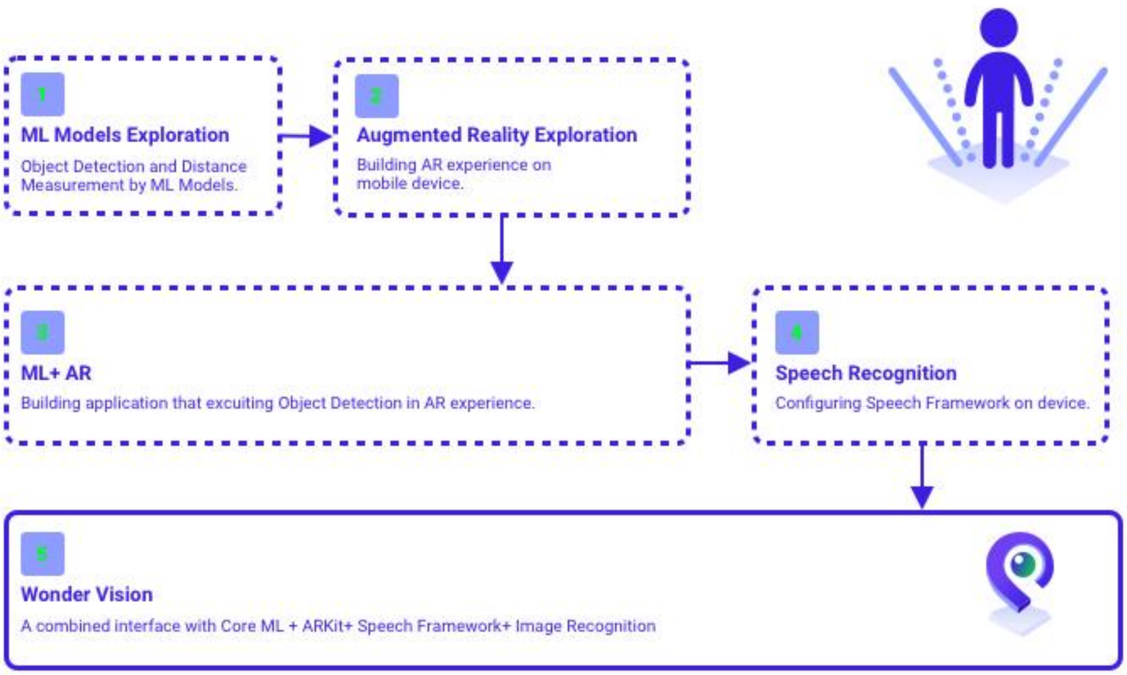


Figure 16. Map of iterative prototyping

In the selection of devices, HoloLens[26] and iPhone 13 [27]are two ideal platforms for building the application. Although HoloLens has superior processing power and is already being used in medical education, it is still too costly in the early phases of development.

---

26 HoloLens is a cable free wearable MR independent holographic device,

27 iPhone 13 is one the mobile devices launched by Apple in 2021

The prior online survey indicates the iPhone is one of the most popular smartphones with the most advanced hardware. The iPhone 13 is features three cameras, along with LIDAR28 sensor. The advantage of LIDAR is that it measures distance by reflecting lasers off objects and measuring the time it takes for the light to return to the sensor. This technology becomes mature and accurate, so it is applied everywhere, from self-driving cars to assisted driving, robotics, and drones. Due to the iPhone's affordable price and useful sensors, the prototypes were built on an iPhone.

5.2.1 Prototype 1 Object Detection and Distance measurement

Goal

A basic operational flow of AR applications is to find plane and then place virtual objects. Follow this, Vision's first objective is to identify objects and measure their physical distance.

Process & Outcomes

MediaPipe [29] is a Machine Learning (ML) platform built by Google, and its models are open-sourced for non-commercial use. There are two functions for detecting 2D objects and 3D objects separately. In the first stage, an object detector finds a 2D crop of the object. The latter is more advanced by using a two-stage pipeline (Figure 17). In the second stage, the image is cropped and 3D bounding boxes are calculated. Ideally, 3D prediction (Figure 18) would perform more accurately if it were able to capture additional information about an object, such as its size, position, and orientation.

---

28 LIDAR stands for "light detection and ranging"

[29] https://google.github.io/mediapipe/

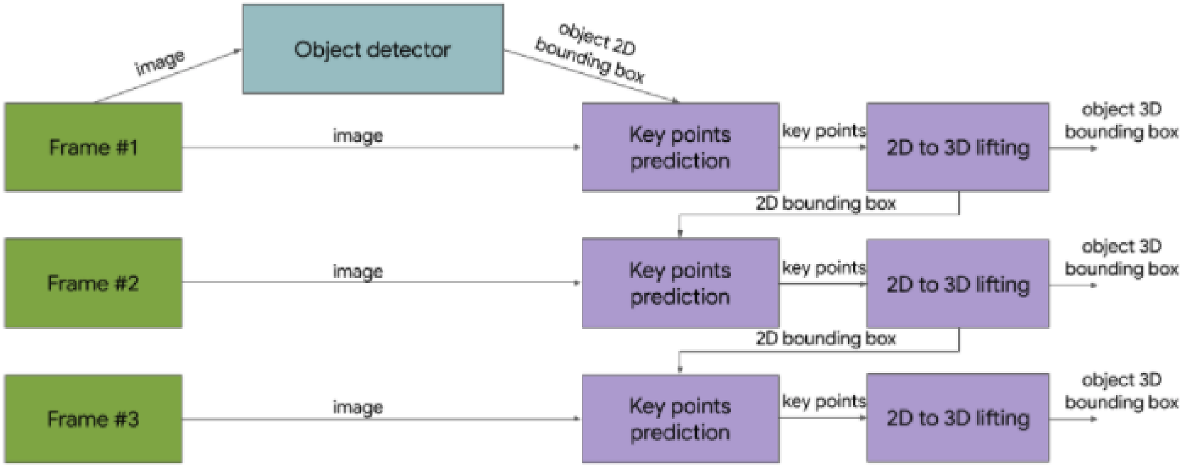


Figure 17. Network architecture and post-processing for two-stage 3D object detection("MediaPipe Objectron," 2021)

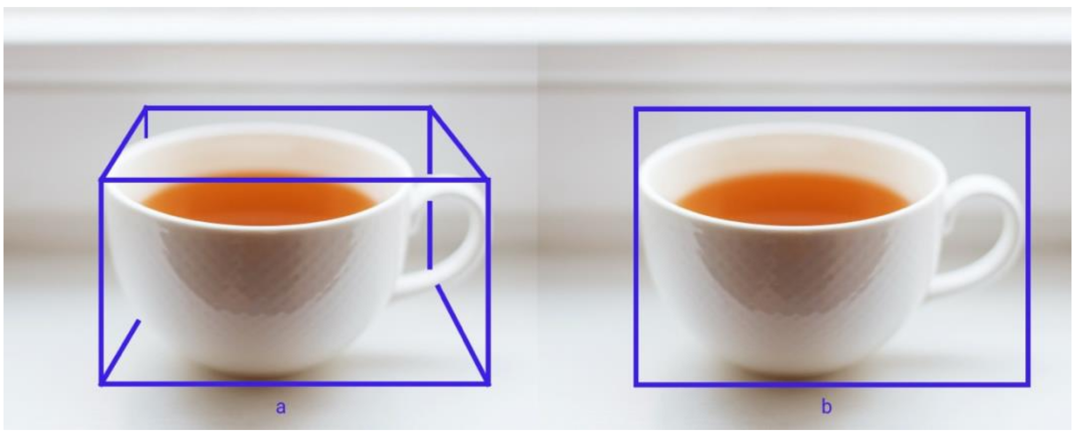


Figure 18(a). A representation of 3D bounding box (b).A representation of 2D bounding box

Though 3D prediction is more advanced, but it cannot be integrated on iOS system. As so, I built 2D prediction on iPhone and test it in an outdoor space. Figure 19 display the performance of 2D prediction:

- The application runs smoothly on the device
- The system can detect moving people and static objects
- Distance measurements beyond one meter are inaccurate

Figure 19. Prototype 1 Demo : (a)(b)The performance of 2D object detection run iPhone

Reflections

The 2D prediction performs well in detection, but is less accurate in distance measurement. This is because their coordinates are different (Figure 20). A 3D prediction has coordinates (x, y, z) and the center of the bounding box is (0,0,0). The image coordinate has only (x, y), and the origin point is different from the object coordinate. To have accurate distance data, it is essential to find a 3D object detection model.

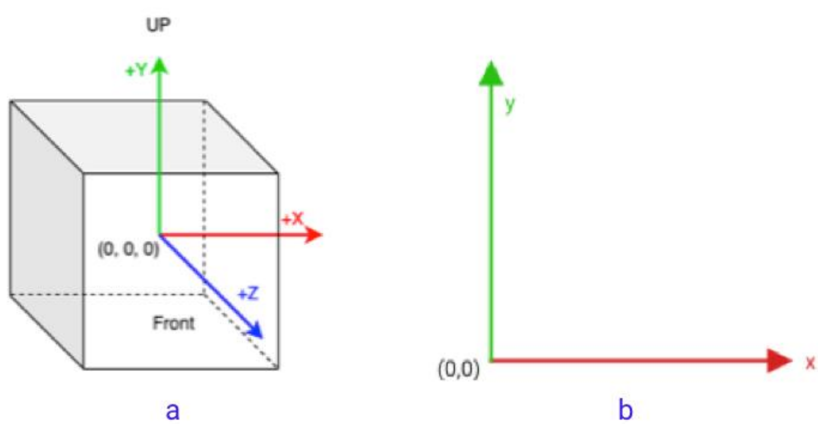


Figure 20 (a). 3D object coordinates((“MediaPipe Objectron,” 2021) (b)2D object coordinates.

5.2.2 Prototype 2 Augmented Text in Real Space

Goal

The second prototype intends to harness the power of ARKit by enabling the user to create 3D anchors and mounting virtual models on top of them. This will be helpful when we combine it with object detection, but for this prototype only, the virtual models will be created by tapping on the screen.

Process & Outcomes

The development tool is ARKit[30](Figure 21(a)). It combines a high-performance rendering engine with an API for importing, editing, and rendering 3D content. Based on a default SceneKit Demo (Figure (b)), I created a prototype that can anchor a 3D object on the tapping point. It uses the ScenicKit framework for creating 3D anchors and an event handler function to implement the model loading and placement. The 3D text is a pre-generated model. Figure (c)) The performance shows that 3D text looks like a real sign fitting into the surroundings, rotating and changing size in response to user perspective and distance.(Figure 22)

---

[30] AR developing tool from Apple

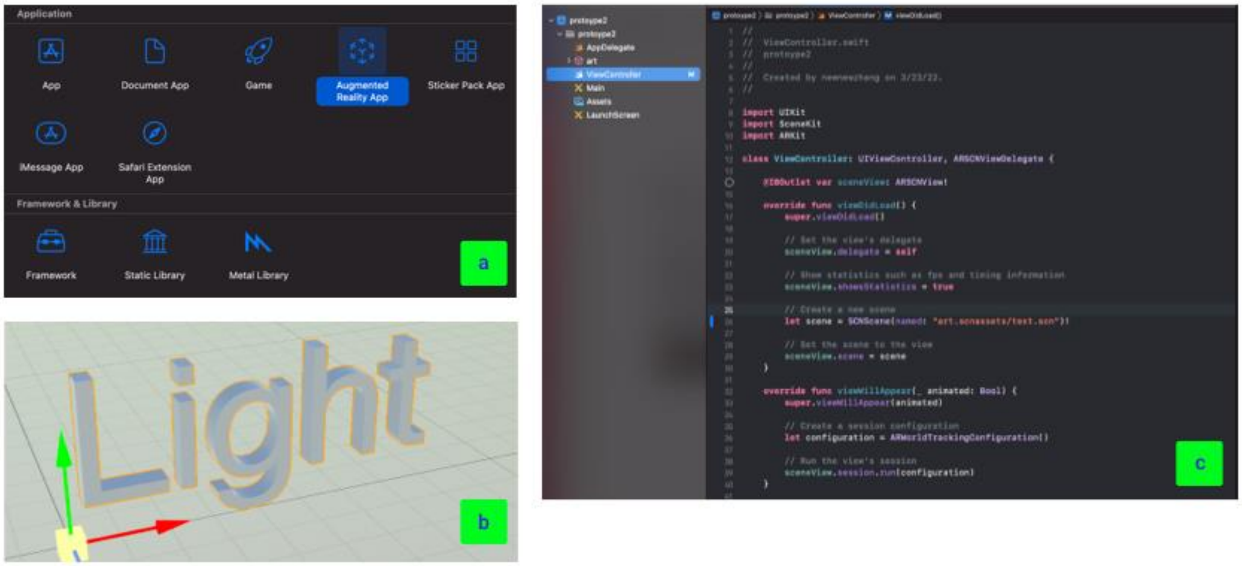


Figure 21. Developing process of prototype 2(a)creating workplace on ARKit,(b)designed 3D text (c)workplace on ARKit.

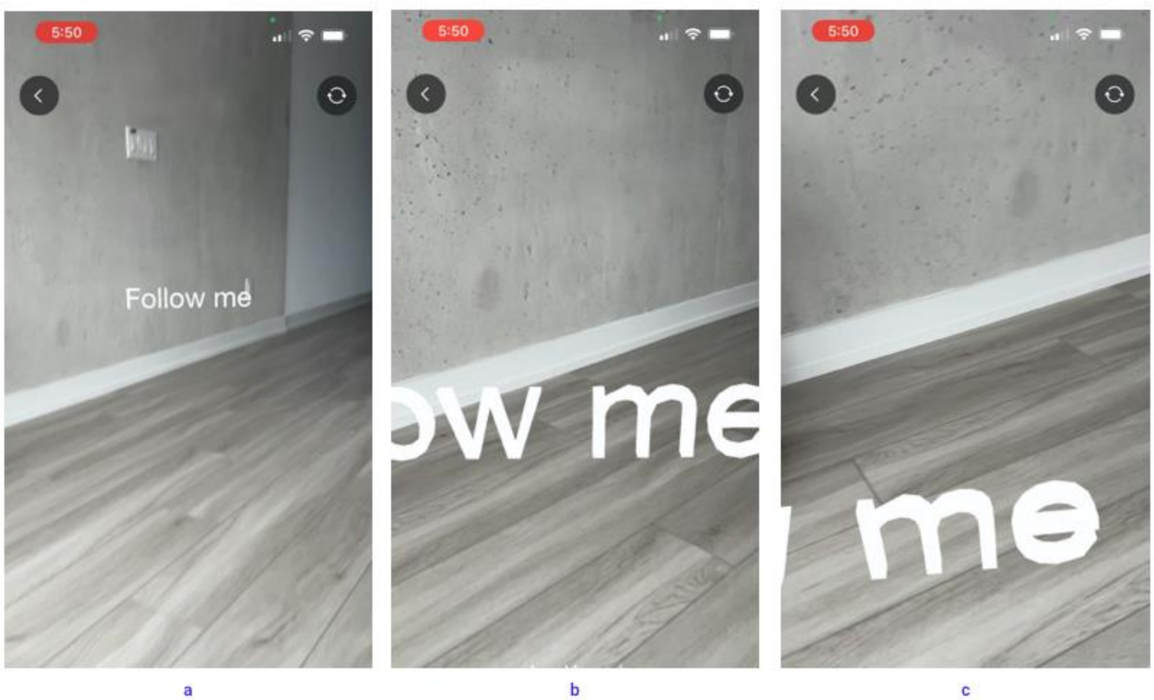


Figure 22(a)(b)(c). Screenshot of playing with prototype, the text becomes bigger in approaching.

Figure 23 (a). AR generated text displayed in dark scene, (b)A Physical sign displayed in dark scene,(c)AR generated and physical sign displayed in same scene

Reflections

The second prototype shows how AR objects can be seamlessly embedded into real space. Even in dark, the content can be clearly seen. For people with low vision, AR content is easier to access and is less affected by the environment lighting (Figure 23(a)), on the other hand, the physical sign might be difficult to read in darkness (Figure 22(b)). (Figure 22(b))

5.2.3 Prototype 3 Exploration of Core ML

Goal

The purpose of this prototype is to find object detection and classification model that can be integrated on iOS system.

Process & Outcomes

As Media Pipe's ML models did not support the iPhone, in this prototype, Apple's CoreML[31] was used. The research community has developed many pre-configured machine learning models for image classification, object detection, pose detection, drawing classification, text analysis, etc. I tested with object detection. Figure 24 shows the working behavior: identifying classified objects with confidence levels.(Figure 24)

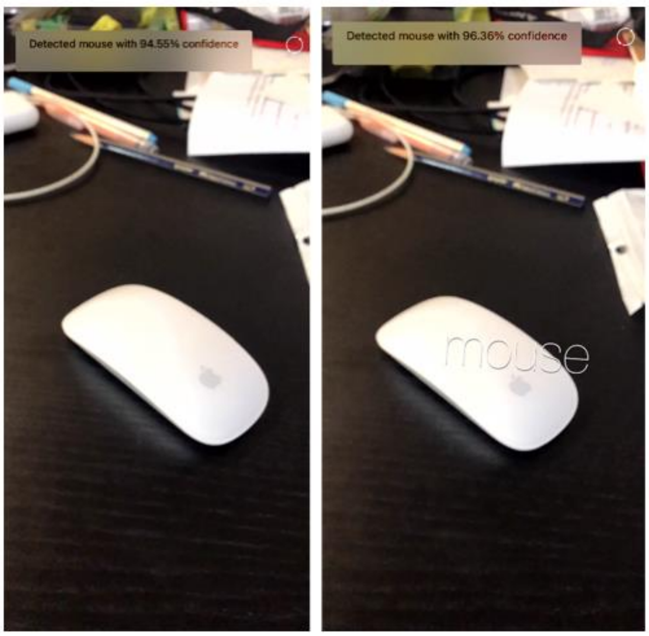


Figure 24. Performance of prototype 3: using Core ML to detect the mouse and an augmented title and confidence of prediction are displayed.

Reflections

This model works fast and accurately recognize ordinary things such as furniture, food, and drinks. For a typical way-finding task, it is crucial to identify traffic, signs, walking people,

---

[31] https://developer.apple.com/machine-learning/models/

and roadside dumpsters. Unfortunately, these classes are not provided in pre-configured models. Training a model for detecting objects in way-finding requires a large amount of data and a high-performance GPU. To make a functional prototype in a short time, I need to define the task more specifically. So I narrowed down the object of interest to be a fixed but common sign: the exit symbol. In the final product, I set up the system to find the 'EXIT' symbol.

5.2.4 Prototype 4 Audio Recognition Exploration

Goal

This prototype aims to experiment speech-to-text feature provided by the Speech framework. I created an application that could transcribe users' voice input lively.

Process & Outcomes

Apple offers a Speech Framework [32]that can perform speech recognition on live or pre-recorded audio, receive transcriptions, alternative interpretations, and a confidence rating for the results. Since I am not familiar with audio development, I started with an online tutorial that built an application that uses speech recognition to transcribe some pop song recordings. A live audio transcriber dictates each word the user speaks and appends it to the current text.(Rames, 2017) The outcomes (Figure 25) show the system can convert speech to text with almost no lagging, and these text inputs could be fed into a state machine to be translated into commands that the system could understand.

[32] https://developer.apple.com/documentation/speech

Figure 25. Prototype 4:testing with live audio recognition, the audio input of the camera capture peoples' speech and converting to captions.

Reflections

In this prototype, I experiment with speech recognition. I found that the maximum input audio is one sentence. This framework only supports text to speech without any commands. Each sentence should include keywords so that the computer can recognize the request and act accordingly.

5.2.5 Prototype 5 Wonder Vision

Goal

This prototype is to integrate all the features experimented within previous prototypes together and create a functional prototype that helps users detect the Exit signs and provides comprehensive instructions.

Process & Outcomes

The application that I created detects the 'Exit' sign, creates a virtual anchor, and mounts an magnified 3D exit sign. With the help of the virtual anchor, the system is able to measure the distance between the user and the anchor. The working flow of the system is illustrated by a diagram (Figure 26). First, the system requests permissions to access sensitive data from cameras and microphones. Then, the camera searches for the Exit sign with a trained CoreML model. Once found, the system creates the virtual anchor and enters a guiding mode, providing voice guidance for the user to approach the exit. In order to prevent the user from being overwhelmed by the voice instructions, the distance and location are articulated at the beginning, and later, the distance will be encoded into beeps. The beep sound increases in volume and shortens the interval as the user approaches the target to provide an extra hint. Finally, when the user arrives at the exit, the system ends the guidance mode, and waits for the next command.

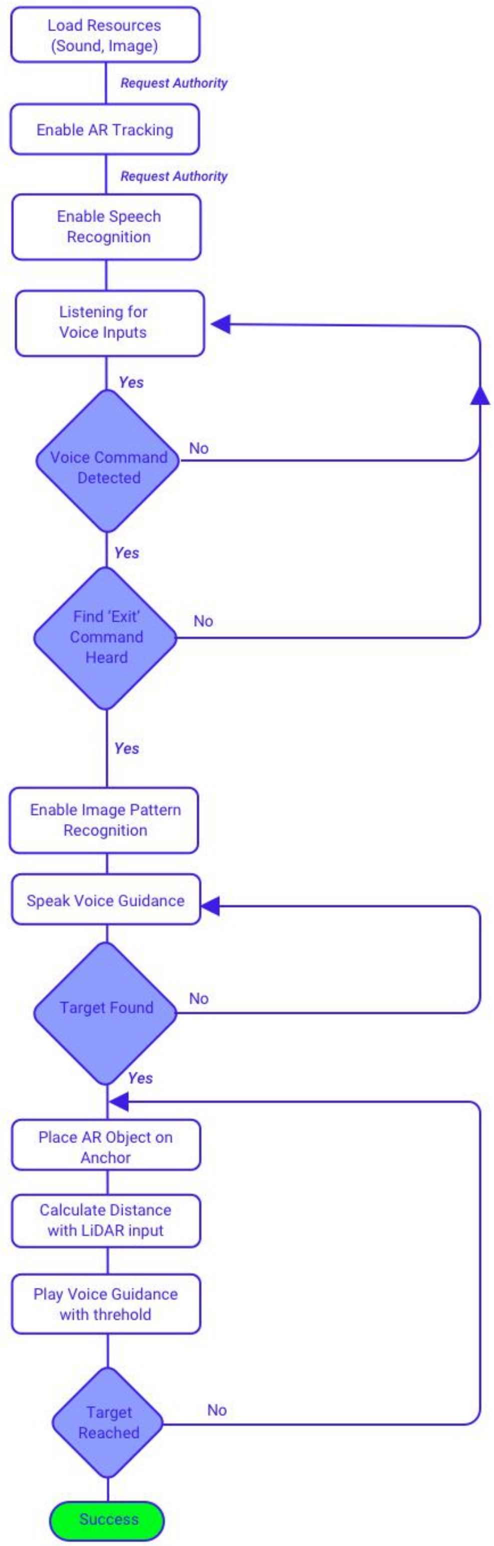


Figure 26. A system working flow of Wonder Vision.

Voice interfaces follow a plan-based scheme. Conversations begin with a request, such as finding a specific object. The system will recognize the keywords in preliminary and convert them into commands that the computer can understand. According to the request-based dialog, I designed a conversional scheme(Table 8), “[…]” square brackets are used for placing pre-defined keywords, the system extracts these keywords to take actions. A series of user interfaces were then designed. (Figure 27)

Table 8. Key words recognition based dialogue of searching 'Exit' sign

Step1. Using ‘Siri’ to open the application.
User: Hey, **[Wonder]**, where is my **[Exit]**? Wonder: okay, please look around and let me search for it.
Step 2. Searching the environment and locate the object.
Wonder: The **[Exit]** is in **[front]** of you, about **[0.1-1.0]**meters away. Do you want to get there?
User: **[Yes/No].**
If the system failed to locate the target.
Wonder: I cannot find the **[…],** please move your head and let me do the searching again. Wonder: I cannot find the **[…],** please turn to the **[right/ left]** side slowly.
Step 3. Starts to move: tracking movements and playing special sound
User is approaching The sound of ‘Beep’ gets louder, and frequency quickens
User is deviating from the target(the system can’t see the target)
Wonder: You're **[deviating/still deviating]**, please turn to the **[right/ left]** side slowly.

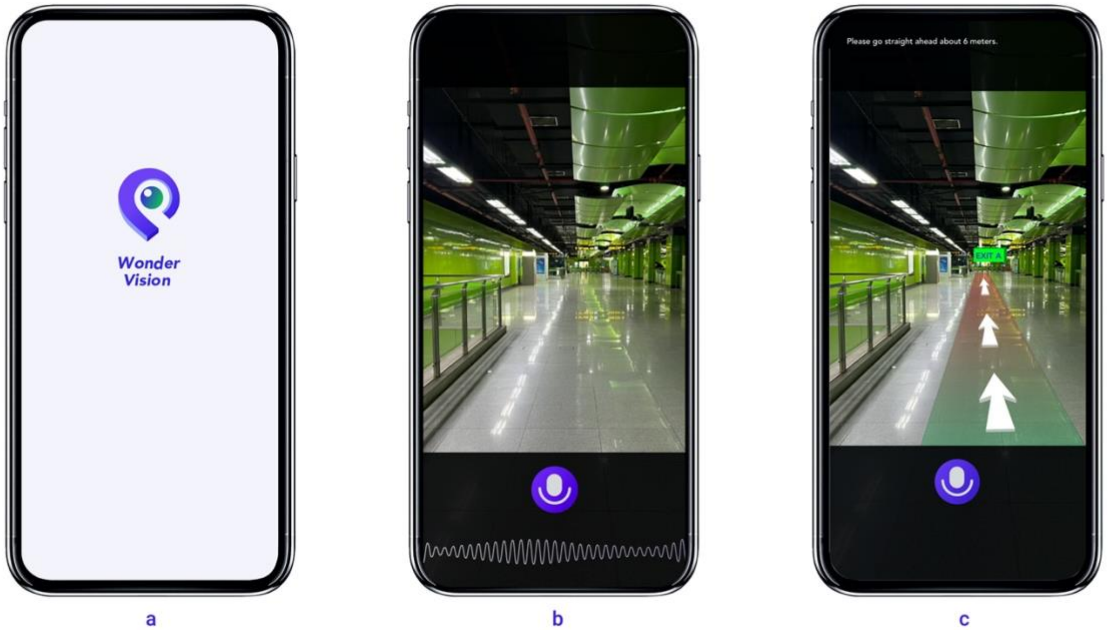


Figure 27. The visua interfacel design of the Wonder Vision:(a)loading page, (b)system has been activated, user holds the button to speak to the system at the bottom,(c)A virtual Exit has been placed in the real space, along with route visualization.

I made this prototype on ARKit. (Figure 29(a))The default SceneKit Demo (Figure (b)) provides a framework for displaying virtual objects in real space. A 3D text is pre-generated. Figure (c))

Figure 29 shows the phases of using Wonder Vision. Initially, I activated 'Siri' with a voice command to open Wonder Vision, and the camera started searching. Once the physical 'Exit' sign was found, a virtual 'Exit' model would be displayed in the real space. The virtual object is not affected by the environmental lighting, so it is highly recognizable. Once collected the distance information from LiDAR, the system started to give voice instructions. And later distance is encoded into a repeated beep sound when the user moves near the target. The system only gave instructions whenever it thinks the user is losing his/her track. The Wonder Vision focuses on the guidance with text-based signs in public transportation, the current version was inspired by the example of Detecting Images in an AR experience[27]. This example app looks for any of the several reference images, and once successfully detected, it can tringle the next step of placing a virtual anchor.

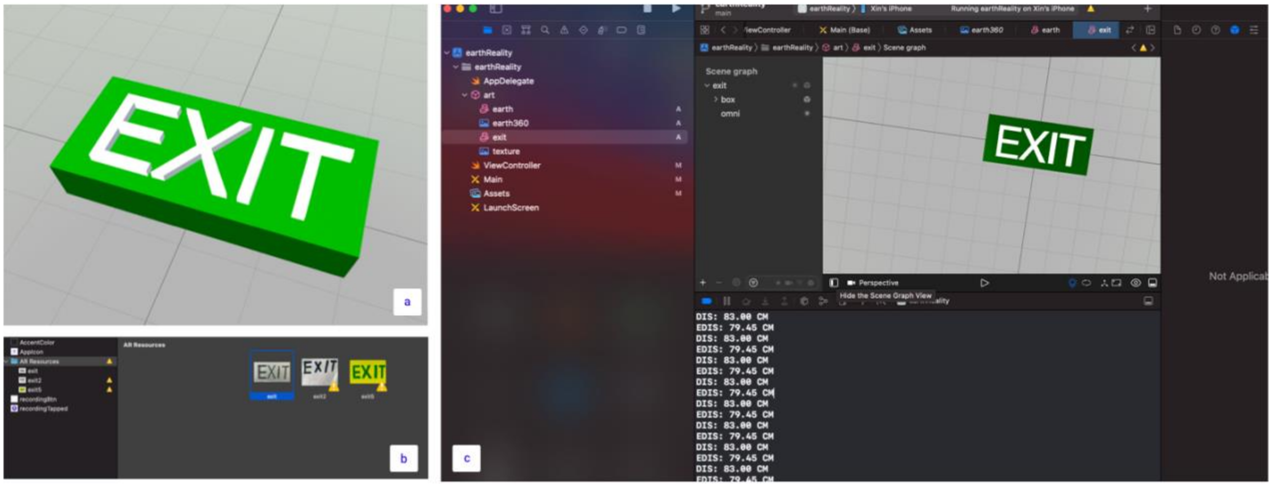


Figure 28. Development of Wonder Vision:(a) 3D model of Exit,(b)uploaded reference images,(c)distance monitoring.

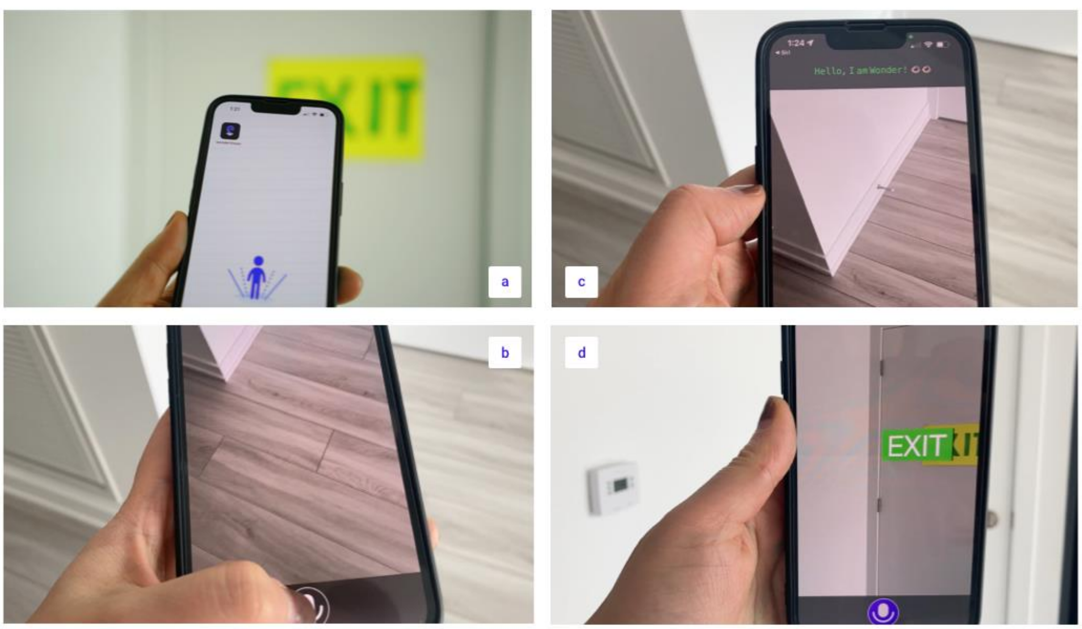


Figure 29. Play with the Wonder Vision[33]:(a) logo displays,(b)Wonder has been waked up,(c)clicked the button to start live audio input, (c) target has been founded, the exit visualization appeared.

---

[33] Online video of Wonder Vision: https://vimeo.com/675275559

Reflections

The prototype successfully incorporated the detection function and verified the feasibility of LIDAR ranging. Even though the LIDAR sensor can only measure within a meter at this point, the advantage is adaption in both indoors and outdoors, and no need of deployment. It is more convenient than Bluetooth beacon in terms of flexibility.

5.3 Summary

The iterative journey of the prototyping process paves the way for my final project Wonder Vision. Each of the first four prototypes validates one crucial technology, and design decisions around it. Thinking step by step through the iterations helps me to identify possible problems, and come up with a comprehensive solution: the Wonder Vision. And along the way of exploring, I enriched my knowledge about mobile application development, and frameworks supporting it, such as AR utilization, speech conversion, and machine learning classification.

6. Evaluation

This chapter introduces the evaluation method that I used in this work. Based on a descriptive method, I comparatively analyze Wonder Vision with selected projects to assess the design contributions. Also, the Wonder Vision has been introduced to target users, and their feedback confirms the potential usage from the user's perspective. Finally, I applied a time-line based perspective to evaluate the foresight of this work.

6.1 Descriptive Evaluation for Wonder Vision

Design is not only about innovation but also problem-solving with purpose, especially among digital products. (Rikke Friis Dam & Teo Yu Siang, 2020). Evaluation of the design comes in observation, analysis, experimenting, testing, and description. (Table 10) Specifically, these methods attempt to evaluate the system in terms of functionality, completeness, consistency, accuracy, performance, reliability, usability, fit with the organization, and other relevant quality attributes(Hevner et al., 2004).

1. Observational	Case Study: Study artifact in depth in business environment
	Field Study: Monitor use of artifact in multiple projects
2. Analytical	Static Analysis: Examine structure of artifact for static qualities (e.g., complexity)
	Architecture Analysis: Study fit of artifact into technical IS architecture
	Optimization: Demonstrate inherent optimal properties of artifact or provide optimality bounds on artifact behavior
	Dynamic Analysis: Study artifact in use for dynamic qualities (e.g., performance)
3. Experimental	Controlled Experiment: Study artifact in controlled environment for qualities (e.g., usability)
	Simulation – Execute artifact with artificial data
4. Testing	Functional (Black Box) Testing: Execute artifact interfaces to discover failures and identify defects
	Structural (White Box) Testing: Perform coverage testing of some metric (e.g., execution paths) in the artifact implementation
5. Descriptive	Informed Argument: Use information from the knowledge base (e.g., relevant research) to build a convincing argument for the artifact's utility
	Scenarios: Construct detailed scenarios around the artifact to demonstrate its utility

Table 9. Design Evaluation Methods (Hevner et al., 2004)

Making user testing requires in-person participation. However, due to the ongoing pandemic, it is impossible to conduct in-person testing in the lab during this project being integrated iteratively. I am using knowledge-based information to build arguments and assess the design work with a descriptive method. It addresses the information argument and analysis within constructed scenarios (Table 10). The primary purpose of UCD is to deal with user-related problems, and my study has a specific user context in way-finding.

Table 10. Comparative studies in evaluating way-finding related problems

Tpes of navigational tools	**Reading Text**	**Recognizing Faces**	**Identifying Objects**	**Finding Destination**	**Estimating Distance**	***Function Scores***
White Cane						0
Sighted People	✓	✓	✓	✓	✓	5
Guided Dog				✓		1
Google Maps				✓	✓	2
Be My Eyes	✓	✓	✓			3
Soundscape			✓			1
Seeing AI	✓		✓			2
RightHear				✓	✓	2
WonderVision	✓		✓		✓	3

Apparently, sighted people can provide comprehensive assistance in searching, reading, and measuring distance and it is also a preferred option for most of users. 'Be My Eyes' [34] can provide human assistance remotely. Through video calls and online networking, users are connected with sighted volunteers. This feature makes the application perform better among selected tools. In contrast, Wonder Vision depends less in human assistant, but comes up with the

[34] https://www.bemyeyes.com/

same function scores as ‘Be My Eyes’. It took advantage of the iPhone13 platform with advanced sensors, CoreML framework, and a high penetration rate. These features of the iPhone make the Wonder Vision application much more accessible physically and technically, supporting more use cases, both indoors and outdoors. Additionally, the AR visualization deals with the needs of low vision and auditory information that can prevent overwhelmed speech information.

Although user testing was not possible, a brief collection of feedback from selected interviewees has been gathered to clarify specifica user scenarios and to assess the functionality of Wonder Vision, so I recorded the operation process and introduced it to some participants who joined the prior audio interviews.

Interviewee D responded: *‘When I walk outside, I can figure out the approximate appearance of the object, but it is tough to read text. This often made me step in the wrong way. The designed function could help me a lot if it could work in practice. In a real scenario, I believe users like me may easily walk in the wrong way. A distinctive prompt or instruction will be essential for correcting the right direction.’*

Interviewee E responded: *‘When I was having myopic surgery, a sound like a beep was very effective for me. It is easy for me to understand distance, just like how we listen to the reversing radar. It gives a prompt effectively. In a real environment, a sound has a spatial effect. Similarly, the beep might sound weaker when the phone is farther away from an exit and louder when closer.’*

The feedback is encouraging related to Wonder Vision's target functionality, especially for low vision users and users who have temporary vision challenges. Meanwhile, their responses suggested some improvements to the audio design. Audio cue will provide more comprehensive guidance when associated with spatial variation. Enhancing the sound characteristics will give a more realistic feel to virtual content. When Wonder Vision is used in a natural environment, it is essential to consider several details. Develop different sound effects so that the system can prompt with precision, such as correcting the user's direction and returning to the correct path if the user veers off course. Overall, based on this short user feedback, there is some indication that the prototype directions may be beneficial to the user group, but this remains for future testing.

6.2 Towards Foresight for Wonder Vision

The speculative design presents a future lens to assess the design contribution to sustainability throughout the entire lifecycle. Future Cone is a framework for speculative design (Figure 30). It reveals the power of speculation-driven design. The essential metric of this framework is time changing. Within a timeline, the Future Cone is formed into six classes: preposterous, possible, plausible, the 'projected' Future, probable, and preferable.

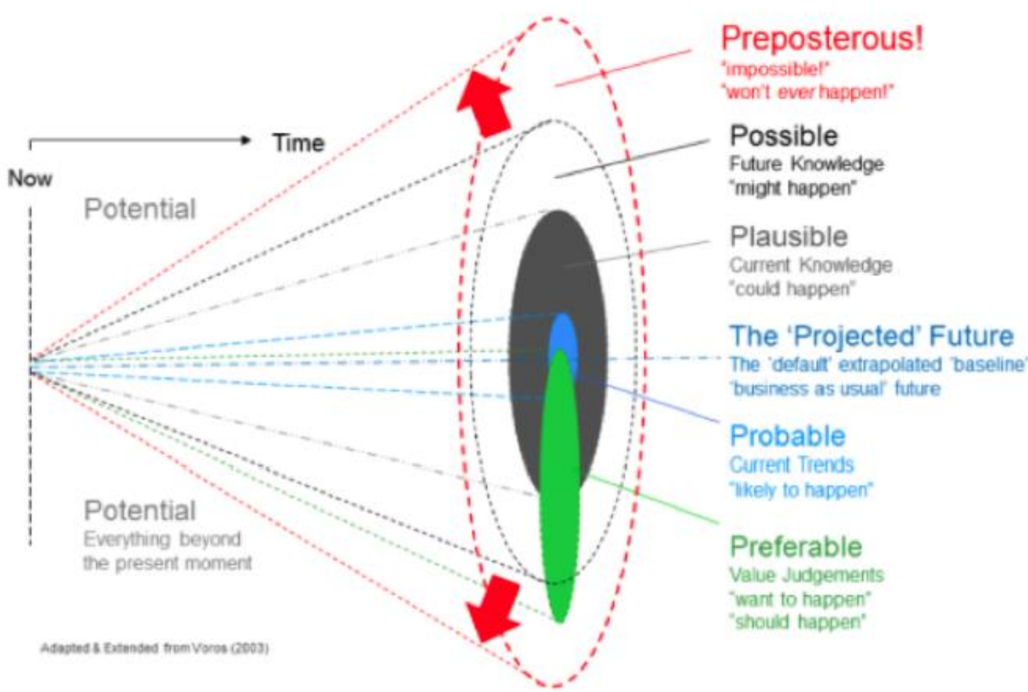


Figure 29. Adapted & Extended Futures Cone illustration(Voros,2017)

The ‘projected future’ is a core of the Cone, in other words, this part of the future could come true. Wonder Vision is an experimental project that extends the boundaries of different technologies. The iterative working process is also consistent with this foresight lens. With the ongoing timeline, the Wonder Vision will include more elements. (Figure 31)

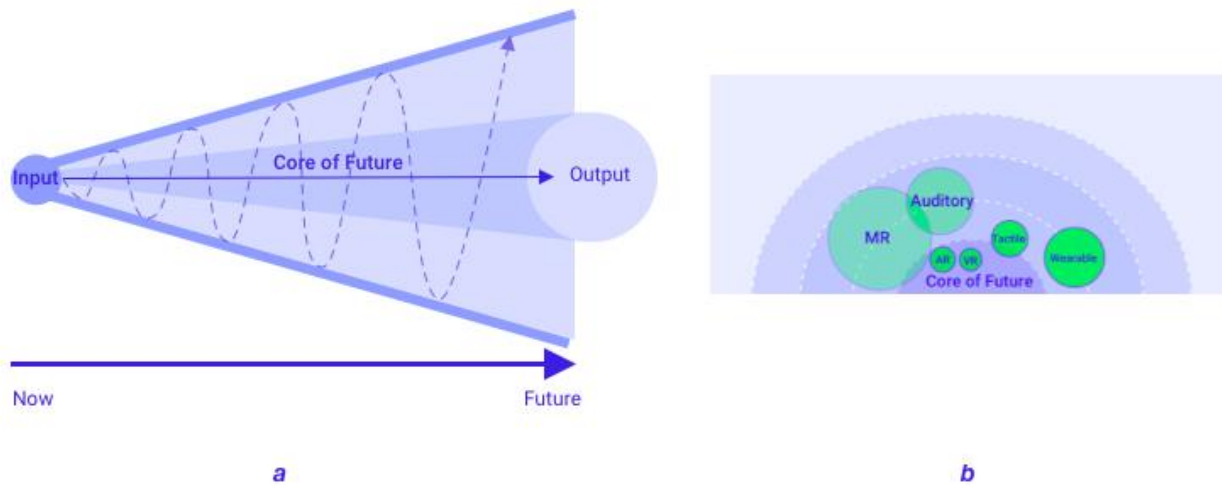


Figure 31(a). The Loop of iterated prototyping according to Future Cone (b)The core of future is expanding and including more novel technology.

Based on this framework, Wonder Vision would be further designed speculatively. Figure 32 maps out the images of Wonder Vision in different kinds of Futures. In the probable future, MR will be mature, along with the development of wearable technologies, some lightweight devices are coming to the industry. People can wear it in most places and use it as external eyes. As material science develops, Wonder Vision comes up in contact lenses, and it is a big forward step that embeds the external devices into the physical body. Finally, Wonder Vision could be a pair of cyborg eyes(Niketeghad & Pouratian, 2018). It was connected to the human nervous system. It became part of the physical body, and enabled users to have the same visual perception as sighted people.

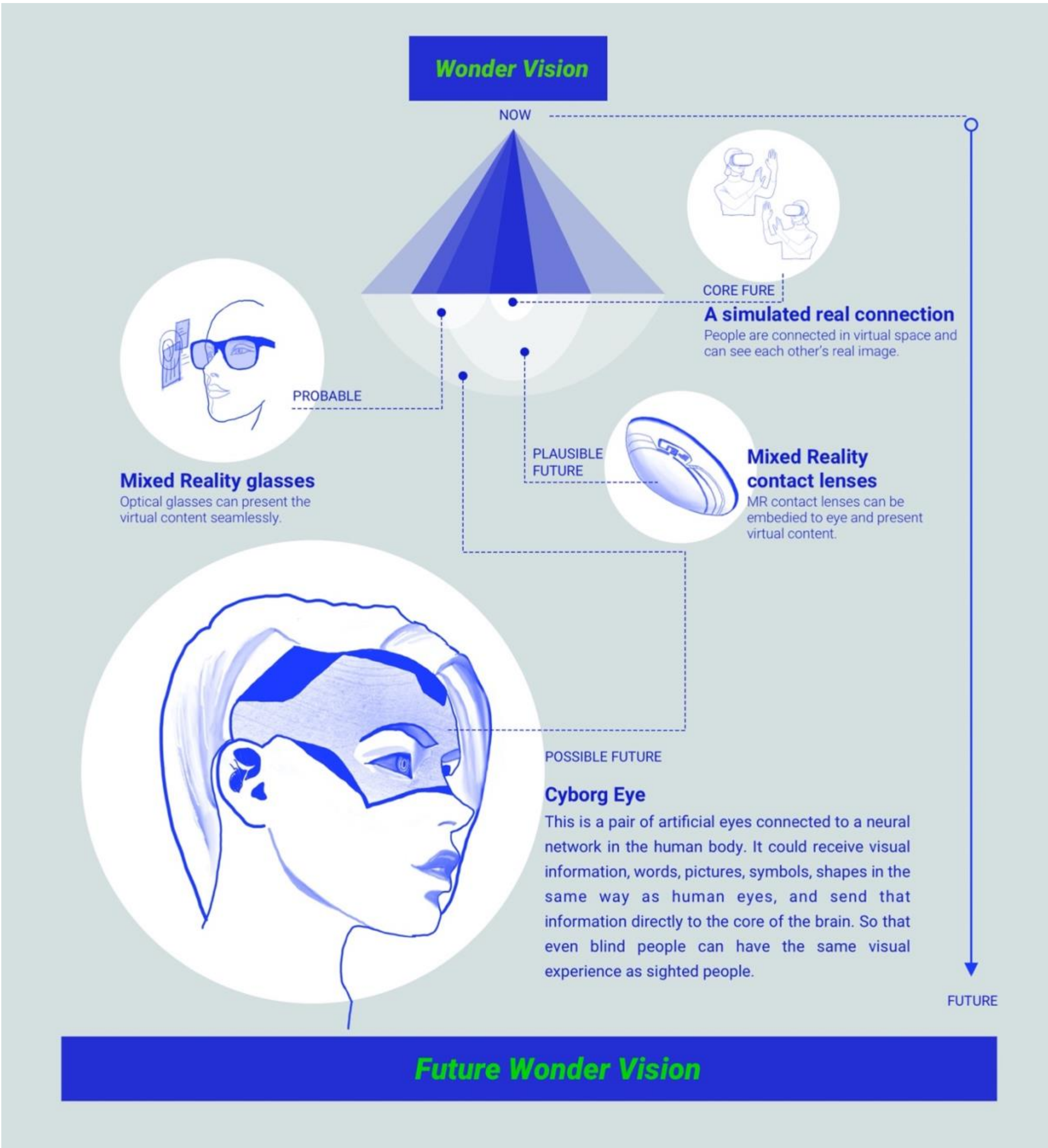


Figure 30. A future development of Wonder Vision basing on 'Futures Cone'.

6.3 Summary

The descriptive method suggests Wonder Vision's advances in multiple functions in way-finding scenarios. The feedbacks from target users affirm the functionality of this hybrid systems. Also, it also comes up with a series of further improvement in audio design. The creation of hybrid interfaces covers the user needs from of low vision and blind people. We are at the intersection of AR, MR, and reality, the MR glass will enhance human vision and allow humans to access more information from their surroundings. This device will be propelled by wearable technology, from a lightweight machine to a close part of the body. At that time, the accessibility issues will be minimized.

Chapter 7. Conclusions

7.1 Contribution

This research aims to clarify the challenges in way-finding that people with visual impairment face and identify potential solutions. Navigational behaviors indicate that our eyesight plays an important role and that much of the spatial information is absorbed through our eyes. This study aims to find design space and develop solutions that can provide users with comprehensive information about how to find their way around. This study investigates the impact of varying degrees of vision and the limitations of existing navigation tools. Iterative prototypes demonstrate how they adapt unique features from existing tools to address in-situ way-finding problems. AR and VUI navigation demonstrate cross-framework compatibility (ARKit and Speech Framework). Audio and speech instructions are also provided through the hybrid interface. Furthermore, Wonder Vision stresses the importance of innovative technologies in enhancing the power and reach of this platform through speculative and descriptive design.

7.2 Limitation and Future Work

Improving in learning curve

People have different levels of familiarity with mobile UI, especially with AR applications. People with visual impairment may have a limited understanding of AR, not knowing the operation flow. Additionally, since the voice interface is less obvious, a short audio tutorial is essential to help users bootstrap with ease.

Training Models

I plan to train an ML model with a more extensive database of signs we see in daily life in the future iteration. The model can recognize more categories of signages, such as store billboards, street names, and entrance signs. Moreover, an ML model functioning to detect obstacles will also be essential. As in a real street corner, obstacles like traffic, green belts, trees, and steps could be identified. Training model to detect these could let the system know ahead of time and prevent users from collision and falling risks.

Tweaking the Voice Interface

In the latest version, the voice interface can only communicate mechanically. For future tasks, I will address the design of interfaces in terms of human-like responses. As in a more complex environment, there might be situations where the system could not give a definitive answer. I plan to tweak the voice interface to provide helpful guidance even when the system fails to detect the target object.

Sound Characteristics

Auditory feedback may be perceived differently by different users. High volume or frequency may cause discomfort. Therefore, there is a need to do experiments in the range of variance in volume and frequency. To find the most effective sound delivery, I plan to conduct experiments on sound in real space, and to design effective sound feedback for different user operations.

User Testing & Observation

This version of Wonder Vision was only tested by me, which limits the research to find more convincing evidence. The further user experiment will be arranged at different times of the day and recruit participants to perform the searching task with designed application in a simulated experiment. This process aims to test the reliability and usability under different circumstances and optimize interactive experience through observation of user behaviors.

In conclusion, with the continuous improvement in artificial intelligence and mixed reality, our physical world has been integrated with various hybrid interfaces. Wonder Vision is an exploration of the domain of AR and VUI. With a fundamental in accessibility design, it provides potential solutions for dealing with the visual search tasks in way-finding.

BIBLIOGRAPHY

1. Blindness and vision impairment. Who.int; World Health Organization: WHO. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment
2. Anuprayaasadmin. (2020, July 10). Blind vs Visually Impaired: Do you know the Difference? | Anuprayaas. Anuprayaas - Making India Accessible. https://anuprayaas.org/blind-vs-visually-impaired/
3. Chen, Q., & Liu, Z. (2019). Idiopathic Macular Hole: A Comprehensive Review of Its Pathogenesis and of Advanced Studies on Metamorphopsia. Journal of Ophthalmology, 2019, 1–8.https://doi.org/10.1155/2019/7294952
4. What determines our navigational abilities? Trends Cogn Sci | Request PDF. (2022). ResearchGate. https://doi.org/10.1016\/j.tics.2010.01.001
5. Approach Taxonomy of Human Way-finding Tasks: A Knowledge-Based Approach. (2022). ResearchGate. https://doi.org/10.1080\/13875860902906496
6. Mirowski, P., Grimes, M. K., Malinowski, M., Hermann, K. M., Anderson, K., Teplyashin, D., Simonyan, K., Kavukcuoglu, K., Zisserman, A., & Hadsell, R. (2018). Learning to Navigate in Cities Without a Map. ArXiv.org. https://arxiv.org/abs/1804.00168
7. Parker, A. T., Swobodzinski, M., Wright, J. D., Hansen, K., Morton, B., & Schaller, E. (2021). Way-finding Tools for People With Visual Impairments in Real-World Settings: A Literature Review of Recent Studies. Frontiers in Education, 6. https://doi.org/10.3389/feduc.2021.723816
8. The Effectiveness of Visual and Audio Way-finding Guidance on Smartglasses for People with Low Vision. (2020). ResearchGate. https://doi.org/10.1145\/3313831.3376516
9. Human Factors of Way-finding in Navigation. (2022). ResearchGate. https://doi.org/10.1201\/9780849375477.ch394
10. Qiu, X., Wen, L., Wu, C., Yang, Z., Wang, Q., Li, H., & Wang, D. (2020). Impact of Learning Methods on Spatial Knowledge Acquisition. Frontiers in Psychology, https://doi.org/10.3389/fpsyg.2020.01322
11. Kalantari,. (2017). Cumincad : CUMINCAD Papers : Paper caadria2017_145:Human-Computer Interaction in the Form-Making Process. Cumincad.org. https://doi.org/http://papers.cumincad.org/data/works/att/caadria2017_145.pdf

12. Bosch, S. J., & Gharaveis, A. (2017). Flying solo: A review of the literature on way-finding for older adults experiencing visual or cognitive decline. Applied Ergonomics, 58, 327–333. https://doi.org/10.1016/j.apergo.2016.07.010
13. Vision | Introduction to Psychology. (2012). Lumenlearning.com. https://courses.lumenlearning.com/wmopen-psychology/chapter/outcome-vision/
14. Mirowski, P., Grimes, M. K., Malinowski, M., Hermann, K. M., Anderson, K., Teplyashin, D., Simonyan, K., Kavukcuoglu, K., Zisserman, A., & Hadsell, R. (2018). Learning to Navigate in Cities Without a Map. ArXiv.org. https://arxiv.org/abs/1804.00168
15. Rikke Friis Dam, & Teo Yu Siang. (2020, August 25). Design Thinking: Getting Started with Empathy. The Interaction Design Foundation; UX courses. https://www.interaction-design.org/literature/article/design-thinking-getting-started-with-empathy
16. What is UX Research? (2018). The Interaction Design Foundation; UX courses. https://www.interaction-design.org/literature/topics/ux-research
17. Formplus Blog. (2020, February 26). Structured Interviews: Definition, Types + [Question Examples]. Formplus; Formplus. https://www.formpl.us/blog/structured-interview
18. Wikipedia Contributors. (2022, January 18). Quantitative research. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Quantitative_research
19. Wikipedia Contributors. (2021, September 17). Categorical variable. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Categorical_variable
20. View of Cultivating the Under-Mined: Cross-Case Analysis as Knowledge Mobilization | Forum Qualitative Sozialforschung / Forum: Qualitative Social Research. Qualitative-Research.net. https://www.qualitative-research.net/index.php/fqs/article/view/]
21. The role of sound cues in the spatial knowledge of blind individuals. (2012). ResearchGate; ResearchGate. https://www.researchgate.net/publication/232221279_The_role_of_sound_cues_in_the_spatial_knowledge_of_blind_individuals
22. Some General Principles for Conceptual, Integration, and Enterprise Data Models. Developing High Quality Data Models, 63–78. https://doi.org/10.1016/b978-0-12-375106-5.00006-3

23. What Are User Personas and Why Are They Important? | Adobe XD Ideas. (2021). Ideas. https://xd.adobe.com/ideas/process/user-research/putting-personas-to-work-in-ux-design/#:~:text=What%20is%20a%20user%20persona,see%20in%20the%20example%20below).
24. Wikipedia Contributors. (2022, January 19). Microsoft HoloLens. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Microsoft_HoloLens
25. Wikipedia Contributors. (2022, February 10). iPhone 13 Pro. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/IPhone_13_Pro
26. Cipriani, J. (2021, February 19). iPhone 12 users: Impress your friends with this cool LIDARr trick we discovered. CNET; CNET. https://www.cnet.com/tech/services-and-software/iphone-12-users-impress-your-friends-with-this-cool-LIDARr-trick-we-discovered/
27. Apple Developer Documentation. (2022). Apple.com. https://developer.apple.com/documentation/arkit/content_anchors/detecting_images_in_an_ar_experience
28. Apple Developer Documentation. (2022). Apple.com. https://developer.apple.com/documentation/arkit/aranchor
29. Wikipedia Contributors. (2022, January 14). Speech recognition. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Speech_recognition
30. Apple Developer Documentation. (2022). Apple.com. https://developer.apple.com/documentation/speech
31. Jain, D., Junuzovic, S., Ofek, E., Sinclair, M., R. Porter, J., Yoon, C., Machanavajhala, S., & Ringel Morris, M. Towards Sound Accessibility in Virtual Reality. Proceedings of the 2021 International Conference on Multimodal Interaction. https://doi.org/10.1145/3462244.3479946
32. What are Voice User Interfaces? (2022). The Interaction Design Foundation; UX courses. https://www.interaction-design.org/literature/topics/voice-user-interfaces
33. Accessible Mixed Reality - Microsoft Research. (2021, July 7). Microsoft Research. https://www.microsoft.com/en-us/research/project/accessible-mixed-realityanast/
34. MILGRAM, P., & Fumio KISHINO. (1994). A Taxonomy of Mixed Reality Visual Displays. IEICE TRANSACTIONS on Information and Systems, E77-D(12), 1321–1329. https://doi.org/https://search.ieice.org/bin/pdf_link.php?fname=e77d_12_1321&category=D&lang=E&year=1994&abst=

35. What is mixed reality? - Mixed Reality. Microsoft.com. https://docs.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality
36. Mott, M., Cutrell, E., Franco, C., Holz, E., Ofek, R., Stoakley, M., & Morris. (n.d.). Accessible by Design: An Opportunity for Virtual Reality. Retrieved February 24, 2022, from
37. Forlizzi, J., Zimmerman, J., & Evenson, S. (2008). Crafting a Place for Interaction Design Research in UCD. Design Issues, 24(3), 19–29. https://doi.org/10.1162/desi.2008.24.3.19
38. Diagraam Design. (2019, November 5). How to conduct research, through design - Diagraam Design - Medium. Medium; Diagraam Design. https://medium.com/diagraam/how-to-conduct-research-through-design-e3a5d19fb79fWikipedia Contributors. (2022, January 27). Human-centered design. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Human-centered_design
39. Gu, Y., & Warren, J. (2017, February 27). Chapter 11 Methods for Descriptive Studies. Nih.gov; University of Victoria.https://www.ncbi.nlm.nih.gov/books/NBK481606/#:~:text=Descriptive%20evaluation%20studies%20describe%20the,purposes%20%E2%80%94%20the%20specific%20healthcare%20organization.Design Science in Information Systems Research. (2022). ResearchGate; ResearchGate. https://www.researchgate.net/publication/201168946_Design_Science_in_Information_Systems_Research
40. Blue Oak Interactive. (2014, March 23). What is Way-finding? | SEGD. Segd.org. https://segd.org/what-way-finding
41. Blind, L. (2019). See through My Eyes | A Low Vision Simulation [YouTube Video]. In YouTube. https://www.youtube.com/watch?v=xG7d-kIlnT8
42. Council, W. (2020). A Day in the Life of a Person Who is Blind or Visually Impaired [YouTube Video]. In YouTube. https://www.youtube.com/watch?v=L-AUf0Ky1GM&t=378s
43. Digitalcraftsman. (2017). Inclusive and Accessible Design Guidelines. Humanrights.ca. https://id.humanrights.ca/visitor-supports/way-finding/
44. Human Factors of Way-finding in Navigation. (2022). ResearchGate. https://doi.org/10.1201\/9780849375477.ch394
45. Wikipedia Contributors. (2022, February 10). Cognitive map. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Cognitive_map

46. Hegarty, M. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, *30*(5), 425–447. https://doi.org/10.1016/s0160-2896(02)00116-2
47. What Is Low Vision? (2021, April 23). VisionAware. https://visionaware.org/your-eye-condition/eye-health/low-vision/
48. Low Vision and Vision Rehabilitation. (2022). Aoa.org. https://www.aoa.org/healthy-eyes/caring-for-your-eyes/low-vision-and-vision-rehab?sso=y
49. Low Vision and Blindness. (2022). Nfb.org. https://nfb.org//sites/default/files/images/nfb/publications/bm/bm06/bm0610/bm061005.htm
50. Designing AR Visualizations to Facilitate Stair Navigation for People with Low Vision | Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology. (2019). ACM Conferences. https://dl.acm.org/doi/10.1145/3332165.3347906
51. Wikipedia Contributors. (2022, February 28). Google Maps. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Google_Maps
52. Wikipedia Contributors. (2022, February 20). Microsoft HoloLens. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Microsoft_HoloLens
53. Understanding In-Situ Use of Commonly Available Navigation Technologies by People with Visual Impairments | The 22nd International ACM SIGACCESS Conference on Computers and Accessibility. (2020). ACM Conferences. https://dl.acm.org/doi/abs/10.1145/3373625.3416995
54. Guided Meditation VR on Oculus Rift | Oculus. (2022). Oculus.com. https://www.oculus.com/experiences/rift/998016373574782/?intern_source=blog&intern_content=relaxing-vr-games-to-help-you-unwind
55. Oculus | VR Headsets, Games & Equipment. (2022). Oculus.com. https://www.oculus.com/
56. What is User Interface Design? (2018). The Interaction Design Foundation; UX courses. https://www.interaction-design.org/literature/topics/ui-design
57. Accessibility needs of extended reality hardware: A mixed academic-industry reflection | ACM Interactions. (2021). Acm.org. https://interactions.acm.org/archive/view/may-june-2021/accessibility-needs-of-extended-reality-hardware

58. Siu, A. F., Sinclair, M., Kovacs, R., Ofek, E., Holz, C., & Cutrell, E. (2020). Virtual Reality Without Vision: A Haptic and Auditory White Cane to Navigate Complex Virtual Worlds. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3313831.3376353
59. Designing AR Visualizations to Facilitate Stair Navigation for People with Low Vision. (2019). ResearchGate. https://doi.org/10.1145\/3332165.3347906
60. McTear, M., Callejas, Z., & Griol, D. (2016). The Conversational Interface. Springer International Publishing. https://doi.org/10.1007/978-3-319-32967-3
61. Conversational User Interfaces: In-Depth Guide [2022 Update]. (2018, March). AIMultiple. https://research.aimultiple.com/conversational-ui/
62. Real-Time 3D Object Detection on Mobile Devices with MediaPipe. (2020, March Google AI Blog. https://ai.googleblog.com/2020/03/real-time-3d-object-detection-on-mobile.html
63. Objectron (3D Object Detection). (2021). Mediapipe. https://google.github.io/mediapipe/solutions/objectron.html
64. Apple Developer Documentation. (2022). Apple.com. https://developer.apple.com/documentation/arkit
65. Apple Developer Documentation. (2022). Apple.com. https://developer.apple.com/documentation/arkit/arworldtrackingconfiguration
66. Apple Developer Documentation. (2022). Apple.com. https://developer.apple.com/documentation/arkit/arsession
67. Wikipedia Contributors. (2022, March 11). Unity (game engine). Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Unity_(game_engine)
68. Apple Developer Documentation. (2022). Apple.com. https://developer.apple.com/documentation/coreml
69. Apple Developer Documentation. (2022). Apple.com. https://developer.apple.com/documentation/speech
70. Rames, J. (2017, June 13). Speech Recognition Tutorial for iOS. Raywenderlich.com; raywenderlich.com. https://www.raywenderlich.com/573-speech-recognition-tutorial-for-ios
71. Design Science in Information Systems Research. (2022). ResearchGate; ResearchGate. https://www.researchgate.net/publication/201168946_Design_Science_in_Information_Systems_Research
72. A generic foresight process framework. (2017). ResearchGate. https://doi.org/10.1108\/14636680310698379

73. What is Theory of Change? - Theory of Change Community. (2021, January 30). Theory of Change Community. https://www.theoryofchange.org/what-is-theory-of-change/
74. The Futures Cone, use and history. (2017, February 24). The Voroscope; The Voroscope. https://thevoroscope.com/2017/02/24/the-futures-cone-use-and-history/
75. Myopia (Nearsightedness): What It Is, Causes & Treatment. (2020). Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/8579myopinearsightedness#:~:text=High%20myopia%3A%20A%20rare%20inherited,refractive%20error%20greater%20than%20%2D6.
76. Downtown Pedestrian Mall: Transportation and Parking | City Of St. John's. (2022). Stjohns.ca. https://www.stjohns.ca/media-release/downtown-pedestrian-mall-transportation-and-parking
77. GPS.gov: GPS Overview. (2022). Gps.gov. https://www.gps.gov/systems/gps/
78. What is GIS?,Geographic Information System Mapping Technology. (2022). Esri.com. https://www.esri.com/en-us/what-is-gis/overview
79. Wikipedia Contributors. (2022, March 11). Automotive navigation system. Wikipedia;Foundation.https://en.wikipedia.org/wiki/Automotive_navigation_system
80. Wikipedia Contributors. (2022, March 9). Intelligent transportation system. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Intelligent_transportation_system
81. Loh KY;Ogle J. (2020). Age related visual impairment in the elderly. The Medical Journal of Malaysia, 59(4). https://pubmed.ncbi.nlm.nih.gov/15779599/#:~:text=The%20most%20common%20causes%20of,depression%2C%20social%20isolation%20and%20dependency.
82. How Situational Disabilities Impact Us All. (2020). Userway.org. https://userway.org/blog/how-situational-disabilities-impact-us-all
83. Turn on and practice VoiceOver on iPhone. (2022). Apple Support. https://support.apple.com/enca/guide/iphone/iph3e2e415f/ios#:~:text=With%20VoiceOver%E2%80%94a%20gesture%2Dbased,app%20your%20finger%20is%20on.
84. Be My Eyes - See the world together. (2022). Bemyeyes.com. https://www.bemyeyes.com/
85. Microsoft Soundscape - Microsoft Research. (2022, February 25). Microsoft Research.https://www.microsoft.com/en-us/research/product/soundscape/

86. Wikipedia Contributors. (2022, March 20). Seeing AI. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Seeing_AI

87. Right-Hear. (2013). RightHear - Blind and Visually Impaired Assistant. Google.com. https://play.google.com/store/apps/details?id=com.righthear&hl=en_CA&gl=US

88. Clew: Revolutionary Indoor Navigation for iOS. (2022). Clewapp.org. http://www.clewapp.org/#:~:text=Clew%20is%20an%20indoor%2C%20path,after%20exploring%20your%20surrounding%20environment.

89. Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. Computers in Human Behavior, 75, 461–468. https://doi.org/10.1016/j.chb.2017.05.045

90. Xu, Y., Wang, D., Collins, P., Lee, H., & Warschauer, M. (2021). Same benefits, different communication patterns: Comparing Children's reading with a conversational agent vs. a human partner. Computers & Education, 161, 104059. https://doi.org/10.1016/j.compedu.2020.104059

91. The Futures Cone, use and history. (2017, February 24). The Voroscope; The Voroscope. https://thevoroscope.com/2017/02/24/the-futures-cone-use-and-history/

APPENDICES

Appendices A: Semi-Structured Interview Questions

1. Your Gender: Male/Female/ Others/ Not to answer

2. Your Age: Under 20/ 20-30/30-40/More than 40

3. Which of the following mobile system do you use: iOS/ Android/ Other

4. Degree of eyesight: Normal/ Low vision/ Legally Blind/ Totally Blind

5. Which of the following area do you live in? Urban Center/ City Suburbs/ Other

6. Which of the following transportation did you usually take? Bus/ Metro/ Shared Riding/ Others, please specify

7. Which of the following landmarks do you use for the mental route? Bus stop/ Metro Entrance/ Stores/ Others, please specify

8. How many tools do you use for way-finding? Only1/ 2-3/ More than 3

9. Which of the following tools do you use mostly for navigation? You may select more than one. The White Cane/ Dog/ Mobile Application/ Others, please specify

10. On a scale of 1-10, Where 1 is "not at all" and 10 is "extremely", how often do you use the tool you chose in Question 9?

1	2	3	4	5	6	7	8	9	10

11. On a scale of 1-10, where 1 is "not at all" and 10 is "extremely", do you think the tool you chose in Question 9 can provide complete assistance in way-finding?

1	2	3	4	5	6	7	8	9	10

12. Which of the following information is important to you in way-finding? You may select more than one. Smell/ Sound/ Tactile information/ Other, please specify

13. On a scale of 1-10, where 1 is “not at all” and 10 is “extremely”, how useful is the information you chose in Question 12?

1	2	3	4	5	6	7	8	9	10

14. On a scale of 1-10, where 1 is “not at all” and 10 is “extremely”, how confident do you feel when you go out independently?

1	2	3	4	5	6	7	8	9	10

Appendices B: Collection From Online Questionnaires

Part I Need-finding Data Collection in online questionnaire

Number of total participants: 53, percentage of participantsin for each question: 100%

Data Collection Source:https://wj.qq.com/stat/overview.html?sid=9547466

Your Gender:	Raw Number	Percentage
Male	32	57.10%
Female	24	42.90%
Not to answer	0	0
Your Age:		
Under 20	3	5.70%
20-30	11	20.80%
30-40	13	24.50%
More than 40	26	49.10%
Not to answer	0	0
Which of the following mobile system do you use:		
iOS	26	49.10%
Android	26	49.10%
Other	1	1.90%
Degree of eyesight:		
Normal	1	1.90%
Low vision	17	32.70%
Legally Blind	24	45.20%
Totally Blind	11	26.20%
Which of the following area do you live in?		
Urban Center	29	73.80%
City Suburbs	13	21.40%
Suburbs	9	17.00%
Other	2	3.80%
Which of the following transportation did you usually take?		
Bus	41	77.40%
Metro	24	45.30%
Taxi	28	53%
Shared Riding	15	28.30%
Others, please specify	9	17.00%
Which of the following item did you usually usa as landmarks? You may select more than one		
Bus Stop	39	73.60%
Metro Exit	31	58.50%
Store	28	52.80%
Street Corner	27	50.90%
Others	11	20.80%

How many tools did you use for wayfinding?		
Only 1	27	51%
2~3	22	41.50%
More than 3	4	7.50%
Which of the following tool do you use for wayfinding? You may select more than 1		
White Cane	20	40%
Guided dog	2	2.40%
Sighted people nearby	34	66.70%
Face-time with sighted people	14	33%
Mobile application	38	69%
Other	2	3.80%
On a scale of 1-10, Where 1 is "not at all" and 10 is "extremely", how often do you use the tool you chose in Question 9?		
1	3	5.70%
2	6	11.30%
3	7	13.20%
4	2	3.80%
5	9	17.00%
6	2	3.80%
7	4	7.50%
8	9	17.00%
9	4	7.50%
10	7	13.20%
On a scale of 1-10, where 1 is "not at all" and 10 is "extremely", do you think the tool you chose in Question 9 can provide complete assistance in way-finding?		
1	4	7.50%
2	1	1.90%
3	1	11.30%
4	6	15.10%
5	8	15.10%
6	8	15.10%
7	7	13.20%
8	6	11.30%
9	3	5.70%
10	9	17.00%
Which of the following information is important to you in way-finding? You may select more than one.		
Smell	20	37.70%
Sound	49	92.50%
Tactile information	17	32.10%
Other, please specify	7	13.20%

On a scale of 1-10, where 1 is “not at all” and 10 is “extremely”, how useful is the information you chose in Question 12?		
1	4	7.50%
2	1	1.90%
3	1	1.90%
4	6	11.30%
5	8	15%
6	8	15.10%
7	7	9.50%
8	6	11.30%
9	3	5.70%
10	9	17.00%
On a scale of 1-10, where 1 is “not at all” and 10 is “extremely”, how confident do you feel when you go out independently?		
1	4	7.50%
2	1	1.90%
3	2	3.80%
4	4	7.50%
5	8	15.10%
6	7	13.20%
7	11	20%
8	3	5.70%
9	6	11.30%
10	7	13.20%
Would you like to be contacted about participation in a follow-up audio interview to talk about problems you found in wayfinding?		
Yes	26	49%
No	21	50%

Appendices C: Transcript of Audio Interview

What kinds of assistant tools do you often use when going out? What kinds of assistance do these tools provide? Do you often explore new places independently? What preparation will you do before going to a new place? What kinds of information about your destination do you need to know before the trip? When you explore new places, what kinds of landmarks will you use to remember the route? What kinds of navigational applications do you often use? What kinds of assistance does the application provide? What kinds of information does the application not provide? What kinds of challenges have you encountered in navigation? Did you experience situational disability, such as getting lost in overwhelmed information in public space, or can't find the doorway after arriving at the destination When in situational disabilities, how did you find the way? Have you ever used the Augmented Reality application which uses computer technology to superimpose virtual content in the real world, display it through mobile phones, tablets, and other devices? (If yes, what kinds of AR technologies did you use before?) Do you need a screen reader when using the mobile application? If yes, do you think the screen reader can help you obtain enough information from the interface? Do you use intelligent voice assistants, like Siri, Google Home? When do you use them?

Appendices D: Summary of Audio Interview

Interviewee A

Description of daily navigation

A is a blind person and relies on sighted people and a white cane to go out. Meanwhile, he uses Google Maps primarily for planning, such as checking traffic transportation and memory roadmap in advance. For safety considerations, A seldom uses Google Maps when walking outside. The synchronized instruction popping up prevents him from focusing on road detection. Seeking help from sighted people is always A's priority. From his perspective, people can provide comprehensive support, describing surroundings, and telling additional prompts, which help build a mind map. In addition, sighted people are more trustworthy than digital tools, which have technical or networking issues. So when heading to new places, he usually had someone accompanied. Or he calls the people there to request accessible assistance when he arrives.

In-situ challenges in navigation

A is skillfully using the white cane. It helps him to detect obstacles in front. Sometimes, he was still hit by branches or low ceilings. Because the white cane could not detect the things above his head. Another safety issue was in walking on stairs as the white cane prevented him sensing the steps directely. He had to move very carefully and slowly. When in some shopping malls, he felt challenging to find the specific store as he could not read signs and text. So, he sometimes used smell as a cue to track the location.

Interviewee B

Description of daily navigation

B is a blind person, he primarily uses a white cane to detect obstacles. Usually, he consults with others about traffic information before going to a new place. He took the subway a lot, and people around him sometimes helped him read the text information of choosing the right station. Sometimes he uses WeVoice service, which is a platform that connecting visual impairmed people with sighted people by face-time. The volunteers can help them find things or read information by see through video camera.

In-situ challenges in navigation Though taking the metro is a daily routine for B, he is familiar with the route and structure of the station he takes to work. It was difficult to find an exit or entrance in some new statations. Sometimes he may board the wrong train or step off the wrong platform, as he couldn't read the station information.
Interviewee C **Description of daily navigation** Since high school, C has been suffering from hereditary amblyopia (also called 'Lazy Eye,' a degenerative eye condition that runs in families). This problem gets worse with aging. He is in low vision, and he can see the object in his central picture but has difficulties seeing obstacles from two edges. For safety reasons, he will not go out independently. Instead, he hired a driver and always had people accompany him, like his wife or children. **In-situ challenges in navigation** He cannot read or see detailed content with limited vision, such as text or miner obstacles from side vision. Without an accompanying person, he may felt tripped over by obstacles. He uses Google Maps often but is less confident and afraid of accidents without an accompanying person.
Interviewee D **Description of daily navigation** D is a young adult and joins community social activities quite often. The most public transportation he takes is the metro. Sometimes, he takes taxi when going to new places. He mentioned that the mobile GPS application helped him with traffic information. White cane and sighted people are two primary approaches that allow him to arrive particular place. **In-situ challenges in navigation**

The city where D lives is very busy with traffic, and the subway station is often very large. Sometimes, it takes him a long time to find the elevator or the exit, because he can't see the text sign.
Interviewee E **Description of daily navigation** D had 500-degree myopia plus 500 astigmatisms since high school. She did myopia surgery last year. After surgery, her eyesight was covered with gauze and she could identify people and objects only by sound. During that times, her mother usually accompanied her to go out. Otherwise, she took taxi quite often. **In-situ challenges in navigation** When D could not see things, she learned to use voiceOver to operate the mobile GPS, but it even brought more challenges to finding the places. She could not understand the interface with the screen reader, and it became difficult to identify different voices outside. She also felt that the GPS did not work well indoors, especially in some metra stations or on a train. Weak GPS signals caused a system delay and misleading the direction. These difficulties made her struggle in navigation.

Appendices E: Code of Wonder Vision

https://github.com/xinzhang-jessy/wonderVision-exitReality.git