Augmented Reality for Visually Impaired People (AR for VIPs)

University of California, Berkeley | School of Information

MIMS Capstone Project Report

May 2019

Team Members: Dylan Fox, Alyssa Li, Anu Pandey, Rohan Kar

Advisor: Professor Kimiko Ryokai

Contents

Acknowledgments

We would like to thank all the people who helped us in completing this project. We thank our advisor, Professor Kimiko Ryokai. We would also like to thank Professor John Chuang and Professor Coye Cheshire for their insightful comments, help and guidance throughout the semester on the project. Thanks to Professor Emily Cooper for her guidance and support, as well as for her fundamental research exploring the HoloLens as an assistive device and generous lending of HoloLens hardware for testing purposes.

Major thanks as well to the undergraduate members of our prototype development team: Rajandeep Singh, Manish Kondapolu, Elliot Choi, and Teresa Pho. We couldn't have done it without you! Thanks too to Jay Amado for assistance with the economic analysis and routes to potential enterprise implementation.

We would also like to thank Web Accessibility Evangelist Lucy Greco, Professor Georgina Kleege, Lighthouse for the Blind and Visually Impaired, East Bay Center for the Blind and the National Federation of the Blind. Each of them helped us understand the pain points of the blind and visually impaired community, directed us towards a solution, and finally helped us validate our approach by being a part of the usability testing exercise.

Thanks to Robin Held and Joe Kelly, as well the others at Microsoft who provided us invaluable guidance on HoloLens design and development. Thanks to the Big Ideas at Berkeley competition for encouraging us and giving us excellent feedback on the project.

We would like to thank the members of Virtual Reality at Berkeley and the MIMS community who generously lent us equipment and space, as well as helped us on prototype development and worked with us through multiple iterations of usability testing.

Finally, thanks to our friends, family, and partners who have patiently supported us through these intense few months, and encouraged us to keep going when times got tough.

Abstract

For visually impaired people, tasks such as navigating unfamiliar environments or reading text can become difficult or impossible without external aid. In this paper, we focus on solving the “five-meter problem”. That is, once a blind person has navigated to within five meters of their destination using tools such as GPS, they often have trouble locating their goal, especially if that goal can only be positively identified by reading (non-braille) text.

Our project, Augmented Reality for Visually-Impaired People (AR for VIPs), aims to improve their ability to navigate unfamiliar environments and give them the independence of performing these tasks unaided. We do so with two capabilities of modern augmented reality headsets: spatial mapping and machine vision. First, our application scans the environment around the user, creates a map of nearby obstacles, and selectively sonifies those obstacles so that even totally blind users are aware of them. Second, it uses machine vision to recognize text in front of the user, providing them vital information that would normally be inaccessible.

For our research, we conducted an in-depth literature review of different solutions, including navigational devices, using spatialized sound in assistive devices, human assisted devices, and augmented reality assistants. This shaped our understanding of the space and informed our subsequent generative user research, where we interviewed academic researchers, visually-impaired people and organizations, as well as others in the assistive technology space. With that, we were able to narrow down our scope from sonifying everything in the environment to focus on providing contextual information about the environment where necessary.

After a design and development phase, we conducted 7 in-person usability tests with blind users, who were then asked to rate the tasks and provided qualitative feedback on the experience. Overall, the results were positive, and users found the learning process easy and intuitive. We then discuss further applications and research, and finally conclude the paper with current limitations and future work.

Although technical constraints currently limit the extent to which we can develop this project for real-world use, we believe our research has helped lay the groundwork for the future of using augmented reality technology to aid visually impaired people.

Introduction

Problem

About 36 million people worldwide are blind (Bourne et al., 2017). The World Health Organization estimated that approximately 1.3 billion people lived with some form of distance or near vision impairment in 2018.

Independence and mobility can pose a major challenge for blind and visually impaired individuals (Varma et al., 2016). Navigation is a complex task, for which people typically rely heavily on vision (Maguire et al. ,1999; Wolbers & Megarty, 2010). Both wayfinding and obstacle avoidance can pose challenges for individuals with vision loss (Legge, 2013). Guide sticks and dogs are commonly used by blind and visually impaired people to assist them in navigation and to lead them around various obstacles in their surroundings. However, **both cane and dog are incapable of communicating semantic information in the environment** (Liu et al, 2016). Thus, navigating independently in unfamiliar environments still remains a major challenge for blind and visually impaired people.

Focus on the 5 Meter Problem

In this paper, we focus on what's known as the "5 meter problem." Large-scale tools like GPS are often enough to guide blind people to the approximate location of their target, and within 1 meter they can use their cane or other techniques to zero in on it; however, navigating the ~5 meters in between those two can often be a serious problem. A classic example of the 5 meter problem is finding the correct bus stop. GPS can guide someone to approximately the right location on the street, at which point a sighted person would have no problem seeing the stop and checking which busses stop there. However, a blind person faces serious difficulty in locating the signpost and ensuring it corresponds to their desired bus. For them, 5 meters can be the difference between catching the bus and hearing it drive right past them because they were waiting in the wrong place.

Thus, a significant problem for visually impaired people is that there's no obvious way to easily locate and acquire semantic information from silent objects in their surroundings. A blind person must instead rely on sighted people or on touch or context clues obtained by their cane or dog to learn about their surroundings, neither of which are reliable methods in such situations.

Hypothesis

Our hypothesis is that by using Augmented Reality (AR) technology to give a "sound" to silent objects in the environment, we can help blind and visually impaired users compensate for their lack of visual cognition. Combining this ability to locate obstacles with text recognition should enable blind users to navigate more independently and confidently in the 1-5 meter range by increasing their awareness of their environment and empowering them to locate and access semantic information.

Why Augmented Reality

For people with visual impairment, wearable displays such as Augmented Reality (AR) devices hold the potential to enhance visual function. As these technologies advance, it is essential to explore their promise as visual aids.

Unlike camera-based systems that rely on two-dimensional (2D) information, these devices build up a 3D map of the user's environment, as represented by a mesh of surfaces. The hands-free nature of these headsets, coupled with their ability to capture texts and generate the 3D mesh of the user's surroundings, gives these devices a huge potential to be used as a tool for navigation as well as text analysis for visually-impaired and blind people. Additionally, unlike many single-purpose accessibility technologies sold today, the multi-purpose computational power of modern AR headsets could provide extra value to users and purchasers.

Research

Different approaches

Offering Positional Information

To address challenges in navigation and supplement the functional vision of this population, a variety of assistive tools have been developed, including visual aids and sensory substitution devices. Many researchers and companies have proposed and developed tools that utilize devices such as **smartphones to provide the user with positional information** and assist in wayfinding guidance (Zhang et al., 2013; Asad, 2012). Soundscape, developed by Microsoft, is also a mobile application that runs in the background in combination with other navigation apps such as google maps to provide users with additional context about their environment and strives to build a richer awareness of surroundings for the blind. While existing aids such as the GPS technology can assist with obstacle avoidance and outdoor wayfinding, **the delivery of semantic information by such devices is limited** (Roentgen et al, 2011; BlindSquare, 2019; Humanware, 2019).

Heightening Visual Information

The impressive advances in mobile computing and machine vision technologies are facilitating new devices and apps that offer one or another assistive function for the vision impaired. A range of companies are building ***head-mounted display systems for generals consumers of virtual and augmented reality applications*** such as Microsoft (Hololens) and Magic Leap (Lightwear). These devices (Microsoft Hololens, 2019; Magic Leap One, 2019) generate a 3d mesh of the surrounding and thus have the potential to augment vision by conveying this information to its wearer. Also, what makes these commercial devices attractive is that they are mass produced and have the advantage of being constantly improved (a new version is released after every few weeks). These devices are resilient to support a range of software applications and allow the affordability of developing new features on the top of the existing device to other developers. Many efforts have been made to help visually impaired people using such devices over the last few years.

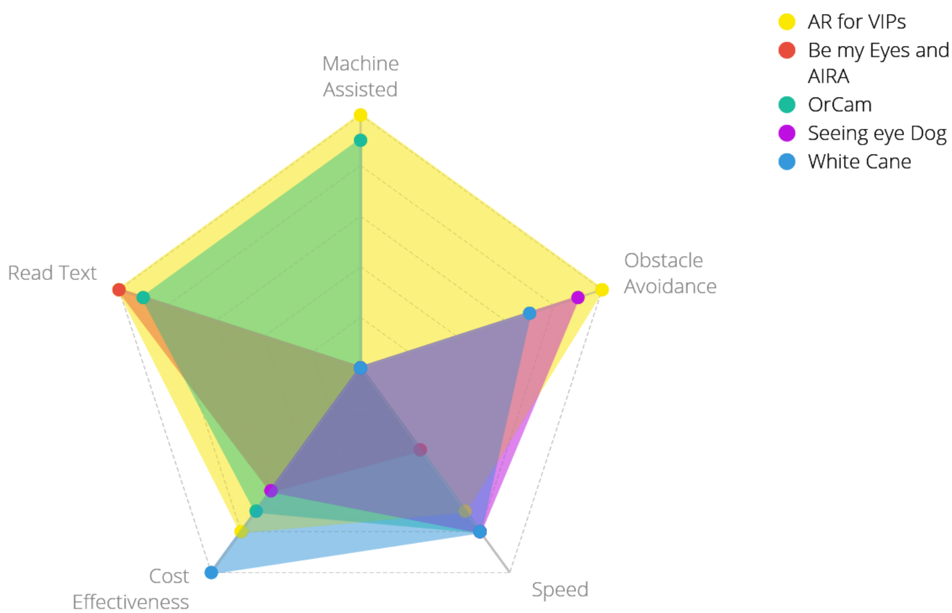
One potential approach to assist visually impaired individuals with finding their way to unfamiliar locations is to heighten the visual information about the location and content of existing signage in the environment. Huang et al. developed a prototype software application, which runs on a consumer head-mounted augmented reality (AR) device, to assist visually impaired users with sign-reading. The sign-reading assistant identifies real-world text (e.g., signs and room numbers) on command, highlights the text location, converts it to high contrast AR lettering, and optionally reads the content aloud via text-to-speech. While this approach is **remarkably helpful for individuals with reduced vision, it does not specifically assist people who are blind.**

Use of Spatialized Sound in Assistive Devices

American psychologist Winthrop Niles Kellogg's research showed that both blind and sighted subjects wearing blindfolds could learn to detect objects in the environment through sound, and that, with some training, both blind and sighted individuals can precisely determine certain properties of objects, such as distance, size, shape, substance and relative motion from sound alone. Spatialized sound in assistive devices have been considered for some time (Spagnol et al., 2018). For the purpose of outdoor navigation, Loomis et al. (2005) tested a system that guides the user along a path by virtual sounds emitted from waypoints. Subjects seem to prefer speech sounds over simple tones in some scenarios and performed better under those conditions. Another device called vOICe (vision technology for the totally blind) aims to provide the experience of live camera views through image-to-sound renderings. Images are converted into sound by scanning them from left to right while associating elevation with pitch and brightness with loudness. In theory, this could lead to synthetic vision with truly visual sensations, by exploiting the neural plasticity of the human brain through training. A company called OxSight claims to built and is now testing augmented reality glasses to help the visually impaired recognize and navigate objects in their environment. They propose it as a **hearing aid for the blind and a potential replacement for canes and seeing-eye dogs.**

A group of researchers built a sonar-like algorithm that utilizes segmented 3D scene images, personalized spatial audio, and musical sound patterns to highlight on obstacle avoidance and spatial orientation for blind and visually impaired people by providing them spatial audio cues (Bujacz, Pec, Skulimowski, Strumillo & Mater, 2011). Their trials demonstrated that **it is possible to quickly learn and efficiently use such a sonification algorithm to aid spatial orientation and obstacle avoidance.**

Assistive Technology Competitive Landscape



A competitive landscape analysis of assistive devices shows that no existing solution currently solves both obstacle avoidance and text recognition in a single solution.

Human Assisted Devices

Tools like Aira and Be My Eyes leverage the assistance of sighted humans (crowd-sourcing techniques) by connecting the user with a volunteer who can receive video from them and assist them with a particular task. While the approach of these devices is very effective, they require available human labor and thus lack scalability as the human agents are not only expensive but are limited in supply (Coughlan & Miele, 2017). Plus these devices can only help if the blind individual can take a well-framed picture of the desired target. Thus, this technology poses a fundamental challenge: how can a user with little or no vision take a well-framed picture of the desired target object? Similarly, machine vision systems like OrCam can autonomously provide information about text, faces, colors, and objects. While this technology again has an astonishing capability, **it does not support the user navigate around obstacles or in helping them find relevant objects to scan.**

Augmented Reality Assistants

Researchers at Massachusetts Institute of Technology (Liu, Stiles, Meister, 2018) have used augmented reality to build a visual-assistive device called CARA (Cognitive Augmented Reality Assistant for the blind) that reads the objects in the environment on the user's command. The system has an intuitive user interface as each object in the environment had a voice and communicated with the user on command. They found that the system was able to help blind subjects in traversing an unfamiliar multi-story building on their first attempt. While such a system can prove to be very useful while one is navigating, our user interviews indicated that such a system has a potential for cognitive overload by presenting a constant outflow of verbal information from the system.

Our Approach

We believe there is a **better way of presenting this semantic information in the surrounding** to users and thus our prototype uses a combination of audio clues and voice outputs. The information is delivered to the user using existing augmented reality devices. Studies have shown that blind people are more sensitive than sighted people to binaural sound-location cues, particularly inter-aural level differences(Nilsson & Schenkman, 2016). We believe that the combination of sound design with augmented reality technology can provide people who have visually impaired a greater sense of awareness of their surroundings and supplement them with information that can not be provided by their cane or a dog. Our proposed solution uses a **mix of spatial audio cues** (to help our users not only avoid but also locate the interested objects) **and speech sounds** to communicate semantic information.

User Research

We performed generative user research to identify and formulate our goal of solving the 5 meter problem. Our interviewees included renowned tech accessibility evangelists, disability experts, advocacy groups for the blind and particularly **visually impaired users who have in-depth experience with assistive technologies**. Our interview results revealed four major themes.

1. Users Want To Augment Their Present Capabilities

Through interviews with visually impaired users of assistive tech, it was clear that the proposed solution would need to work alongside and not replace a white cane and/or a guide dog. Many visually impaired people needed to use multiple assistive mobile apps to solve different problems at different times. One user revealed how apps such as AIRA were useful within the last 10 feet of their destination while OrCam was used to scan print text. There was clearly no all-in-one solution.

2. Users Don't Want To Be Overwhelmed With Information

Blind users rely primarily on their hearing to navigate. Hence it is important to avoid cognitive overload that can stem from presenting the user with too much auditory information. Hearing is a much lower bandwidth source of information than sight, so trying to play too many sounds at once would give users a hard time in separate and processing them. The application would need to selectively sonify and avoid redundant information. Covering up sounds in the environment with others that are too loud or distracting could be hazardous.

3. Smartphones Are Not Ideal Assistive Devices For Blind People

While the form factor of a smartphone provided users ease-of-use, it was difficult to operate with a cane in one hand and a dog leash in another. Additionally, it is hard to orient the phone camera to the exact target as even a slight wrist movement could offset the right angle. AR headsets were hence viewed favorably as alternatives although concerns of having an expensive AR Device while out in public were voiced by many participants. Apart from the heft and discomfort from prolonged usage, some participants shared concerns about theft or loss, aesthetics, and drawing attention to themselves. There was also a possibility of the visor from AR devices interfering with hearing and echolocation.

To improve device adaptability, users requested support for natural language commands in addition to gestures in the proposed solution.

4. No Current Solutions Effectively Address The 5 Meter Problem

GPS works well in outdoor situations and helping users wayfind to the general location. However, blind users mentioned that they face particular difficulty in getting assistive technology to help them when they are within 5 meters of their target or destination. As one user mentions "knowing how **close** something is to us in a space is the top need of the hour".

Some scenarios that were identified as part of the 5 meter problem during our user research were knowing what store the user is passing in a mall, identifying a restroom, a particular door in hallway, a specific gate, the right ticket counter etc. Additionally, some users expressed challenges performing tasks such as queue following.

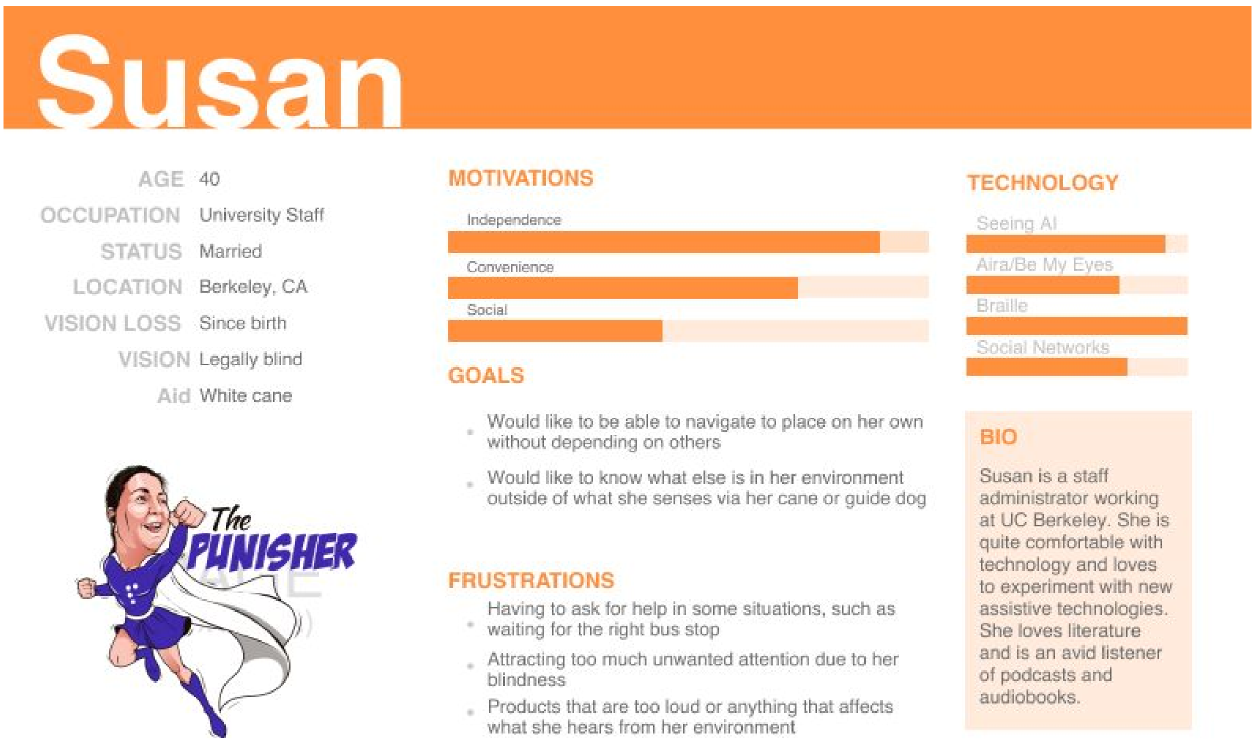
After brainstorming we identified that the "5 meter problem" could potentially be solved via mix of obstacle avoidance as well as text recognition capabilities - this can help deliver semantic information from the environment to blind and visually impaired.

Prototype Design & Development

Personas

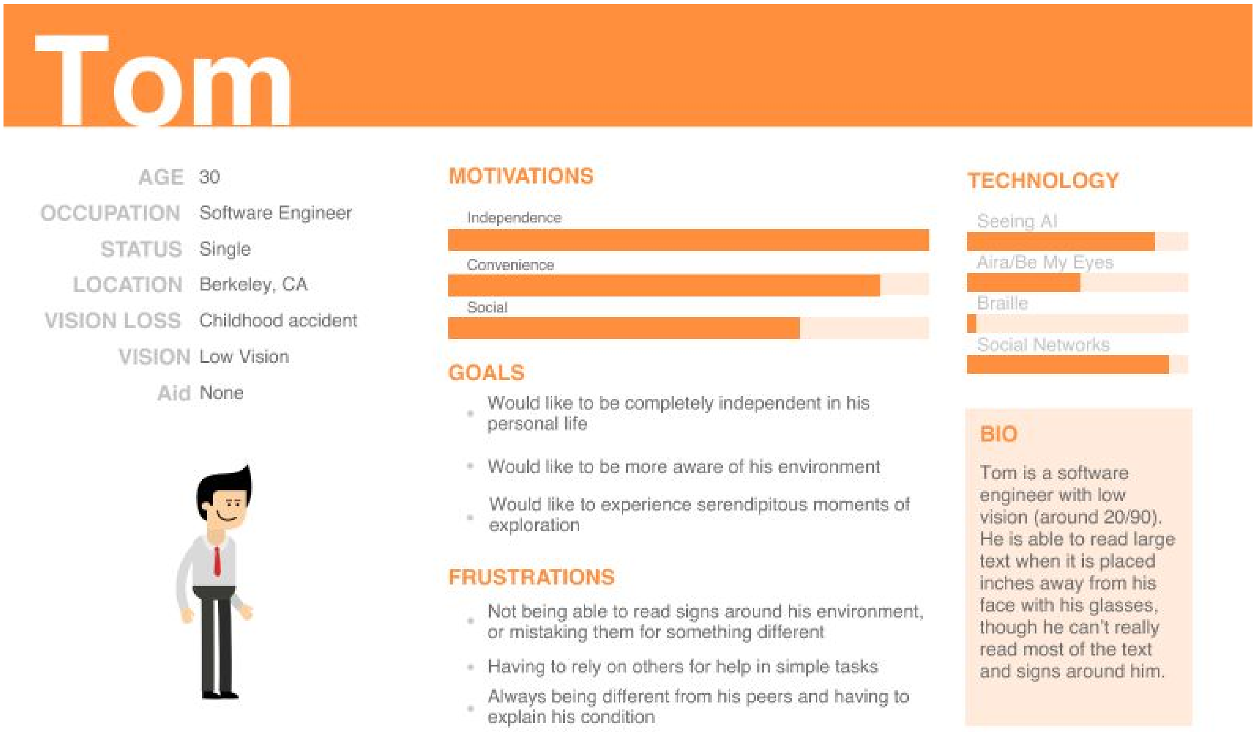
Based on our user research, we developed two personas to help guide us during the design and development process.

Primary Persona: Susan



Susan is a staff administrator working at UC Berkeley. She is quite comfortable with technology and loves to experiment with new assistive technologies. She loves literature and is an avid listener of podcasts and audiobooks.

Secondary Persona: Tom



Tom is a software engineer with low vision (around 20/90). He is able to read large text when it is placed inches away from his face with his glasses, though he can't really read most of the text and signs around him.

Hardware

As our original idea relied on the ability to create spatial maps of the user's surroundings, we had two options: the Magic Leap One or the Microsoft HoloLens Development Edition. We opted for the latter based on availability and development resources.

The HoloLens (see image to the right) is an augmented reality device. Unlike virtual reality devices which totally block out the user's senses, it is intended to augment their physical environment with digital information. Typically, that means displaying visual information such as holograms for the user to interact with; however, we used it primarily for audio information.

The HoloLens is equipped with many sensors essential to our application, including multiple cameras, an inertial measuring unit (IMU), microphones, and multiple infrared illuminators that help it detect the depth of objects. It’s also equipped with a powerful onboard processing unit and a spatial audio system that uses a head-related transfer function (HRTF) to simulate sounds coming from any direction around the user. Finally, it can connect to the internet via wifi and access important APIs.

An important aspect of the HoloLens is its ability to construct spatial maps by meshing the user’s environment, as seen in the picture below. It uses information from the sensors to create a triangular mesh that lets applications approximate the positions of real objects. We use this to help us sonify the environment and place text information within it.

A spatial mapping mesh created by the HoloLens

Development Strategy

In preparing to develop AR for VIPs, we decided to split into three subteams. One team would work on each of our core features, **obstacle sonification** and **text recognition**; and a third team would focus on **sound design**, ensuring all the sounds of the application worked well together.

After some delays in acquiring the necessary resources to begin prototyping, we had approximately 8 weeks in which to develop our prototype from start to finish. Our code can be found here: https://github.com/arvips/Mesh-Manipulation

Obstacle sonification

Our initial goal for the project was to sonify the entire spatial mesh around the user. In theory, this would give the user the echolocation-style ability to hear the objects around them, and know without looking the locations of all nearby obstacles. Ultimately, this proved overwhelming (see “Spotlight Mode” in sound design, below) but this idea drove our initial development.

Spatial Mesh

In our initial model, each object around the user would make sounds letting the user know of their presence. Closer objects would sound louder, and we attempted to experiment with the pitch of objects based on height as well. However, we had an important constraint: **we could not tell where one object started and another stopped just from the spatial mesh.** Without advanced object recognition, we couldn’t tell for example that a desk and a laptop lying on it were two separate objects; instead, all of these blended together into one shape in the mesh. However, even if we couldn’t differentiate objects, anything that was meshed would be an obstacle that the user would have to either navigate around or be interested in focusing on.

Obstacle Beacon Distribution

Our solution was based on “obstacle beacons” **-** point sources of audio that would spread out over the mesh and produce sound. We set up a command that, when triggered, would perform several hundred raycasts outward from the user; wherever these hit the mesh, we would instantiate an obstacle beacon. This had the result of “spraying” beacons from the user onto obstacles, as seen below.

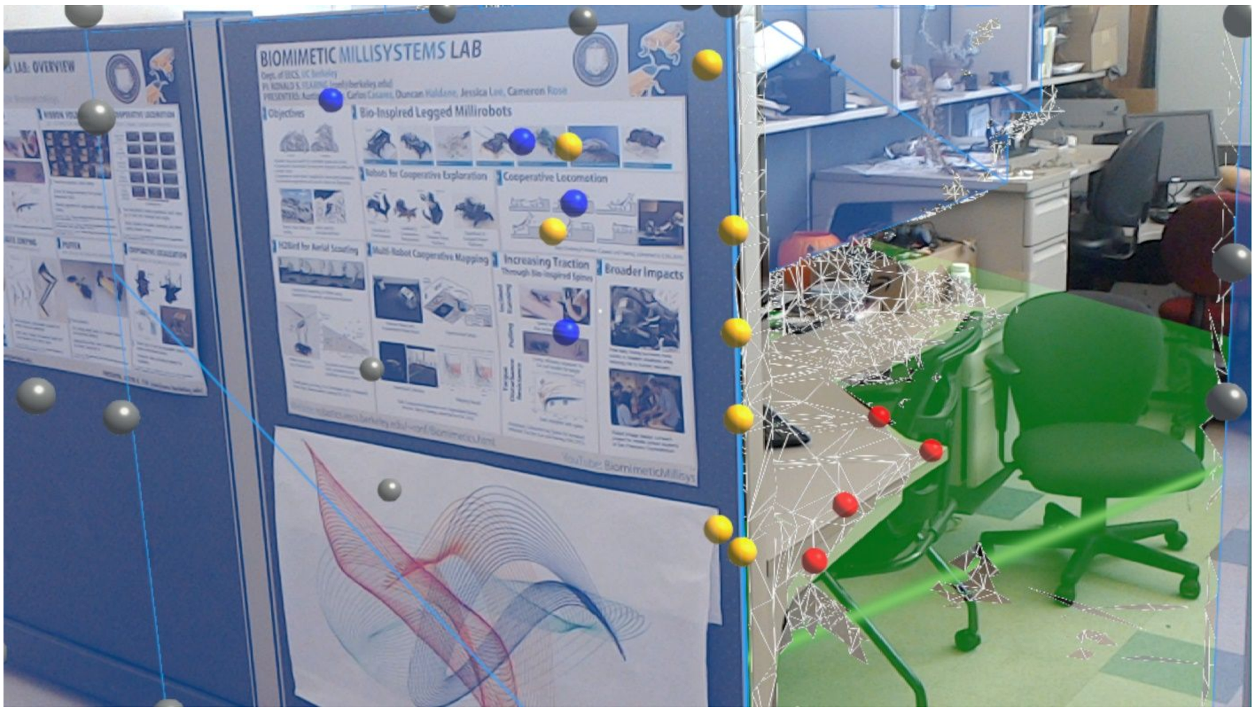
The spatial mesh with obstacle beacons. In Spotlight Mode, only the beacons near the user's gaze would produce sound (as seen here in red).

One issue we ran into when raycasting is that narrow objects like poles were unlikely to receive beacons - unless they happened to be sprayed exactly at the pole, they would go past it and stick to further objects. We resolved this by changing the raycasts to spherecasts - instead of functioning like laser pointers, it was more akin to shooting out cannonballs (whose size we could adjust at will) and sticking to the first obstacle that any portion of the cannonball hit. That way, we could shoot fewer beacons but be more assured of them hitting even narrow objects.

The next issue was that as the user walked further from where they had performed the last spray, obstacles were decreasingly likely to have beacons on them. This was especially true if the user turned a corner or entered a new room, such that the HoloLens didn't have line of sight to these new obstacles. To counter this, we changed "obstacles" from a singular command to a toggle: the user could say "obstacles on," and until they turned the feature off, the app would periodically spray new beacons. While we experimented with respraying every few seconds, we found more success in respraying every time the user moved a certain distance away from the last spot where a spray was performed. There was a slight interruption in sound and small performance lag when beacons were sprayed, so we didn't want to do it too often; we found that 1.5-2 meters was a sweet spot. With improvements to the hardware and sound design, this distance could shrink to have more consistent beacon placement.

Mesh Processing & Wall Beacons

Finally, during development we noticed that all of the beacons sticking to walls, floors, and ceilings were drowning out those attached to more relevant obstacles. To solve this, we used mesh processing to identify portions of the mesh that were floors, ceilings, and walls, respectively. We prevented beacons from instantiating on floors and ceilings, and caused hit detections on walls to instantiate a “wall beacon” instead of an “obstacle beacon,” with a much more subtle sound. The net result was that it became much easier for users to differentiate relevant obstacles from only semi-relevant walls.



An example of a processed mesh. The blue planes represent walls, and the green floors or ceilings. Typical obstacles would use obstacle beacons (red), whereas those identified as walls would use wall beacons (yellow) with a much more subtle sound. Neither obstacle nor wall beacons would attach to floors or ceilings. Text beacons (blue) could attach to any surface.

Unfortunately, performing this mesh processing currently demands stopping the ongoing scan, meaning that we would have to scan an area fully before enabling this feature; and additionally, obstacles close to walls could sometimes have their mesh accidentally deleted.

Future Improvements

Gaze-based audio - In addition to the obstacle localization provided by the audio beacons, we could also directly sonify the user’s gaze, allowing them to hear more instantaneously when their gaze passes from a close object to one far away.

Dynamic mesh processing - Currently, in order to analyze the mesh for walls, ceilings, and floors, the app needs to stop scanning the environment. A means of processing the mesh without halting the scanning process would allow much more flexibility.

Object recognition - Combining the mesh map with an object recognition algorithm could help give the user a better idea of what exactly is in the space around them and perhaps even divide the mesh into discrete objects that could be sonified separately. This would open up the opportunity to use representative sounds, e.g. a flushing sound to represent a toilet.

Mesh matching - If we have access to pre-existing meshes of an entire space (perhaps a museum mesh, including annotations for bathroom and exhibit locations), we could compare the user's current mesh to the larger mesh to understand where in the space the user is.

Text recognition

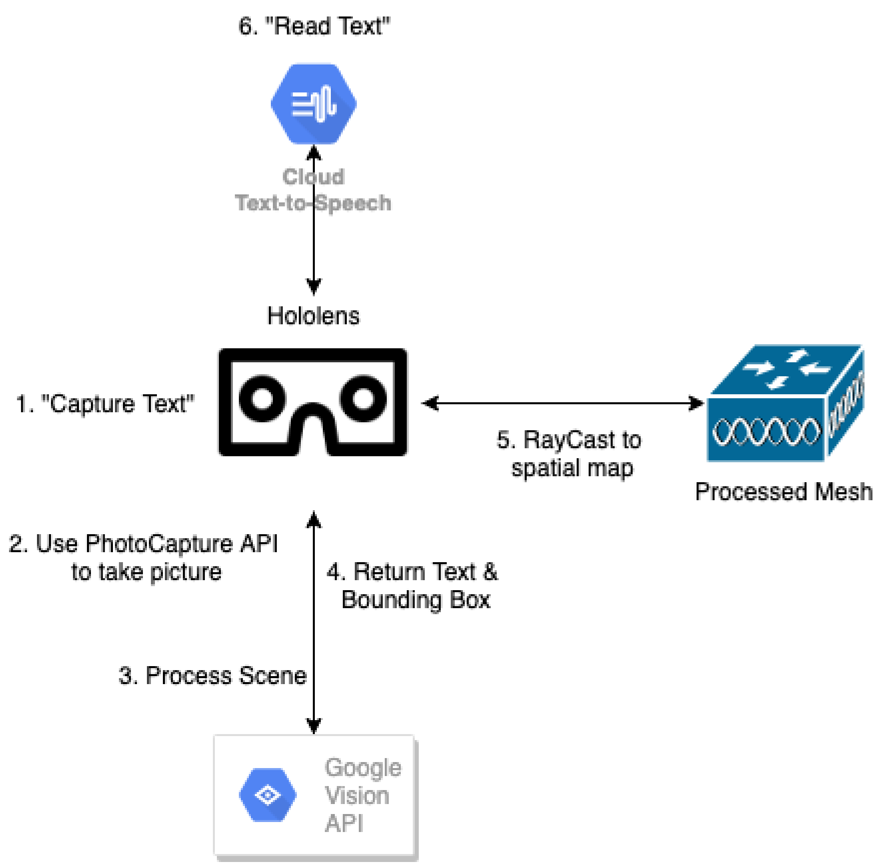
One of the key goals of our project was to solve the 5 meter problem. To effectively and completely capture contextual information from the environment the device needed to support text recognition. A major shortcoming of visual aid devices for the blind is in knowing when and where to look out for text in their surrounding so that they may then use text-to-speech; our obstacle sonification system is designed to help address this issue.

Manual Mode

To test the combination of spatial navigation for wayfinding along with text recognition, we designed for 2 modes- manual and automatic. In the current manual version the user would be made aware of a obstacle such as a bus stop sign in the environment using the obstacle detection commands stated earlier. Using voice commands such as “capture text”, “read text”, and “read all text” the user could then prompt the device to read out any text within 5 meters in their field of view. Since users trigger text recognition features manually, their hearing will be largely unaffected until the user feels they are in a safe enough space to use the application. If detected, the text will be read out one at a time- using 3D stereo sound of the hololens to indicate the positioning of the text in the present field of view of the user. Playback commands such as “pause”, “stop”, “next” and “repeat” gave more control to the user when encountering multiple texts in the environment.

Application Flow

To perform text recognition, the in-built camera on the hololens paired with the PhotoCapture API in Unity, takes a photo of the environment. After sending the image to Google Text Recognition API a response is returned with the text string along and its bounding box. Google Vision APIs proved to be the most accurate of the Optical Character Recognition (OCR) approaches in test environments because of its consistent performance analyzing natural scenes.

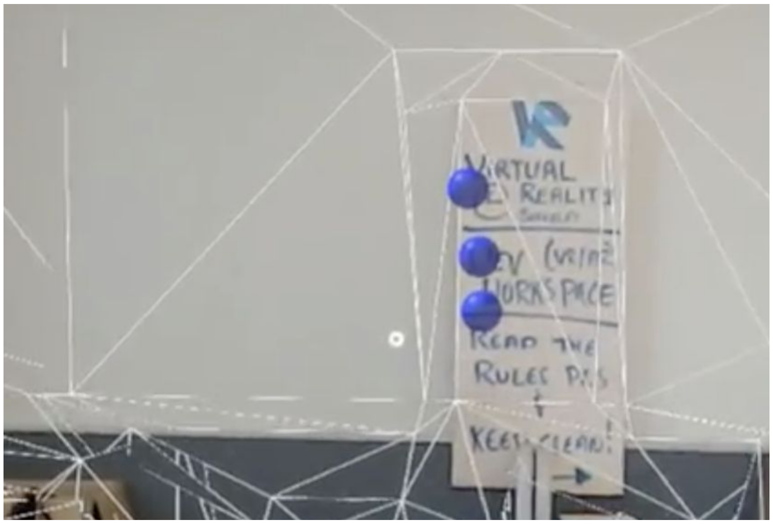


Architecture diagram of text recognition process.

The application detects text located in front of the user sprays/spherecasts spherical beacons where text exists on the spatial mesh. These beacons are anchored to the location of the target so that when users move around, the beacons stay in the same location as the original target. A large part of the text recognition functionality was based on the text detection state as well as thresholds, parameters and other design decisions from prior research in TextSpotting (Huang, 2017).

Design Tradeoffs

We also designed the application to avoid capturing duplicate images by ensuring no new images were captured on top of existing beacons in the mesh. This saved a significant amount of time sending an image back and forth to Google's API as well as computational resources for the Hololens as each beacon placed into the mesh was process intensive. This meant making an important design tradeoff of accuracy vs efficiency. Bigger but fewer spherical beacons meant there was some difficulty in capturing 2 different signs that were very closely adjacent to each other.



An example of spherical beacons placed over text in 3D mesh

However, the key challenge of an out of the box Text Recognition API was the latency issues that occurs from capturing and sending the image, performing OCR over the cloud and returning the text string along with bounding box. The process had to conclude with then placing beacons over the text in its 3D environment and reading out using Text-to-speech capabilities of the Hololens. Nonetheless, we reduced the entire process to an average 4 seconds but possibly longer due to internet connectivity issues. Low light conditions and catching text at an angle were some observed limitations based on the present hardware capability of the Hololens.

Future Improvements

In our next iteration, we hope to implement an automatic mode, wherein the application will continuously look for text by capturing photos every few seconds. This would enable users to command the device to “look for [certain text]”; it would then keep scanning until it finds the text and then guide the user towards it. This feature was highly requested in our user testing.

The same concept can also be extended to facial recognition to let blind users know that there are people nearby, or pose detection and sentiment analysis giving blind users more context around the non-verbal cues that sighted people depend on when interacting with others. In this way, the AR headset would not only act as an external vision system for users, allowing them to find relevant objects (such as specific doors), and gain awareness of the presence of people around them, but could also be a way for them to “read a room”.

Sound design

From our initial user research, we found that unlike visual input, most people are very sensitive to too much auditory stimuli. This was further validated with existing literature and research. We also found that blind users do not want any devices that could interfere with how they perceive their environment, including listening to environmental cues (such as the sound of traffic). These insights directed us in the sound design process for obstacle sonification and text location.

Sound Characteristics

In the sound design process, we then experimented with various sound characteristics:

Foreground vs background - Foreground sounds are typically alerts or notifications (aims to inform the user), or alarms and warnings (usually convey adverse or urgent situations). Background sounds, in comparison, are less urgent or important, and typically convey additional non-crucial information. In our case, we went with the obstacle beacons as the foreground, and wall beacons as the background.

Simple vs complex - We experimented with simpler tones, like sine waves and triangle waves, and more complex tones such as a combination of musical instruments. We found that simpler tones work better due to the presence of transience, which communicates localization. Simpler tones also helped mitigate issues with phasing (which impacts spatial localization) as a result of multiple voices playing at the same time.

Natural vs artificial - We explored the spectrum of sounds in the natural environment (such as raindrops or footsteps, or sounds composed artificially). We opted to go for more artificial sounds so that it was easier to vary other auditory dimensions such as frequency.

Musical vs non-musical - We explored both tonal but non-musical sounds (like sine waves), as well as short musical phrases for a more aesthetic experience. We decided to go for non-musical tones, as our research indicated that repetitive musical phrases tend to sound annoying after a prolonged period.

Length of audio clip - We varied the length of the audio clip attached to each obstacle beacon to find the optimal length. We decided to go with a shorter clip for the obstacle beacon to better accommodate for users' head and body movements.

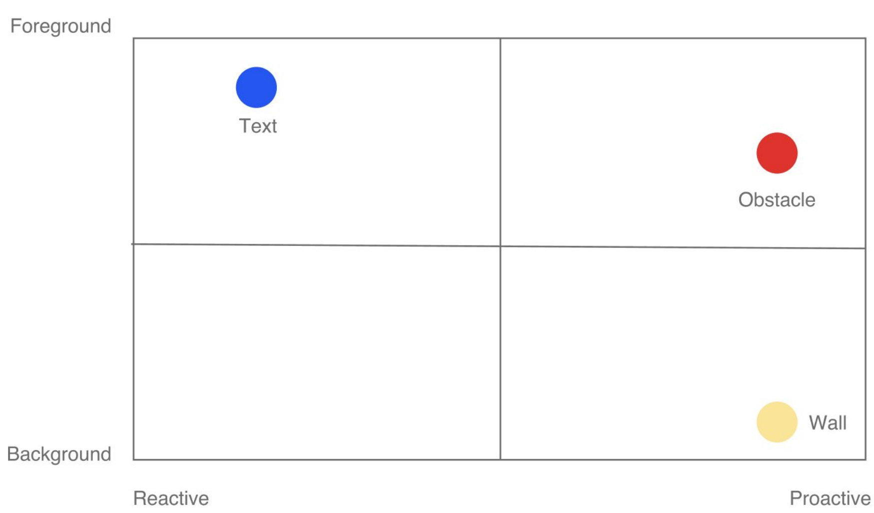
Auditory Dimensions

While there have been research in sensory substitution that attempts to convert raw images to corresponding soundscapes (such as the vOICe), a high amount of training and effort is required to train the brain in allowing the brain to decode the soundscapes in visually meaningful terms. Even then, the result is uncertain. To avoid a steep learning curve, we steered away from direct visual-to-auditory experiences, and used more intuitive auditory dimensions to represent useful real-world information.

Volume - Used to indicate distance; the louder the beacon, the closer it is (see below for Proximity Mode). The range of the variance in volume was kept within the vocal loudness of no more than 70 db.

Frequency - We explored varying the frequency according to the relative height of the obstacle, hence the higher the obstacle relative to the user, the higher the frequency. However, we faced a number of challenges: most people are unable to perceive a small difference in frequency. In addition, the overall soundscape of obstacle beacons with varying frequencies was jarring. This was further limited by how change in frequency would be implemented in the Unity Game Engine: any change in frequency would also impact the length of the audio clip.

Final Sound Design



This sound design diagram shows our final choices regarding beacon characteristics.

Obstacle beacon - earcon of obstacle beacons, which are attached to any obstacles identified by the mesh. While the obstacle beacon sound should be in the foreground, because it is repetitive, we decided to go for a more subdued tone instead of harsher warning sounds. The final sound design was a short, simple triangle wave attached to each obstacle beacon.

Wall beacon - earcon of wall beacons, which are attached to walls or any large, flat, and vertical surface areas identified by the mesh. Since this should be background noise, we decided to go for white noise on the wall beacons.

Text beacon - earcon of text teacons, the user hears it when text is successfully captured. We used a bell (ding) sound effect, designed to be distinctive so that the user could hear it above wall and obstacle sounds.

Sound Modalities

Spotlight Mode - when turned on, spotlight mode only sonifies the beacons that the user is looking at, while all other beacons remain silent. This was implemented as we found that it is very difficult for users to differentiate how far the obstacles are when all the beacons are being sonified. Spotlight mode is now on by default.

The picture below visualizes how spotlight mode works. The obstacle beacons highlighted red are part of the “spotlight” where the user is looking, and that is the part being sonified. The other obstacle beacons colored grey are silent.

Visualization of Spotlight Mode

Proximity Mode - refers to how the volume of obstacle beacons change as the user approaches the obstacle. When turned on, the volume of obstacle beacons remain generally constant, and then increase by a steep amount when a user gets closer to the obstacle, as an indication to the user of how far he/she is. Whereas, when turned off, the change of volume remain relatively constant as the user approaches the obstacles.

Future Work

We have laid a good foundation for future work in sound design. Further exploration of embedding semantic information into sound could be helpful. Examples include encoding relative text size to a particular auditory dimension to communicate visual hierarchy. Other examples include translating common icons to earcons. For text recognition, further user research into the tradeoffs between reading text immediately or playing a earcon would be helpful.

User interface

Based on our user research, we knew our interface would have to meet several key criteria:

1. **Minimize sudden or unexpected sounds** that could disrupt our users' hearing at an important moment.
2. **Keep core commands simple**, so users can focus on their surroundings instead of the interface; avoid cognitive overload.
3. **Be usable with at most one hand** so the other could hold a traditional white cane or guide dog lead.

Core Commands

Our core voice commands, the ones we would expect all users to learn, are as follows:

Obstacles on/off - By saying this command, users can toggle the obstacle and wall beacons on or off. As mentioned above, the beacons continuously refresh while the user walks around as long as obstacle mode is on. We also set our single unassigned gesture & remote button to this function, as being able to turn off this repeating source of sound quickly would be important.

Capture Text - This command causes the app to take a picture and begin analyzing it for text. After a few seconds, the user hears either a "No text found" message, or a ding that indicates the presence of text.

Read Text and **Read All Text -** These commands read off captured text. If "read text" is used, the user will only hear the text in a cone near their gaze, similar to the "spotlight mode" for obstacle beacons. If "read all text" is used instead, the app will read off all text captured in the current session. The former is useful when the user has just captured text and wants to read it quickly; the latter when the user is not certain where a piece of text they captured is located.

Next/Stop/Repeat - These commands let user adjust text playback: skipping to the next piece of text, stopping playback, and repeating the last played text, respectively.

Increase/Decrease Speed - These commands increase or decrease the speed of text playback. Combined with the above playback controls, these tend to be expected in most audio-only accessibility applications.

For a full list of commands and setup instructions, see Appendix A.

Based on technical limitations (see below), the majority of our commands are voice only. Future versions would ideally offer gesture- and controller-based alternatives to the voice commands listed above.

Audio Confirmation

To help the application feel responsive, we added **audio confirmations** to the majority of our commands. After the user issues a command, the system will confirm it by replying "obstacles on," "Capturing image; analyzing…" or similar. We felt it important to quickly give the user feedback on whether they had been heard correctly, and this was later confirmed by our user testing feedback.

Hardware-based Limitations

The HoloLens development edition, while an amazing step forward in augmented reality hardware, is yet to reach the level of functionality that would be required to make this kind of application viable in real-world scenarios. Limitations, in order of relevance, include:

Poor ergonomics - the HoloLens is too heavy and places too much pressure on the nose to wear for long periods of time. Every user we tested with listed this as a serious complaint with the device. Additionally, its large profile can interfere with users' natural sense of sound and their ability to wear glasses.

Meshing speed and accuracy - currently, the HoloLens can only mesh a relatively small area in front of the user at once, and often requires the user to slowly walk around while gazing in every direction in order to acquire a usable mesh. Additionally, the user must often rescan areas to fill in gaps or erase phantom mesh elements. To be viable in the real world, the device would have to mesh a large area in front of the user while maintaining walking speed.

Dynamic wall detection - instead of stopping the scanning process in order to identify walls, a viable product would have to be capable of identifying walls while updating the mesh.

Battery life - the HoloLens' battery can last for a maximum of about three hours of use, or less if it is constantly scanning. This may be enough for some scenarios but poses a serious limitation to those who would rely on it for everyday use.

Sunlight interference - because the HoloLens relies on infrared mapping to create the mesh, it performs poorly in areas lit with an abundance of infrared light - i.e. anywhere with natural sunlight. To be useful outdoors, it would have to compensate for this with additional depth perception or other methods.

Improved voice command reception & background noise reduction - the HoloLens currently struggles with understanding voice commands in noisy areas or those voiced by people with accents. Reliable voice commands will be important to the user experience. It will also have to get better at recognizing freeform voice commands not previously registered with the application, such as "find [situational text phrase]."

Improved gesture recognition & controller options - currently, outside of voice commands, most HoloLens interactions are gaze-based: the user gazes at a target object and uses the "select" gesture to interact. As this is not feasible for blind people, additional gesture-based commands (hand and/or head) or controller options with additional buttons or touchpads will be necessary to have an alternative for voice commands, an oft-requested feature.

Lesser internet connection dependency - finally, most of the machine vision capabilities of the HoloLens are based on internet access via a wifi network. Connection losses and slowdowns can therefore severely impact usability. Improved integration of machine vision systems would improve its reliability as an assistive device.

Fortunately, the HoloLens 2, on sale later this year, should solve some of these issues. Nonetheless, much as phones improved gradually over decades to be the devices they are today, augmented reality headsets will likely follow the same curve of progress.

Usability Testing

Goals

We wanted to gauge the **effectiveness** of our prototype in communicating the semantic information present in the surrounding to the user and in turn **assessing the usefulness of the prototype in assisting blind and visually impaired people** in navigation. We conducted two rounds of usability tests to evaluate our approach.

Flowchart of process followed by our team. The process included 4 main phases - Learn, Build, Adjust and Test.

Testing Process

Round 1: Trial Run With Sighted Participants

The first round of user testing was conducted with 2 graduate students with normal or corrected-to-normal vision (1 female and 1 male). Both participants wore the Hololens equipped with our designed software. The participants were instructed to close their eyes and only rely on audio cues in their surroundings (natural and those produced by the augmented reality device) while performing the tasks.

Recruitment: We recruited two UC Berkeley students as participants for this usability test. We communicated that the prototype is still “Work in Progress” and that it is not a finished version to make it easier for them to express their honest opinion.

A picture of a participant performing tasks in Round - 1 of Usability testing

Round 2: With Blind and Visually Impaired Participants

The second round of user testing was conducted with a total of 7 blind participants (5 female and 3 male). The participants had a varying degree of visual impairments - ranging from low vision to complete blindness. The participants wore the Hololens equipped with our application and were allowed to use their cane or guide dog while performing the usability testing tasks.

Recruitment: For recruiting blind subjects, we approached various organizations for the blind and visually impaired individuals (such as the Lighthouse for the Blind and Visually Impaired, East Bay Center for the Blind, National Federation of the blind) and blind individuals. We shared our hypothesis and the purpose of our usability test and asked them to evaluate our prototype.

Rationale Behind the Choice and Number of Participants for Usability Testing

Research has shown that testing with 5 people lets you find almost as many usability problems as you'd find using many more test participants and that best results come from testing no more than 5 users (Macefield, 2009; Nielsen & Landauer, 1993). Tom Landauer and Jakob Nielsen have shown that the number of usability problems found in a usability test with n users is: $N(1-(1-L)^n)$, where N is the total number of usability problems in the design and L is the proportion of usability problems discovered while testing a single user.

Since our system is specially developed to help blind people in navigation, we decided to test our prototype with 5-7 blind users. Usability testing with blind users helped us in assessing the potential of a system like ours that uses both audio cues and speech sounds to convey the semantic information in the surroundings. Testing with blind and visually impaired users helped us in testing our hypothesis, evaluating our identified problem in the initial round of user interviews and in turn understanding the effectiveness and limitations of our approach.

Since our target population was hard to reach, we decided to conduct a round of usability testing with sighted participants beforehand. Testing the device with sighted individuals was done to comprehend:
a) if the prototype was behaving in an intended way,
b) if the information was delivered to the user in an easy to understand manner and
c) if the participants were able to understand and subsequently complete the tasks used in usability testing.

As the **goal of this round of usability testing was to perfect the setup for the interested population,** only 2 sighted participants were recruited for the first round of usability testing. The first round of usability testing helped us in polishing the tasks and the prototype for our target group.

Logistics

We aimed to create a fair testing environment and a similar setup and experience for each participant in the usability testing.

The tests were conducted at the following two locations:

South Hall, UC Berkeley - Usability tests for 6 participants (2 sighted and 4 blind and visually impaired participants) was conducted in a room in South Hall. The room was emptied so that users would have enough space to physically perform the tasks. We used a combination of chairs, poles and cardboard signs to mimic bus stop signs with text (bus stop numbers). Soft plastic and cardboard objects were used in the set up to minimize any risk to our participants and ensure safety.

The setup in South Hall gave us the affordability of controlling for external variables such as noise in the background, good wifi connection.

A picture of set up used in the usability testing conducted at South hall, UC Berkeley

East Bay Center for the Blind, Berkeley - Usability testing for 3 participants (blind and visually impaired participants) was conducted in a room in the East Bay Center of the Blind. We tried creating a similar environment to the one used for usability testing in South hall by clearing the room so that users had enough space to physically perform the tasks. We used a similar combination of chairs, poles and cardboard signs to mimic bus stop signs with text (bus stop numbers), but because of space constraints, **only 2 bus stop signs** were used in this setup. As this was a public place, we had little to no control over external variables such as noise, wifi and surroundings in this particular setup.

A picture of setup used in usability testing conducted at the East Bay Center for the Blind

The bus stop signs were roughly placed at a distance of 8-10 m away from the participants, in each case. The bus stop signs were placed in random direction and a random distance from each other.

Methods and Study Design

Our usability testing consisted of two parts for each participant: a prototype testing session and a user feedback session.

Prototype testing

In this part, participants heard a short explanation of what to expect, then donned the HoloLens and launched into a series of the four tasks described below. *No training sessions were provided, and all the data was gathered live during a 1-2 hr visit.*

Tasks Scenarios for Usability Testing:
We used a task scenario for usability testing (attached in Appendix B) to provide context to the

users so that users can engage with the interface in the intended way. The users were asked to imagine that *“they want to catch bus number 52 to reach their doctor’s office and that they are dropped by a friend somewhat near the bus stop.”*

Task 1: Object Localization
Locate all the bus stops (3 in each case) in the space around you using the sound cues generated by the device
The participant could ask the device to go on a specific mode *(obstacle mode on*) in which the obstacles around them emitted audio cues. The participant was asked to point toward the direction of the 3 bus stop signs in their vicinity.

Task 2: Text Recognition
Identify the bus stop sign with stop 52
The participant could command the device to *capture and read the text* in a given direction to help the device guide them in the tasks. The participant was asked to point toward the direction of the bus stop sign 52.

Task 3: Direct Navigation
Walk to bus stop sign 52
The participant could leverage the combination of the sound cues and the text recognition of the capabilities of the device to perform the task. The task was over once the participant confirmed they have found the bus stop 52

Task 4: Text Recognition and Direct Navigation in Proximity Mode
Identify the bus stop sign 80 and *walk towards it*
The participant was asked to perform this task with the “*proximity mode on”* setting - this mode was designed to make the signs a little harder for them to hear from far away but assist them in navigating to a given object. They could use a combination of commands (such as obstacle mode on, capture text and read text) to complete this task. The task was over once the participant confirmed they have found the bus stop 80.

Providing feedback

After completing the above four tasks, the participants were asked to share their feedback and subsequently they were asked questions about the following:

- Experience of performing the tasks
- Specific questions about each task focusing on their experience, things they liked and disliked.
- Rate each task experience and the overall experience of performing the 4 tasks,
- Elaborate on their likes and dislikes
- Share scenarios where a similar device can help them in their daily life if any?
- Comments on our approach
- Suggestions for future iterations

The second part of the usability testing exercise (collected feedback using a semi-structured questionnaire) was only performed with blind and visually impaired individuals.

Rationale for Using a Qualitative Approach in Evaluation

We used a qualitative approach in our evaluation as the objective of user testing was to *understand the experience of the user and to reveal new information* such as thoughts, beliefs and experiences. Face-to-face, semi structured interviews was considered the most suitable primary data collection tool to evaluate the prototype and to enable flexible, in-depth exploration of the issue. A semi-structured questionnaire was used to guide the discussion in the process (attached in Appendix C).

The hypothesis that we were trying to test via the above described tasks was “ Can Augmented Reality devices help people with vision loss in navigating better by conveying the semantic information in the surroundings such as text in the form of audio clues and voice commands?”

A summary of Tasks used in the usability testing.

Results and Findings

Object Localization (Task 1)

Here, we tested the user's ability to localize the sound source. The participants were asked to identify the direction of the bus stops placed randomly at an 8-10 m distance from them. The bus stops like other objects in the room made a sound when the user activated the "*obstacle mode" on* setting.

Round 1**:** Both the sighted participants were able to correctly indicate the direction of the three bus stops, however, the process of finding the right direction took some time. They stated that the audio cues were faint and suggested that the sound generated by the device should be louder to make it more discernible. The feedback was incorporated in the prototype version used for blind and visually impaired subjects

Round 2**:** 5 out of 7 blind and visually impaired participants were able to correctly identify the direction of at least two of the three bus stops. Four of these seven participants were tested in a controlled setting in a room in South Hall where we had control over the noise and the wi-fi signals received by the device was better and one was tested at the East Bay Center for the Blind. In case a user pointed in the wrong direction - no attempt was made to correct them.

One participant mentioned, "*It was good - I could tell the difference between 1, 2 and 3*". Two blind participants were unable to identify the right direction of any of the 3 bus stops. They both performed the task at the East Bay center of the blind set up - where wifi connection was poor and there was a lot of background noise in the surrounding. Though most participants were able to identify the direction of the bus stop signs, many mentioned that it was hard for them to identify it was a pole. One participant mentioned that he was able to perceive the height of the poles and indicated the height of the poles with his hands. Most participant mentioned the **sound was not intrusive and provided desirable cues** - "*It gets nice and loud as I approached the bus stops*." These results show that users could accurately localize the sound cues generated by HoloLens, even though the time each user took in the process varied.

A picture of a female participant performing tasks in Round 2 of usability testing.

Text Recognition (Task 2)

In task 2, we asked the participant to identify a particular bus stop - the participant needed to take a picture in a particular direction followed by commanding the device to read the text to them to complete this task.

Round 1: Both participants were able to identify the right bus stop but the process was not seamless. The process enabled us to identify a bug in the prototype - the device was reading all past text captured by the device to the user rather than reading a specific text. The process enabled us to identify this issue and correct it for usability testing round 2 .

Round 2: Most participants were able to use the two commands "*capture text*" and *"read text*" and find the right bus stop. The two users who could not identify the direction of the bus stops in the first task didn't do well in this task as well. It appears the latency was a major issue for these two participants because of a patchy network connection. A few users were quick in identifying the right bus stop whereas a few struggled.

A common pattern amongst those who struggled was that they were looking at a height of ~5 feet (eye level) and then asking the device to find the text. The device thus gave them an output *"No text found*" in few instances. It should be noted that the bus stop sign was placed at a height of 7 feet from the ground to the bottom of the sign in our setup similar to what is true in real-world settings. These users were subsequently nudged to look up by mentioning that a regular bus stop is around 7 feet from the ground. Many users tried confirming the bus stop once they felt they were at the right bus stop by asking the device to capture text again. The device did not perform very well in these situations as the user did not know which direction to focus their gaze on at such a close distance. One user commented, "*Pointing ears more of a habit than looking in a particular direction, it took me some time since looking is not intuitive to me".* Most users, however, commented that the learning process was easy and intuitive.

Most participants felt that **commands were easy to learn and self-explanatory;** however one participant felt that the commands were poorly chosen and that things like "proximity mode" made no sense to a user who does not know what that mode does. The participants had mixed views on the input method, order of commands and output received. Almost half of them hinted that the process mimics a real-life scenario in which you need to look in different directions until you find what you want. A user remarked " *It is Interesting that I turned my head and the device read to me... I could see getting used to it*". However a few felt that process though worked properly, took a lot of time. *"I wish I could just ask it to find 52 and it could guide me to that bus stop .. rather than using the commands - capturing text and reading the text again and again... This is something that you should look to include in your future iterations."*

Most participants liked the *"read all text*" command (which reads all captured text in the surrounding and the sound appears to come from the location of the text in the physical world). A few participants mentioned that "read all text" command can be used to replace both the "capture text" and "read text" command - *"I wonder if the device can keep capturing pictures on its own rather than me asking to do capture text again and then eventually read the entire thing to me once .. that would be grea*t."

Direct Navigation (Task 3)

Here, the participant was instructed to walk to the random bus stop, located at 8-10 m away at a random location (which they identified in the text recognition task) . In the obstacle mode, the bus stops like other obstacles in the surroundings generated audio cues and the subject could use these cues along with other commands to detect and read the text and subsequently walk to a specific bus stop.

Round 1: Both the sighted participants were able to navigate to the right bus stop. They, however, hinted that audio cues were too subtle to detect. The feedback was incorporated in future usability testing sessions.

Round 2: All blind and visually impaired participants who were able to successfully complete task 2 were also able to complete task 3 by walking towards the right bus stop. They used the text recognition task to localize the bus stop followed by walking towards it to complete the task of direct navigation.

Text Recognition and Direct Navigation in Proximity Mode (Task 4)

In the last task the participants were instructed to walk to a random bus stop, located at 8-10 m away at a random location (different from the bus stop that they identified in Task 2). The participants were asked to perform these tasks with the "proximity mode" on setting in our application on augmented reality device. The mode was created with the intention of helping participants hone in to a given obstacle/object by providing more audio signals in their proximity while making objects that were farther away from them a little harder to hear.

Round 1: Both the sighted participants were able to navigate to the right bus stop.

Round 2: Five of the seven blind and visually impaired participants were able to successfully complete this task. The two participants who were unable this task were the same two participants who had issues in performing task 1, 2 and 3. It appears that poor wifi and noise were the main reason behind their failure to complete the task.

Most users followed a **two-phase strategy** to complete this task: first, localize the bus stop by using the “capture text” and “read text” command and then swiftly walk towards it. Though most of the users were able to complete this task, they were unable to differentiate the audio cues provided by the device in the “proximity mode - on” setting from that provided in the “proximity mode - off” setting.

In the feedback sessions, most blind and visually impaired participants mentioned that they see a *real utility* to this device. The participants liked the idea that their hands were free and that they did not have to struggle to aim the device or find the relevant objects to point their smartphone at. However, all participants mentioned that the device used to deliver the instructions i.e. the **Microsoft Hololens was heavy and uncomfortable.** Some blind users also mentioned that they are not accustomed to having anything on the face other than the sunglasses and most of the users said that they are *more likely to use the device if the device could resemble a sunglasses or was lighter.*

They subsequently shared situations and **scenarios in which they felt this device can help** them. Most of them mentioned that a similar device (with a few modifications) can help them in the following situations:

- At the airports when they are trying to find a particular gate,
- In the shopping mall when they want to find a particular shop/ Walking down the street to find a particular store
- In the grocery store, when they want to find a specific item
- Checking the license plate of Uber vehicles before boarding the cab
- Identifying a particular bus stop in a series of bus stops
- Finding a doctor’s office while visiting a clinic/hospital
- Serendipity of information - For simply exploring their surrounding and being aware of what’s in the environment that is not accessible to them at present
- Recognizing faces in party (future iterations of the prototype using machine vision to recognize faces)

Most users commented that the ***learning process was easy and intuitive***. They all seemed pretty comfortable in using the commands by the time they were performing Task 4. However, one blind user remarked that the commands were not user-friendly and did not convey anything meaningful to the user. While a few participant mentioned that technology or machine learning still can't replace human intervention or technologies that utilize human assistance such as "Be My Eyes". One participant on the other hand felt that "*Technology being anonymous yields a personal touch and a sense of privacy that lacks in human based approach".*

Out of seven, three blind and visually impaired participant were also part of our initial round if user interviews and they commented that we are *focusing on the right issues, have the right granularity and are on the right track*. Half of the participants felt that automatic text recognition and lock-on would be a good contender for future iteration. They wanted to be able to ask the device to find a specific bus stop on its own rather than going through a loop of capture text and read text commands. The other half enjoyed the *serendipity of information and felt that the current process was more organic.*

Though most users were comfortable using the voice commands and commented that they were already accustomed to voice inputs, a few users expressed concerns that using voice commands in a public space can draw unnecessary attention to them. They mentioned they would prefer an option to input commands both by using voice commands and gestures. Similarly, a few users mentioned that they would prefer the option of receiving both haptic/vibration and audio cues - and allowing them to choose the kind of cues they want to receive. Almost all users expressed value in the text recognition capabilities of the device without the pain of deciphering the right place and direction to point the device. A few users enquired if the technology can be integrated with other technologies such as Aira.

"Aside from the weight, I did like that my ears were exposed, it wasn't as obtrusive as some other devices...I am intrigued by the possibilities."

- *Participant 1*

"I liked the simplicity of interface.. the commands easy to recognize and remember "

- *Participant 2*

Overall, the usability testing proved that our hypothesis that Augmented Reality devices can help blind and visually impaired people in navigating better by conveying the semantic information in the surroundings in the form of audio cues and voice commands.

Further Discussion

Study Limitations

While we were able to test the prototype with a good sample size of participants (~7 blind and visually impaired participants), we were limited by certain aspects of the study:

Indoor environment - Due to the limitations of the Hololens 1, we could only use the Hololens in an indoor environment. Hence, we were unable to test the tasks for finding the bus stop in a realistic outdoor environment.

Controlled setting - The user testing sessions were also done in a controlled environment, where we made our own printed versions of bus stop signs. Users did not have to deal with the messy abundance of text and obstacles in the wild

Slight differences in test settings - Even though the user testing sessions were controlled, they occurred at different times of the day, and in different indoor settings, and hence were impacted by the different lighting and text around the user

Tight constraints - Even though we were able to define the scope to be somewhat manageable, the development timeline was still rather tight

Qualitative user study - This was intended as a purely qualitative user study to collect feedback on the prototype. However, it is worth noting that it was not a quantitative study, and we were not comparing the the prototype relative to another solution. Hence, we are unable to conclude on the specific efficiency or performance of our prototype.

Future Work

From our user testing sessions, we found that providing contextual information was highly requested. This includes enhanced text recognition, such as allowing a user to search for objects in the environment by searching for a specific text, or object recognition. We have grouped future work into four categories: Orientation Tools, Enhanced Machine Vision, Obstacle Warning, and Tutorial.

Orientation Tools

Orientation Tools are meant to help the user know their place in their physical environment, which would aid navigation. This includes features such as:

Perimeter scanning - We use sound to spatially scan around the room to allow users to feel the size of the room.

Historical path tracking - We track where a user has been historically to allow them to trace their path back to where they came from. An example would be in grocery stores, where there are multiple exits and it is easy to get disoriented.

Anchoring sound - This acts like a compass in the background, where there is an ambient sound to indicate the direction of north. It is also intended to help users navigate in confusing new environments.

Enhanced Machine Vision

Enhanced Machine Vision Tools provide users with more semantic information about their environment, which includes

Improved text recognition - allowing the user to find specific text like a spatial search; reading out text on demand

Facial recognition and pose detection - allows for people detection and sentiment analysis

Object recognition - providing a description of the environment the user is in, either proactively or reactively; describing objects when the user points to one

The AR headset would act as an external vision system for users, allowing them to find relevant objects (such as specific doors), and gain awareness of the presence of people around them, which could perhaps be a way for them to “read a room”. Features such as people detection aims to let blind users know that there are people nearby, if they are looking for help, or in the case of shopping at a grocery store, it allows them to know when they are at a cashier. Similarly, pose detection and sentiment analysis gives blind users more context around the non-verbal cues that sighted people depend on when interacting with others..

Enhanced Obstacle Warning

Enhanced Obstacle Warning would build on the core obstacle detection capabilities, and provide a more dynamic warning system. This leverages the machine vision capabilities to detect certain objects as obstacles, and provide sound as warning for each type of object. An example would be a warning for a cliff, which would be a different warning sound that that for a tree. In addition, we would also like to **build capabilities to avoid moving obstacles**, such as cars and people. While blind people can probably hear moving objects that are nearby, those that are further would be harder to detect, and could still be potentially dangerous. Knowing that there is a busy road ahead would allow the user to plan ahead and perhaps find a different route.

Onboarding Tutorial

Due to the novelty of this experience, building an Onboarding Tutorial would be helpful for most users. The tutorial would educate users on the capabilities of the product and would comprise of multiple audio simulations, bringing the user to different virtual environments, such as in a virtual grocery story, museum, or airport. New users will also learn to be acquainted with certain sounds, such as the warning sounds for obstacle detection, and learn to navigate these virtual environments while using the orientation and machine visions features.

For future features beyond next year, we would like to explore the **ability to integrate the all-purpose AR device with other human assistance applications,** such as Be My Eyes or Aira. Currently, these applications are not platform-agnostic, and users would have to use a certain platform just to access these services. With our product, we could offer a "one-stop-shop" for visual-impairment assistance. In addition, we could also expand to vision-based enhancements for people who are visually-impaired but not blind. Examples include OxSight, or utilizing Professor Emily Cooper's work in enhancing visual contrast (Kinateder, Max, et al). We would also like to explore the ability to annotate and share AR maps amongst users, similar a crowd-sourced map of notes. Users could share relevant information for each location, which can be seen by other users. While this could be interesting for blind users of our product, it would also have a broader use case beyond this product, and could be an exciting enhancement for augmented reality devices in general.

Economic Opportunities for Implementation

Supposing a mature version of the application featuring audio, visual, and human assistance could be developed and distributed over an app store, there remains the question of how blind users would gain access to it. While a few VIPs may be able to afford their own device, the most feasible solution we see in the near term is an enterprise accessibility solution.

Much like motorized wheelchairs are often lent to mobility impaired customers when they enter a grocery store, **visually impaired customers could be lent an AR device for the duration of their visit.** This type of distribution could be suitable to many public and semi-public facilities such as grocery stores, malls, airports, and museums. It can also make up for technical limitations such as low battery life, as devices could be charged between uses, and provides relatively controlled environments that would be easier to manage than "anywhere" use.

Benefits to Enterprise Customers

Enterprise customers could gain many advantages by utilizing such a system, including:

Improve blind customer user experience - Many facilities rely on visual information and signage to communicate important information to customers. Gaining access to this information could drastically improve usability and the overall user experience of blind customers.

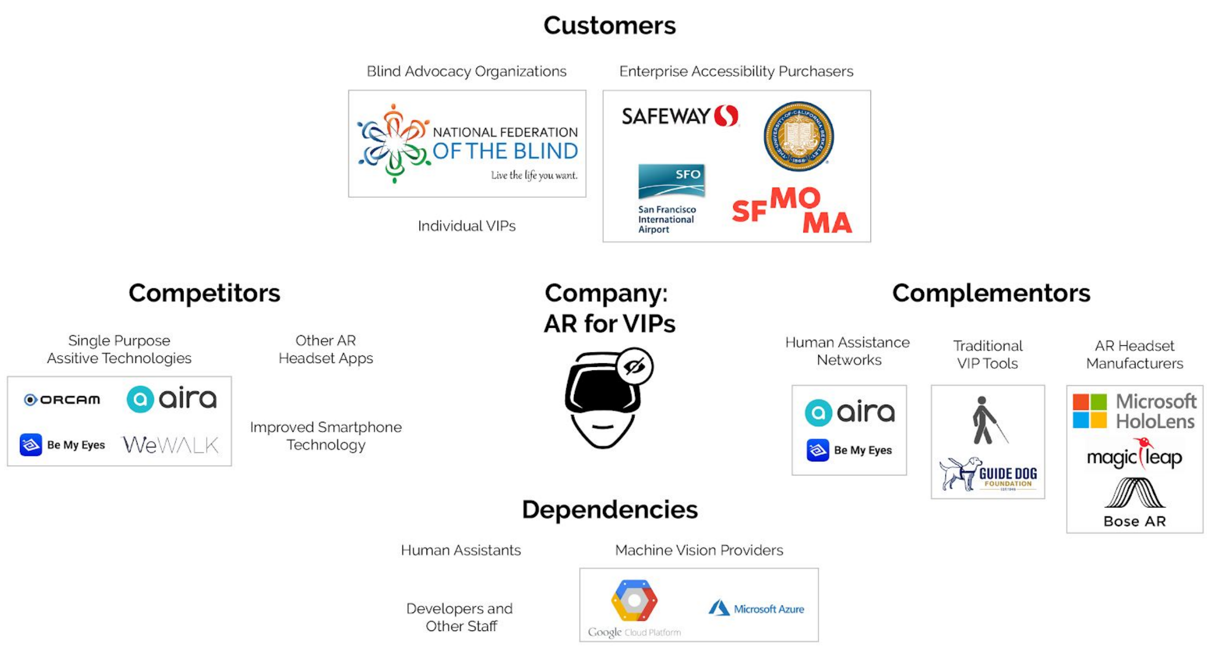
Enhanced reputation for accessibility - Making hi-tech devices available for use could increase the reputation of enterprise customers and draw more blind users to their facilities.

Free up employee time - Often, organizations simply assign employees to chaperone blind customers. An accessibility device like this could free up employee time while granting customers an increased sense of independence.

Gather navigational data - As AR devices track their location over time, organizations could get excellent data on the routes users take through their facilities (given the users' informed consent to gather such data, of course).

General-purpose devices - Unlike single-purpose competitors aimed specifically at blind users, these AR platforms are multipurpose. When not in use as accessibility devices, they could be lent out to employees for productivity purposes or to sighted visitors to enhance their experience at the facility.

Economic Landscape



This diagram shows the economic landscape surrounding a mature AR for VIPs app.

To understand where AR for VIPs could fit in the economic landscape, consider these surrounding players:

Customers - As outlined above, the primary customers would be enterprise accessibility purchasers such as grocery stores, museums, schools, and airports. Blind advocacy groups and individual blind people may also purchase the headset and application.

Dependencies - In addition to development staff, AR for VIPs would rely on machine vision providers such as Google Cloud or Microsoft Azure, as well as a human assistance network. This could be borrowed from Eyes for the Blind or Aira, or be formed specifically for AR for VIPs.

Competitors - AR for VIPs would be competing primarily with a variety of smartphone-based applications, as well as single purpose assistive technologies such as OrCam. It would also of course compete with any other assistive augmented reality applications released in the future.

Complementors - Headset manufacturers such as Microsoft, Magic Leap, and Bose would form AR for VIP's key complementors, as a headset is required for the app to function at all. Traditional tools like white canes and guide dogs function as complementors in that the software was designed to work alongside them, and these reliable tools can help cover the weaknesses of the application. Finally, Aira and Be My Eyes could be complementors were they to allow AR for VIPs to make use of their human assistance networks.

Conclusion

Despite the current limitations, this projects demonstrates the potential of augmented reality technology in helping with visual impairment by providing contextual environmental information. We believe that as the hardware and machine learning technology improve, assistive technology using augmented reality will become more commonplace. We hope this project will serve as inspiration to further the advancement and application of augmented reality technology in assisting people with disabilities.

References

1. Bourne, R. R., Flaxman, S. R., Braithwaite, T., Cicinelli, M. V., Das, A., Jonas, J. B., . . . Zheng, Y. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: A systematic review and meta-analysis. *The Lancet Global Health, 5*(9). doi:10.1016/s2214-109x(17)30293-0

2. (2019, January 15). Retrieved from https://www.who.int/blindness/en/

3. Visual Impairment and Blindness in Adults in the United States: Demographic and Geographic Variations from 2015 to 2050. JAMA Ophthalmol 2016;134:802–9.

4. Maguire EA, Burgess N, O'Keefe J. Human spatial navigation: cognitive maps, sexual dimorphism, and neural substrates. Curr Opin Neurobiol. 1999; 9: 171–177. PMID: 10322179

5. Wolbers T, Megarty M. What determines our navigational abilities? Trends Cogn Sci. 2010; 14: 138–146. https://doi.org/10.1016/j.tics.2010.01.001 PMID: 20138795

6. Legge GE. Prentice medal lecture 2013: visual accessibility: a challenge for low-vision research. Optom Vis Sci. 2014; 91: 696–706. PMID: 24978657

7. Liu, Q., Li, R., Hu, H., & Gu, D. (2016). Using semantic maps for room recognition to aid visually impaired people. *2016 22nd International Conference on Automation and Computing (ICAC)*. doi:10.1109/iconac.2016.7604900

8. Zhang D, Lee DJ, Taylor B. Seeing Eye Phone: a smart phone-based indoor localization and guidance system for the visually impaired. Mach Vis Appl. 2013; 25: 811–822

9. Asad M, Ikram W. Smartphone based guidance system for visually impaired person. IEEE IPTA. 2012: 442–447

10. Microsoft Soundscape. (n.d.). Retrieved April 8, 2019, from https://www.microsoft.com/en-us/research/product/soundscape/

11. Roentgen UR, Gelderblom GJ, de Witte LP. Users' evaluations of four electronic travel aids aimed at navigation for persons who are visually impaired. J Vis Impair Blind. 2011; 105: 612–623.

12. BlindSquare. (n.d.). Retrieved March 8, 2019, from http://www.blindsquare.com/

13. Humanware. (n.d.). Retrieved May 7, 2019, from http://www.humanware.com/en-usa/products/blindness/talking_gps/gps

14. Microsoft HoloLens | Mixed Reality Technology for Business. (n.d.). Retrieved March 31, 2019, from https://www.microsoft.com/en-ie/hololens

15. Magic Leap One: Creator Edition | Magic Leap. (n.d.). Retrieved March 31, 2019, from https://www.magicleap.com/magic-leap-one

16. Huang, J., Kinateder, M., Dunn, M. J., Wojciech Jarosz, X. Y., & Cooper, E. A. (2019, January 16). An augmented reality sign-reading assistant for users with reduced vision. Retrieved from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0210630

17. Marr, D. (2010). Vision. *A Computational Investigation into the Human Representation and Processing of Visual Information*. doi:10.7551/mitpress/9780262514620.001.0001

18. Kellogg, W. N. (2008, June 26). Sonar System For The Blind. Retrieved from https://www.sciencedaily.com/releases/2008/06/080625153404.html

19. Spagnol, S., Wersényi, G., Bujacz, M., Bălan, O., Martínez, M. H., Moldoveanu, A., & Unnthorsson, R. (2018). Current Use and Future Perspectives of Spatial Audio Technologies in Electronic Travel Aids. *Wireless Communications and Mobile Computing, 2018*, 1-17. doi:10.1155/2018/3918284

20. Loomis, J. M., Marston, J. R., Golledge, R. G., & Klatzky, R. L. (2005, January). Personal Guidance System for People with Visual Impairment: A Comparison of Spatial Displays for Route Guidance. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2801896/

21. The vOICe - New Frontiers in Sensory Substitution. (n.d.). Retrieved from https://www.seeingwithsound.com/

22. Simone, Wersényi, György, Martínez, H., Marcelo, Alin, . . . Oana. (2018, March 21). Current Use and Future Perspectives of Spatial Audio Technologies in Electronic Travel Aids.Retrieved from https://www.hindawi.com/journals/wcmc/2018/3918284/

23. Shivakumar, F., & Shivakumar, F. (2017, February 16). OxSight uses augmented reality to aid the visually impaired. Retrieved March 2, 2019, from https://techcrunch.com/2017/02/16/oxsight-uses-augmented-reality-to-aide-the-visually-impaired/

24. Oxsight Products. (n.d.). Retrieved from https://www.oxsight.co.uk/products/

25. Aira Home Page. (n.d.). Retrieved from https://aira.io/

26. Bringing sight to blind and low-vision people. (n.d.). Retrieved from https://www.bemyeyes.com/

27. Liu, Y., Stiles, N. R., & Meister, M. (2018, November 27). Augmented reality powers a cognitive assistant for the blind. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6257813/

28. J. M. Coughlan and J. Miele, "AR4VI: AR as an Accessibility Tool for People with Visual Impairments," *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, Nantes, 2017, pp. 288-292. doi: 10.1109/ISMAR-Adjunct.2017.89. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8088507&isnumber=8088417

29. Orcam Home Page(n.d.). Retrieved from https://www.orcam.com/en/

30. Nilsson, M. E., & Schenkman, B. N. (2016). Blind people are more sensitive than sighted people to binaural sound-location cues, particularly inter-aural level differences. *Hearing Research, 332*, 223-232. doi:10.1016/j.heares.2015.09.012

31. Bujacz, M., Pec, M., Skulimowski, P., Strumillo, P., & Materk, A. (2011). Sonification of 3D Scenes in an Electronic Travel Aid for the Blind. *Advances in Sound Localization*. doi:10.5772/15311

32. Huang, J. (2017). A HoloLens Application to Aid People who are Visually Impaired in Navigation Tasks

33. Macefield, R. (2009, November). How To Specify the Participant Group Size for Usability Studies: A Practitioner's Guide. Retrieved from http://uxpajournal.org/how-to-specify-the-participant-group-size-for-usability-studies-a-practitioners-guide/

34. Nielsen, J., & Landauer, T. K. (1993, April 29). A mathematical model of the finding of usability problems. Retrieved from https://dl.acm.org/citation.cfm?id=169166

35. Kinateder, Max, et al. “Using an Augmented Reality Device as a Distance-Based Vision Aid-Promise and Limitations.” Wojciech Jarosz, cs.dartmouth.edu/~wjarosz/publications/kinateder18using.html

Appendix A: Application Setup and Interface Guide

Purpose

This guide is designed to help users and experimenters get everything ready for testers to try out AR for VIPs on a HoloLens device. It will tell you everything you need to know to get things going. The current code for AR for VIPs can be found on Github: https://github.com/arvips/Mesh-Manipulation

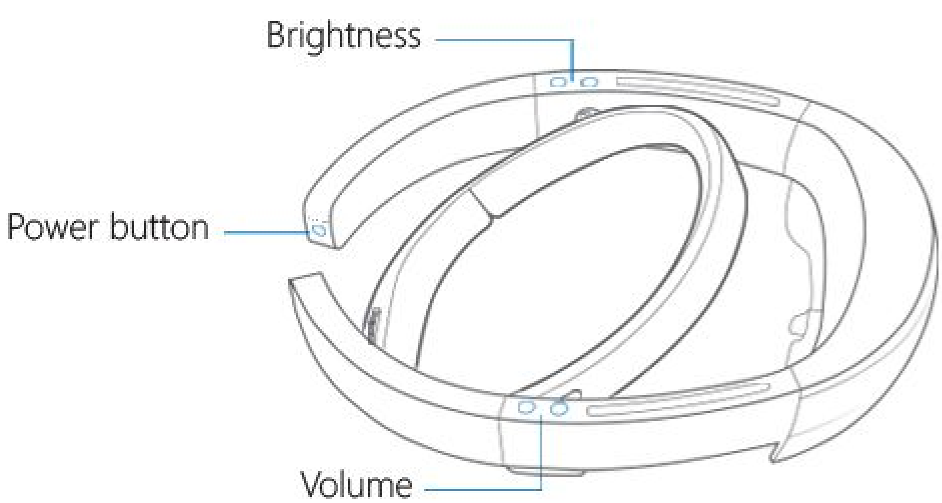
HoloLens Build Setup

For instructions on installing the latest build of AR for VIPs, see this document. Currently the HoloLenses in Cory 307 should be equipped with the latest builds.

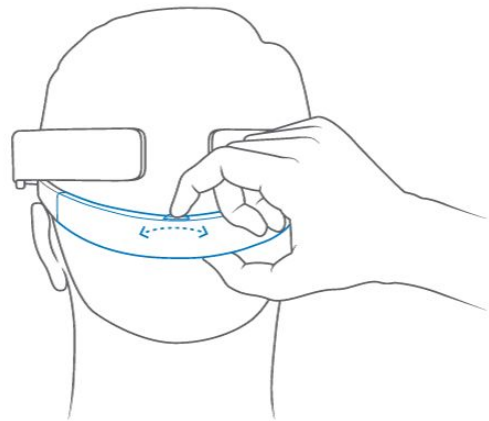
Physical Device

When packing up the HoloLens, be sure to include the device, charging cable, USB power adapter, and clicker. The case may also include instruction booklets, cleaning cloths, and fitting accessories - be careful not to lose any.

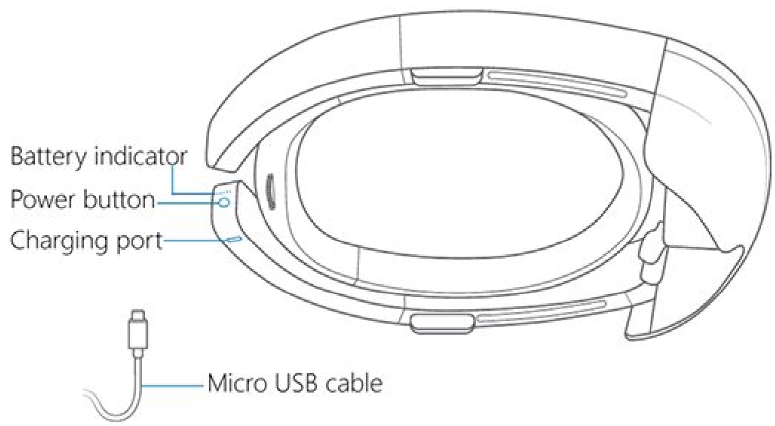
First, power on the HoloLens using the button on the end of the headband. Note that the buttons on top of the headband are used to control volume (right side) and brightness (left side).



You can adjust the fit of the device by following these instructions. Note especially the adjustment wheel on the back of the headband.



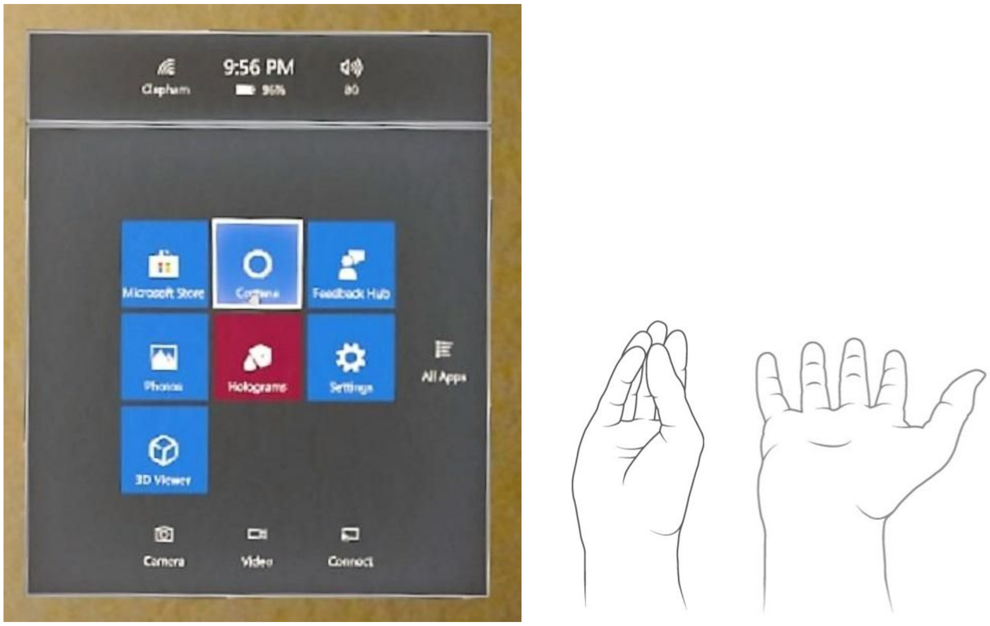
Be sure to charge the device between sessions.



HoloLens UI Basics

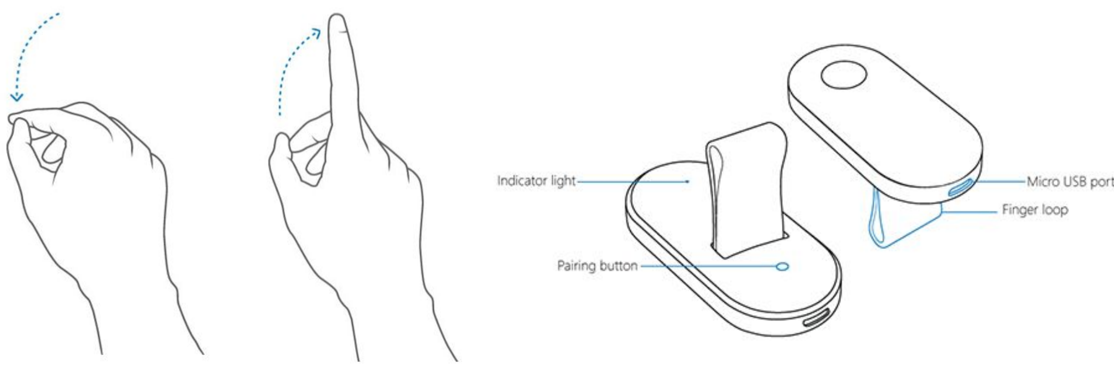
Start Menu

When you start up the HoloLens, it should scan your environment and show you the start menu (below left). You can access this menu at any time (and quit any application) with the bloom gesture, performed by pinching all your fingers and thumb together with your palm facing up, then opening them while keeping your palm facing up. (below right)



Start Menu and Bloom Gesture

To start AR for VIPs, tap on the “All Apps” section using the air tap gesture or the clicker, then select it from the apps list. By tapping and holding it, you can pin it to the Start menu for easy access.



Air Tap Gesture (left), HoloLens clicker (right)

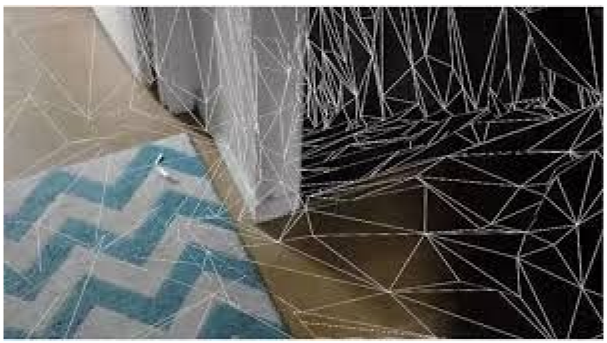
WiFi Setup

Make sure you have wifi, or none of the text recognition functions will work! You can confirm wifi status in the upper left corner of the start menu. To join a network,, launch the Settings command from the start menu (or All Apps if necessary). Tap to place the window it brings up, then select “Network and Internet,” choose a network, and sign in.

Setting Up AR for VIPs

Scan Environment

When you start up AR for VIPs, you should initially see a wireframe mesh start to form around the environment. This indicates that the HoloLens is in scanning mode. To start, slowly walk around the environment and gaze at the objects within it until you can see that the mesh is reasonably accurate.



Sample environment mesh.

Process Mesh

Use the **Process Mesh** command to have the HoloLens stop scanning. It will then convert walls, floors and ceilings into colored planes. After a few seconds, it will also delete the mesh underlying those planes. After this, you're ready for the next step. If you need to re-scan the environment, use the **Restart Scanning** command.

Core Commands

Obstacles On/Off

This enables obstacle sonification, meaning the obstacles around the user will begin making sounds. The user should be able to hear obstacles around them making sounds. The default modes for this are as follows (see below for explanation). Note that walls will sound different only if the mesh has been processed, as per above.

- Spotlight mode on
- Proximity mode on

Capture Text

This causes the HoloLens to take a picture, send it to the Google image to text API, and place sound beacons where text was found. The user should hear the words "Capturing Image. Analyzing…", followed in a few seconds by the "ding" sound of a text beacon being placed if text was found or a message saying "No text found."

Read Text

This causes the HoloLens to read text in the user's cone of view. (Say "Stop" to stop playback.) The command "**Read All Text"** will instead read all text that the user has captured this session.

Full Command List

Obstacles

- Obstacles On/Off
  - See "core commands."
- Spotlight Mode On/Off
  - If spotlight mode is on, only obstacles the user is looking at will sonify. Imagine a spotlight coming straight out of the HoloLens, and only the obstacles this light falls on will make sound. (Default: on)
- Proximity Mode On/Off
  - If proximity mode is on, obstacles will make significantly more noise when the user gets close to them, and less far away. This is good for letting users know when they are getting closer to an object, but bad for letting them find faraway objects. (Default: off)
- Shoot beacon
  - This shoots a single test obstacle beacon straight ahead of the user.
- Clear Beacons
  - This clears all text beacons.
- Max Beacons Up/Down
  - This increases the number of beacons used in obstacle sonification. (NOTE: may cause crashes if turned too high!)
- Deviation Up/down
  - This increases the spread of beacons used in obstacle sonification. Higher means more evenly spread in a sphere around the user; lower means spread in a tight cone in front of the user.

Mesh

- Process Mesh
  - This stops scanning and processes the mesh, turning walls, floors, and ceilings into planes. The planes will appear first, followed a few seconds later by the underlying mesh disappearing.
- Restart Scanning
  - This deletes existing planes and restarts the object scan.

Text

- Capture Text
  - See “core commands.”
- Read Text
  - See “core commands.”
- Read All Text
  - See “core commands.”
- Stop
  - Stops text playback.
- Next
  - Skips to next phrase.
- Repeat
  - Repeats current phrase.
- Clear Text
  - Clears all current text beacons.
- Shoot Text 1-6
  - Shoots a single text beacon in front of the user, preloaded with certain text. E.g. “Shoot Text 1” will put out a text beacon with “mango,” “Shoot Text 2” will put out a beacon with “banana,” etc. Sample phrases can be set in the Unity inspector.
- Increase/Decrease Speed
  - Increases or lowers speed of text playback.

Debugging

- Clear Debug
  - Clears floating debug menu.
- Toggle Debug
  - Toggles floating debug menu on or off.

Appendix B: Tasks for Usability Testing

Welcome to South Hall and thank you for your help on our project. For our prototype testing, we would like to first ask you to wear the Microsoft hololens headset and then perform four tasks. We will share more details on what the tasks are and how to perform them using the device once you have the device comfortably on. After completing these tasks, We will ask you a few questions related to your experience with the prototype.

"We want you to imagine a scenario where you want to catch bus number 52 to reach your doctor's office and you are dropped by a friend somewhat near the bus stop.
You now want to find your way to your the right bus stop- bus stop from where you can catch bus number 52 but there are a number of bus stops and signs in your vicinity (a total of 3 in this room)."

Hand them the device with obstacles off, proximity mode off, spotlight mode on

Task 1: Identify the bus stops
You have 3 chairs with poles attached to them around you placed in random direction at random distance from each other. Each pole has a signboard with text on them. The set up somewhat resembles bus stops in real-life. These 3 bus stops are at a distance of 8-10 m from you.

Your first task is to locate these bus stops signs

- *When you are ready for the task you can ask the device to help you hear obstacles by saying the command - "obstacles on". This should allow you to locate the obstacles (bus stops in our case) in the space around you as the sound would appear to come from where these bus stops are located in the physical setting.*
- *Once you feel you have an idea about the position of this bus stop signs, point with your right index finger towards the direction of these 3 stop signs*

Task 2: Your next task is to identify the bus stop sign with stop 52

- *You can look in any direction and say "capture text"*
- *The camera in the device will take a picture and You will hear a ping in a few seconds - indicating the device is ready to process your command.*
- *You can then ask the device to read the text to you by saying "read text". The device would then either read the text on the stop sign in from if you are looking in the right direction or "No text found." if there is no text in the direction you are looking at.*
- *We want to you identify the direction for the bus stop sign 52*
- *Point your right index finger in the direction of the bus stop sign 52 once you feel you have identified it.*

Task 3: Walk to bus stop 52

- *Now that you have identified the bus stop sign for 52 we now want you to walk towards it.*
- *You should be able to hear the beeping get louder as you start walking towards the sign.*
- *If you need to hear the sign in front of you again, you can say "read text" and it will read out the piece of text on the bus stop in front of you.*
- *If you want to hear all signs you can say "read all text" and you will hear all 3 signs and the sound should appear to come from the physical location of these bus stop signs.*
- *Once you feel you are at the correct position, You can say "capture text" followed by "read text" again to confirm if you're at the right bus stop sign.*
- *Say "complete" when you feel you have reached the bus stop 52 and completed this task.*

Task 4: Walk to bus stop 80

- *We now want you to walk to bus stop sign 80 with the proximity mode on setting*
- *The Proximity mode will make the signs a little harder for you to hear from far away but will make the beeping louder as you get closer.*
- *You can start the mode by saying "proximity mode on"*
- *You can confirm if you are at the right stop by asking the device to take a new picture for you using the commands capture text and Read text feature to hear what text lies in front of you.*
- *Say "complete" when you feel you have reached the bus stop 80 and completed this task*.

Appendix C: Interview Guide for User Testing

1. How was the experience of performing the task?

    1: Identify stop signs

    2: Identity sign 52

    3: Walk towards sign 52

    4: Walk towards 80

How was the Learning curve for the tasks and commands:

2. Rate each task on a 5 point scale   0 is the least score and 5 is the highest score?

3. What did you like or dislike about each task?

4. Can you elaborate on your likes (to help us understand the potential advantages of using a system like this in real life)?

5. Do you think a similar device can help you in your daily life while navigating? If yes, Under what conditions do you see yourself using this device?

6. Can you elaborate on your dislikes?

7. What do you think about the approach to solving this problem?
    a. What do you like about it?
    b. What do you think is lacking in our approach? How can we do better? What are the things that we should keep in mind for future iterations?

8. What modification can we do to our prototype to make the experience we intend to provide better for you?